

Allosteric regulation of SARS-CoV-2 protease: towards informed structure-based drug discovery

Khaled Abdel-Maksoud,^{1,*†} Mohamed Ali al-Badri,^{2,*‡} Christian D. Lorenz,² and Jonathan W. Essex¹

¹School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

²Department of Physics, King's College London, Strand, London WC2R 2LS, UK

**Contributed equally to this work*

† k.abdel-maksoud@soton.ac.uk, ‡ mohamed.al-badri@kcl.ac.uk

The Coronavirus Disease of 2019 (COVID-19) is caused by a novel coronavirus known as the Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2). Despite extensive research since the outset of the pandemic, definitive therapeutic agents for the treatment of the disease are yet to be identified. The main protease (M^{Pro}) of SARS-CoV-2 is an enzyme essential for virus replication through viral proteolytic activity and subsequent generation of infectious virus particles. Current computational efforts towards SARS-CoV-2 M^{Pro} inhibitor design have generally neglected an allosteric mechanism linked to His41-Cys145 catalytic dyad disruption and thus fail to target the open conformational state. We identify the rare event associated with the allosteric regulation of M^{Pro} activity in the orientation of the His41 imidazole side chain away from Cys145. In this work, we show that molecular dynamics and metadynamics simulations are fundamental for performing computer-aided M^{Pro} inhibitor design where the sampling of this allosteric mechanism within a computationally feasible

timescale is essential. We calculate a 4.2 ± 1.9 kJ/mol free energy difference between the open and closed states of the SARS-CoV-2 M^{Pro} active site, indicating that favourable ligand interactions with His41 over the Cys145-His41 dyad interaction can stabilise the open state.

Coronaviruses have proven to be a challenge for drug discovery since Severe Acute Respiratory Syndrome Coronavirus 1 (SARS-CoV-1), better known as SARS (2002-2004) and the Middle East Respiratory Syndrome (MERS-CoV) (2012-2013). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the coronavirus disease 2019 (COVID-19). It is a positive-sense single-stranded RNA virus that is contagious in humans, with zoonotic origins and genetic similarities to bat coronaviruses.¹ The virus mainly enters human cells through the receptor angiotensin converting enzyme 2 (ACE2).²

To date, more than 21 million cases of COVID-19 have been confirmed, with more than 760,000 deaths worldwide.³ SARS-CoV-2 has a combination of high transmissibility, longer incubation period and a much shorter interval between symptom onset and maximum infectivity when compared with SARS and MERS-CoV.⁴ As a result, even with a relatively low mortality rate, COVID-19 is proving much harder to eradicate and will therefore remain an epidemiological problem until a therapeutic agent is developed.

Broad spectrum antiviral medication such as remdesivir have been proposed to decrease SARS-CoV-2 RNA production, however remdesivir was not associated with statistically significant clinical benefits.⁵ Recent clinical research has shown that administration of dexamethasone⁶ or systemic corticosteroids,⁷ were found to be associated with lower 28-day all-cause mortality when

compared with usual care or placebo.

The most promising tools for the cessation of the epidemic spread of COVID-19 are vaccines, with many in latter stages of clinical trials, that are expected to be available in late 2020 or early 2021.⁸ The majority of these vaccines are based on platforms such as inactivated viral vectors or RNA sequences encoding the spike glycoprotein of SARS-CoV-2 that trigger an immunogenic response. The SARS spike protein has been identified as the major target of selective pressure in the adaptive evolution of SARS coronaviruses.⁹ Recently, molecular dynamics (MD) simulations have been used to study the druggability of the SARS-CoV-2 spike.¹⁰ They have identified vulnerabilities in the spike glycan shield — utilised to frustrate an immune response — that can be harnessed for vaccine development.

Protease inhibitors, however, do not depend on an immunogenic response to elicit immunity. Unlike immunogenic approaches, any inhibitor identified for the SARS-CoV-2 M^{Pro} would very likely also serve as an inhibitor of further evolution of this virus as the sequence and structure of the M^{Pro} are closely related to those from other betacoronaviruses.¹¹ Elsewhere, protease inhibitors have been extensively used for the treatment of HIV-AIDS^{12,13} and hepatitis-C¹⁴. Molecular dynamics (MD) simulations were paramount in identifying the dynamic bound and free states of the HIV-1 protease flaps (two glycine-rich β -hairpins) that cover a large substrate-binding pocket used as a target for antiviral drugs.¹⁵ Also, MD simulations were integral in identifying a cryptic trench within the HIV integrase, which became the target for the first FDA approved HIV integrase inhibitor (raltegravir).^{12,16}

The SARS-CoV-2 M^{Pro} is largely responsible for the proteolytic processing of the polyproteins transcribed by the SARS-CoV-2 genome, which are responsible for viral transcription and replication. M^{Pro} cleaves the polyproteins at 11 conserved sites using the catalytic dyad.^{17,18} A catalytic dyad is a set of two coordinated amino acids, common to some enzyme active sites. Within a histidine-cysteine (His-Cys) dyad, the His amino acid will act as a base and activate the Cys mercaptan as a nucleophile for polypeptide cleavage.

Owing to the high sequence similarity, the conservation of primary structure about the catalytic sites and the high degree of tertiary structure similarity between both proteases, it has been surmised that the His41-Cys145 catalytic dyad plays the same role of regulating protease activity within both SARS-CoV M^{Pro} proteins.^{11,18} Disrupting this dyad, then, disables M^{Pro} activity and subsequent virus replication. In order to be able to cleave the SARS-CoV-2 polyprotein, hydrolysis must be facilitated by priming the M^{Pro} Cys145 mercaptan group for nucleophilic attack via deprotonation by the His41 imidazole group. The proteolytic mechanism in the SARS-CoV M^{Pro} active site regulates protease activity through activation of the Cys145 mercaptan. The rotation of the His41 imidazole towards Cys145 serves as an allosteric trigger to inducing proteolytic activity in SARS-CoV M^{Pro}. One way to disrupt the proteolytic mechanism is to induce a conformational change in the more flexible His41 imidazole side chain of the dyad and prevent it from abstracting a proton from the Cys145 mercaptan group. An inhibitor would serve to stabilise this disrupted conformation and therefore inhibit M^{Pro} activity.

Recently, MD simulations have been used to study the druggability of the SARS-CoV-2

M^{Pro} .^{19,20} However, conventional MD fails to disrupt the M^{Pro} catalytic dyad and thus deactivate M^{Pro} activity within computationally achievable timescales. As such, screening potential ligands from vast drug libraries to inhibit M^{Pro} is inaccurate and could fail to identify potent inhibitors. Instead, metadynamics (MetaD) enhances the sampling of rare events to reconstruct the free energy landscape by discouraging revisiting of sampled states. It is a useful tool for studying mechanisms of drugs binding to flexible targets where conventional MD may otherwise fail to ergodically sample the free energy landscape. Defining the free energy landscape of a complex simulation is non-trivial and depends on a choice of a few collective variables (CV).²¹ Other enhanced sampling techniques (Gaussian accelerated MD²²) have been applied in elucidating cryptic pockets not detectable from the M^{Pro} crystal structure, identifying additional pockets for studying M^{Pro} inhibition beside the active site.²³

In this work, we apply MetaD to identify the allosteric mechanism of SARS-CoV-2 M^{Pro} using the inhibition of SARS-CoV-1 M^{Pro} with a previously identified potent inhibitor (D3F)²⁴ (Fig. 1A). In order to comprehensively study whether a contender ligand successfully inhibits the SARS-CoV-2 M^{Pro} , the M^{Pro} allosteric mechanism must be sampled within feasible timescales. This sampling is difficult to achieve using conventional computational methods as there is a high free energy cost to dyad disruption that these approaches cannot account for. We show that the allosteric mechanism can be sampled using MetaD simulations through biasing rotations about a single dihedral within the dyad, and illustrate this by incorporating a contender ligand (LIG) (Fig. 1B) to SARS-CoV-2 M^{Pro} .

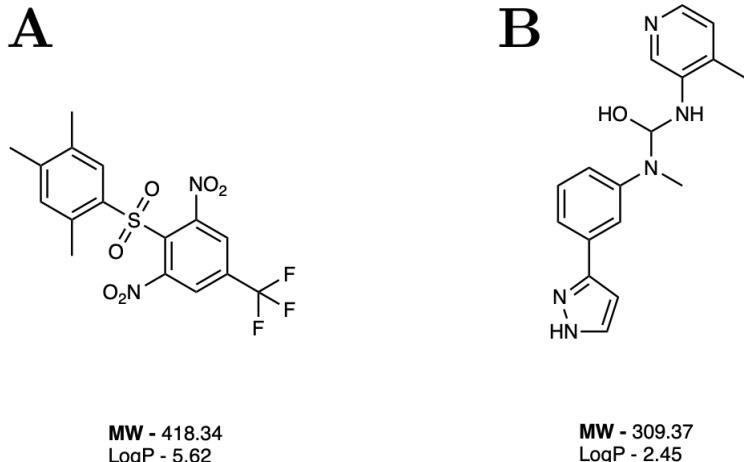


Figure 1: Inhibitor candidate ligand structures with molecular weight (MW) and calculated lipophilicity (LogP). (A) D3F, a strong binder and inhibitor of SARS-CoV-1 M^{Pro} . (B) The drug candidate (code named LIG herein) considered for binding and inhibition of SARS-CoV-2. LogP values were calculated using the ChemDraw LogP estimation tool.

Results

Allosteric regulation of SARS-CoV-1 and SARS-CoV-2 M^{Pro} activity is linked to the His41-Cys145 interaction

Co-solvent MD simulations²⁵ are a useful tool for identifying binding hotspots on protein surfaces by simulating proteins in a solution of water and co-solvent molecules. This dynamic approach to identifying binding sites incorporates proteins' inherent flexibility, allowing the cosolvent molecules to compete with water to bind to the protein surface.²⁶ Our co-solvent MD simulations were initiated by placing five LIG molecules randomly surrounding SARS-CoV-2 M^{Pro} . Of the five LIG molecules, only one entered the binding site of SARS-CoV-2 in order to stabilise a

binding mode (Fig. 2).

A short simulation of SARS-CoV-1 M^{Pro} with the inhibitor D3F bound in the active site was also performed and was used to evaluate the success of the observed LIG binding in inhibiting SARS-CoV-2 M^{Pro}, as the initial D3F binding mode has been proven to successfully inhibit SARS-CoV-1 M^{Pro} in previous studies.²⁴ When observing the bound state of D3F to SARS-CoV-1 M^{Pro}, the nitrate group most proximal to His41 forms a strong electrostatic interaction between a D3F nitro O atom and the N1-H of the His41 imidazole (D3F N-O · · · H-N1 His41) (Fig. 2A). This interaction stabilises the orientation of His41 away from Cys145 which comprises the open (“holo”) conformation of M^{Pro}. Despite the ligand appearing as bound within the SARS-CoV-2 M^{Pro} active site, His41 is not seen to interact with the bound ligand at all (Fig. 2B). The primary difference between the binding modes of D3F to SARS-CoV-1 M^{Pro} and LIG to SARS-CoV-2 M^{Pro} is the orientation of the His41 imidazole with respect to Cys145. In the latter case, the His41-Cys145 catalytic dyad is maintained and thus the M^{Pro} is within a closed (“apo”) conformation. The disruption of the catalytic dyad via reorientation of the His41 imidazole or interaction with the Cys145 mercaptan side chain¹⁸ therefore serves as a prospective diagnostic tool for successful inhibition.

Generating the SARS-CoV-2 M^{Pro} active state via His41 side chain reorientation

The disruption of the catalytic dyad to promote ligand binding was further explored through the use of enhanced sampling simulations. The binding mode of D3F to SARS-CoV-1 M^{Pro} indicates an ability to achieve inhibition of SARS-CoV-2 M^{Pro} through promoting interactions with the His41

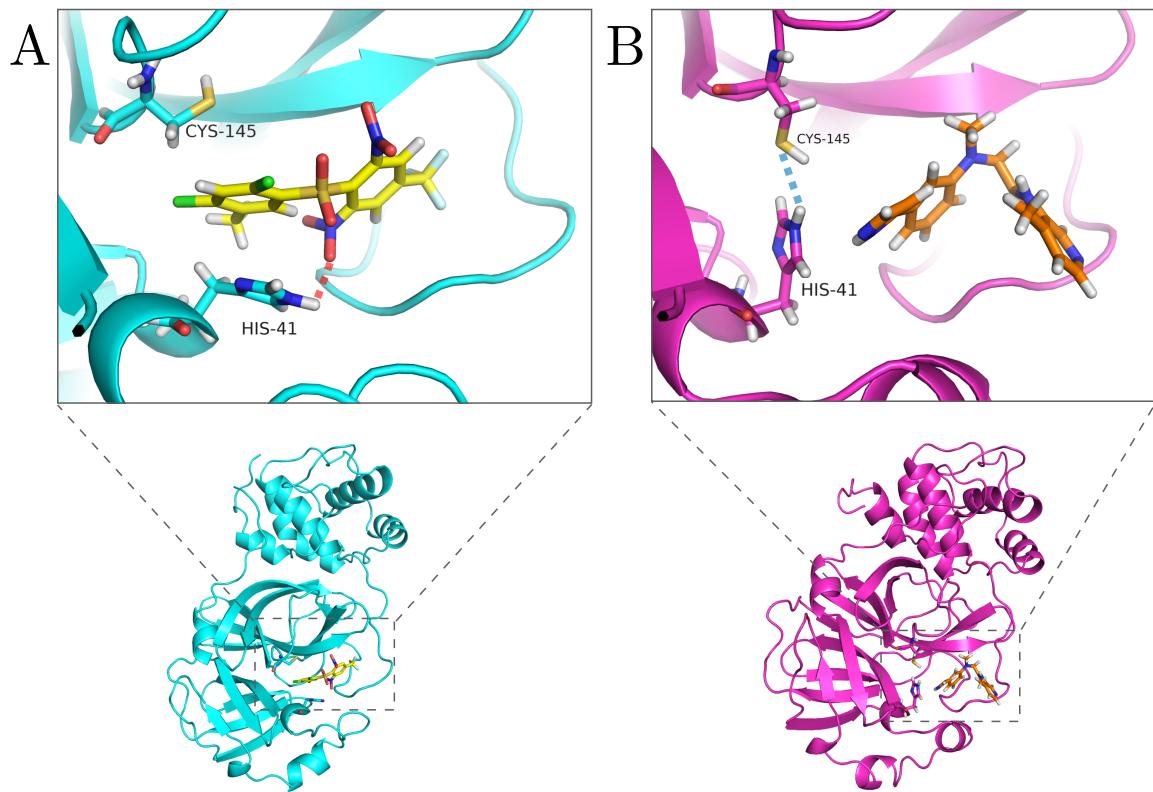


Figure 2: The observed binding modes of (A) D3F (yellow) within the SARS-CoV-1 M^{Pro} catalytic binding site (cyan), showing the disrupted catalytic dyad (“holo”) and the strong D3F interaction with His41 (dashed red line) and (B) LIG (orange) within the SARS-CoV-2 M^{Pro} catalytic binding site (pink) with a maintained catalytic dyad (“apo”). The dyad residues and their interactions (dashed blue line) are labeled His41 and Cys145.

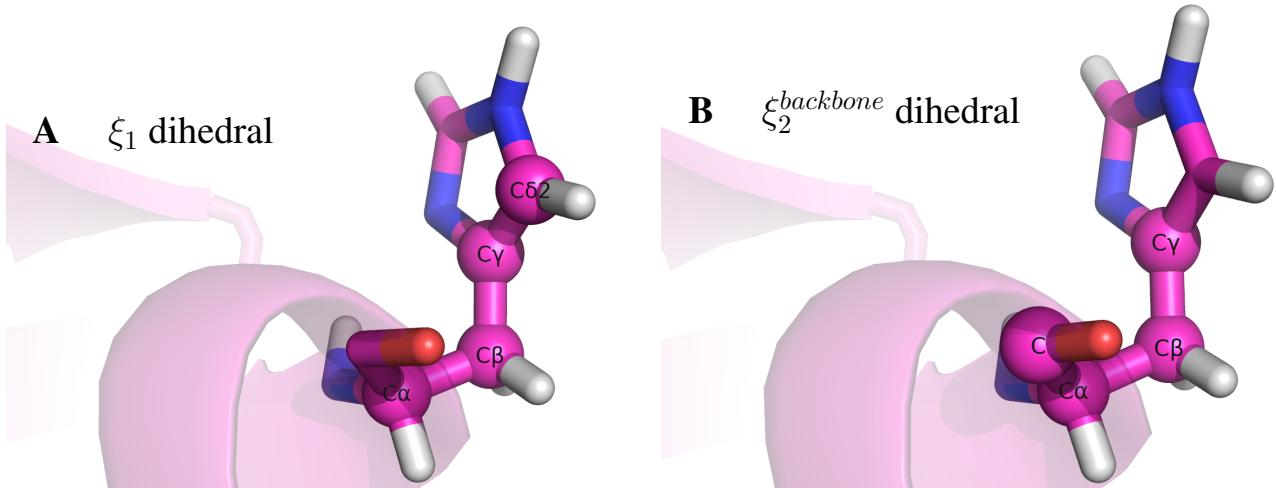


Figure 3: His41 imidazole side chain dihedrals. (A) ξ_1 dihedral used for construction of the MetaD bias. (B) ξ_2^{backbone} dihedral considered within analysis of the performance of the bias and characterising the disassociation of His41 from Cys145.

imidazole in the holo form. Metadynamics (MetaD) is herein employed with a bias potential over the ξ_1 dihedral of His41 in order to induce rotation of the His41 side chain imidazole group about the ξ_2^{backbone} dihedral. The ξ_2^{backbone} dihedral cannot be used as a CV due to an intrinsic conformational restriction, since the dihedral is part of the protein backbone. Instead, we have applied a bias to the ξ_1 dihedral torsion, which allows us to investigate the free rotational orientation of the His41 side chain imidazole (Fig. 3).

Ergodic sampling along the ξ_1 torsional CV space was achieved using MetaD simulations. The free energy surface with respect to the ξ_1 dihedral was calculated and is shown in (Supplementary Fig. 1, top). This surface was used to verify metadynamics convergence by evaluating the relative free energy differences between pairs of free energy minima, which stabilised after 70 ns

(Supplementary Fig. 1, bottom).

M^{Pro} catalytic dyad conformational changes

In order to verify that our choice of CV of ξ_1 torsion couples to and changes the degree of ξ_2^{backbone} torsion, we compare the probability densities found from our MetaD simulation to the probability distributions observed in our co-solvent MD simulations of LIG-SARS-CoV-2 and the D3F-SARS-CoV-1, which we use as the “apo” (closed conformation) and “holo” (open conformation) states, respectively. In (Fig. 4A), the MetaD probability distribution clearly samples both “apo” and “holo” states. In (Fig. 4B), this choice of ξ_1 torsion CV also biases the the catalytic dyad into an open state by orienting the His41 imidazole away from Cys145, measured through the (His41-N_{ε2})-(Cys145-S_γ) distance.

To demonstrate the interconnection of the ξ_2^{backbone} torsion and (Cys145-S_γ)-(His41-N_ε) distance, we can represent the sampling of co-solvent MD simulations (“holo” - (Fig. 5A) and “apo” - (Fig. 5C)) with the interchangeable MetaD sampling of both “apo” and “holo” states in a 2D space (Fig. 5B). Comparing against the three density functions, it is noted that the region about $\xi_2^{\text{backbone}} = -0.5$ rad overlaps with the “holo” state, while the region about $\xi_2^{\text{backbone}} = 0.5$ rad overlaps with the sampled density of the “apo” in which the catalytic dyad remains intact throughout. This comparison allows for the characterisation of each of the respective regions in the metadynamics density function above to be considered as “apo” and “holo” conformational states (Fig. 5B).

In (Fig. 5B), the bimodal probability density in the 2D CV space includes points at shorter

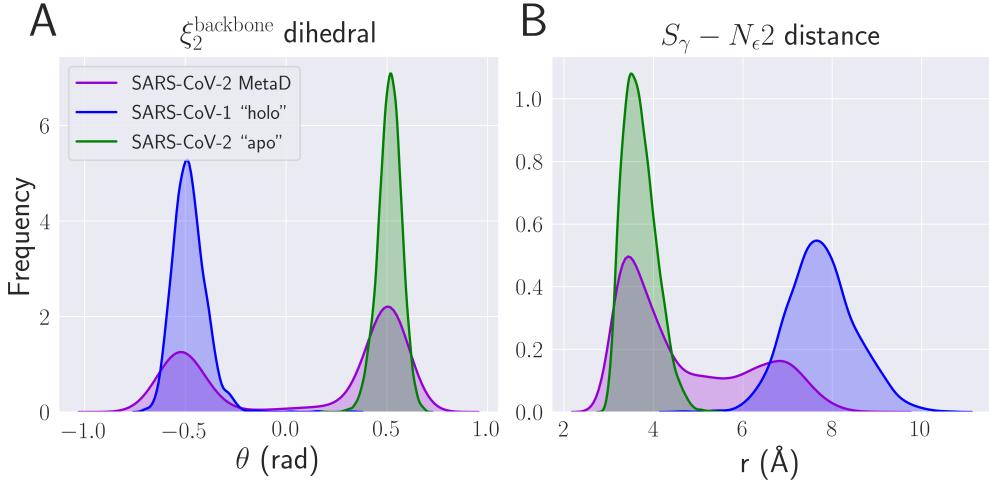


Figure 4: 1D probability density functions obtained from the co-solvent MD simulations of SARS-CoV-1 “holo” (blue), SARS-CoV-2 “apo” (green) and SARS-CoV-2 MetaD simulation (purple) over the (A) ξ_2^{backbone} dihedral space and (B) ($\text{Cys145-}S_\gamma\text{)-(His41-}N_e$) distance.

($\text{Cys145-}S_\gamma\text{)-(His41-}N_e$) distances at $\xi_2^{\text{backbone}} = -0.5$ rad. This region indicates that, in the absence of a potent inhibitor, the “holo” state cannot be stabilised by biased ξ_1 sampling alone. As such, the dyad will reorient to maintain the Cys145-His41 side chain interaction.

To confirm that the choice of a ξ_1 bias samples the “apo”–“holo” transition, the time-dependent behaviour of both the ξ_2^{backbone} dihedral and the ($\text{Cys145-}S_\gamma\text{)-(His41-}N_e$) distance was evaluated using K-means clustering.²⁷ Three distinct clusters were found (Supplementary Fig. 2) corresponding to the “apo” conformation ($\xi_2^{\text{backbone}} = 0.5$ rad), “holo” conformation ($\xi_2^{\text{backbone}} = -0.5$ rad) and transition region. The time dependent behaviour of ξ_2^{backbone} dihedral and ($\text{Cys145-}S_\gamma\text{)-(His41-}N_e$) distance show the choice of CV freely samples the reversible “apo”–“holo” transition throughout the 100 ns trajectory.

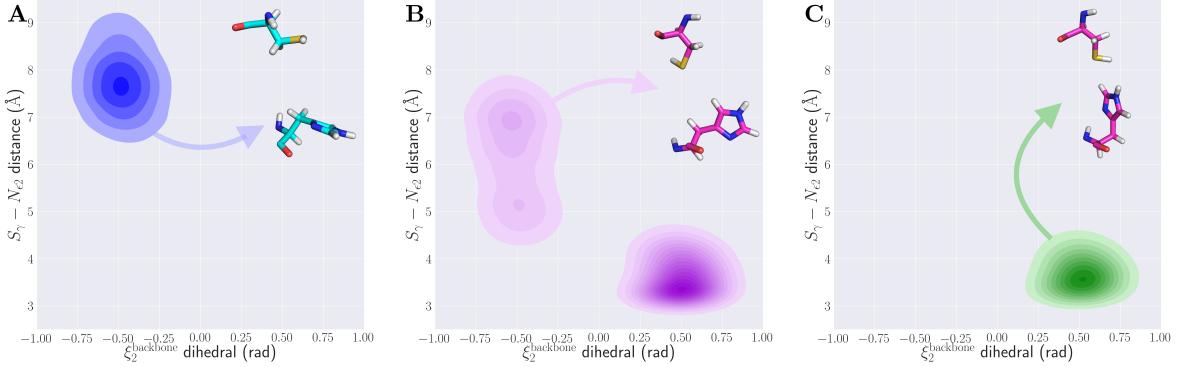


Figure 5: Probability density functions of the (A) SARS-CoV-1 “holo” , (B) SARS-CoV-2 MetaD (inset showing MetaD “holo”) and (C) SARS-CoV-2 “apo” MD simulation defined within a 2D CV space of the His41 ξ_2^{backbone} dihedral and the (Cys145- S_γ)-(His41-N ϵ) atomic distance.

M^{Pro} free energy surfaces

In order to verify the reproducibility of the constructed bias, four replicas of the MetaD simulation were performed until convergence was observed in each. The free energy profiles obtained in each of the replicas show agreement, with deviation from the average free energy values observed chiefly at the transition states and at the minimum at $\xi_1 = 1.9$ rad. The deviation at this minimum is due to the corresponding state being sparsely sampled in all replicas compared to the other system states as a result of the high free energy barrier restricting sampling from other observed minima (Supplementary Fig. 3).

As convergence of the bias was observed for all four replicas, it can be assumed that any sampling beyond the point of convergence is within the desired thermodynamic ensemble. Thus, the mean free energy surface with respect to the His41 ξ_2^{backbone} dihedral over the 4 replicas was

computed (Fig 6). The mean surface is bimodal, with positions of the minima at 0.5 and -0.5 rad corresponding to the sampled “apo” and “holo” states of the SARS-CoV-2 M^{Pro} catalytic dyad, respectively. The “apo”–“holo” relative free energy difference of 4.2 ± 1.9 kJ/mol indicates that the “holo” state is less energetically stable than the “apo” state in the absence of any ligands interacting with His41. Considering this free energy difference between states and the high free energy barrier (15.0 ± 1.3 kJ/mol) between them, it can be assumed that the dyad “apo”–“holo” transition is unlikely to be sampled using unbiased conventional MD simulations alone.

The free energy surface of SARS-CoV-1 “holo” (dashed blue line) was computed from the reference simulation of the D3F-SARS-CoV-1 M^{Pro} complex using population analysis (Fig. 6). This surface is used as a reference state to compare against the mean MetaD free energy surface as the “holo” state. The SARS-CoV-1 “holo” conformation is more energetically stable than the SARS-CoV-2 MetaD surface “holo” minimum by approximately 4.2 ± 1.9 kJ/mol. This energy difference highlights how the energetic penalty associated with dyad disruption in the “apo”–“holo” transition may be recovered by ligand interactions in the active site.

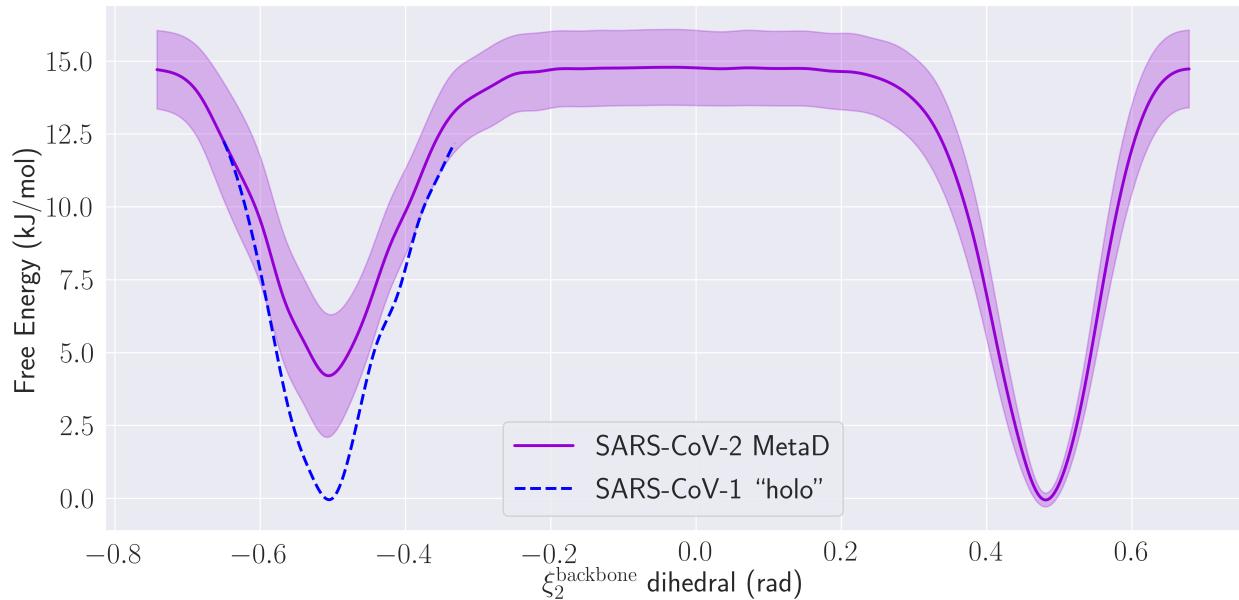


Figure 6: Mean free energy surface obtained for co-solvent MD simulations of D3F-SARS-CoV-1 “holo” (blue) and four replicas of His41 torsional MetaD simulations (purple), defined within the ξ_2^{backbone} dihedral space. The shaded region corresponds to the standard deviation about each free energy point calculated over the set of MetaD replicas.

Conclusions

We have presented the application of metadynamics to molecular dynamics simulations of the SARS-CoV-2 M^{Pro} to better inform the drug discovery efforts including virtual screening, molecular docking and unbiased molecular dynamics simulations. We show that the proteolytic mechanism of M^{Pro} is contingent on the integrity of the His41-Cys145 catalytic dyad. and that the rotation of the His41 imidazole side chain to Cys145 acts as an allosteric trigger to regulating this proteolytic activity.

Using metadynamics, we find that promoting ligand binding and sampling of the active site of the protease is achieved through disrupting the catalytic dyad by biasing over the His41 ξ_1 dihedral to subsequently sample over the ξ_2^{backbone} dihedral. Using the (Cys145-S γ)-(His41-N ϵ) distance and the ξ_2^{backbone} dihedral to define the collective variable (CV) space, we identify clusters of unbound, intermediary and bound conformations of the flexible M^{Pro} active site throughout the simulation. We show that repeated replicas of the His41 torsional metadynamics reproduce the free energy surface along the 1D ξ_1 and ξ_2^{backbone} spaces. Our results detail the allosteric regulation of the SARS-CoV-2 M^{Pro}, irrespective of the choice of ligand. The candidate ligand LIG acts as a toy model and its inadequacy in disrupting the catalytic dyad was established by drawing comparisons of its interaction in the M^{Pro} active site against the ligand D3F with SARS-CoV-1 M^{Pro}.

The application of our analysis uncovers the changes on the receptor structure as a result of an allosteric mechanism resolved using enhanced sampling. This observation can assist in selecting

an optimal strategy for screening ligands from drug libraries. The free energy comparison between the D3F unbound and bound “holo” state minima suggests that the energy expended in disrupting the dyad can be readily recovered using ligand interactions in the active site, and so this open dyad state should be considered alongside the closed state as a target for virtual screening.

Methods

Molecular Dynamics

All MD simulations were performed in GROMACS version 2019.4 on the ARCHER Cray XC30 supercomputer on a single 2.7 GHz, 12-core E5-2697 v2 (Ivy Bridge) series processor node, NVIDIA GeForce RTX 2060 or GeForce GTX1080Ti GPUs with the CUDA 10.2 toolkit. The AMBER14SB forcefield²⁸ was used to model the system, where ligand molecule forcefield parameters were generated using the General Amber ForceField v2 (GAFF2) with charges calculated using the AM1-BCC semi-empirical method.²⁹.

The structure of the candidate ligand LIG was derived from a virtual fragment expansion, docking, and screening exercise provided by Gabriel Grand, Elana Simon, Michael Bower, Bruce Clapham and Jonah Kallenbach of Reverie Labs.³⁰ Protein X-Ray diffraction (XRD) structures for SARS-CoV-1 M^{Pro} holo structure (PDB code: 2GZ7)³¹ and SARS-CoV-2 M^{Pro} apo structure (PDB code: 5RE4) were used for these simulations. Missing residues from the SARS-CoV-2 XRD structure were modelled using MODELLER³² and subsequently solvated in a 90 Å cubic box of TIP3P water molecules.

For co-solvent MD simulations, five of the candidate ligand (LIG) were placed in random coordinates within the box and the system net charge was neutralised by adding 4 sodium ions into the system. The co-solvent MD simulations was performed for 100 ns. The simulation of D3F bound in SARS-CoV-1 M^{Pro} was performed for 10 ns. The net charge in the system was neutralised by adding 3 sodium ions. In all structures prepared for simulation, the His41 side chain was maintained at the N1-H tautomeric state, as this state is primed for nucleophilic attack from the Cys145 mercaptan in the active catalytic dyad.

The system was relaxed energetically using steepest-descent energy minimisation for 50000 steps with an energetic step size of 0.01 kJ/mol. The minimisation was terminated after the maximum energetic contribution was lower than a threshold of 10.0 kJ/mol. NVT and NPT equilibration was performed for 1 ns using two separate velocity-rescaling thermostat coupling temperature to velocities for protein, drug and solvent molecules (NVT), where a temperature of 300 K was maintained and 1 bar using the Parrinello-Rahman barostat (NPT).³³ The Verlet cut-off scheme was employed to generate pair lists and the electrostatic interactions were calculated using the Particle-Mesh Ewald algorithm.³⁴ Both electrostatic and van der Waals interactions were cut off beyond 1.2 nm. All bonds involving hydrogen atoms were constrained using the LINCS algorithm.³⁵ Production simulations ran with an integration stepsize of 2 fs. MDAnalysis was used to postprocess the MD trajectories for analysis.^{36,37}

Metadynamics

Four independent replicas of the non-tempered MetaD simulation were run to convergence.³⁸ The bias potential was setup to sample energetically hindered rotations about the ξ_2^{backbone} dihedral of His41 by biasing the sampling along the ξ_1 torsional profile directly. The bias was accumulated with a Gaussian deposition rate $\tau = 1$ ps. The deposited Gaussians had a fixed height of 0.1 kJ/mol and employed an adaptive width scheme in which the correlation between the biased CV space and the microscopic configuration space is utilised to recalculate the covariance matrix. A correlation length of 0.5 Å was used.³⁹ MetaD simulations were performed using Plumed 2.5.4.⁴⁰

1. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nature Medicine* **26**, 450–452 (2020).
2. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
3. World Health Organization. Coronavirus disease (COVID-19): situation report, 209. URL <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>. Accessed: 02.09.2020.
4. Petersen, E. *et al.* Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet Infectious Diseases* **20**, e238–e244 (2020). URL [https://doi.org/10.1016/s1473-3099\(20\)30484-9](https://doi.org/10.1016/s1473-3099(20)30484-9).
5. Wang, Y. *et al.* Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet* **395**, 1569–1578 (2020).
6. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *New England Journal of Medicine* (2020). URL <https://doi.org/10.1056/nejmoa2021436>.
7. and Jonathan A. C. Sterne *et al.* Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19. *JAMA* (2020).
8. World Health Organization. DRAFT landscape of COVID-19 candidate vaccines. URL <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines>. Accessed: 02.09.2020.

9. Zhang, C.-Y., Wei, J.-F. & He, S.-H. Adaptive evolution of the spike gene of SARS coronavirus: changes in positively selected sites in different epidemic groups. *BMC microbiology* **6**, 88 (2006).
10. Casalino, L. *et al.* Beyond Shielding: The Roles of Glycans in SARS-CoV-2 Spike Protein (2020). URL <https://doi.org/10.1101/2020.06.11.146522>.
11. Ullrich, S. & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorganic & Medicinal Chemistry Letters* **30**, 127377 (2020).
12. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC biology* **9**, 1–9 (2011).
13. Deeks, S. G. *et al.* HIV RNA and CD4 cell count response to protease inhibitor therapy in an urban AIDS clinic: response to both initial and salvage therapy. *Aids* **13**, F35–F43 (1999).
14. Lamarre, D. *et al.* An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**, 186–189 (2003).
15. York, D., Darden, T., Pedersen, L. & Anderson, M. Molecular dynamics simulation of HIV-1 protease in a crystalline environment and in solution. *Biochemistry* **32**, 1443–1453 (1993).
16. Schames, J. R. *et al.* Discovery of a novel binding trench in HIV integrase. *Journal of Medicinal Chemistry* **47**, 1879–1881 (2004).
17. Hegyi, A. & Ziebuhr, J. Conservation of substrate specificities among coronavirus main proteases. *Journal of General Virology* **83**, 595–599 (2002).

18. Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).
19. Macchiagodena, M., Pagliai, M. & Procacci, P. Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. *Chemical Physics Letters* **750**, 137489 (2020).
20. Ghahremanpour, M. M. *et al.* Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2 (2020). URL <https://doi.org/10.1101/2020.08.28.271957>.
21. Laio, A. & Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **71**, 126601 (2008).
22. Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **11**, 3584–3595 (2015).
23. Sztain, T., Amaro, R. & McCammon, J. A. Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 protease (2020). URL <https://doi.org/10.1101/2020.07.23.218784>.
24. Lu, I.-L. *et al.* Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *Journal of Medicinal Chemistry* **49**, 5154–5161 (2006).

25. Seco, J., Luque, F. J. & Barril, X. Binding site detection and druggability index from first principles. *Journal of Medicinal Chemistry* **52**, 2363–2371 (2009).
26. Ghanakota, P. & Carlson, H. A. Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics: Miniperspective. *Journal of Medicinal Chemistry* **59**, 10383–10399 (2016).
27. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
28. Maier, J. A. *et al.* ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**, 3696–3713 (2015).
29. Case, D. *et al.* AMBER 2015. *University of California: San Francisco, CA, USA* (2015).
30. Grand, G., Simon, E., Bower, M., Clapham, B. & Kallenbach, J. Reverie Labs PostEra COVID Moonshot submission GAB-REV-df6. URL <https://discuss.postera.ai/t/submission-gab-rev-df6/339>. Accessed: 10.09.2020.
31. Lu, I.-L. *et al.* Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *Journal of Medicinal Chemistry* **49**, 5154–5161 (2006).
32. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234**, 779–815 (1993).

33. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182–7190 (1981).
34. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092 (1993).
35. Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463–1472 (1997).
36. Richard J. Gowers *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In Sebastian Benthall & Scott Rostrup (eds.) *Proceedings of the 15th Python in Science Conference*, 98 – 105 (2016).
37. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**, 2319–2327 (2011).
38. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **99**, 12562–12566 (2002).
39. Branduardi, D., Bussi, G. & Parrinello, M. Metadynamics with Adaptive Gaussians. *Journal of Chemical Theory and Computation* **8**, 2247–2254 (2012). PMID: 26588957.
40. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. "PLUMED 2: New feathers for an old bird". *Computer Physics Communications* **185**, 604 – 613 (2014).

Acknowledgements M. A. B. and K. A. M. are grateful to Salvatore F. Pedicona for discussions. This work was supported by BBSRC (grant BB/M009513/1) and EPSRC (grant EP/L015722/1). This work used the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>). K. A. M. acknowledges the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

Author contributions M. A. B. and K. A. M. conceived and planned the research. K. A. M. and M. A. B. performed the calculations. K. A. M., M. A. B., C. D. L. and J. W. E. analysed the data and contributed to the final manuscript.

Competing Interests The authors declare no competing interests.