

## Supporting Information

# Dynamic Profiling of the $\beta$ -Coronavirus 3CL M<sup>pro</sup> Protease Ligand Binding Sites

Eunice Cho,<sup>1</sup> Margarida Rosa,<sup>1</sup> Ruhi Anjum,<sup>2</sup> Saman Mehmood,<sup>3</sup> Mariya Soban,<sup>2</sup> Moniza Mujtaba,<sup>4</sup> Khair Bux,<sup>5</sup> Syed T. Moin,<sup>5</sup> Mohammad Tanweer,<sup>1</sup> Sarath Dantu,<sup>6</sup> Alessandro Pandini,<sup>6</sup> Junqi Yin,<sup>7</sup> Heng Ma,<sup>8</sup> Arvind Ramanathan,<sup>8,9</sup> Barira Islam,<sup>10</sup> Antonia S. J. S. Mey,<sup>11</sup> Debsindhu Bhowmik,<sup>12</sup> and Shozeb Haider<sup>1\*</sup>

<sup>1</sup> UCL School of Pharmacy, London WC1N 1AX, United Kingdom

<sup>2</sup> Department of Biochemistry, Aligarh Muslim University, Aligarh, 202002, India

<sup>3</sup> Department of Zoology, Aligarh Muslim University, Aligarh 202002, India

<sup>4</sup> Herricks High School, New Hyde Park, New York, 11040 USA

<sup>5</sup> Third World Center for Science and Technology, H.E.J. Research Institute of Chemistry, International Centre of Chemical and Biological Sciences, University of Karachi, Karachi, 75270 Pakistan

<sup>6</sup> Department of Computer Science, Brunel University, Uxbridge, UB8 3PH, United Kingdom

<sup>7</sup> Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

<sup>8</sup> Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439, USA

<sup>9</sup> Consortium for Advanced Science and Engineering, University of Chicago, Chicago, IL 60637, USA

<sup>10</sup> Department of Bioscience, University of Huddersfield, Huddersfield, HD1 3DH, United Kingdom

<sup>11</sup> EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom

<sup>12</sup> Computer Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

\* Corresponding author: Shozeb Haider

Email: [Shozeb.haider@ucl.ac.uk](mailto:Shozeb.haider@ucl.ac.uk)

ORCID: Shozeb Haider: 0000-0003-2650-2925

## Mapping the binding sites

Most of the ligands that have been crystallised bind to the substrate binding site (site A in Figure 2A). We identified 185 ligands that bind either covalently or non-covalently, exploiting various subsites of this pocket. The substrate binding site is a deep cleft between domains I and II. C145 is spatially positioned in the deep distal end of the pocket. For inhibitors to covalently attach to C145, they would have to penetrate deep inside the binding site. The non-covalent inhibitors occupy one of the subsites within the substrate binding site. PDB 6LU7 was chosen as the representative structure to highlight this binding pocket <sup>1</sup>.

Sites B (PDB 5RFC; <sup>2</sup>) and C (PDB 5RE6; <sup>2</sup>) are present in domain I and form part of the antiparallel  $\beta$ -barrel structure (Figure 2B/C). The two sites lie adjacent to one another, separated by a part of the  $\beta$ 4 strand (I78-M82). The N- group in methyl (2-methyl-4-phenyl-1,3-thiazol-5-yl)carbamate interacts with the backbone carbonyl oxygen of H80 and the pyrimidine group in N-{4-[(pyrimidin-2-yl)oxy]phenyl}acetamide in site C is positioned above H80. Five ligands have been co-crystallised in site B, while two have been identified to bind in site C.

Site D (PDB 5RF8; <sup>2</sup>) is a shallow, surface exposed binding site formed as a part of a large hairpin loop that connects domain II and III (residues P184-Q192) (Figure 2D). 4-amino-N-(pyridin-2-yl)benzenesulfonamide (Z271004858) is anchored in this binding site via hydrogen bonds with R188 on one end and backbone carbonyl oxygen of A191 on the other (Figure S2D). It is noteworthy that site D shares residues R188-Q192 with the substrate binding site (site A). Comparing the conformation of the linker loop that connects domain II with III in site D, with the structure of the loop when the ligands are present in their respective binding sites in site A, suggests that only one binding site can be occupied at a time. Ligand binding in site A shifts the conformation of the guanidinium side chain of R188 such that binding site D is lost. However, ligand binding to site D has no effect on subsite S4 of the substrate binding pocket. Only one ligand has been reported to bind in site D.

Site E (PDB 5RF5; <sup>2</sup>) is a pocket formed between the  $\beta$ 7- $\beta$ 8 loop and the proximal end of the linker loop in domain II (Figure 2E). The pocket is so small that it can accommodate only a small fragment. Larger fragments also bind to this site by anchoring inside this pocket. Four fragments have been identified that bind to this pocket including 1,1

bis(oxidanylidene)thietan-3-ol (Z3241250482) that has been used as its representative structure.

Site F (PDB 5REG; <sup>2</sup>) is formed from residues from three different loops including  $\beta$ 1- $\beta$ 2 (R105, Q107),  $\beta$ 3- $\beta$ 4 (F134) in domain II and the linker loop (G183). The binding site is small and can only accommodate half of the (2- $\{S\}$ )- $\{N\}$ -(4-aminocarbonylphenyl)oxolane-2-carboxamide (Z1545313172), with the oxolane-carboxamide group inside the binding pocket and the aminocarbonylphenyl group on the outside. The backbone of G183 makes hydrogen bond and F134 interacts with the ligand by forming  $\pi$ -stacking interactions. The binding pocket is also connected to another, larger pocket via a narrow passage. However, the passage is not wide enough for a phenyl group to pass through it (Figure 2F).

Site G (PDB 5RF0; <sup>2</sup>) is a pocket, present just below the substrate-binding site (site A), at the interface between the two protomers. The site is sandwiched between K137 and G138 (protomer A) and the N-finger residues (S1-A7) of protomer B. A part of this binding site is also formed by a loop between  $\beta$ 11- $\beta$ 12, which makes the lower lip of binding site A (L166-V171). L282, G283 and N214 from the  $\alpha$ -helix B in domain III also contribute to this pocket (Figure 2G). Binding of a ligand in the substrate-binding pocket (site A) alters the conformation of the  $\beta$ 11- $\beta$ 12 loop at the proximal end of the binding site near S4 subsite. This change in conformation of the loop has an effect on the dimensions of site G. Only one ligand POB0073 ([1-(pyridin-2-yl)cyclopentyl]methanol) has been identified that binds to this pocket.

Site H (PDB 5RGS; <sup>2</sup>) is present in the  $\alpha$ -helical dimerization domain III (Figure 2H). The site is formed from two subpockets that are separated by the side chains of T198 and E240. Residues from  $\beta$ 7- $\beta$ 8 loop (P132, N133) from domain II, the linker loop (G195, T196, T198), the loop between  $\alpha$ -helix C and D in domain III (N238, Y239, E240) and A234, M235 from the terminal end of  $\alpha$ -helix C form this binding site. The methanol group makes hydrogen bonds with the carbonyl oxygen atom from the backbone of M235 and the morpholino oxygen makes interactions with the side chain of N238. Four ligands have been identified that can bind to this pocket, while Z1259086950 has been used as a representative ligand to describe this site.

In the representative structure for Site I (PDB 5RGR; <sup>2</sup>), the fragment N,1-dimethyl-N-(propan-2-yl)-1H-pyrazolo[3,4-d]pyrimidin-4-amine (Z328695024) binds at two sites on M<sup>pro</sup>. The first site is the same site as site E, where the N-propanyl group is tucked inside the pocket and the pyrazolo-pyrimidine stacks on the  $\beta$ 7 strand of the 2<sup>nd</sup> antiparallel  $\beta$ -barrel in domain II. The second site (I2) is formed by residues from  $\alpha$ -helix C (N228, L232, M235) and the loop between  $\alpha$ -helix C and D (P241) in the dimerization domain III (Figure 2I). The binding site is an exposed, shallow pocket. Only one ligand has been identified to bind to this site.

Site J (PDB 5RE7; <sup>2</sup>) is formed on the distal end, at the entrance of the antiparallel  $\beta$ -barrel in domain I. This site is a part of a long channel that runs between domains I and II (Figure 2J). Residues from  $\beta$ 1 (W31), loop  $\beta$ 4- $\beta$ 5 (A70, G71),  $\beta$ 5 (V73, L75), and the loop between  $\beta$ 6 from domain I and  $\beta$ 7 from domain II (T93, A94, P96, K97) lie within 4 Å from the ligand. The N-sulfamoyl group and the phenyl ring in N-[(4-sulfamoylphenyl)methyl]acetamide (Z30932204) are positioned in the site while the acetamide group protrudes into the solvent.

Site K (PDB 5RFD; <sup>2</sup>) is a shallow pocket formed between the loops that connect  $\beta$ 2 and  $\beta$ 3 in antiparallel  $\beta$ -barrel in domain I and the loop that connects domains I and II (Figure 2K). It lies adjacent to site J and residues A94, N95, P96 ( $\beta$ 6- $\beta$ 7 loop) are shared between the two sites. The ligand 2-[(methylsulfonyl)methyl]-1H-benzimidazole (Z126932614) in site K is solvent exposed. Ligand binding to site K has no effect on the dimensions of site J. Three ligands have been identified that are bound to site K.

Site L (PDB 5RF9; <sup>2</sup>) is a long, elongated channel like site located immediately downstream from site K, along the loop that connects domain I and II (residues P96-K100). One end of the site is a deep cavity formed from K12, K100, D155, G156, V157 and binds the pyrazol ring in 1-[(2- $\{S\}$ )-2-methylmorpholin-4-yl]-2-pyrazol-1-yl-ethanone (Z217038356). The ketone oxygen anchors the inhibitor in the binding site by making interactions with K12. The methyl group substituted of the morpholino ring projects away from the binding site towards site K (Figure 2L). Three ligands have been identified to bind in this site.

Site M (PDB 5RFA; <sup>2</sup>) is a deep cavity formed at the interface between the two protomers. The site is formed between domain II and III on protomer A, including residues from the N-terminal (M6, A7, F8, P9),  $\alpha$ -helix F (D295, R298, Q299),  $\beta$ 8 (T111, S113) and  $\beta$ 9 (Q127).

Residues Y118 (loop  $\beta$ 8- $\beta$ 9), S123, G124 ( $\beta$ 9) and L141 (loop  $\beta$ 9- $\beta$ 10) from protomer B contribute to this site. When 1-methyl-N- $\{[(2S)\text{-oxolan-2-yl}]\text{methyl}\}$ -1H-pyrazole-3-carboxamide (Z2643472210) binds, the pyrazole ring penetrates the binding site between the N- and C-terminal, while the amide group and the methyl oxolane ring interacts with residues from protomer B. The amide group makes hydrogen bond with the backbone carbonyl oxygen of S123 from protomer B. Two ligands have been identified that bind to this site.

Site N (PDB 5REF; <sup>2</sup>) is sandwiched between  $\alpha$ -helix D and F. The site forms a deep cleft that branch off an extended channel of cavities formed between the antiparallel  $\beta$ -barrel in domain II and the  $\alpha$ -helices in dimerization domain III (Figure 2N). The methylsulfonylamino group in methyl 3-(methylsulfonylamino)benzoate (Z24758179) faces the channel, while the benzoate group is sandwiched between F294 and I249. Residues from  $\alpha$ -helix F are shared between Site M and Site N. Ligand binding to site M alters the conformation of the side chain of F294, which completely occludes site N. When the ligand is bound in site N, the phenyl side chain rotates away from the binding site and is positioned above  $\alpha$ -helix F, providing enough space for the ligand to bind. Another ligand, AT7519 (PDB 7AGA; <sup>3</sup>) also binds to site N, not in the cleft but in an orientation that is along the channel. When this happens, the side chain of F294 adopts a similar conformation when a ligand is bound to site M, which obstructs the deep cleft.

Site O (PDB 5REA; <sup>2</sup>) is a solvent exposed, pseudo-ligand binding site where (azepan-1-yl)(2H-1,3-benzodioxol-5-yl)methanone (Z31432226) forms interactions with N277, G278 and R279 (Figure 2O). These residues are part of the loop between  $\alpha$ -helices E and F. There is no cavity in which the ligand binds but is anchored to the exterior of M<sup>PRO</sup> via hydrogen bonds with the side chain of N277 and R279. The ligand in site O is positioned in site J, in its symmetry-related protomer (Figure S4).

Site P (PDB 7AXO; <sup>4</sup>) is a pocket formed by residues at distal ends of  $\alpha$ -helices C and E and the loop that leads to  $\alpha$ -helix F, in the dimerization domain (Figure 2P). The phenyl ring in N-hydroxy-4- $\{[(2S)\text{-3-methyl-2-phenylbutanoyl}]\text{amino}\}$ benzamide (AR-42) is tucked in a hydrophobic subpocket formed by residues M276, A285, L276 and L287. The central amide anchors the ligand to the backbone carbonyl oxygen of L272, while the benzamide group is exposed to the solvent <sup>3</sup>. Two ligands have been identified that bind to this pocket.

Site Q (PDB 7AMJ; <sup>4</sup>) is formed at the interface of the dimerization domains, with contribution from the residues at the distal ends of  $\alpha$ -helices B, D and F in domain III in one protomer and the loop between  $\beta$ 9 and  $\beta$ 10 (residues L141, N142) in the other protomer (Figure 2Q). It is worth noting that N142 also forms a part of the substrate binding site (site A). The piperazine ring and its dimethylphenyl ring substituent in PD168568 occupy a hydrophobic pocket formed from the three-helix bundle present in the dimerization domain. The isoindolone group interacts with N142 at the entrance of the substrate binding site in the adjacent protomer. Five ligands have been identified that interact with this binding site.

Site R (PDB 7ARF; <sup>4</sup>) is present in domain II. Residues from  $\beta$ 7 (K100, Y101, K102) and  $\beta$ 11 (D155, C156) contribute to this shallow, solvent exposed binding site (Figure 2R). Small fragments bind to this site by covalently attaching to the surface exposed C156 via a disulphide bond. Fragments like thioglucose, can make additional interactions with the backbone carbonyl oxygen of Y101 and K102. Six fragments have been identified that covalently bind in this binding site.

Site S (PDB 7AWR; <sup>4</sup>) is a shallow, surface binding pocket between domain I and II, formed by residues D33, T98-F103 (Figure 2S). Tegafur binds to the site via  $\pi$ -stacking interactions between Y101 and the pyrimidine ring of the ligand. Site R and S are separated by common residues from the  $\beta$ 7 strand. A glutathione isopropyl ester (PDB 7AX6) straddles  $\beta$ 7 strand and interacts in both site R and S simultaneously.

Site T (PDB 7KVL; <sup>5</sup>) is sandwiched between  $\alpha$ -helices C (V233, Y237) and E (K269, Q273) in the dimerization domain (Figure 2T). The nitrogen in the pyridine ring makes interactions with the hydroxyl group in the Y237 and the K269 side chain forms a hydrogen bond with oxygen in the carboxamide group of the fragment. This site lies adjacent to site P and is separated by the side chain of Y237.

Site U (PDB 7KVR; <sup>6</sup>) is a pseudo-ligand binding site formed at the interface of two symmetry-related protomers (Figure 2U). The dimethylpyridine-2,4-diamine fragment stacks on side chains from residues in the loop between  $\beta$ 5- $\beta$ 6 strands (Q83, N84). A hydrogen bond between the side chains of Q83 and K88 ( $\beta$ 6 strand) further stabilizes this stacking platform. A water molecule bridges interactions between the fragment and the backbone

oxygen atom of M82. The fragment also makes hydrogen bond with the side chain of D48 of the symmetry-related protomer (Figure S4).

Site V (PDB 7LFP; <sup>7</sup>) is present at the interface of the two protomers at the central axis of the enzyme (Figure 2T). Residues in the N-finger (R4, K5, A7) and  $\beta$ 9 strand (V125, Y126, Q127) from both protomers surround the ligand within the binding site. The side chains of R4 and from one protomer and Q127 from the second protomer make hydrogen bonds with the urea nitrogen atoms.

### **CVAE-based Deep learning analysis**

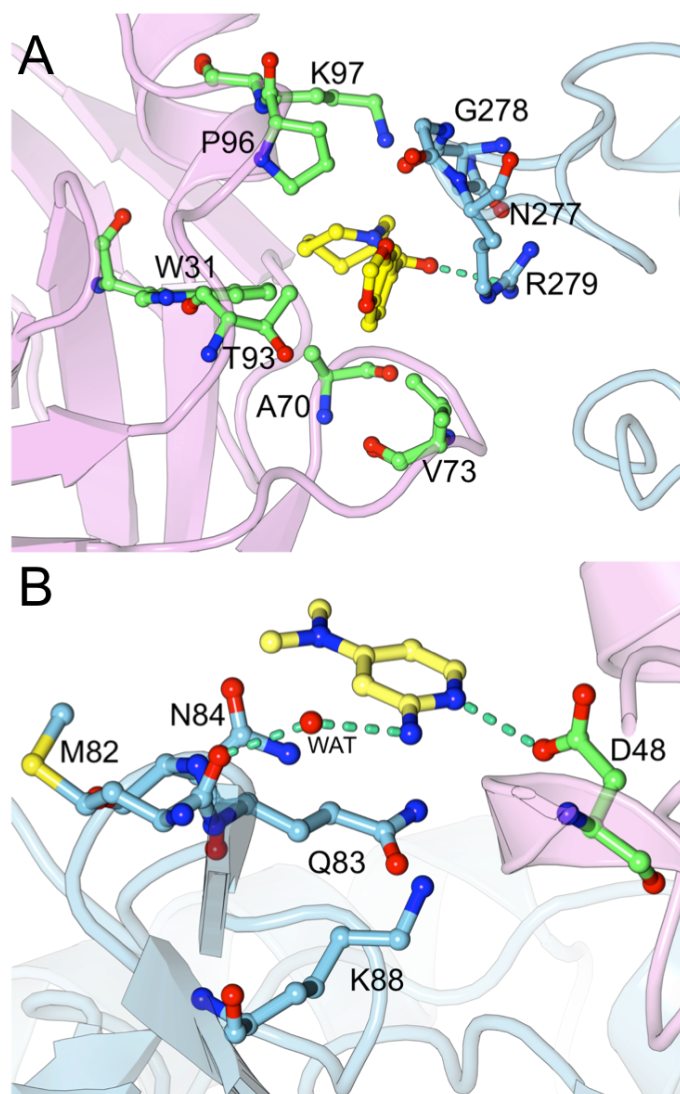
A CVAE can organize the conformational landscape in terms of a small number of biophysically relevant conformational coordinates from long timescale simulations.<sup>8,9</sup> In protein folding trajectories, the CVAE-inferred reduced conformational coordinates correspond to separating folded and unfolded states (and potentially transitions involving these states)<sup>8</sup> and for equilibrium simulations, the CVAE provides biophysically relevant information related to conformational transitions induced by changes in hydrogen-bonds/hydrophobic interactions.<sup>9,10</sup> In this work, the CVAE was used to identify differences in the collective conformational fluctuations within  $M^{\text{pro}}$  simulations from SARS-CoV2, SARS-CoV and MERS-CoV.

First, the CVAE model quality from the simulation trajectories was evaluated by examining the training and validation loss for various models, with latent dimension size ranging from 3 to 11 (Figure 6B). Initially, as the dimension size increases, the corresponding model compresses less and hence has more representation capability. When the latent dimension becomes too large, the model may over fit local features and introduce extra noise, and the regularizing term (Kullback-Leibler divergence) of the loss will play a bigger role. The overall loss approaches an optimal value in between those two extremes. For this dataset, the CVAE model is quite stable and robust, considering the validation loss stays close across various latent dimensions.

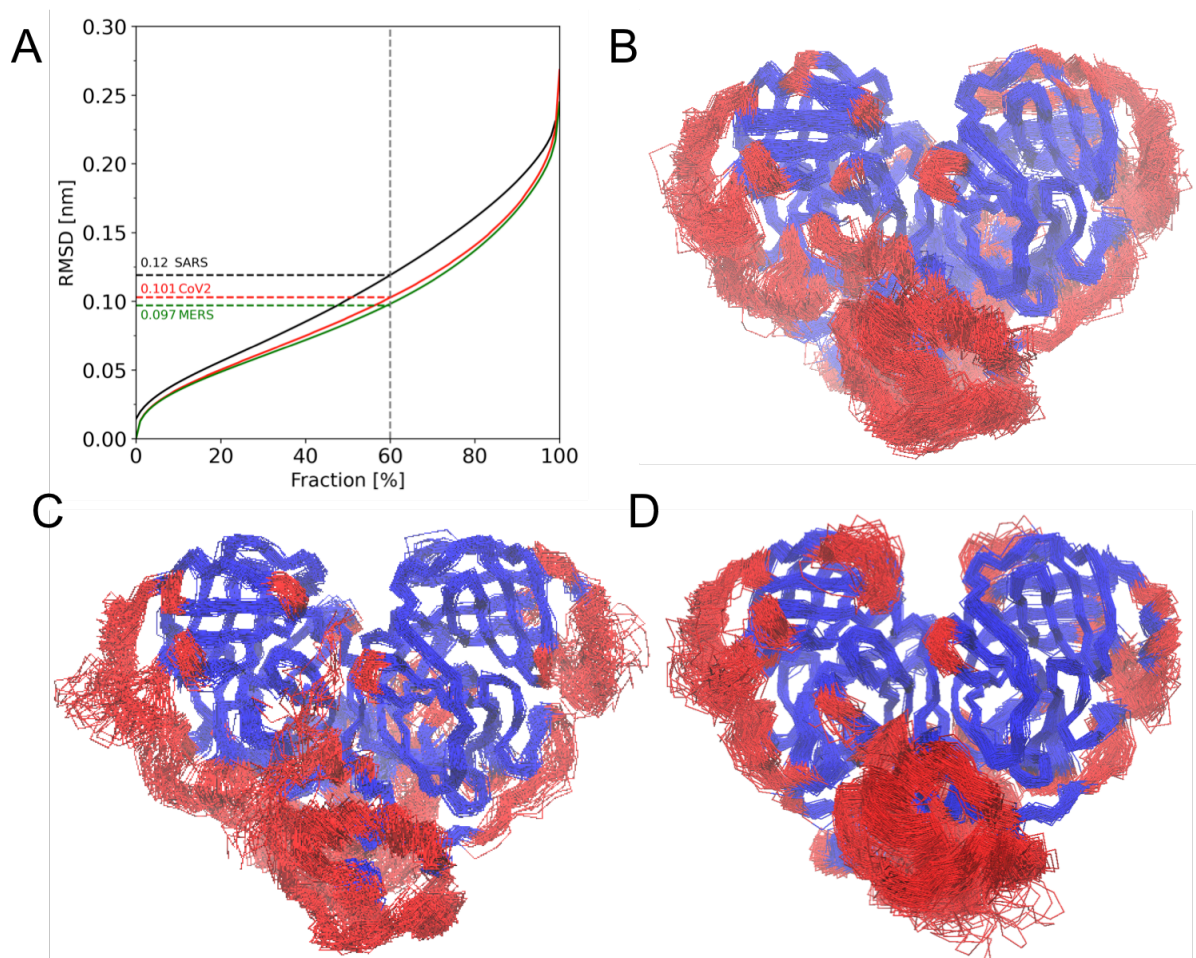
A latent dimension of 7 was selected based on the reconstruction loss as well as the uncertainty in the validation set. Next, t-distributed stochastic neighbor embedding (t-sne) was performed on this compressed lower-dimension data to visualize in two dimensions.<sup>8</sup>

Figure 3A shows the two-dimensional t-sne representation of the CVAE low-dimension data, while Figure 3B depicts three dimensions of the compression CVAE low-dimensional data. The CVAE is able to completely cluster the three different  $\beta$ -coronavirus M<sup>pro</sup> types based on the local and global conformational dynamics (Figure 3). This is more visible in the two-dimensional representation. Here, SARS-CoV2 and SARS-CoV behave similar to each other while MERS-CoV is very different. One must note that this clustering is not evident when using traditional features (such as RMSD or native contacts) to distinguish among these three types of closely related  $\beta$ -coronavirus homologs, proving the sensitivity of the CVAE implementation.

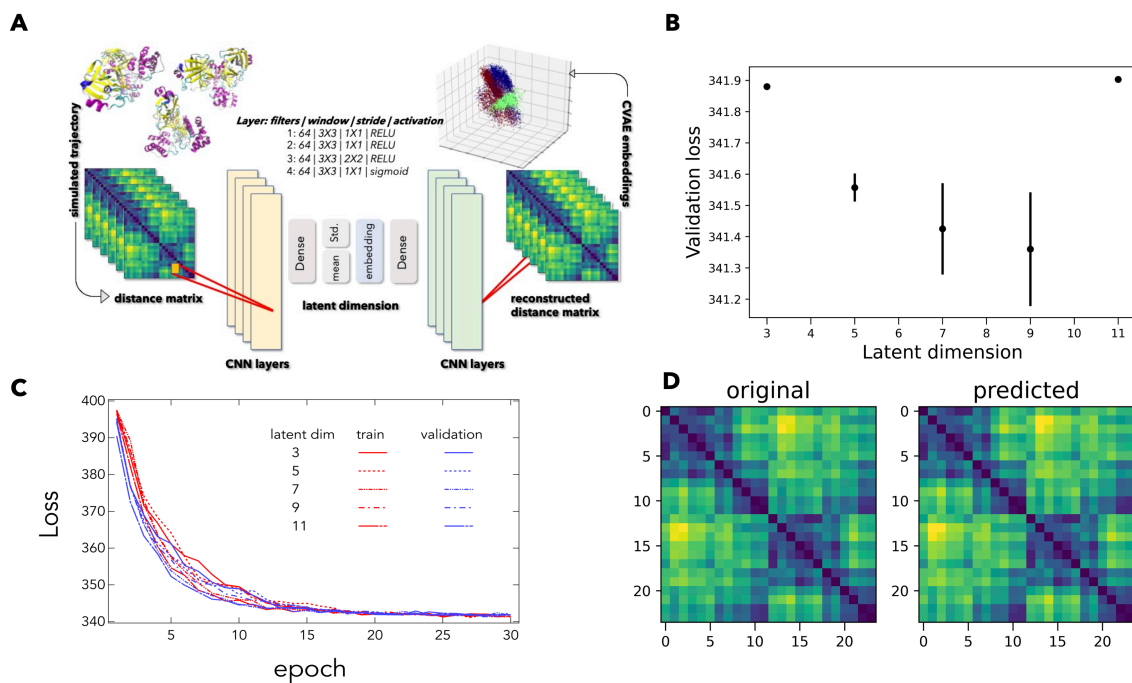




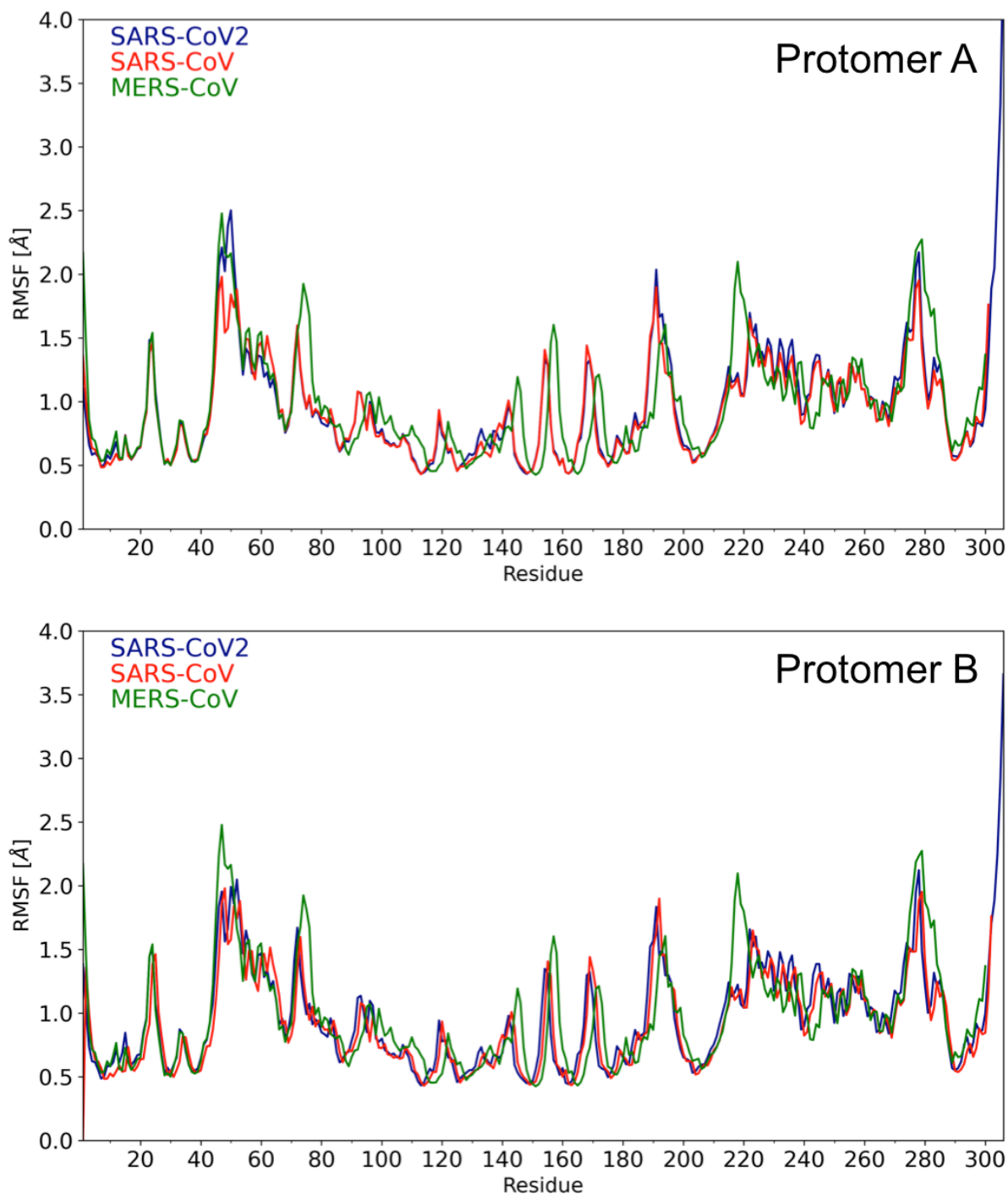
**Figure S1:** Interactions between the SARS-CoV2 M<sup>pro</sup> (cyan) in the pseudo-ligand binding sites. (A) The ligand (yellow sticks) in site O (cyan sticks) interacts with residues in site J in its symmetry-related protomer (pink ribbon/green sticks) in PDB 5REA.<sup>2</sup> (B) The ligand in site U is stabilized by hydrogen bond interaction with the backbone of M82 via a water molecule and with D48 from its symmetry-related protomer (pink ribbon/green sticks) in PDB 7KVR.<sup>6</sup>



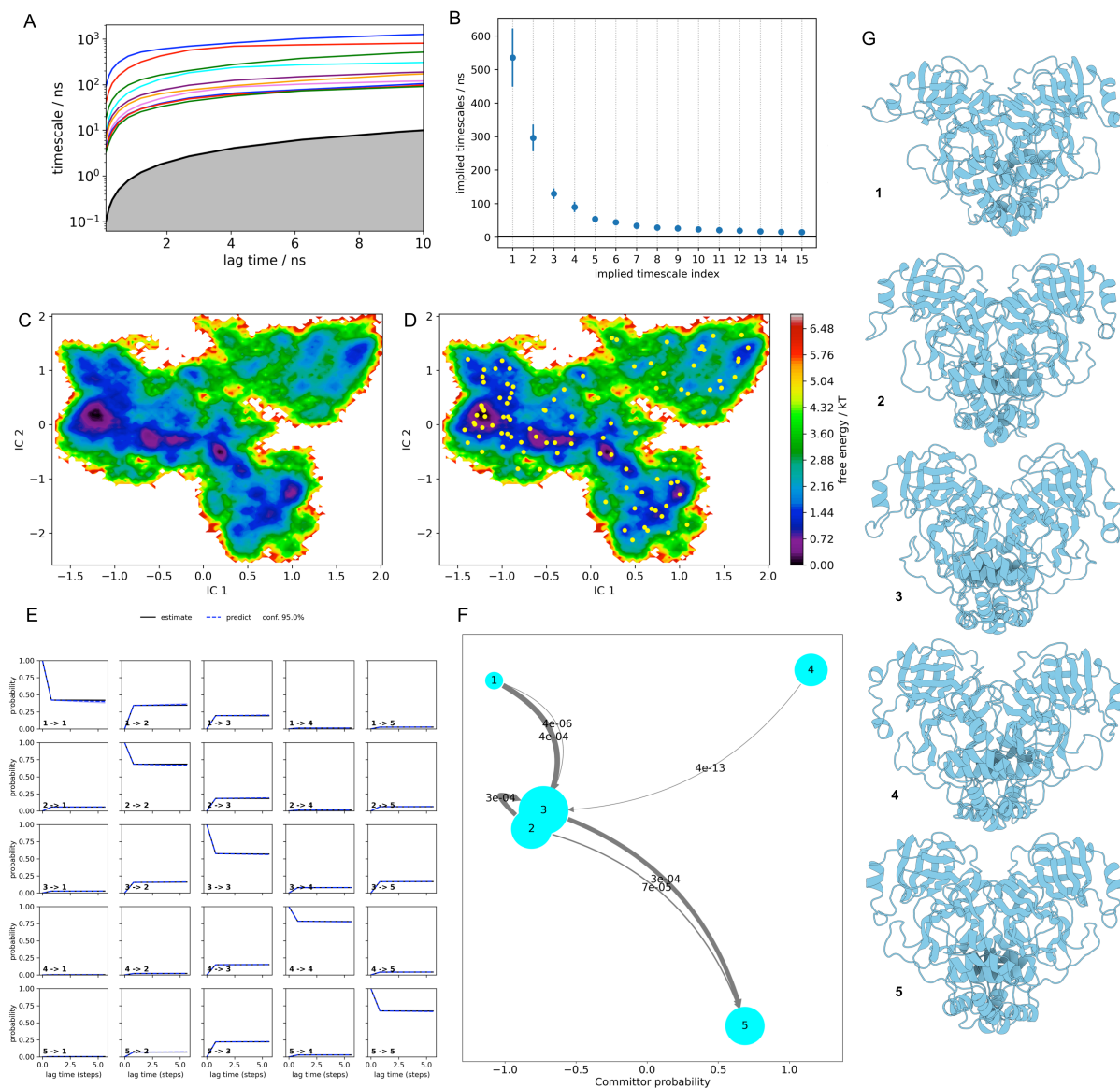
**Figure S2:** Conformational drift in  $M^{\text{pro}}$  enzymes. (A) 60% of residues in SARS-CoV2 (red), SARS-CoV (black) and MERS-CoV (green), can be aligned to below 0.10 nm, 0.12 nm and 0.09 respectively. Conformations of (B) SARS-CoV2 (C) SARS-CoV and (D) MERS-CoV, extracted at every 100 ns are superimposed to highlight the stable core (blue) and highly dynamic (red) regions in the  $M^{\text{pro}}$ . Domain I and II are most stable, except for the loops and the linker regions. The dimerization domain III is the most dynamic region of the enzyme.



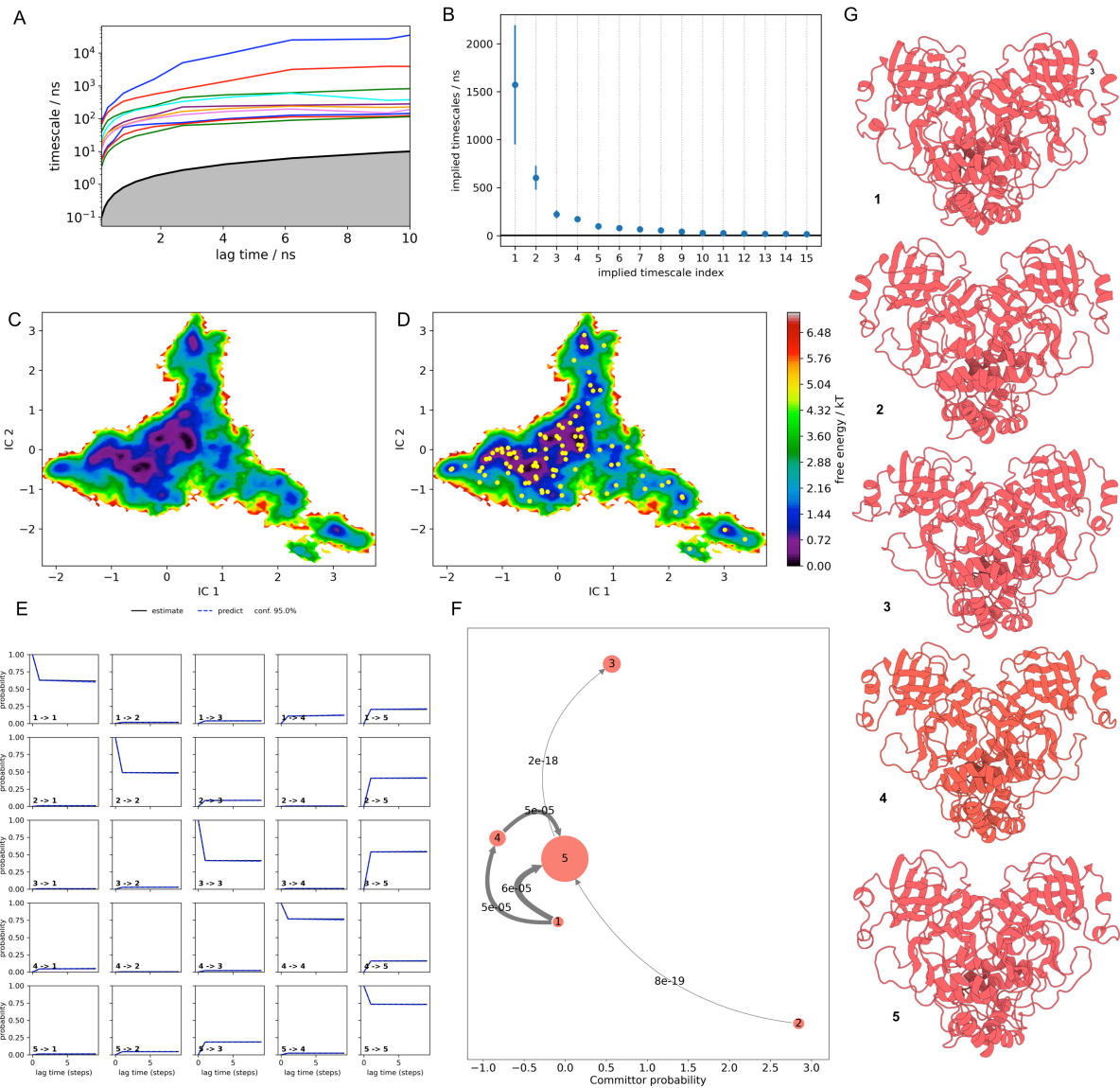
**Figure S3:** CVAE based Deep Learning Implementation. (A) Complete CVAE architecture where distance matrix is used as input data to feed into the CVAE for generating low-dimensional representation. The input data is then trained where the training quality can be followed by training and validation loss at different dimension over the epochs. (B) The validation loss is plotted at different latent dimension to determine optimum values of the low dimension. (C) Simultaneously the training and validation loss is assessed over consecutive epochs at various dimensions. (D) A comparison between original input and predicted data to ensure no loss of information during compression process.



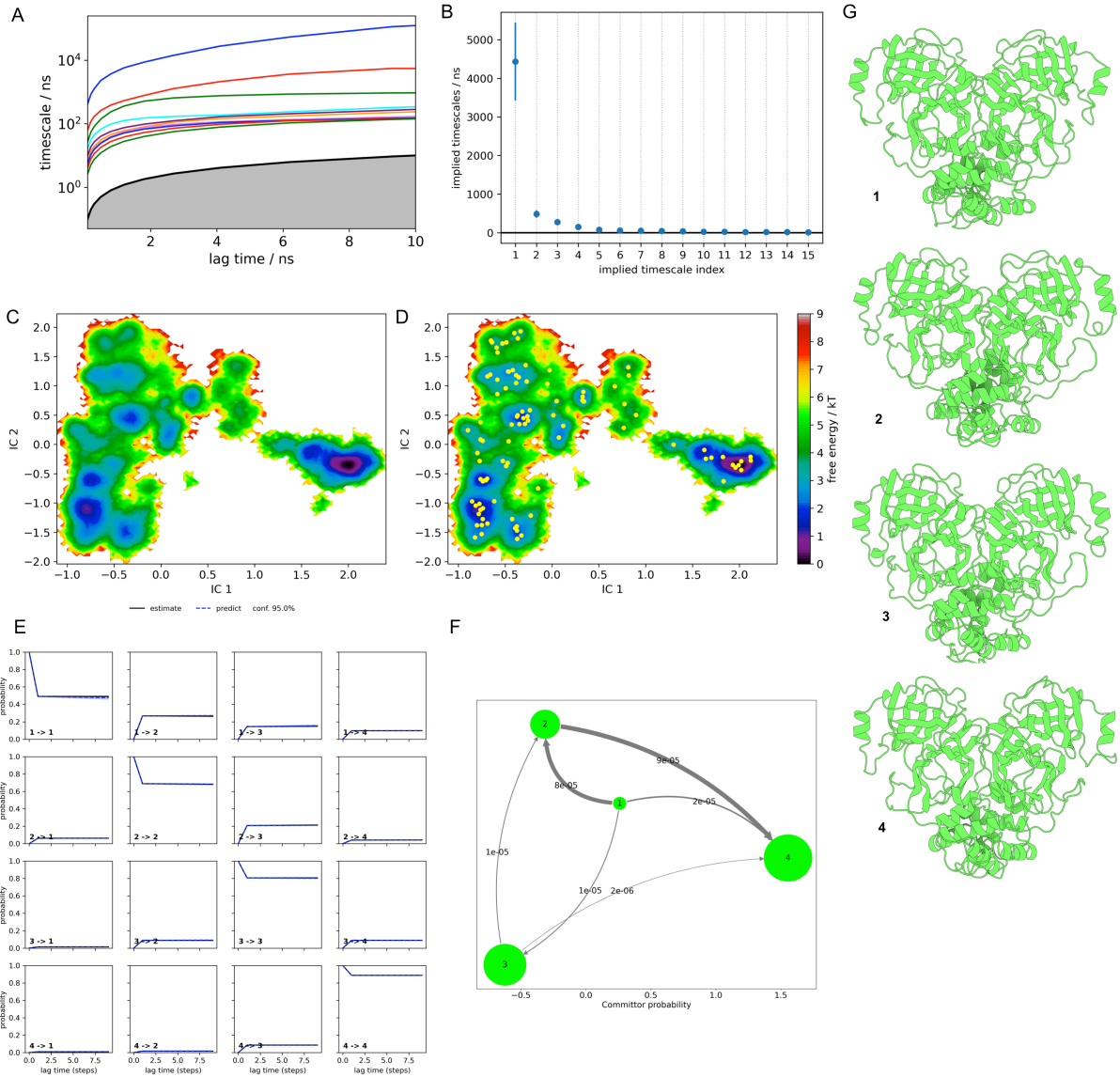
**Figure S4:** Conventional root mean square fluctuation (RMSF) analysis. The RMSF plot was generated after averaging 500 x 50 ns simulations. The RMSF of SARS-Co2 and SARS-CoV systems show similar C $\alpha$  fluctuations.



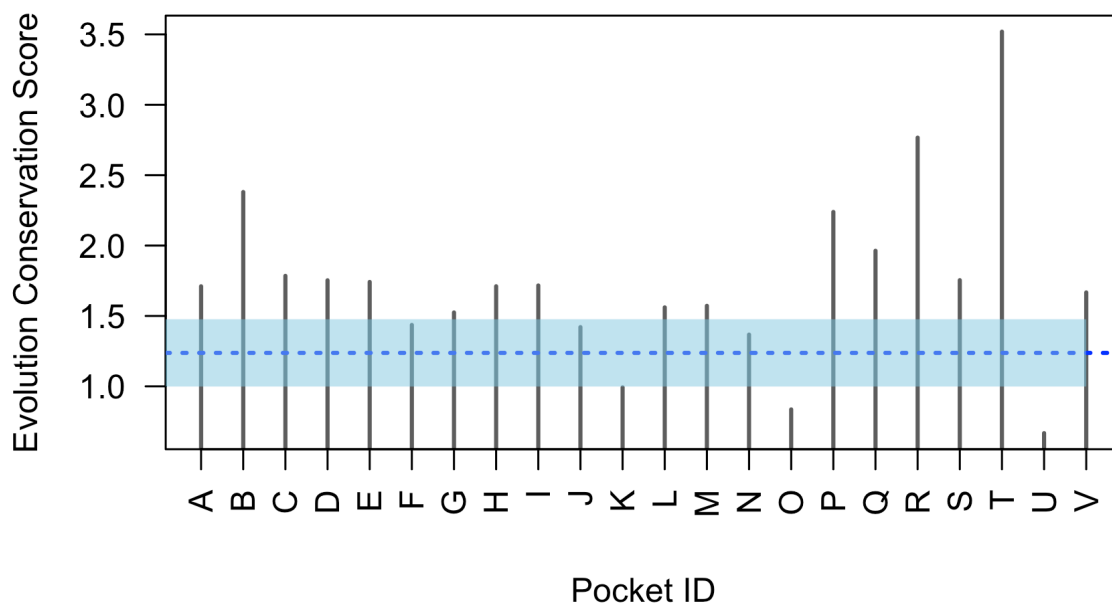
**Figure S5:** Markov State Model of SARS-CoV2 M<sup>pro</sup> enzyme. (A) Implied timescales (B) Ev/ITS index (C) Free energy landscape (D) Cluster centers plotted on the free energy landscape (E) Chapman-Kolmogorov plot (F) Transition path theory analysis. The state number 1 is the state to which all the starting frames of the trajectories have been aligned. We assume this macrostate to be the crystal structure-encompassing state. The size of the circles is proportional to their population, and identified as the dominant state. The fluxes (reactive transition probabilities) are shown as rates (probability per time unit) and shown on the paths. (G) Metastable conformations. The ensemble of backbone geometries contained in each state is illustrated by displaying overlays of the most probably structures of the state (cartoon) on top of samples of the entire state (transparent lines) to highlight both the intrastate conformational variability and the interstate conformational differences. The most probable position of the backbone in each metastable state is illustrated as a cartoon.



**Figure S6:** Markov State Model of SARS-CoV M<sup>PRO</sup> enzyme. (A) Implied timescales (B) Ev/ITS index (C) Free energy landscape (D) Cluster centers plotted on the free energy landscape (E) Chapman-Kolmogorov plot (F) Transition path theory analysis. The state number 1 is the state to which all the starting frames of the trajectories have been aligned. We assume this macrostate to be the crystal structure-encompassing state. The size of the circles is proportional to their population, and identified as the dominant state. The fluxes (reactive transition probabilities) are shown as rates (probability per time unit) and shown on the paths. (G) Metastable conformations. The ensemble of backbone geometries contained in each state is illustrated by displaying overlays of the most probable structures of the state (cartoon) on top of samples of the entire state (transparent lines) to highlight both the intrastate conformational variability and the interstate conformational differences. The most probable position of the backbone in each metastable state is illustrated as a cartoon.

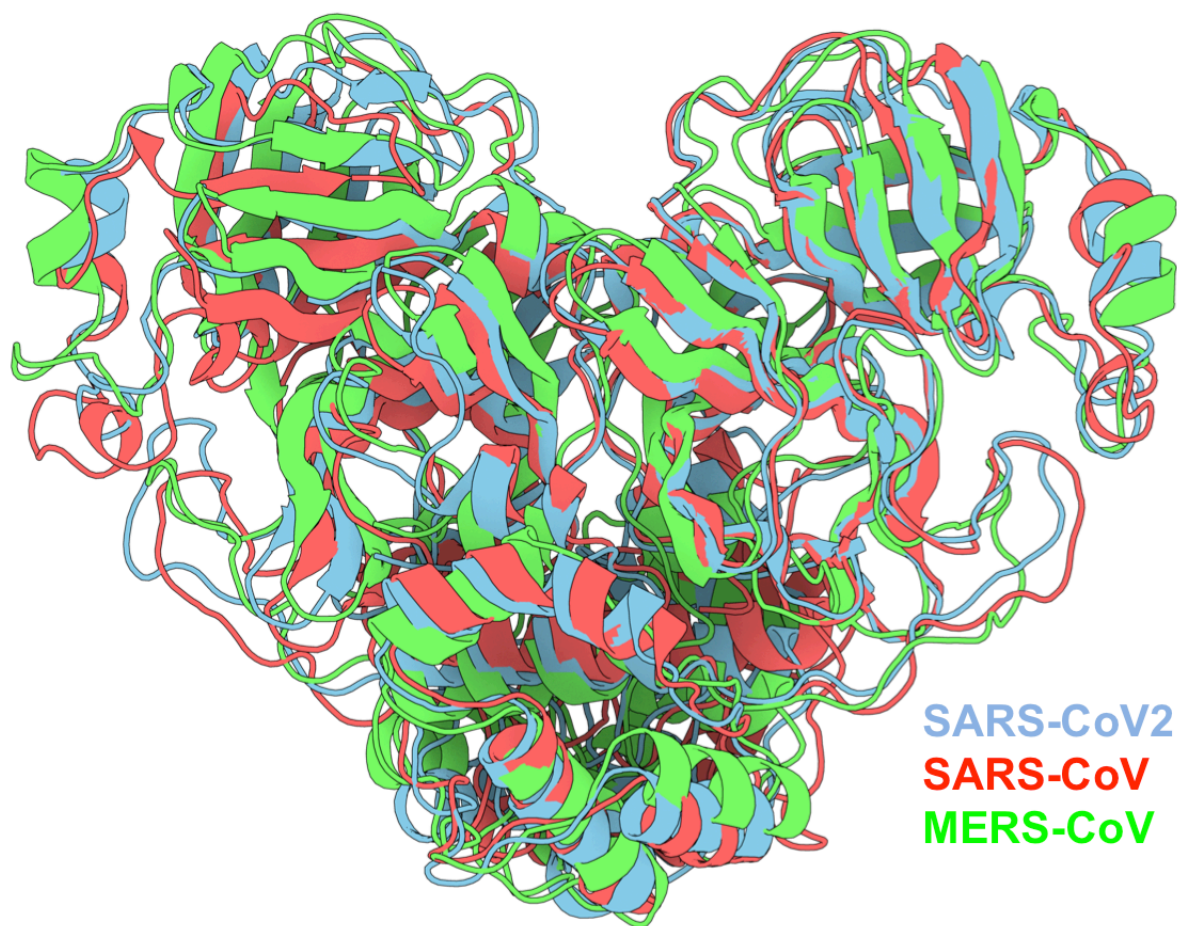


**Figure S7:** Markov State Model of MERS-CoV M<sup>pro</sup> enzyme. (A) Implied timescales (B) Ev/ITS index (C) Free energy landscape (D) Cluster centers plotted on the free energy landscape (E) Chapman-Kolmogorov plot (F) Transition path theory analysis. The state number 1 is the state to which all the starting frames of the trajectories have been aligned. We assume this macrostate to be the crystal structure-encompassing state. The size of the circles is proportional to their population, and identified as the dominant state. The fluxes (reactive transition probabilities) are shown as rates (probability per time unit) and shown on the paths. (G) Metastable conformations. The ensemble of backbone geometries contained in each state is illustrated by displaying overlays of the most probably structures of the state (cartoon) on top of samples of the entire state (transparent lines) to highlight both the intrastate conformational variability and the interstate conformational differences. The most probable position of the backbone in each metastable state is illustrated as a cartoon.

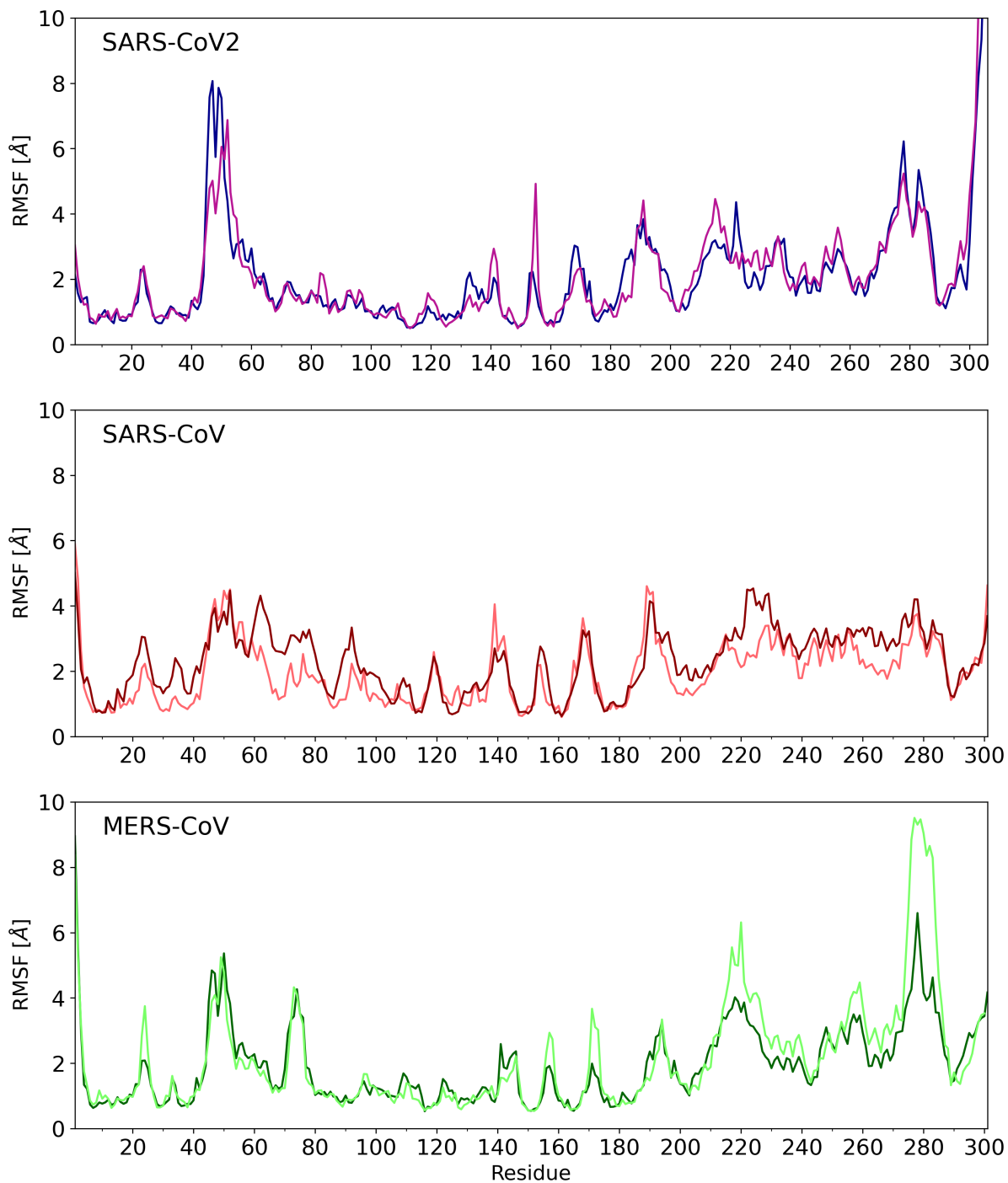


**Figure S8:** Evolution conservation score calculated for each pocket described in Table 1. The score estimates the evolutionary constraints on the pocket as per-residue average of the pairwise correlations. Pairwise correlations are calculated only among pocket residues in the viral MSA (see Methods in the main text). The blue dotted line is the median conservation score calculated from all the surface residue pairs and the blue shaded region highlights the upper and lower limits of the standard deviation.

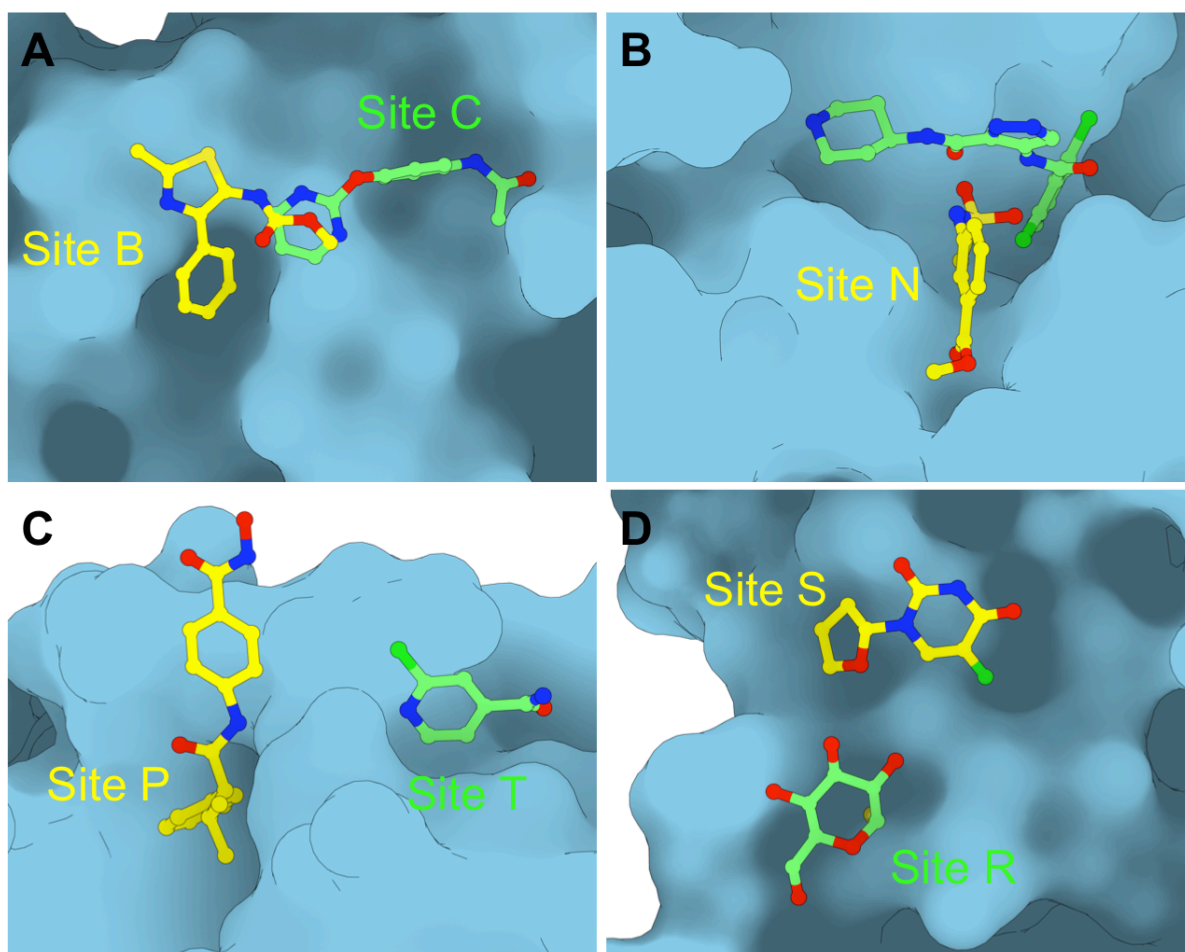




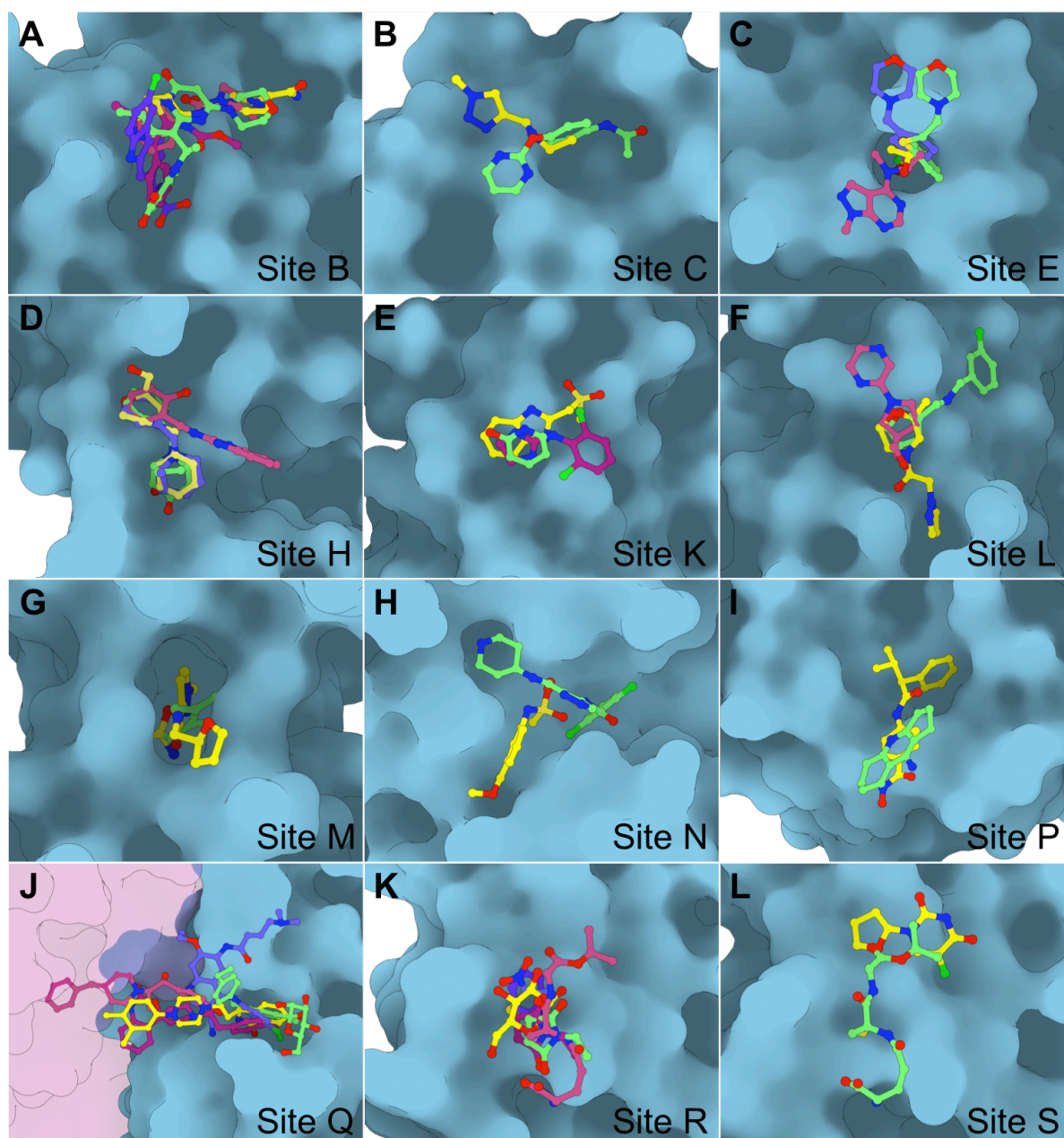
**Figure S9:** Representative structure of the most populated metastable state. State 3 (SARS-CoV2; cyan), state 5 (SARS-CoV; red) and state 4 (MERS-CoV; green) highlight the dynamic regions in different domains.



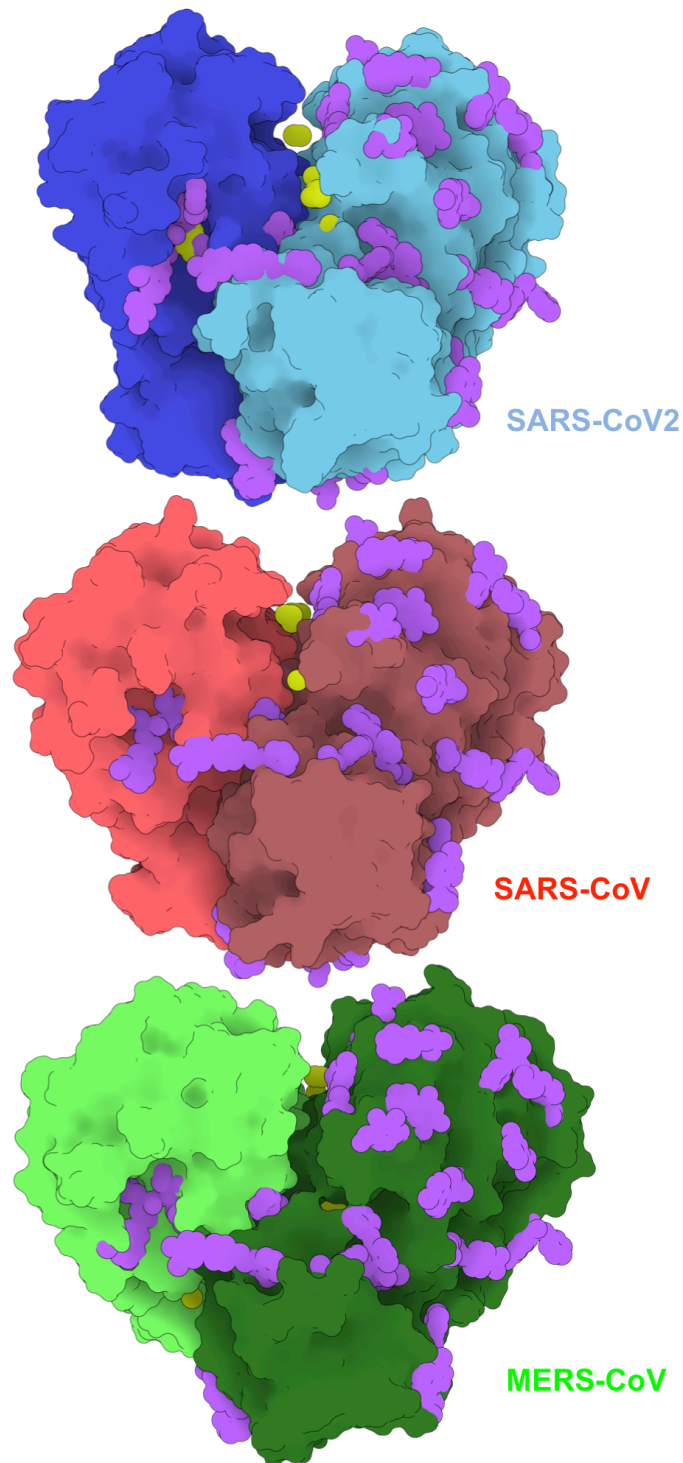
**Figure S10:** Root Mean Squared Fluctuation (RMSF) plots of  $M^{\text{pro}}$  enzymes after 60% fractional alignment. The fluctuations in the both protomers are shown separately.



**Figure S11:** Adjacent binding sites in SARS-CoV2. The ligands binding to these sites can be potentially linked to form novel chemical entities.



**Figure S12:** Multiple ligands have been identified that bind to (A) Site B, (B) Site C, (C) Site E, (D) Site H, (E) Site K, (F) Site L, (G) Site M, (H) Site N, (I) Site P, (J) Site Q, (K) Site R and (L) Site S. Details of the ligands and their representative PDB are tabled in Table S2. Site A has not been illustrated for clarity.



**Figure S13:** Potential binding sites on SARS-CoV2, SARS-CoV and MERS-CoV. Most pockets identified in the apo structure have a ligand bound in it. However, there are some potential binding sites at the interface of the protomers in domain I (yellow spheres). The two protomers are colours differently and the position of the binding sites described in this study is illustrated by surface representation of ligands occupying the binding site (purple). Ligands binding to only one protomer are illustrated for clarity.

## References

- (1) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of M pro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582*, 289–293.
- (2) Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi-Balogh, P.; Brandão-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keserü, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; von Delft, F.; Walsh, M. A. Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat Commun* **2020**, *11*, 5047.
- (3) Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H. M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; Yefanov, O.; Meier, S.; Lorenzen, K.; Krichel, B.; Kopicki, J.-D.; Gelisio, L.; Brehm, W.; Dunkel, I.; Seychell, B.; Gieseler, H.; Norton-Baker, B.; Escudero-Pérez, B.; Domaracky, M.; Saouane, S.; Tolstikova, A.; White, T. A.; Hänle, A.; Groessler, M.; Fleckenstein, H.; Trost, F.; Galchenkova, M.; Gevorkov, Y.; Li, C.; Awel, S.; Peck, A.; Barthelmess, M.; Schlünzen, F.; Xavier, P. L.; Werner, N.; Andaleeb, H.; Ullah, N.; Falke, S.; Srinivasan, V.; Franca, B. A.; Schwinzer, M.; Brognaro, H.; Rogers, C.; Melo, D.; Zaitsev-Doyle, J. J.; Knoska, J.; Peña Murillo, G. E.; Mashhour, A. R.; Guicking, F.; Hennicke, V.; Fischer, P.; Hakanpää, J.; Meyer, J.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Beccari, A. R.; Bourenkov, G.; Stetten, D. von; Pompidor, G.; Bento, I.; Panneerselvam, S.; Karpics, I.; Schneider, T. R.; Garcia Alai, M. M.; Niebling, S.; Günther, C.; Schmidt, C.; Schubert, R.; Han, H.; Boger, J.; Monteiro, D. C. F.; Zhang, L.; Sun, X.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C. G.; Weiss, M. S.; Schulz, E.-C.; Mehrabi, P.; Karničar, K.; Usenik, A.; Loboda, J.; Tidow, H.; Chari, A.; Hilgenfeld, R.; Uetrecht, C.; Cox, R.; Zaliani, A.; Beck, T.; Rarey, M.; Günther, S.; Turk, D.; Hinrichs, W.; Chapman, H. N.; Pearson, A. R.; Betzel, C.; Meents, A. *Inhibition of SARS-CoV-2 Main Protease by Allosteric Drug-Binding*; preprint; Biophysics, 2020.
- (4) Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H. M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; Yefanov, O.; Meier, S.;

- Lorenzen, K.; Krichel, B.; Kopicki, J.-D.; Gelisio, L.; Brehm, W.; Dunkel, I.; Seychell, B.; Gieseler, H.; Norton-Baker, B.; Escudero-Pérez, B.; Domaracky, M.; Saouane, S.; Tolstikova, A.; White, T. A.; Hänle, A.; Groessler, M.; Fleckenstein, H.; Trost, F.; Galchenkova, M.; Gevorgov, Y.; Li, C.; Awel, S.; Peck, A.; Barthelmess, M.; Schlünzen, F.; Lourdu Xavier, P.; Werner, N.; Andaleeb, H.; Ullah, N.; Falke, S.; Srinivasan, V.; França, B. A.; Schwinzer, M.; Brognaro, H.; Rogers, C.; Melo, D.; Zaitseva-Doyle, J. J.; Knoska, J.; Peña-Murillo, G. E.; Mashhour, A. R.; Hennicke, V.; Fischer, P.; Hakanpää, J.; Meyer, J.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Beccari, A. R.; Bourenkov, G.; von Stetten, D.; Pompidor, G.; Bento, I.; Panneerselvam, S.; Karpics, I.; Schneider, T. R.; Garcia-Alai, M. M.; Niebling, S.; Günther, C.; Schmidt, C.; Schubert, R.; Han, H.; Boger, J.; Monteiro, D. C. F.; Zhang, L.; Sun, X.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C. G.; Weiss, M. S.; Schulz, E.-C.; Mehrabi, P.; Karničar, K.; Usenik, A.; Loboda, J.; Tidow, H.; Chari, A.; Hilgenfeld, R.; Utrecht, C.; Cox, R.; Zaliani, A.; Beck, T.; Rarey, M.; Günther, S.; Turk, D.; Hinrichs, W.; Chapman, H. N.; Pearson, A. R.; Betzel, C.; Meents, A. X-Ray Screening Identifies Active Site and Allosteric Inhibitors of SARS-CoV-2 Main Protease. *Science* **2021**, *372*, 642–646.
- (5) Noske, G. D.; Nakamura, A. M.; Gawrijuk, V. O.; Lima, G. M. A.; Zeri, A. C. M.; Nascimento, A. F. Z.; Fernandes, R. S.; Oliva, G.; Godoy, A. S. SARS-CoV-2 Main Protease Immature from -FMAX Library E01 Fragment. **to be published.**
- (6) Noske, G. D.; Nakamura, A. M.; Gawrijuk, V. O.; Lima, G. M. A.; Zeri, A. C. M.; Nascimento, A. F. Z.; Fernandes, R. S.; Oliva, G.; Godoy, A. S. SARS-CoV-2 Main Protease Immature from -FMAX Library E09 Fragment. **to be published.**
- (7) Noske, G. D.; Nakamura, A. M.; Gawrijuk, V. O.; Lima, G. M. A.; Zeri, A. C. M.; Nascimento, A. F. Z.; Fernandes, R. S.; Oliva, G.; Godoy, A. S. SARS-CoV-2 Main Protease Immature from -F2X Entry Library G05 Fragment. **to be published.**
- (8) Bhowmik, D.; Gao, S.; Young, M. T.; Ramanathan, A. Deep Clustering of Protein Folding Simulations. *BMC Bioinformatics* **2018**, *19*, 484.
- (9) Romero, R.; Ramanathan, A.; Yuen, T.; Bhowmik, D.; Mathew, M.; Munshi, L. B.; Javaid, S.; Bloch, M.; Lizneva, D.; Rahimova, A.; Khan, A.; Taneja, C.; Kim, S.-M.; Sun, L.; New, M. I.; Haider, S.; Zaidi, M. Mechanism of Glucocerebrosidase Activation and Dysfunction in Gaucher Disease Unraveled by Molecular Dynamics and Deep Learning. *Proc Natl Acad Sci USA* **2019**, *116*, 5086–5095.

- (10) Casalino, L.; Dommer, A.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S.-H.; Trifan, A.; Brace, A.; Bogetti, A.; Ma, H.; Lee, H.; Turilli, M.; Khalid, S.; Chong, L.; Simmerling, C.; Hardy, D. J.; Maia, J. D. C.; Phillips, J. C.; Kurth, T.; Stern, A.; Huang, L.; McCalpin, J.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. *AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics*; preprint; Biophysics, 2020.