

Modelling the active SARS-CoV-2 helicase complex as a basis for structure-based inhibitor design

Dénes Berta,^{†,1,2} Magd Badaoui,^{†,2} Pedro J. Buigues,² Sam Alexander Martino,^{1,2} Andrei V. Pisliakov,^{3,*} Nadia Elghobashi-Meinhardt,^{4,*} Geoff Wells,^{5,*} Sarah A. Harris,^{6,*} Elisa Frezza,^{7,*} Edina Rosta^{1,2,*}

[†] Equal contributions

¹ Department of Physics and Astronomy, University College London; London WC1E 6BT, UK

² Department of Chemistry, King's College London; London SE1 1DB, UK

³ Computational Biology, School of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, UK

⁴ Department of Chemistry, Technische Universität Berlin, 10623 Berlin, Germany

⁵ UCL School of Pharmacy, University College London, 29/39 Brunswick Square, London WC1N 1AX, UK

⁶ School of Physics & Astronomy, University of Leeds, Leeds LS2 9JT, UK

⁷ Université de Paris, CiTCoM, CNRS, F-75006 Paris, France

ABSTRACT

Having claimed over 1 million lives worldwide to date, the ongoing COVID-19 pandemic has created one of the biggest challenges to develop an effective drug to treat infected patients. Among all the proteins expressed by the virus, RNA helicase is a fundamental protein for viral replication, and it is highly conserved among the *coronaviridae* family. To date, there is no high-resolution structure of helicase bound with ATP and RNA. We present here structural insights and molecular dynamics (MD) simulation results of the SARS-CoV-2 RNA helicase both in its apo form and in complex with its natural substrates. Our structural information of the catalytically competent helicase complex provides valuable insights for the mechanism and function of this enzyme at the atomic level, a key to develop specific inhibitors for this potential COVID-19 drug target.

INTRODUCTION

Only a few approved drugs currently repurposed for treating COVID-19, a disease caused by the human coronavirus SARS-CoV-2, despite that its close relatives, SARS-CoV and MERS-CoV are responsible for multiple outbreaks earlier this century. As of September 2020, Remdesivir and Dexamethasone are used in clinical practice.^{1,2} Therefore, drugs need to be developed that can be used against the viral replication to help patients overcome the disease in the most severe cases.

Here we focus on determining the catalytically active complex structures of the SARS-CoV-2 RNA helicase, labelled as Non-structural Protein (NSP) 13 (Fig. 1). This protein is part of the Orf1ab polyprotein, which gets spliced to produce the enzymes required for viral replication. The RNA helicase performs two essential functions for the viral replication making it an ideal drug target. It is thought to perform the first step in the 5'-capping of the viral RNA by its

triphosphatase function hydrolyzing the 5'-triphosphate group to form diphosphate-RNA.^{3,4} Furthermore, its main helicase function is to enable RNA translocation and unwinding in an ATP-dependent mechanism during viral replication.

Accordingly, numerous studies have already demonstrated that it is possible to develop potent inhibitors of viral helicases as antiviral agents.⁵

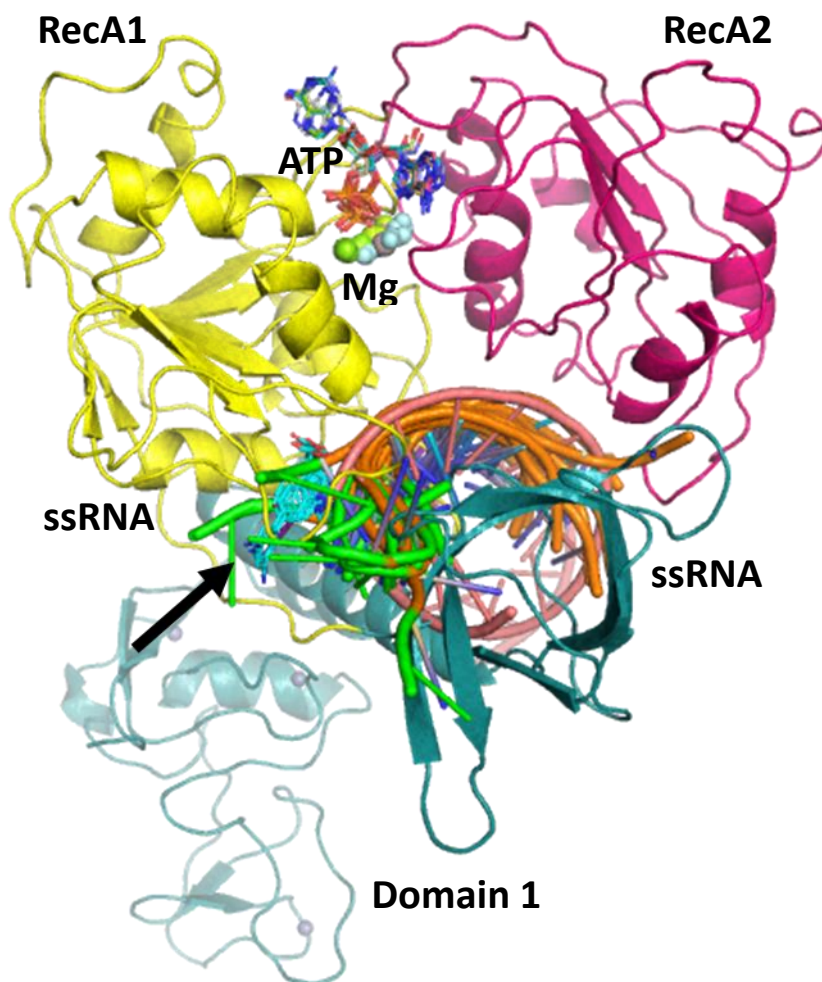


Figure 1. Cartoon representation of the RNA helicase NSP13 of SARS-CoV-2 monomer model composed of three domains: RecA1 (yellow), RecA2 (magenta), and Domain 1 (aquamarine). ATP analogues (sticks) along with Mg (green sphere) and single stranded nucleic acids are depicted from aligned homologous structures (complete list in the SI). 3' ends of the nucleic acids present the same orientation in all chains (highlighted in green). Specific helicase inhibitor binding region with allosteric inhibitors displayed in cyan (black arrow).

Coronaviral RNA helicases share high similarity. 600 out of the 601 residues of the SARS-CoV2 RNA helicase are identical to those of the SARS-CoV virus, and 70% match that of the MERS-CoV NSP13, demonstrating that these proteins are highly conserved within the coronaviridae family. Despite the importance of this target protein, currently only the apo structure is available crystallographically. There are no structures currently known of the RNA helicase bound with either inhibitors or nucleic acids, severely limiting the structure-based mechanistic understanding and the design of more potent drugs. Currently the only experimentally available information on SARS-CoV-2 helicase comes from a recent study by

Shi et al., that offer a low resolution cryo-EM structure with ATP bound in the active site.⁶ They fit the crystal structure of the APO helicase from 6jyt to their cryo-EM density maps and refined using several softwares.⁷ Recent works mainly focusing on the RNA dependent RNA polymerase (RdRp) NSP12,^{7,8} which is expressed in the polyprotein sequence just before the helicase, also yielded structures of the replication machinery, including low resolution cryo-EM images of the helicase. Unfortunately, the level of resolution is too low in this structure to model the ATP pocket in a catalytically competent conformation.

Helicase Structures and Models

The recent July 2020 SARS-CoV-2 helicase structure (PDB ID 6zsl) and the almost identical SARS-CoV helicase structure from 2019 (PDB ID 6jyt)⁹ were both resolved as crystallographic dimers (Fig. 2a-b). Interestingly, the dimerization interface is different in the two cases, leading to structurally dissimilar complexes. In the cryo-EM structure of the RdRp complexed with the RNA helicase (and co-factors NSP7 and NSP8), the two helicase protomers are non-interacting (Fig. 2c). Therefore, the catalytically active form of a monomer is of interest, and a dimer may not be the biologically functioning unit.

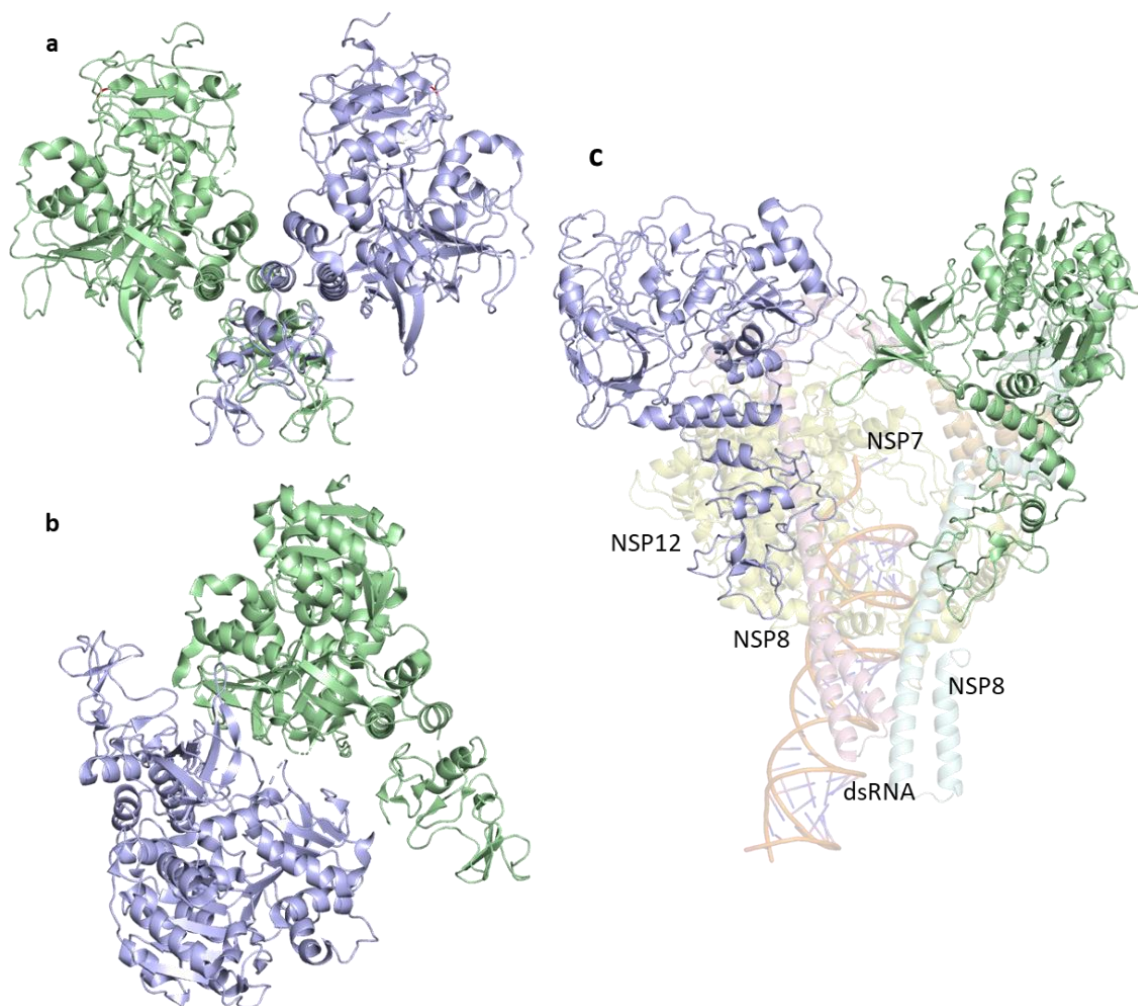


Figure 2. Structural comparison of the deposited PDB structures of the helicase dimer in SARS-CoV-1 (PDBID: 6jyt), SARS-CoV-2 (PDBID: 6zsl) and SARS-CoV-2 in complex with NSP7 NSP8 and NSP12 (PDBID: 6xez). The

interaction between the two helicase monomers differs depending on the experimental method used to resolve the structures.

Here we present the first computational models of the SARS-CoV-2 RNA helicase with ATP and RNA substrates bound. We have performed sequence similarity searches to identify key domains and homologous sequences suggesting structurally important conserved motifs. We also performed structural alignments of available homologous helicase crystal structures to help position the bound RNA and ATP substrates (Fig. 1). A previous computational model by Hoffmann et al.¹⁰ described the proposed helicase ATP interactions. However, there is no RNA incorporated in these modeled structures. Here, we also present long timescale MD simulations of both the apo and ligand bound states to address the flexibility and the stability of our catalytically competent structures. Our results will help guide ongoing drug development with the identification of novel pockets.

METHODS

Homologous Sequence Analysis

Sequence alignments were done using BLAST with default settings and BLOSUM62 distance matrix, requesting the most similar 1000 hits from UniProtKB.¹¹ Pairwise alignments for the obtained sequences were used to count identities and similarity for each residue in the SARS-CoV-2 helicase.

Homology Modelling

Proteins with crystal structures were aligned with mustang for a combined structural-sequence alignment.¹² The apo structure was based on PDB ID 6jyt.⁹ Missing residues and the I570V mutagenesis were constructed in pymol. The position of the ATP was determined using the coordinates of PDB ID 2xzo,¹³ as a template, modifying residues around the ATP pocket, except for Arg443 which was modeled based on PDB ID 6jim.¹⁴ The single stranded RNA (ssRNA) was placed based on 2xz1.¹³

Molecular Dynamics

The helicase model was used as starting point for MD simulations. The system consists of the helicase, three Zn²⁺ ions, ATP, Mg²⁺ and ssRNA with eight uracil bases. The MD simulations were performed using NAMD 2.13,¹⁵ using CHARMM36 force field.¹⁶

The system was solvated by 50,000 – 70,000 TIP3P water molecules resulting in a box of 120 Å per side. To neutralize the system and account for a 0.15 M KCl solution we added 171 K⁺ and 189 Cl⁻ ions.¹⁷ Periodic boundary conditions (PBC) were used in all the simulations and the particle mesh Ewald (PME) method was used for long-range electrostatic interactions. SHAKE algorithm was deployed to constraint the covalent bonds involving hydrogen atoms. A cutoff 12 Å was used to treat non-bonding interactions.

The energy of the system was minimized using a standard protocol via steepest descent algorithm for a total number of 10,000 steps, followed by 50 ns equilibration with restrained heavy atoms (heavy atom of the backbone of the protein and the nucleic acid with an isotropic force of 1000 kJmol⁻¹nm⁻¹) in constant pressure and temperature (NPT) and constant volume

and temperature (NVT; up to 1ns) at 303.15 K via standard MD procedure with a time step of 2 fs.

To help equilibrate the complexes, we used a harmonic constraint on selected contacts with a force constant of 10 kcal/mol for 15 ns in our preliminary MD simulations to maintain relevant contacts. These constraints were subsequently progressively reduced and removed during the next 20 ns, using the colvar function implemented in NAMD. For the apo structure, we performed five independent unbiased MD simulations, each 1 μ s long, for a total of 5 μ s simulations.

To compare simulation results obtained with MD, we also carried out MD simulations using GROMACS 2018¹⁸⁻²¹ with the Amber ff99+ parmbsc0+chioL3 force field^{22,23} for ssRNA and Amber14SB²⁴ for the helicase. To maintain the coordination of the Zn²⁺ ions, the ZAFF model was used²⁵. The molecular systems were placed in a cubic box and solvated with TIP3P water molecules.¹⁷ The distance between the solute and the box was set to at least 14 Å. The solute was neutralized with potassium cations and then K⁺Cl⁻ ion pairs were added to reach the salt concentration of 0.15 M. We used the ion corrections of Joung et al.²⁶ as this force field has been shown to produce stable RNA structures.²⁷ The parameters for Mg²⁺ are taken from Ref. 28. Long-range electrostatic interactions were treated using the particle mesh Ewald method^{29,30} with a real-space cut-off of 10 Å. The hydrogen bond lengths were restrained using P-LINCS,^{19,31} allowing a time step of 2 fs.³² Translational movement of the solute was removed every 1000 steps to avoid any kinetic energy build-up.³³ After energy minimization of the solvent and equilibration of the solvated system for 10 ns using a Berendsen thermostat ($\tau_T = 1$ ps) and Berendsen pressure coupling ($\tau_P = 1$ ps),³² simulations were carried out in an NTP ensemble at a temperature of 300 K and a pressure of 1 bar using a Bussi velocity-rescaling thermostat³⁴ ($\tau_P = 1$ ps) and a Parrinello-Rahman barostat ($\tau_P = 1$ ps).³⁵ During minimization and heating, all the heavy atoms of the solute were kept fixed using positional restraints. The restraints on the RNA and the protein backbone were relaxed slowly during the equilibration from 1000 kJmol⁻¹ nm² to 10 kJmol⁻¹ nm².

Additional MD simulations, constructed using the AmberTools20 building package, were performed with the GPU version of Amber18 using the ff14SB force field to represent the protein,³⁶ the ff99OL3 force field for the RNA,^{37,38} ATP parameters from Meagher et al³⁹ and parameters for Mg²⁺ are taken from Ref. 28. The tetrahedral co-ordination state of the zinc was maintained using the ZAFF bonded force field.²⁸ Note that additional parameters were required for the HIS-33 that interacted with the zinc via its epsilon nitrogen by reference to comparable parameters in the ZAFF using a hybrid of the center ID 4 and 6 models.⁴⁰ For structures where ATP is bound, the octahedral co-ordination of the Mg²⁺ (which involves bonds to the ATP β and γ phosphate oxygen atoms, one with oxygen of the Ser289 hydroxyl group and three structural water molecules) was constructed using the Chimera metal center builder.⁴¹ The solute was neutralized with potassium cations, then the protein was immersed in a box of TIP3P water molecules extending a minimum of 10 Å from the protein surface, and K⁺Cl⁻ ion pairs were added to achieve a salt concentration of 0.14 M. MD simulations were performed in the NTP ensemble, with Berendsen temperature and pressure coupling. SHAKE was applied to all bonds involving hydrogen, allowing an MD integration timestep of 2 fs.

Long-range electrostatic interactions were treated using the particle mesh Ewald method^{29,30} with a real-space cut-off of 12 Å. To equilibrate the protein and nucleo-protein complexes, the systems was initially energy minimized with positional restraints placed upon the solute, followed by minimization of both solvent and solute. The system was then heated to 300 K in the presence of positional restraints upon the solute, which were gradually reduced from 50 kcal/mol Å² to 1.0 kcal/mol Å² over a timescale of 100 ps. For the apo-helicase structure, all restraints were then removed. For the ATP-RNA helicase complex which included the coordinated Mg²⁺ ion, an additional 50 ns of equilibration was performed with harmonic distance restraints (set at 2.1 Å with a spring constant of 20 kcal/mol Å²) to maintain the positions of coordinated atoms, and angle restraints imposing the octahedral geometry around the Mg²⁺ ion. An additional restraint was imposed to maintain the orientation of Asp374 and Glu375 to the adjacent coordinated water molecule, as observed in the MutS-ATP complex (PDB ID 1w7a,⁴²). Three 1 μs simulations of the apo-structure at a salt concentration of 140 mM, and one 1.5 μs simulation in neutralizing salt were performed. We have also obtained 1 μs simulations of the ATP-helicase (two replicas), the RNA-helicase and the ATP RNA-helicase complex. For all coordinated ATP Mg²⁺ metal centers, these equilibration protocols provide stable octahedral geometries, including the complexed water molecules, during unrestrained MD over 1 μs timescales.

Pocket analysis

For the analysis of the ATP pocket size, we used the open-source cavity detection software, Fpocket.⁴³ From our MD simulations of the apo and ATP-RNA helicase, we extracted 100 equally time-spaced protein structural snapshots from the last 200 ns simulation of both systems. We performed the pocket size analysis using several parameter sets using Fpocket and compared the results for our system (see SI). We then calculated the average volume of the ATP pocket for the most optimal parameter sets and compared these between the apo and the ATP-RNA complex simulations. The parameters we modified were: -m: minimum α-sphere size; -M: maximum α-sphere size -D: first clustering α-sphere size (see SI).

RESULTS

Helicase domains and their sequence homology

Following previous studies, the single chain SARS-CoV-2 helicase can be divided into three domains as depicted in Fig. 1. The sequence starts with Domain 1 (residues 1-260), which features: a Zinc-binding domain (ZBD, residues 1-100), known to facilitate nucleic acid recognition; a Stalk region shaped by 2 contiguous alpha helices (residues 100-150) which functions as an interface connecting the ZBD with the rest of the Domain 1 (residues 150-260, also known as Domain 1B) that interacts with the RNA. The rest of the chain splits into RecA1 and RecA2 domains, which are well characterized in the superfamily 1B type helicases and bind ATP at their interface.⁴⁴

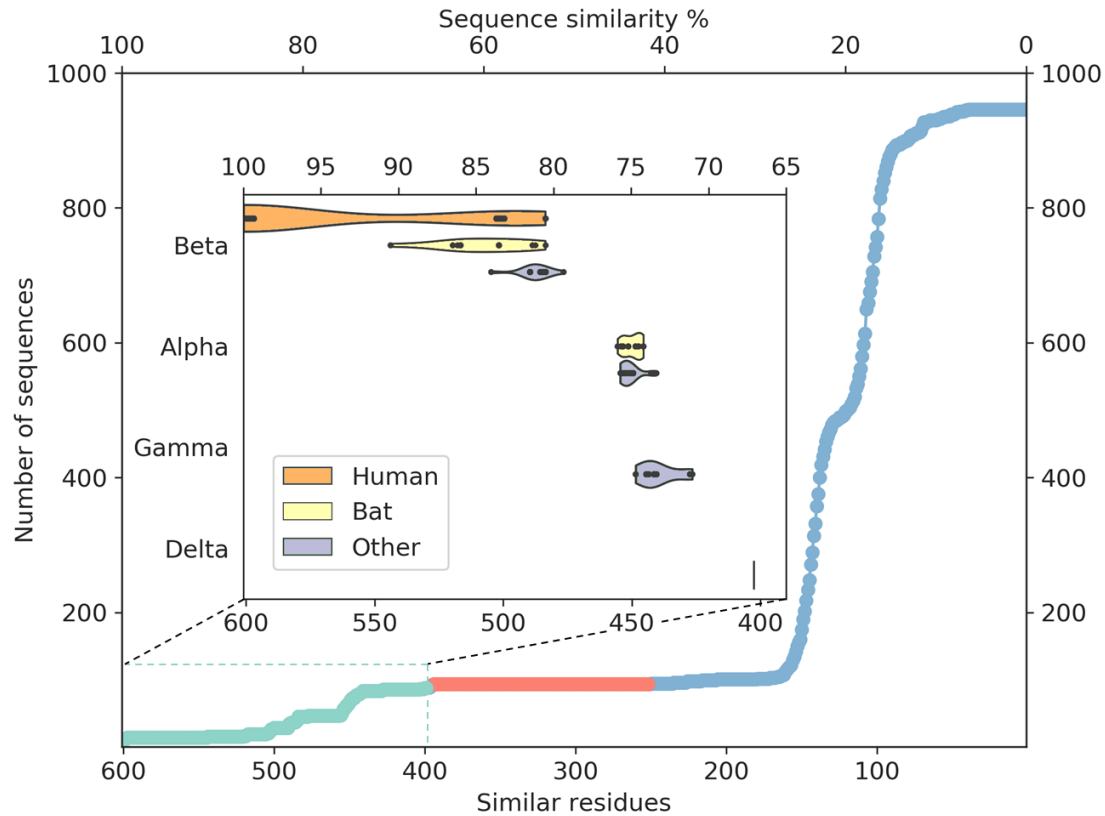


Figure 3. Distribution of the pairwise sequence alignments to the SARS-CoV-2 helicase. There are only members of *coronaviridae* above 398 matching residues (66%, lime circles, 95 entries). There are no sequences with medium similarity (235-394 similar residues, red circles). The closest relatives (95 sequences highlighted in lime dashed frame) are grouped in coronavirus subfamilies with principal hosts highlighted in the inset.

We have obtained the most homologous 1000 non-redundant sequences and their alignments from the UniProtKB library. About 10% of these sequences show similarities across the whole helicase sequence and the best 95 has 400 or more positives or similarities in the sequence alignment (Fig. 3). These are all coronaviruses mainly derived from bat virome (beta and alphacoronaviruses), and affecting various hosts in the animal kingdom, including humans. Intriguingly, the next best sequence alignment only covers 235 amino acids as similar residues; all of these and subsequent aligned regions are specific to the RecA domains and range through all types of organisms.

To evaluate any similarities to Domain 1 only, we also performed a search using only the first 230 residues. This search for sequences that match at least 70 residues resulted in the exact same 95 sequences as before, exclusively belonging to *coronaviridae*. An additional only 21 sequences match shorter segments of this domain between 1-230 residues, corresponding to a 22% sequence identity or below.

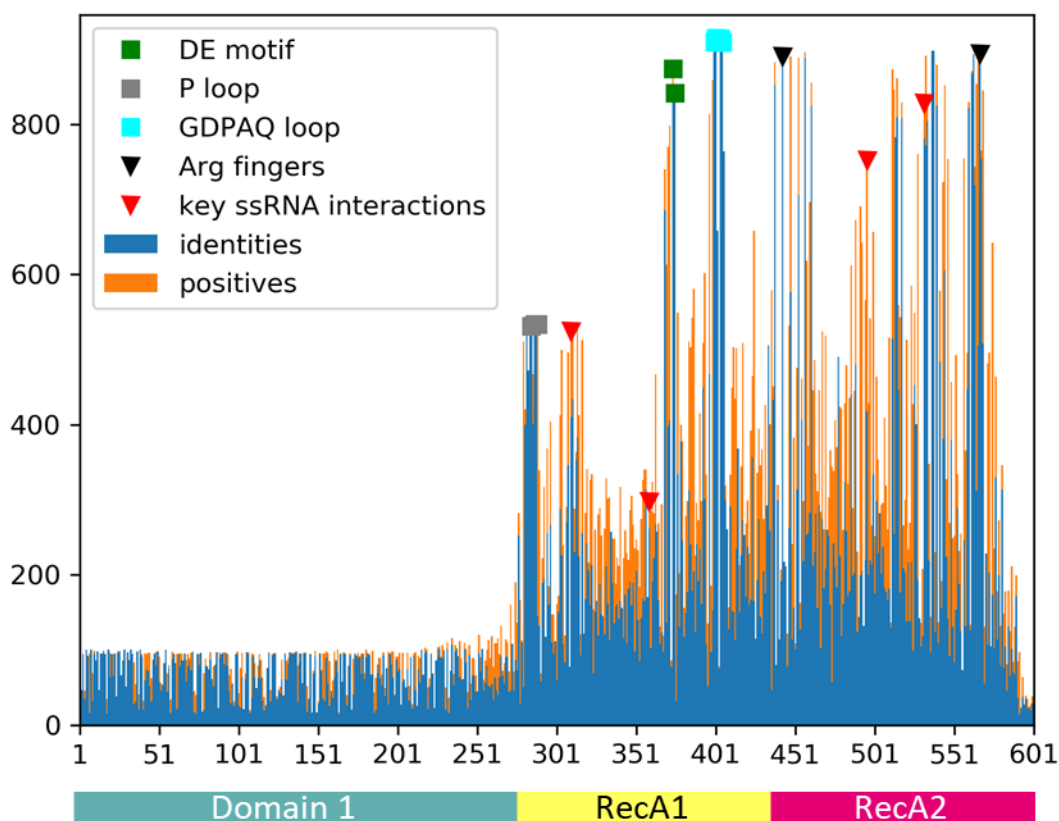


Figure 4. Sequence similarity (orange) and identity (blue) of the closest 946 sequences from UniProtKB using BLAST pairwise alignments to the 601-residue long SARS-CoV-2 RNA helicase. Domain 1 shows similarity only to the close relatives (95 sequences), while the RecA1 and RecA2 domains are more common across ATPase sequences. Key structural motifs are highlighted using symbols (P-loop: grey square, DE motif: green square, arginine fingers: black triangle, ssRNA interactions: red triangles).

Amongst crystal structures containing ATP analogs, most helicases have a very low sequence similarity to NSP13. The closest homologues are 2xzo, 5mzn, and 6jim with 11.00%, 10.20%, and 8.40% sequence identity, respectively. Despite the low sequence identity, most residues in the ATP binding pocket are conserved. At the same time, the closest human sequence homologue based on our homology search, ZGRF1, a putative RNA helicase, shares only 22% sequence similarity, restricted to the RecA1 and RecA2 domains. This relatively narrow bandwidth of sequence similarity can be harnessed to design specific inhibitors against the coronavirus RNA helicases that do not inhibit human proteins.

Structural model of the ATP binding site

We modelled the ATP-bound active site using the 2xzo structure as a template (Figure 7). The essential Mg^{2+} ion cofactor coordinates both the β and γ -phosphates and a conserved Ser288. The active site contains a DE of the DEAD-motif of RNA helicases. The conserved Asp374 H-bonds with the Ser288 and one of the Mg-coordinating water molecules, whereas the Glu375 is positioned as the proton acceptor.^{10,45,46} The γ P is stabilized via H-bonds with Arg567 Lys289 and Gln404 through a water molecule, that are found in respectively 95, 56 and 57%

of the homologous sequences analyzed, while β P forms a H-Bond with Arg443. The sugar region of the ATP likely interacts also with Glu540 and Lys320 as seen in ten and four pdb structures, respectively. Unlike the highly conserved residues recognizing the triphosphate pocket, the environment of the sugar and purine moieties (Figure 7, right) shows a greater diversity. The purine ring is stabilized through multiple π stack interactions, from one side with Arg442, while in some helicases with a tyrosine, and from the other side with His290 and Phe261. Additionally, there is a H-Bond between the amino group of the purine ring with Asn265, a residue which is more typically served by a glutamine in similar sequences.

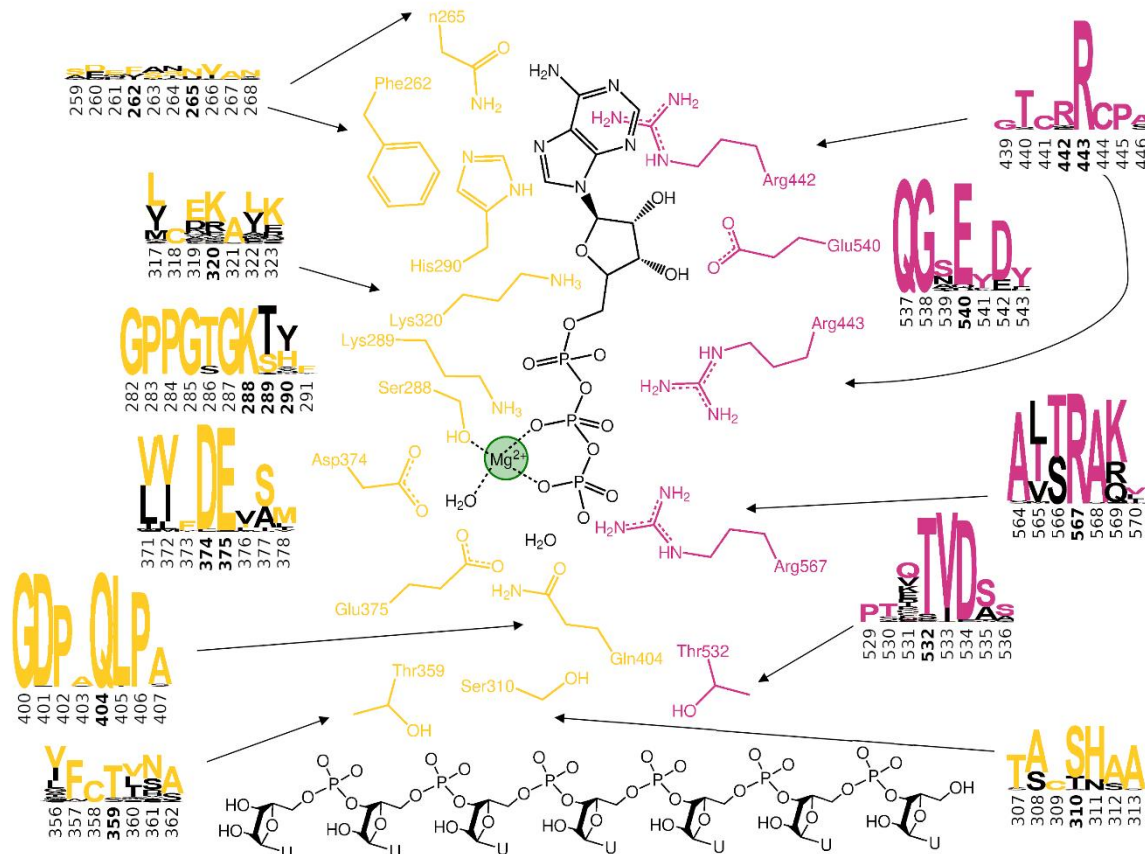


Figure 5. Conserved residues coordinating the ATP and RNA substrates of the SARS-CoV-2 helicase. Sequence conservation for RecA1 (orange) and RecA2 (magenta) domains are depicted in logos for each residue and its neighbors (data from Figure 4). Colored letter represent the residues in the SARS-CoV-2 helicase sequence, depicted residue indices are bold in the logos.

A lack of specificity towards the purine group is likely due to the dual function of the SARS-CoV-2 helicase to aid the 5'-capping of the RNA by the triphosphate hydrolysis of most NTP substrates.³ Due to these major differences, this area of the nucleotide binding pocket can serve in the design of SARS-CoV-2-specific antiviral drugs.

Structural model of the RNA binding site

To identify the main contact points with the ssRNA, it is first important to understand the unwinding function of the helicase. Filtering the related crystal structures to the ones containing RNA, we noticed that the RNA directionality relative to the ATP pocket is well defined (Fig. 1). The unwinding is driven by a domain-wise translocation process, which

moves the RNA one base in two steps.^{13,14,47} Upon ATP hydrolysis, domain RecA2 translocate 1 base towards the 5' end of the RNA which is then followed by the RecA1 when a new molecule of ATP binds to the enzyme.

Domain 1, being in contact with the sidechain of the RNA, does not feature specific motifs, thus allowing different RNA bases to translocate. A long loop transitions into the RecA1 and 2 domains sandwiching the ATP pocket on the side of RNA backbone. This region, equipped with the necessary functionalities to perform the ATP hydrolysis, has a higher degree of conservation along the helicases. Both RecA domains have specific residues reliable for the contact with the RNA phosphates, depicted in Fig. 8. Thr359 in RecA1 and Thr532 are identified as the main anchoring points of the two domains. The base between these two threonine residues is coordinated by a backbone NH of His311, an interaction which is kept in RNA containing crystal structures. Ser310 is also reasonably conserved, although not directly featured in RNA coordination in this state of the enzyme.

Interestingly, the most conserved motif across the sequences is a GDP(A)Q loop interfacing between the RecA1 and RecA2 domains. This motif features Gln404, a residue which we consider to be important in the coordination of the nucleophilic water; moreover, it bridges the γ P and the SH motif discussed earlier. We speculate these moieties play a role in the translocation of RecA2 upon ATP hydrolysis.

MD simulations

Apo structures

All replicas of the apo structure show low flexibility and no major changes of the backbone structure of the dimer. We analyze the overall flexibility of the dimers and compare our results with the experimental b-factor obtained in 6jyt and 6zsl (Figure 6). Our model, in common with the two crystal structures, shows higher flexibility on the external shell of RecA2 domain, while the ATP and the RNA pockets appear to be more conserved. The ZBD shows low flexibility, in agreement with the b-value of 6jyt, but not with 6zsl (especially chain A), in which the temperature factor is higher, due to the different dimerization of the crystal structures.

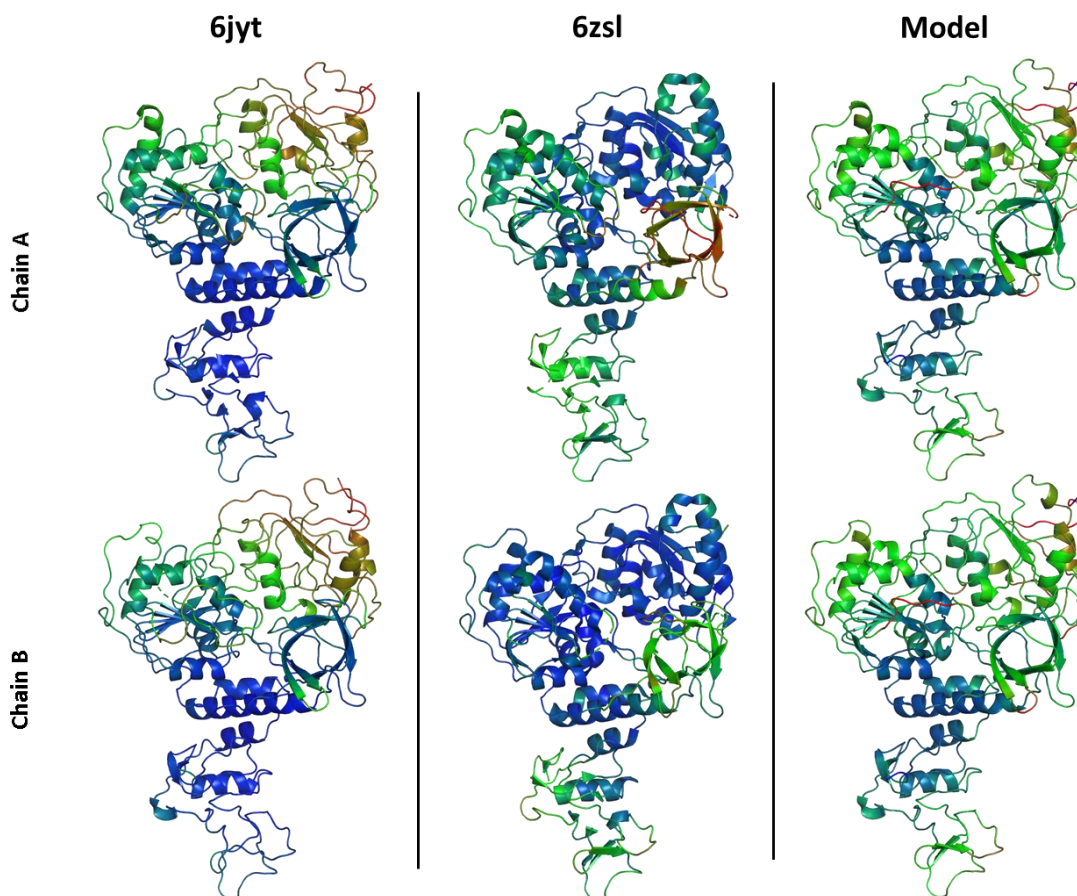


Figure 6. Conformational flexibility of the APO helicase protomers from 6jyt, 6zsl and our model if the apo dimer from the MD simulations. The residues are colored according to the deposited PDB B-factors (6jyt and 6zsl; from blue: low B-factor to red: high B-factor), and by the residue RMSD from the MD trajectory.

ATP-bound pocket

From our unbiased trajectory of the ATP-RNA-helicase complex we extract key distances along the simulations (Figure 7). The hexacoordinate Mg^{2+} shows stable bond with 3 water molecules, the OH of Ser288 and two oxygens of the β and γ phosphate of the ATP, which is essential for the preorganization of the ATP hydrolysis. Further conserved contacts in the pocket are also kept stable during the simulation: including the arginine fingers, Lys288 and the DE motif which both takes part in coordinating the Mg^{2+} and the nucleophilic water. Gln404 shows a greater flexibility deep in the ATP pocket, which might suggest its role in changing the conformation during translocation. Residues participating in the adenosine coordination are less conserved and also exhibit fewer stable contacts.

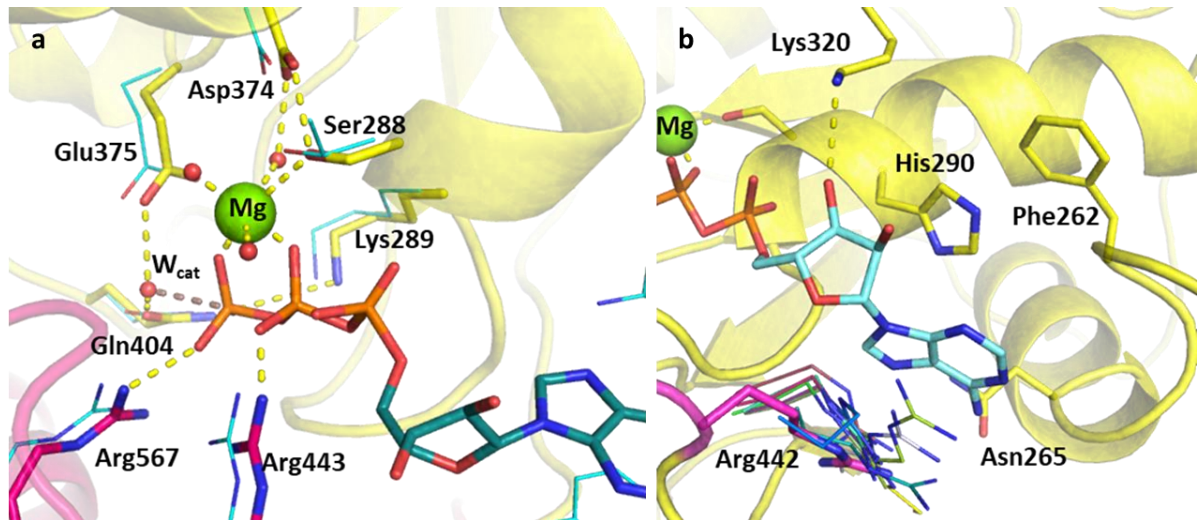


Figure 7. Structure of the ATP pocket aligned with homologous ATP-helicase complexes. RecA1 and RecA2 are shown in yellow and magenta, respectively. **a**) Main protein-substrate interactions of the triphosphate and magnesium ions are compared with alignment for PDB template 2xzo (cyan lines). **b**) Nucleotide binding region focusing on Arg442 (magenta sticks) is aligned with homologous arginine residues (lines, PDB structures 5k8u, 5vhc, 5xdr, 5y4z, 5y6m, 5y6n, 6adx, 6ady, 6c90 and 6jim).

Pocket analysis

We identify the cavity size of the ATP pocket using Fpocket. We tested different combinations of parameters to find a combination that yields more robust results (see SI for details). A comparison of the average pocket volume from the apo trajectory ($394.6 \pm 157.5 \text{ \AA}^3$) with the average pocket volume from the ATP bound trajectory ($367 \pm 146 \text{ \AA}^3$) shows no substantial differences within the error of the analysis. This finding suggests that the ATP cavity structure is not altered significantly by the presence of ATP.

RNA binding site

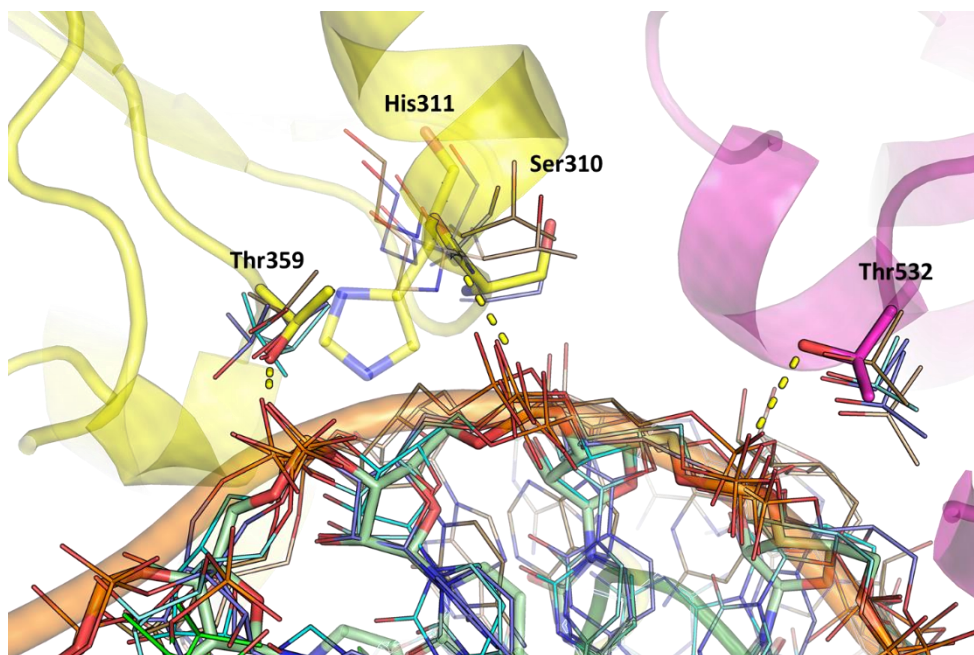


Figure 8. Structures of the RNA binding region aligned with existing RNA-helicase crystal structures complexed with RNA (depicted in lines). RecA1 and RecA2 domains are shown in yellow and magenta, respectively. Key residues (sticks) are labelled and H-bonds are depicted in yellow dashes.

From the structural analysis and the homology modelling, we denote two important and well-conserved interactions between the RNA and the helicase. Both interactions involve a H-bond between a threonine and the O of the phosphate residue on the backbone of the RNA. Specifically, from Thr359 from RecA1 and Thr532 from RecA2. Additionally, another H-bond is made between the central RNA and the N of residue 311, this residue is not highly conserved but the interaction between the backbone of this residue and the OP of the RNA is present in several PDB structures. A key residue close to the RNA pocket is Ser310; this residue is conserved (often present as a threonine) and appears to be important for the communication between the ATP pocket and the RNA pocket.

SUMMARY

We present the catalytically competent computational model of the SARS-CoV-2 NSP13 ATP dependent RNA helicase. Our structure is the first to host ATP and a single strand RNA with detail to the binding modes of the substrates.

The analysis of homologous sequences sheds light upon the specificity of the domain structure of the viral helicase yielding no match over 40% except close relatives from the *coronaviridae* family. Our model features two major improvements compared to the apo protein. The ATP pocket was reconstructed based on the structural conservation of the most similar crystal structures including signature motifs from phosphate binding proteins such as the DE(AD) of helicases, the P loop or the arginine fingers. Furthermore, we identified the main anchoring points of the ssRNA through the helicase, which are essential to understand the translocation driving the unwinding activity of NSP13.

With molecular dynamics, we have verified the stability of conserved interactions in our model as well as improved our initial model to host the nucleic acid. We assess the flexibility of the ATP pocket with and without the nucleotide bound detecting a well-maintained cavity.

Our work provides insight into one of the elements of the viral replication machinery, ideal for targeting by drug developers. Our structure can be the basis of structure-based compound design and screening. Moreover, elaborating the RNA translocations driven by the identified interactions can reveal other targetable states of the helicase.

ACKNOWLEDGEMENT

This project made use of time on ARCHER granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (<http://hecbiosim.ac.uk>), supported by EPSRC (grant no. EP/R013012/1). We would like to thank GENCI for a generous allocation of computing time at CINES within the action "Project contre le COVID 19". We are also grateful for resourced provided by the Leeds ARC HPC service.

REFERENCES

- (1) FDA. Coronavirus (COVID-19) Update: FDA Issues Emergency Use Authorization for Potential COVID-19 Treatment <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-issues-emergency-use-authorization-potential-covid-19-treatment>.
- (2) EMA. First COVID-19 treatment recommended for EU authorisation <https://www.ema.europa.eu/en/news/first-covid-19-treatment-recommended-eu-authorisation>.
- (3) Ivanov, K. A.; Thiel, V.; Dobbe, J. C.; van der Meer, Y.; Snijder, E. J.; Ziebuhr, J. Multiple Enzymatic Activities Associated with Severe Acute Respiratory Syndrome Coronavirus Helicase. *J. Virol.* **2004**, *78* (11), 5619 LP – 5632. <https://doi.org/10.1128/JVI.78.11.5619-5632.2004>.
- (4) Tanner, J. A.; Watt, R. M.; Chai, Y. B.; Lu, L. Y.; Lin, M. C.; Peiris, J. S. M.; Poon, L. L. M.; Kung, H. F.; Huang, J. D. The Severe Acute Respiratory Syndrome (SARS) Coronavirus NTPase/helicase Belongs to a Distinct Class of 5' to 3' Viral Helicases. *J. Biol. Chem.* **2003**, *278* (41), 39578–39582. <https://doi.org/10.1074/jbc.C300328200>.
- (5) Kwong, A. D.; Rao, B. G.; Jeang, K. T. Viral and Cellular RNA Helicases as Antiviral Targets. *Nature Reviews Drug Discovery*. Nature Publishing Group October 23, 2005, pp 845–853. <https://doi.org/10.1038/nrd1853>.
- (6) Chen, J.; Malone, B.; Llewellyn, E.; Grasso, M.; Shelton, P. M. M. M.; Olinares, P. D. B.; Maruthi, K.; Eng, E. T.; Vatandaslar, H.; Chait, B. T.; Kapoor, T.; Darst, S. A.; Campbell, E. A.; Dominic, P.; Olinares, B.; Maruthi, K.; Eng, E. T.; Vatandaslar, H.; Chait, B. T.; Kapoor, T.; Darst, S. A.; Campbell, E. A. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **2020**, *182*, 1–14. <https://doi.org/10.1016/j.cell.2020.07.033>.
- (7) Peng, Q.; Peng, R.; Yuan, B.; Zhao, J.; Wang, M.; Wang, X.; Wang, Q.; Sun, Y.; Fan, Z.; Qi, J.; Gao, G. F.; Shi, Y. Structural and Biochemical Characterization of the Nsp12-Nsp7-Nsp8 Core Polymerase Complex from SARS-CoV-2. *Cell Rep.* **2020**, *31* (11). <https://doi.org/10.1016/j.celrep.2020.107774>.
- (8) Yin, W.; Mao, C.; Luan, X.; Shen, D.-D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; Chang, S.; Xie, Y.-C.; Tian, G.; Jiang, H.-W.; Tao, S.-C.; Shen, J.; Jiang, Y.; Jiang, H.; Xu, Y.; Zhang, S.; Zhang, Y.; Xu, H. E. Structural Basis for Inhibition of the RNA-Dependent RNA Polymerase from SARS-CoV-2 by Remdesivir. *Science* (80-.). **2020**, *368* (6498), 1499–1504. <https://doi.org/10.1126/science.abc1560>.
- (9) Jia, Z.; Yan, L.; Ren, Z.; Wu, L.; Wang, J.; Guo, J.; Zheng, L.; Ming, Z.; Zhang, L.; Lou, Z.; Rao, Z. Delicate Structural Coordination of the Severe Acute Respiratory Syndrome Coronavirus Nsp13 upon ATP Hydrolysis. *Nucleic Acids Res.* **2019**, *47* (12), 6538–6550. <https://doi.org/10.1093/nar/gkz409>.
- (10) Hoffmann, M.; Eitner, K.; von Grotthuss, M.; Rychlewski, L.; Banachowicz, E.; Grabarkiewicz, T.; Szkoda, T.; Kolinski, A. Three Dimensional Model of Severe Acute Respiratory Syndrome Coronavirus Helicase ATPase Catalytic Domain and

- Molecular Design of Severe Acute Respiratory Syndrome Coronavirus Helicase Inhibitors. *J. Comput. Aided. Mol. Des.* **2006**, *20* (5), 305–319. <https://doi.org/10.1007/s10822-006-9057-z>.
- (11) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- (12) Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M. *MUSTANG: A Multiple Structural Alignment Algorithm*.
- (13) Chakrabarti, S.; Jayachandran, U.; Bonneau, F.; Fiorini, F.; Basquin, C.; Domcke, S.; Le Hir, H.; Conti, E. Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. *Mol. Cell* **2011**, *41* (6), 693–703. <https://doi.org/10.1016/j.molcel.2011.02.010>.
- (14) Law, Y. S.; Utt, A.; Tan, Y. B.; Zheng, J.; Wang, S.; Chen, M. W.; Griffin, P. R.; Merits, A.; Luo, D. Structural Insights into RNA Recognition by the Chikungunya Virus NsP2 Helicase. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (19), 9558–9567. <https://doi.org/10.1073/pnas.1900656116>.
- (15) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Héning, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153* (4), 044130. <https://doi.org/10.1063/5.0014475>.
- (16) D. MacKerell, A.; Bashford, D.; Bellott, M.; L. Dunbrack, R.; D. Evanseck, J.; J. Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; T. K. Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; T. Nguyen, D.; Prodhom, B.; E. Reiher, W.; Roux, B.; Schlenkrich, M.; C. Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616. <https://doi.org/10.1021/jp973084f>.
- (17) 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.* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (18) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.
- (19) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (20) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>.

- (21) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**.
<https://doi.org/10.1016/j.softx.2015.06.001>.
- (22) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**.
<https://doi.org/10.1529/biophysj.106.097782>.
- (23) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; E. Cheatham, T.; 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 (9), 2886–2902. <https://doi.org/10.1021/ct200162x>.
- (24) 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 (8), 3696–3713.
<https://doi.org/10.1021/acs.jctc.5b00255>.
- (25) B. Peters, M.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; N. Weaver, M.; M. Merz, K. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, 6 (9), 2935–2947.
<https://doi.org/10.1021/ct1002626>.
- (26) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, 112 (30), 9020–9041. <https://doi.org/10.1021/jp8001614>.
- (27) Bešševová, I.; Otyepka, M.; Réblová, K.; Šponer, J. Dependence of A-RNA Simulations on the Choice of the Force Field and Salt Strength. *Phys. Chem. Chem. Phys.* **2009**, 11 (45), 10701. <https://doi.org/10.1039/b911169g>.
- (28) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion--Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, 8 (4), 1493–1502.
- (29) 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 (12), 10089–10092.
- (30) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, 103 (19), 8577–8593.
- (31) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, 18 (12), 1463–1472.
- (32) Berendsen, H. J. C.; Postma, J. P. M. van; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, 81 (8), 3684–3690.
- (33) Harvey, S. C.; Tan, R. K.-Z.-. Z.; Cheatham, T. E. The Flying Ice Cube: Velocity Rescaling in Molecular Dynamics Leads to Violation of Energy Equipartition. *J.*

- Comput. Chem.* **1998**, 19 (7), 726–740.
- (34) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, 126 (1), 14101.
- (35) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52 (12), 7182–7190.
- (36) 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 (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (37) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Loughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, 92 (11), 3817–3829. <https://doi.org/10.1529/biophysj.106.097782>.
- (38) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; E. Cheatham, T.; 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 (9), 2886–2902. <https://doi.org/10.1021/ct200162x>.
- (39) Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of Polyphosphate Parameters for Use with the AMBER Force Field. *J. Comput. Chem.* **2003**, 24 (9), 1016–1025. <https://doi.org/10.1002/jcc.10262>.
- (40) Li, P.; Merz, K. M. ZAFF Modeling Tutorial <https://ambermd.org/tutorials/advanced/tutorial20/ZAFF.htm> (accessed Oct 29, 2020).
- (41) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, 25 (13), 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- (42) Lebbink, J. H. G.; Fish, A.; Reumer, A.; Natrajan, G.; Winterwerp, H. H. K.; Sixma, T. K. Magnesium Coordination Controls the Molecular Switch Function of DNA Mismatch Repair Protein MutS. *J. Biol. Chem.* **2010**, 285 (17), 13131–13141. <https://doi.org/10.1074/jbc.M109.066001>.
- (43) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinformatics* **2009**, 10 (1), 168. <https://doi.org/10.1186/1471-2105-10-168>.
- (44) Leonait, B.; Han, Z.; Basquin, J.; Bonneau, F.; Libri, D.; Porrúa, O.; Conti, E. Sen1 Has Unique Structural Features Grafted on the Architecture of the Upf1-like Helicase Family. *EMBO J.* **2017**, 36, 1590–1604. <https://doi.org/10.15252/embj.201696174>.
- (45) Briguglio, I.; Piras, S.; Corona, P.; Carta, A. Inhibition of RNA Helicases of SsRNA + Virus Belonging to Flaviviridae, Coronaviridae and Picornaviridae Families. *Int. J. Med. Chem.* **2011**, 2011, 22. <https://doi.org/10.1155/2011/213135>.
- (46) Yang, X.; Chen, C.; Tian, H.; Chi, H.; Mu, Z.; Zhang, T.; Yang, K.; Zhao, Q.; Liu, X.;

Wang, Z.; Ji, X.; Yang, H. Mechanism of ATP Hydrolysis by the Zika Virus Helicase. *FASEB J.* **2018**, *32* (10), 5250–5257. <https://doi.org/10.1096/fj.201701140R>.

- (47) Appleby, T. C.; Anderson, R.; Fedorova, O.; Pyle, A. M.; Wang, R.; Liu, X.; Brendza, K. M.; Somoza, J. R. Visualizing ATP-Dependent RNA Translocation by the NS3 Helicase from HCV. *J. Mol. Biol.* **2011**, *405* (5), 1139–1153. <https://doi.org/10.1016/j.jmb.2010.11.034>.