Biophysical *Journal*
Article

SARS-CoV-2 Nsp16 activation mechanism and a cryptic pocket with pan-coronavirus antiviral potential

Neha Vithani,[1,2] Michael D. Ward,[1,2] Maxwell I. Zimmerman,[1,2] Borna Novak,[1,2,3] Jonathan H. Borowsky,[1,2] Sukrit Singh,[1,2] and Gregory R. Bowman[1,2,*]

[1]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri; [2]Center for Science and Engineering of Living Systems, Washington University in St. Louis, St. Louis, Missouri; and [3]Medical Scientist Training Program, Washington University in St. Louis School of Medicine, St. Louis, Missouri

ABSTRACT Coronaviruses have caused multiple epidemics in the past two decades, in addition to the current COVID-19 pandemic that is severely damaging global health and the economy. Coronaviruses employ between 20 and 30 proteins to carry out their viral replication cycle, including infection, immune evasion, and replication. Among these, nonstructural protein 16 (Nsp16), a 2′-O-methyltransferase, plays an essential role in immune evasion. Nsp16 achieves this by mimicking its human homolog, CMTr1, which methylates mRNA to enhance translation efficiency and distinguish self from other. Unlike human CMTr1, Nsp16 requires a binding partner, Nsp10, to activate its enzymatic activity. The requirement of this binding partner presents two questions that we investigate in this manuscript. First, how does Nsp10 activate Nsp16? Although experimentally derived structures of the active Nsp16/Nsp10 complex exist, structures of inactive, monomeric Nsp16 have yet to be solved. Therefore, it is unclear how Nsp10 activates Nsp16. Using over 1 ms of molecular dynamics simulations of both Nsp16 and its complex with Nsp10, we investigate how the presence of Nsp10 shifts Nsp16's conformational ensemble to activate it. Second, guided by this activation mechanism and Markov state models, we investigate whether Nsp16 adopts inactive structures with cryptic pockets that, if targeted with a small molecule, could inhibit Nsp16 by stabilizing its inactive state. After identifying such a pocket in SARS-CoV2 Nsp16, we show that this cryptic pocket also opens in SARS-CoV1 and MERS but not in human CMTr1. Therefore, it may be possible to develop pan-coronavirus antivirals that target this cryptic pocket.

SIGNIFICANCE Coronaviruses are a major threat to human health. These viruses employ molecular machines, called proteins, to infect host cells and replicate. Characterizing the structure and dynamics of these proteins could provide a basis for designing small-molecule antivirals. In this work, we use computer simulations to understand the moving parts of an essential SARS-CoV2 protein, understand how a binding partner turns it on and off, and identify a novel pocket that antivirals could target to shut this protein off. The pocket is also present in other coronaviruses but not in the related human protein, so it could be a valuable target for pan-coronavirus antivirals.

INTRODUCTION

With the coronavirus disease 2019 (COVID-19) pandemic ravaging communities across the globe, there is a massive ongoing effort to understand the molecular machinery of coronaviruses, which may provide insight into therapeutic opportunities (1–3). The severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) virus responsible for COVID-19 has infected over 60 million and killed over 1.5 million people globally to date (4). Additionally, coronaviruses have caused several past epidemics, including severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) which had fatality rates of ∼10 and ∼34%, respectively (5,6). Therefore, there is likely to be evolution and outbreaks of additional zoonotic coronaviruses in the future (7). Although vaccine trials for COVID-19 are successfully wrapping up, there are still no approved antivirals that reduce mortality to coronavirus

---

Submitted December 11, 2020, and accepted for publication March 25, 2021.

*Correspondence: g.bowman@wustl.edu

Neha Vithani and Michael D. Ward contributed equally to this work.

Editor: Alan Grossfield.

https://doi.org/10.1016/j.bpj.2021.03.024

infections (8–10). Taken together, there is strong incentive to understand the fundamental mechanisms of how these coronaviruses operate in hopes of discovering effective therapeutics. Biophysical studies can provide these details, and a tremendous amount of biophysical work has already been done to understand the virus' 29 proteins. So far, the spike protein, positioned on the outside of the viral envelope, has proven to be a good vaccine candidate (11). Beyond the spike, the 16 "nonstructural" (i.e., accessory) proteins carry out the majority of the virus' essential processes, making them good targets for antiviral therapeutics (12,13).

Among the nonstructural proteins (Nsps), Nsp16 is particularly important to the viral replication cycle because it is essential to coronavirus' immune evasion (14–16). Nsp16 is a 2′-O-methyltransferase (2′-O-MTase) that forms part of the replication-transcription complex (17). It mimics the human protein Cap-specific mRNA (nucleoside-2′-O-)-methyltransferase (CMTr1) to perform a crucial step in capping transcribed mRNA (18). Specifically, Nsp16 facilitates the transfer of a methyl group from its S-adenosylmethionine (SAM) cofactor to the 2′ hydroxyl of ribose sugar of viral mRNA (18,19). This methylation both improves translation efficiency and camouflages the mRNA so that it is not recognized by intracellular pathogen recognition receptors such as IFIT and RIG-I (15,20). Importantly, inhibiting or knocking out 2′-O-MTase activity severely attenuates viral replication and infectivity of coronaviruses (13,20). Thus, developing small-molecule inhibitors of Nsp16 is a promising therapeutic strategy.

Interestingly, whereas all other 2′-O-MTases (eukaryotic and viral) are active as monomers, Nsp16 requires a binding partner, Nsp10, to be active (16–18,21–23). In fact, Nsp16 does not even bind its ligands (SAM and RNA) in the absence of Nsp10. In the experimentally derived structures of the Nsp16/Nsp10 complex, Nsp10 does not form any direct interaction with either ligand (Fig. 1 *A*), suggesting that Nsp10 may allosterically regulate Nsp16 to enable substrate binding (18,19,24–27). Given that there is significant structural variation in the RNA-binding loops of different crystal structures of Nsp16 (Fig. 1 *B*) and structures of monomeric Nsp16 have not been solved, we hypothesized that Nsp16 is highly dynamic in solution, and Nsp10 acts by stabilizing the active state. In contrast, we anticipate that human CMTr1 would be less dynamic because it does not require a binding partner for substrate binding and has been crystalized in its monomeric state. Often, dynamics of proteins reveal allosteric pockets that remain hidden in their crystal structures (i.e., cryptic pockets). If monomeric Nsp16 is more dynamic than CMTr1, it may adopt inactive configurations that reveal allosteric cryptic pockets that can be targeted by small-molecule inhibitors for its selective inhibition.

Here, we use computer simulations to understand the activation mechanism of Nsp16 and identify cryptic pockets that may be valuable antiviral targets. Active site inhibitors, such as sinefungin, have been shown to outcompete SAM binding and render Nsp16 catalytically inactive (28,29). However, there are more than 200 human proteins with known or putative methyltransferase activity that use SAM as a cofactor (30). Therefore, it may be difficult to design antivirals that target the SAM (or RNA) binding sites of Nsp16 without eliciting off-target effects by also binding human methyltransferases. For example, sinefungin has been shown to occupy the SAM-binding pocket of human

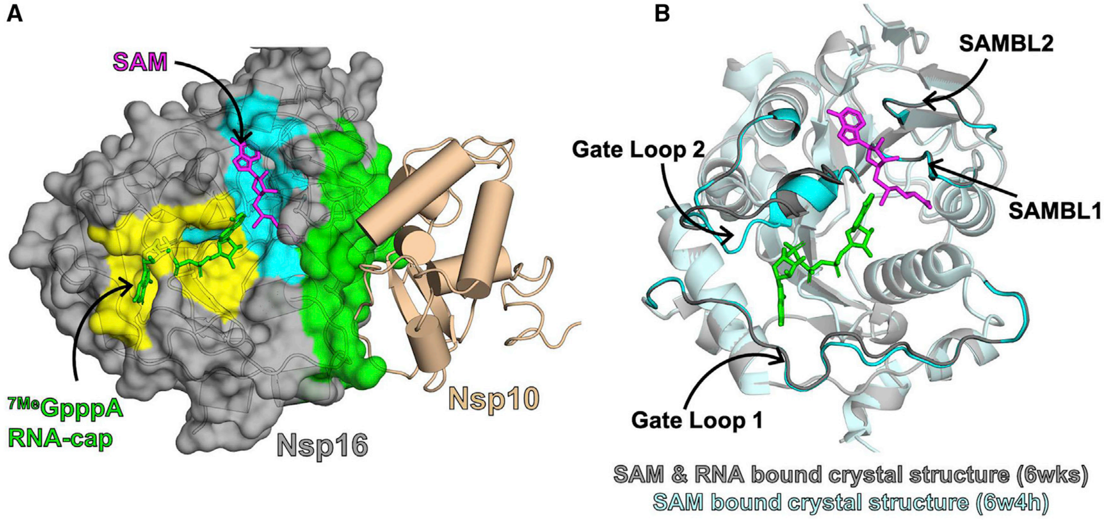


FIGURE 1 Substrate-binding pockets and Nsp10 binding interface of Nsp16 observed in the crystal structure of the Nsp16/Nsp10 complex (PDB: 6wks). (*A*) Surface representation of Nsp16 showing the SAM-binding pocket (*cyan*), RNA-binding pocket (*yellow*), and Nsp10-binding interface (*green*). (*B*) Overlay of Nsp16 structures from structures of the Nsp16/Nsp10 complex with RNA (PDB: 6wks, shown in *gray*) and without RNA (PDB: 6w4h, shown in *cyan*), showing structural heterogeneity in the RNA binding site. Gate loop 1 and gate loop 2 of the RNA-binding pocket, and SAM-binding loop 1 (SAMBL1) and SAM-binding loop 2 (SAMBL2) lining the SAM-binding pocket are highlighted. To see this figure in color, go online.

N7 methyltransferase in a crystal structure (Protein Data Bank, PDB: 3epp). Targeting the Nsp16/Nsp10 interface could be an alternative means to selectively inhibit Nsp16 because CMTr1 lacks a homologous binding partner. Toward this, peptide-based inhibitors that mimic Nsp10 to compete for interactions at the Nsp10/Nsp16 interface have been shown to inhibit Nsp16 activity (31,32). Although this approach seems promising, peptide-based inhibitors face challenges, including limited stability and shelf-life, the possibility of adverse immunogenic responses, and the high cost of production (33). To expand the therapeutic opportunities, we search for other ways to inactivate Nsp16. First, we compare the structure and dynamics of SARS-CoV-2 Nsp16 in the presence and absence of Nsp10 to understand Nsp16's activation. Specifically, we use over 1 ms of molecular dynamics simulation data (2) to characterize how Nsp10 binding shifts Nsp16's conformational ensemble to activate Nsp16. After showing that the resulting model is consistent with a variety of experimental observations, we use it to hunt for cryptic pockets that may provide a means to inhibit Nsp16. Finally, we extend our simulations to SARS-CoV-1, MERS, and human CMTr1 to determine whether targeting such a pocket could provide an opportunity to develop pan-coronavirus antivirals.

METHODS

System preparation

The systems were prepared starting from crystal structures for PDB: 6w4h, 3r24, 5ynf, and 4n49, for SARS-CoV-2, SARS-CoV-1, MERS, and CMTr1, respectively. All ligands, solutes, and water molecules from the crystal structures were removed. For monomeric Nsp16 simulations, Nsp10 was also removed. In the coronavirus homologs, two zinc ions were retained, and the coordinating residues were modified accordingly (CYS- > CYM and HIS- > HID). Missing residues in the crystal structure of CMTr1 were modeled using the Modeler package (34). All systems were solvated in TIP3P water (35) in a rhombic dodecahedral box with periodic boundary conditions and Na^+ and Cl^- ions added to neutralize the system. Systems were then energy minimized with a steepest descent algorithm until the maximal force fell below 100 kJ/mol/nm using a step size of 0.01 nm and a cutoff distance of 1.2 nm for the neighbor list, Coulomb interactions, and van der Waals interactions.

Systems were equilibrated for 1.0 ns in NPT simulations, with all bonds constrained using the LINCS algorithm (36), and virtual sites were used to allow a 4 fs time step. Cutoffs of 1.1 nm were used for the neighbor list with 0.9 for Coulomb and van der Waals interactions. The particle-mesh Ewald method (37) was employed for treatment of long-range interactions with a Fourier spacing of 0.12 nm. The Verlet cutoff scheme was used for the neighbor list. Berendsen barostat was used to control the pressure during the equilibration (38). The stochastic velocity rescaling (v-rescale) thermostat was used to control the temperature at 300 K (39).

Adaptive sampling simulations

The FAST algorithm (40,41) was employed for all four homologs for a total of five FAST simulations (SARS-CoV-2 FAST simulations were performed on both monomeric Nsp16 and the Nsp10/Nsp16 complex). FAST was used here to generally enhance conformational sampling and also to quickly explore cryptic pockets. The procedure for FAST simulations is as follows: 1) run initial simulations, 2) build a Markov state model (MSM), 3) rank states based on FAST ranking, 4) restart simulations from the top-ranked states, and 5) repeat steps 2–4 until ranking is optimized. For each system, MSMs were generated after each round of sampling using a k-centers clustering algorithm based on the root mean-squared deviation (RMSD) between select atoms. Clustering continued until the maximal distance of a frame to a cluster center fell within a predefined cutoff. In addition to the FAST ranking, a similarity penalty was added to promote conformational diversity in starting structures, as has been described previously (42).

For SARS-CoV-2 monomeric Nsp16 and Nsp16/Nsp10, the simulation data were generated in a previous manuscript published by our group. Briefly, FAST-pocket simulations were run at 300 K for six rounds, with 10 simulations per round, in which each simulation was 40 ns in length (2.4 μs aggregate simulation for each system). The FAST-pocket ranking function favored restarting simulations from states with large pocket openings. Pocket volumes were calculated using the LIGSITE algorithm (43). From these simulations, a conformationally diverse set of structures was selected to be run on Folding@home, based on the k-centers clustering algorithm mentioned above. A total of 283 and 770 μs of aggregate simulation time was collected for the Nsp10/Nsp16 complex and monomeric Nsp16, respectively.

FAST-distance simulations were used for SARS-CoV-1 Nsp16, MERS Nsp16, and CMTr1 to sample the $\beta3$–$\beta4$ pocket identified from SARS-CoV-2 simulations. FAST-distance simulations were run at 300 K for 15 rounds, with 10 simulations per round, in which each simulation was 40 ns in length (6.0 μs aggregate simulation for each system). The FAST-distance ranking favored stated with greater distances between the α-carbons of $\beta3$ and $\beta4$.

DiffNets

We used DiffNets, a deep learning-based dimensionality reduction algorithm developed by our group, to highlight biochemically relevant differences between data sets (44). We trained a DiffNet to compare and contrast structure ensembles of monomeric Nsp16 and the Nsp16/Nsp10 complex to find features that discriminate them, highlighting the structural determinants of Nsp16 activation. First, we subsampled the data by a factor of 25 and 68 for the Nsp16/Nsp10 complex and monomeric Nsp16 data, respectively, to have an equal amount of data. We then converted simulation data to DiffNet input, following the data normalization procedure from the original manuscript. Briefly, *XYZ* atom coordinates from simulations were mean-shifted to zero and then multiplied by the inverse of the square root of a covariance matrix that was calculated from simulations. For all DiffNet training and analysis, we used a split architecture (as described previously), in which the classification task was focused on all atoms within 1 nm of SAM or RNA cap based on the 6wks crystal structure. This atom selection was chosen to guide DiffNets to find differences in the active site region of Nsp16, which is inherently linked to its activation. For training, simulation frames are classified as "Nsp16 inactive" or "Nsp16 active," based on initial classification labels of 0 (i.e., Nsp16 inactive) for all monomeric Nsp16 frames and labels of 1 (i.e., Nsp16 active) for all frames from the Nsp10/Nsp16 complex. These labels were iteratively updated in a self-supervised manner described in the original manuscript, in which we choose expectation maximization bounds of [0.1–0.4] for monomeric Nsp16 and [0.6–0.9] for the Nsp10/Nsp16 complex. This allows for more coherent classification labels because monomeric Nsp16 may sometimes adopt structural poses associated with Nsp16 activation and vice versa for the Nsp10–Nsp16 complex. Additionally, we used 30 latent variables, 10 training epochs in which we subsampled the data by a factor of 10 in each epoch, a batch size of 32, and a learning rate of 0.0001.

To analyze the DiffNet output, we calculated 10 representative structures that span from "Nsp16 inactive" states to "Nsp16 active" states (i.e., structures with classification labels spanning 0–1). After training, the DiffNet learns a low-dimensional representation of each simulation frame (i.e., a latent vector) and outputs a classification label for every simulation frame.

We binned the structures into 10 equally spaced bins based on their classification labels, which span from 0 to 1. We then calculated the mean latent vector for each bin and used the DiffNet to reconstruct a structure based on each latent vector. These structures were used as representative structures for each bin. All training and analysis were performed using the open-source package https://github.com/bowman-lab/diffnets.

Markov state models

An MSM is a statistical framework for analyzing molecular dynamics simulations that provides a network representation of a free-energy landscape (45–47). To quantify cryptic pocket opening across the homologs and changes between monomeric Nsp16 and the Nsp10/Nsp16 complex, we performed several measurements that rely on MSMs that are built based on the simulation data. We built a separate MSM for each system, using all simulation data available for that system. All MSMs were constructed with the Enspara python package (48). First, the solvent-accessible surface area (SASA) of each residue side chain was calculated using the Shrake-Rupley algorithm (49) implemented in MDTraj (50) using a drug-sized probe (2.8 A sphere).

We then clustered the data using a hybrid clustering algorithm. First, we used a k-centers algorithm (51,52) to cluster the data. Next, we applied sweeps of k-medoids update steps (three for SARS-CoV-2 data, two for other homologs) that refined the cluster centers to be in the densest regions of conformational space (53). We clustered the simulation data based on the residue-level SASA. For SARS-CoV-2 Nsp16 and Nsp10/Nsp16 complex for which we had massive data sets from Folding@home (283 and 770 μs), we used 5000 cluster centers. For SARS-CoV-1 Nsp16, MERS Nsp16, and human CMTr1 (6 μs of FAST-adaptive sampling data per system), we used 1500 cluster centers. We validated that these produced Markovian models by plotting the implied timescales, from which we chose a lag time of 5 ns (see Fig. S1). To further ensure robustness of the MSMs, we also built models based on alternative clusterings and confirmed that they gave similar results. Specifically, for SARS-CoV-2 Nsp16 and Nsp10/Nsp16 complex, we built MSMs using 1) 5.2 nm^2 cluster-radius cutoff and 2) 5.5 nm^2 cluster radius. For SARS-CoV-1 Nsp16, MERS Nsp16, and human CMTr1, we built MSMs using 1) 4.0 nm^2 cluster radius and 2) 4.5 nm^2 cluster radius. All MSMs were Markovian (see Fig. S1). Moreover, we used these MSMs to recreate the distributions in Figs. 2, 3, and 4 and we find that the results are robust across all MSMs (see Figs. S2–S4). A Markov time of 5 ns was selected based on the implied timescales to build an MSM for each homolog. To build the MSMs, transition probability matrices were produced by counting transitions between states (i.e., clusters), adding a prior count of $\frac{1}{N_{states}}$ and row normalizing, as is described previously (54). Equilibrium populations were calculated as the eigenvector of the transition probability matrix with an eigenvalue of one. For all histograms shown, we calculated the order parameter of distance (e.g., distance between β3 and β4) using cluster centers (i.e., representative structure of the cluster) and weighted the order parameter by the corresponding equilibrium population calculated with the MSM. We also resampled the equilibrium populations 100 times by bootstrapping the MSM, which provided error bars for computing the fraction of SAM- and RNA-compatible states adopted by monomeric Nsp16 and the Nsp16/10 complex.

Distance and SASA calculations

Figs. 2, 3, and 4 include distance and SASA measurements that are explained in more detail here. In Fig. 2 we measure the distance between gate loop 1 and gate loop 2 as the distance between Gln28 and Lys141 because these residues are known to undergo significant changes for RNA binding. We measure the distance between SAM-binding loop 2 and gate loop 2 as the average distance between (Met131, Tyr132, Asp133, Pro134) and (Asp99, Leu100, Asn101, Asp102) because these are key residues that cradle SAM in the bound state. All SASA measurements are performed using Ala79, Thr82, Ala83, Leu86, Thr93, Leu94, Leu95, Val96, Asp97, Ala98, and Asp99 because these are the main components that gets exposed during cryptic pocket opening.

Cryptic pocket detection

Cryptic pockets in SARS-CoV-2 Nsp16 were identified using our previously established approach, called Exposons analysis (55). This analysis was performed using the cluster centers and the equilibrium probabilities derived from the MSMs built on the residue-level SASA described above. The center of each cluster was taken as an exemplar of that conformational state, and residues were classified as exposed if their SASA exceeded 2.0 A^2 and buried otherwise. The mutual information between the exposure/burial of each residue-pair was calculated based on the MSM by treating the SASA values in the cluster centers as samples and weighting them by the equilibrium probability of the representative state. The mutual information was computed using the following equation:

$$MI(X,\ Y) = \sum_{x \in X}\sum_{y \in Y} p(x,y) log\left(\frac{p(x,y)}{p(x)p(y)}\right).$$

Finally, cryptic pockets (Exposons) were identified as groups of residues undergoing cooperative change in SASA by clustering the matrix of pairwise mutual information using affinity propagation.

The β3–β4 cryptic pocket identified in SARS-CoV-2 Nsp16 consists of residues Ala79, Thr82, Ala83, Leu86, Thr93, Leu94, Leu95, Val96, Asp97, Ala98, and Asp99. Total SASA of these residues/homologous residues was measured for detecting cryptic pocket opening in all homologs of Nsp16 (SARS-CoV-2, SARS-CoV-1, and MERS). For measuring equivalent cryptic pocket in CMTR1, total SASA of structurally homologous residues (Gly141, Ser144, Glu145, Val148, Ala155, Lys156, Gly157, His158, Gly159, Met160, and Thr161) was calculated.

Sequence conservation

Protein sequences of Nsp16 from SARS-CoV-2 (YP_009725311.1), SARS-CoV-1 (UniProt: P0C6X7), MERS (UniProt: K0BWD0), NL63 (GenBank: AFD64750.1), HKU1 (NCBI Reference Sequence: YP_460023.1), Turkey coronavirus (CoV) (NCBI Reference Sequence: YP_001941189.1), Bat CoV (NCBI Reference Sequence: YP_008439226.1), and murine hepatitis virus (NCBI Reference Sequence: YP_209243.1) were used for multiple sequence alignment. Sequence alignment was performed on the Clustal Omega server (56). Sequence alignment was visualized, and the sequence conservation score was generated using Jalview 2 software (57).

For sequence comparison of SARS-CoV-2, SARS-CoV-1, MERS, and human CMTr1, shown in Fig. 4, structure-based sequence alignment was performed using UCSF Chimera package (58). For the structure-based sequence alignment, we first aligned the structures of these homologs PDB: 6wks (SARS-CoV-2), PDB: 3r24 (SARS-CoV-1), PDB: 5ynf (MERS), and PDB: 4n49 (CMTr1). The sequences were then aligned based on the structural alignment of the backbone atoms.

RESULTS AND DISCUSSION

Nsp10 promotes opening of Nsp16's SAM- and RNA-binding pockets

Although experimental studies have demonstrated that Nsp16 requires Nsp10 to be functionally active, the structural determinants of Nsp16's activation remain unknown (17,18,23). Chen et al. proposed that Nsp10's stimulatory effects are rooted in its ability to assist Nsp16 in binding

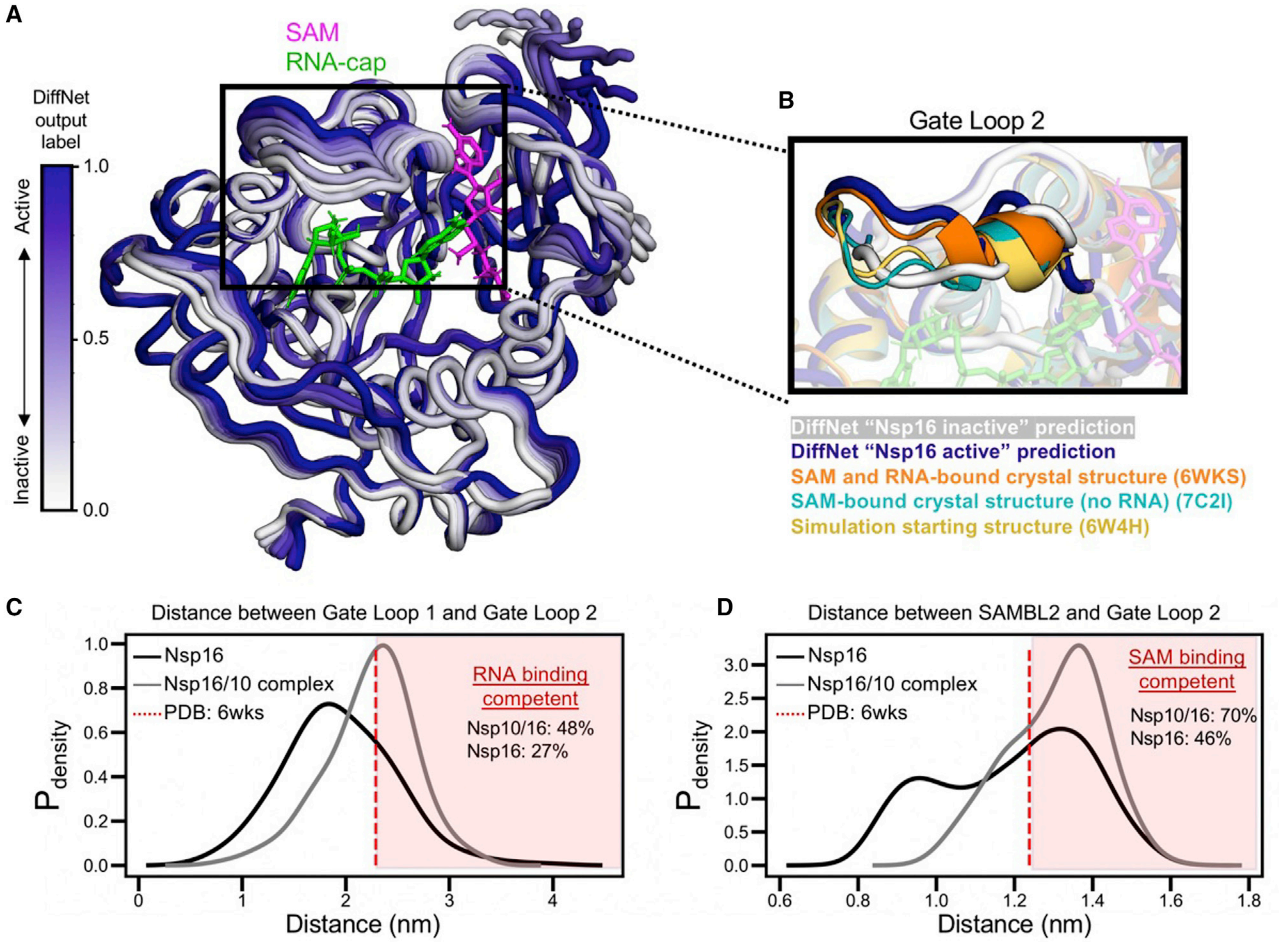


FIGURE 2 Nsp10 binding shifts Nsp16's conformational ensemble, increasing its propensity to adopt structural states that are ligand binding compatible. (*A*) 10 structures of Nsp16 that represent the DiffNet prediction changing from inactive to active (*white* to *purple*). The DiffNet output label scales from 0 to 1 (*white* to *purple*) reflecting the extent the DiffNet predicts a structure to be associated with Nsp16 activation. (*B*) Comparison of the DiffNet-predicted active and inactive states (*purple* plus *white*, respectively) to the starting simulation state (*yellow*), a known SAM- and RNA-bound structural state (*orange*), and a known SAM- (but not RNA-) bound state (*teal*). All structures are aligned to 6wks (*orange*). (*C*) Probability-weighted distance distribution between RNA-binding gate loops 1 and 2, comparing monomeric Nsp16 (*black*) to the Nsp10-Nsp16 complex (*gray*). (*D*) Probability-weighted distance distribution between SAM-binding loop 2 and gate loop 2, comparing monomeric Nsp16 (*black*) to the Nsp10-Nsp16 complex (*gray*). For (*C*) and (*D*), the distance for a SAM- and RNA-bound crystal structure is also plotted (*red dotted line*). To see this figure in color, go online.

SAM and RNA, which is supported by data showing that Nsp16 alone cannot bind SAM or RNA (18). They also propose that Nsp10 manages this by stabilizing or changing the conformation of the SAM-binding pocket based on the fact that Nsp10 contacts SAM-binding loops in their crystal structure (and numerous other structures). However, without assessing the Nsp10-Nsp16 complex's dynamics and comparing it to monomeric Nsp16, this hypothesis is left wanting. It has also been proposed that Nsp10 assists in RNA binding by directly contacting RNA (59). However, a recent crystal structure with RNA bound (PDB: 7jyy) contains a stretch of nucleotides long enough to contact Nsp10, but the RNA curls off into solution instead of interacting with Nsp10. Another recent study compared an RNA- and SAM-bound Nsp10/16 complex structure to one with only SAM bound and found a major opening of RNA-binding gate loops, suggesting that the dynamics of these loops might be important for Nsp16 activation (25). However, it is not clear whether Nsp10 plays a role in those dynamics. Altogether, there is strong evidence that Nsp10 modulates Nsp16's structure and dynamics to assist it in binding SAM and RNA, but the mechanism of these structural changes is unclear.

To explore how Nsp10 activates Nsp16, we analyzed simulations of Nsp16 in the presence and absence of Nsp10 using DiffNets. Recently, our group combined the sampling powers of the FAST-pockets adaptive sampling algorithm (40) and the computational resources of Folding@home to accumulate more than 1 ms of simulation data between simulations of monomeric Nsp16 and the Nsp16/Nsp10 complex (see Methods; (2)). Here, we compare these simulations using a deep-learning-based dimensionality

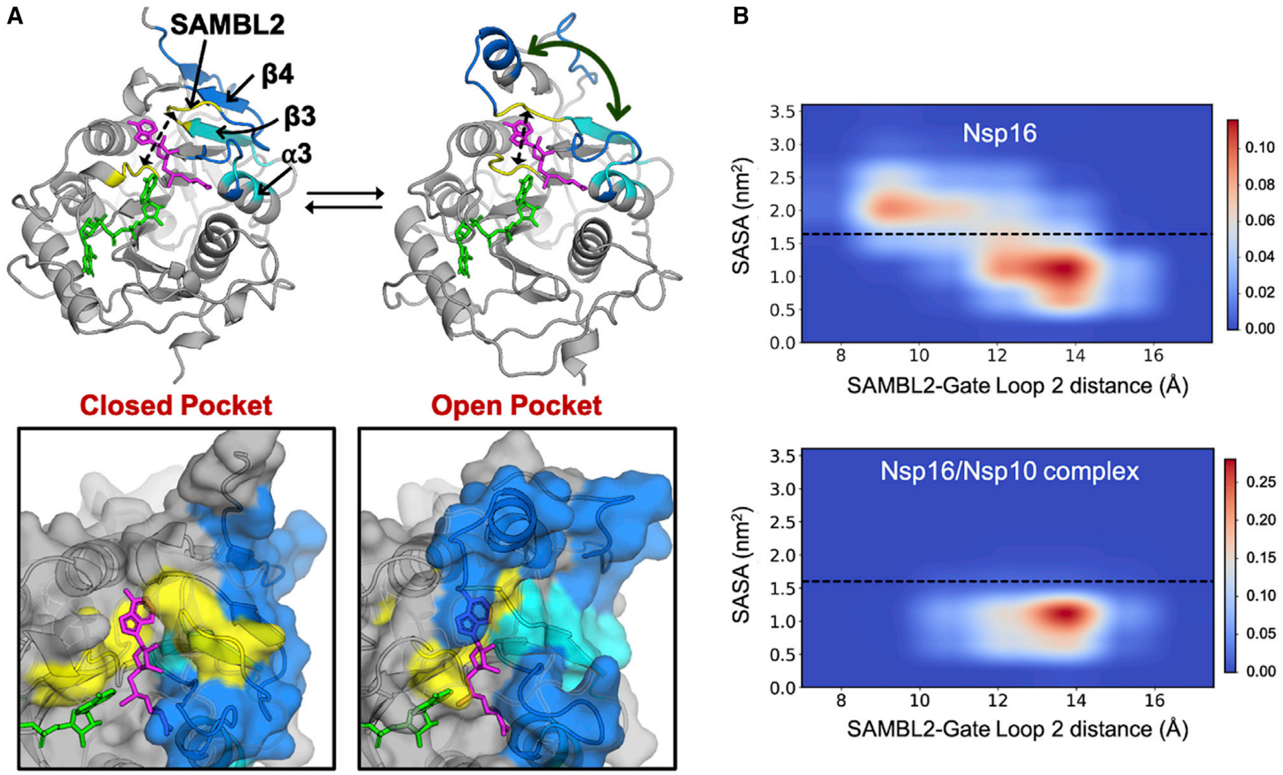


FIGURE 3 Cryptic pocket opening in SARS-CoV-2 Nsp16. (*A*) Structural states with the cryptic pocket closed and open. The insets show surface views of the closed and open pocket. Residues exposed upon pocket opening are shown in cyan, and the regions undergoing the opening motion are shown in blue. Collapse of the SAM-binding pocket is measured as the distance between SAMBL2 and gate loop 2, shown in yellow. (*B*) Equilibrium probability-weighted two-dimensional histograms of solvent-accessible surface area (SASA) of pocket residues (shown in *cyan* in (*A*)) and the distance between SAMBL2 and gate loop 2 in Nsp16 for monomeric Nsp16 (*upper panel*) and the Nsp16/Nsp10 complex (*lower panel*). The black dotted line separates the pocket closed and open states in Nsp16. The Equilibrium probability-weighted distribution scales from 0 to 0.11 (*blue* to *red*) for Nsp16 monomer, and from 0 to 0.26 (*blue* to *red*) for Nsp16/Nsp10 complex. To see this figure in color, go online.

reduction algorithm called DiffNets (44). DiffNets has been shown to accurately capture the structural determinants of biochemical differences between protein variants. Although we are not considering protein variants, our problem is similar because Nsp16 has different biochemical properties when in the presence/absence of Nsp10 (i.e., active/inactive). Therefore, we trained a DiffNet to learn the structural determinants of Nsp16 activation by learning differences between Nsp16's ensemble when in the presence and absence of Nsp10. For each simulation frame, the DiffNet learns a low-dimensional projection of the protein structure and classifies the structure with a label between 0 and 1 that indicates the likelihood that the structure is associated with Nsp16 being active.

Analysis of the DiffNet suggests that Nsp10 shifts Nsp16's conformational ensemble to stabilize more open SAM- and RNA-binding pockets. Using the DiffNet classification labels, we identified 10 structures that are representative of the progression from Nsp16 inactive states to active states (see Methods; Fig. 2). We noticed that RNA gate loop 2 moves away from RNA gate loop 1, making for a more open RNA-binding pocket in active states compared with inactive states (Fig. 2 *A*). Additionally, the SAM-binding pocket also opens up in the active states relative to the inactive states. RNA-binding gate loop 2 and SAM-binding loop 2 move away from each other in the active state, which widens the pocket, creating space for SAM. (Fig. 2 *A*). Strikingly, the structure associated with the highest label (i.e., most strongly associated with Nsp16 activation) matches well to a recently solved crystal structure that is bound to both RNA and SAM (Fig. 2 *B*; (25)). Specifically, when we align the predicted active structure to 6wks, then measure the RMSD of gate loop 2, we find a deviation on par with the typical resolution of crystal structures (1.40 Å). When we perform this calculation for the predicted inactive structure, the RMSD is much higher (3.24 Å). The predicted inactive structure adopts a more collapsed gate loop 2, similar to known structures with being SAM but not RNA bound (i.e., PDB: 6w4h and 7c2i; see Fig. 2 *B*; (26)). This result implies that the DiffNet learned that Nsp10 activates Nsp16, in part, by rearranging the RNA gate loop into an RNA-binding-competent pose. Though it is known that this RNA gate loop needs to open to bind RNA, this is the first evidence, to our knowledge, to suggest that Nsp10

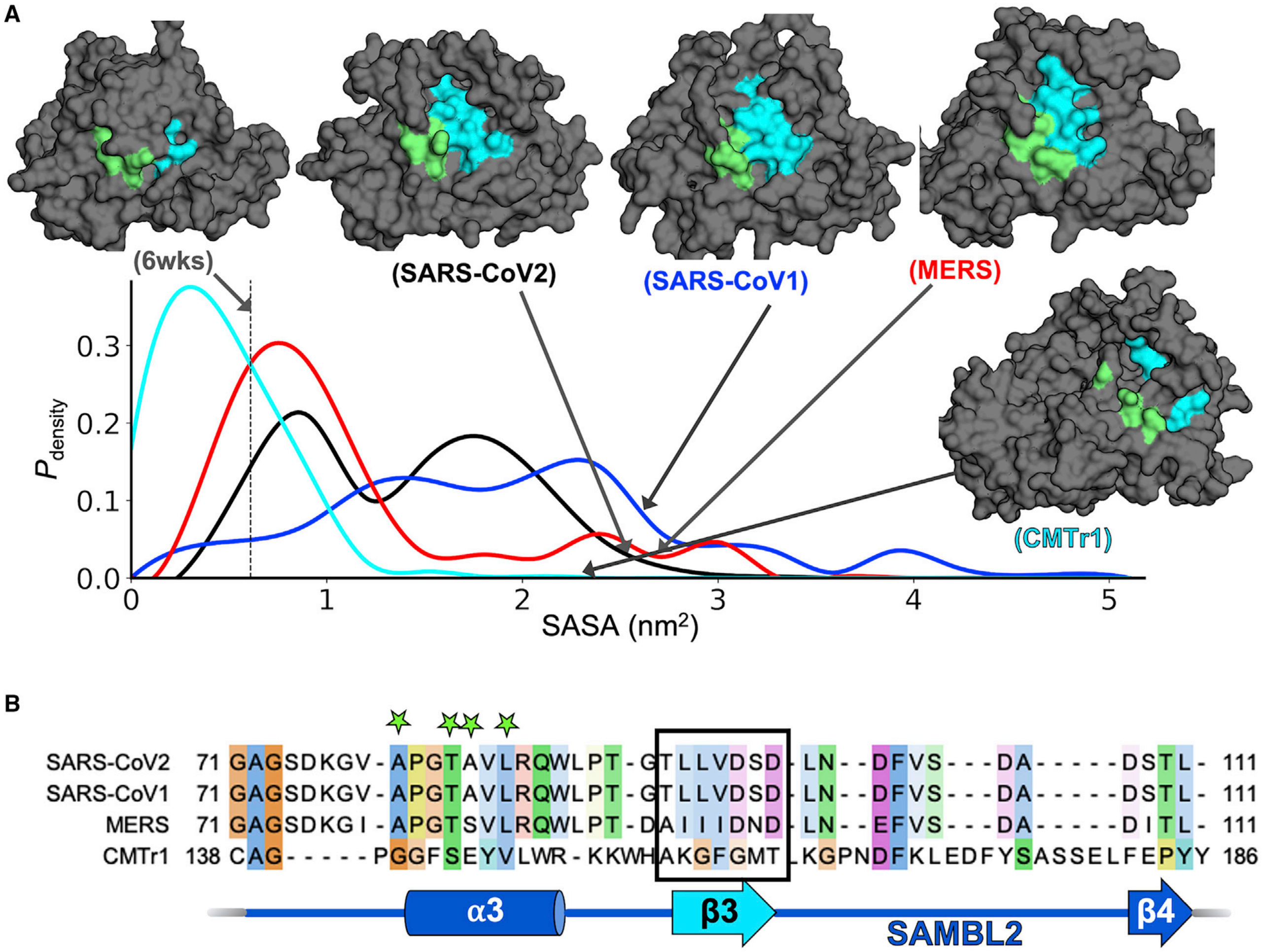


FIGURE 4 Comparison of cryptic pocket opening in Nsp16 homologs and human CMTr1. (*A*) Equilibrium probability-weighted distribution of the solvent exposure of pocket-forming residues for SARS-CoV-2 (*black*), SARS-CoV-1 (*blue*), MERS (*red*), and CMTr1 (*cyan*). Structures representing the open pocket are shown for each homolog, with $\beta 3$ colored in cyan and other pocket-forming residues from $\alpha 3$ colored in green. Black dotted line depicts SASA of pocket residues in the crystal structure of Nsp16/Nsp10 complex (PDB: 6wks). (*B*) Structure-based sequence alignment of Nsp16 homologs (SARS-CoV-2, SARS-CoV-1, and MERS) and human CMTr1 is shown for the cryptic pocket-forming regions. Residues of $\beta 3$ are marked inside the black-colored box, and other pocket-forming residues from $\alpha 3$ are marked by green-colored stars. To see this figure in color, go online.

may activate Nsp16 through increasing its propensity to form a more open RNA-binding pocket. Altogether, these results suggest that Nsp10's presence increases the propensity for both SAM- and RNA-binding pockets to be open.

To quantify the effect of Nsp10 on the SAM- and RNA-binding pockets, we built MSMs for both the complex and monomeric Nsp16. MSMs are a statistical framework for analyzing molecular dynamics simulation data that provide (among other things) a discrete map of structural configurations, an equilibrium population value that corresponds to the proportion of time a protein spends in any given configuration, and the probability of transitioning between any pair of configurations (45). We constructed MSMs for Nsp16 simulations both in the presence and absence of Nsp10.

Our MSMs reveal that Nsp10 binding stabilizes open structures of both the SAM- and RNA-binding pockets that are competent to bind their respective substrates. We first found that the presence of Nsp10 results in a substantial reduction of flexibility in important binding components, including both SAM-binding loops and RNA gate loops (see Fig. S5). This result is somewhat surprising because gate loop 2, which contacts both SAM and RNA, is not in direct contact with Nsp10, suggesting strong allosteric communication. Next, we calculated the distribution of distances for opening and closing of the SAM- and RNA-binding pockets (Fig. 2, *C* and *D*). From these histograms, it is clear that both of these binding pockets have an increased propensity to open when Nsp10 is present. We considered pockets as SAM/RNA-binding competent when the distance between loops in a pocket is at least as open as in the crystal structure that binds both ligands (PDB: 6wks). From this analysis, Nsp16 adopts binding-competent states with higher probability when Nsp10 is present versus when Nsp10 is absent for both SAM (0.70 ± 0.04 vs. 0.46 ± 0.04) and RNA

(0.48 ± 0.04 vs. 0.27 ± 0.03). Altogether, our data suggest that Nsp10 aids SAM and RNA binding by preventing the collapse of SAM- and RNA-binding gate loops. Our analysis also provides structural snapshots of what inactive states look like, which may be useful in targeting Nsp16 with therapeutics.

A cryptic pocket in Nsp16 is a potential therapeutic target

A traditional approach to drug development involves molecules designed to target binding cavities observed in singular structural snapshots of a protein, but this approach often misses "cryptic" pockets that can form in proteins because of thermal fluctuations. Often times, the active site of an enzyme is targeted for drug development to design an inhibitor that can outcompete substrate binding. However, active sites are often conserved among functional homologs. In the case of Nsp16, its human homolog (CMTr1) shares the same overall fold and binds the same substrates. Although there are significant sequence and structural differences in the active site, specificity may be more easily achieved by targeting a less functionally relevant region of the protein. Cryptic pockets can provide both a new target for drug development and the potential to achieve specificity. For example, cryptic pockets that remain closed and invisible in the crystal structure but open in solution because of thermal fluctuations (55) can present unique potential binding sites due to differences in the dynamics of subsets of homologs (e.g., open in coronavirus homologs, but closed in human CMTr1). Therefore, it may be easier to achieve specificity by targeting a cryptic pocket. Importantly, the cryptic pocket must communicate with functional sites to be an effective therapeutic target. Here, we explore whether Nsp16 contains any cryptic pockets that, when open, would stabilize the inactive state identified with DiffNets.

To find cryptic pockets, we applied Exposons, an algorithm (55) that identifies residues with cooperative changes in solvent exposure, to Nsp16 simulation data. Using this method, we found that residues in the β3-strand and α3-helix transition between closed states and open states (i.e., low to high solvent-accessible surface area) (Fig. 3 *A*). Specifically, the β4-strand curls up to form an α-helical structure, which results in surface exposure of β3 and residues from α3 (Fig. 3 *A*). The opening motion of β4 shifts the adjacent SAM-binding loop (SAMBL) 2 against gate loop 2 to collapse the SAM-binding pocket in a closed conformation (Fig. 3, *A* and *B*). This agrees with the DiffNet prediction that the β4-strand moving away from β3 is associated with inactivation (see Fig. S6). Furthermore, several residues forming this cryptic pocket directly contact Nsp10 in crystal structures of the Nsp16/Nsp10 complex (see Fig. S7). The β3–β4 pocket opening displaces these Nsp10 binding residues, which could inhibit Nsp16's association with Nsp10 (see Fig. S7). The Nsp16/Nsp10 binding interface has also been targeted with peptide-based inhibitor design (31,32). Although this flat surface may be amenable to peptide inhibitors, it is a challenging target for small molecules. In contrast, the concave shape of the cryptic pocket identified in this work presents a more viable target for small-molecule inhibitors. Finally, we find that this open pocket structure is commonly visited as part of monomeric Nsp16's conformational ensemble, as measured with MSM equilibrium populations (Fig. 3 *B*). Taken together, we propose that targeting the β3–β4 pocket with a small molecule could inhibit Nsp16's activity by preventing SAM binding or preventing association with Nsp10.

Conservation of the cryptic pocket in Nsp16 makes it a promising target for broad-spectrum inhibitors

To explore the possibility of targeting the cryptic pocket for broad-spectrum inhibition of coronaviruses, we evaluated the conservation of cryptic pocket opening in Nsp16 homologs. Ideally, a therapeutic developed to treat SARS-CoV-2 would also work against other coronaviruses, such as MERS and SARS-CoV-1, and potential future outbreaks. Additionally, the therapeutic target should be sufficiently dissimilar from human CMTr1 such that it would not cause unwanted, off-target effects. Although we identified a promising cryptic pocket in SARS-CoV-2, we wanted to investigate if this pocket is specific to SARS-CoV-2, or specific to coronaviruses in general, or if it is common across homologs, including CMTr1.

First, we analyze cryptic pocket conservation by comparing sequence features and structural features based on the native, folded state. We find that the β3–β4 pocket residues are 100% conserved between SARS-CoV-2 and SARS-CoV-1 (Fig. 4 *B*). Additionally, of the 11 residues that form the pocket, there are only two nonconservative mutations between SARS-CoV-2 and MERS. Based on the sequence similarity, we expect that, if the cryptic pocket forms in all homologs, it may be possible to develop small-molecule therapeutics that targets all three. Furthermore, we find substantial sequence differences between SARS-CoV-2 and CMTr1. 8 out of the 11 pocket residues are nonconservative mutations relative to SARS-CoV-2. Based on sequence differences alone, we reason that selective inhibition could be achieved even if the cryptic pocket is adopted by CMTr1. Moreover, the sequences and structure of SARS-CoV-2 Nsp16 and human CMTr1 are sufficiently different in the β3–β4 pocket region that the human protein may not even have the cryptic pocket (Fig. S8). Based on these sequence and structural differences, combined with the lack of requirement of a stabilizing binding partner, we hypothesized that cryptic pocket opening is not likely to be conserved in CMTr1.

To explore cryptic pocket opening across homologs, we performed FAST-pocket simulations of monomeric Nsp16 for SARS-CoV-1 and MERS, as well as for human CMTr1. We then built an MSM for each homolog and measured the opening of the $\beta3$–$\beta4$ pocket by measuring the equilibrium-weighted solvent exposure of the pocket residues we previously used to define the pocket (i.e., Fig. 3). In these simulations, we find that the $\beta3$–$\beta4$ pocket opens with high probability in both SARS-CoV-1 and MERS Nsp16 (Fig. 4). The timescales for transitioning between the open and closed states of the pocket are given in Table S1. Encouragingly, we find that the $\beta3$–$\beta4$ pocket has a substantially lower probability of opening in CMTr1. Taken together, features of the $\beta3$–$\beta4$ cryptic pocket in coronavirus homologs of Nsp16 appear sufficiently similar to each other and dissimilar to CMTr1 to make for a promising target for pan-coronavirus inhibitors.

CONCLUSIONS

Our work provides mechanistic insight into how Nsp16 is activated and reveals a new opportunity for inhibiting this essential viral component that could provide a target for pan-coronavirus antivirals. First, we elucidate the activation mechanism of Nsp16 by comparing its dynamics in the presence and absence of its activator, Nsp10. Our results are consistent with previous experimental findings that Nsp16 cannot bind its substrates SAM or RNA in the absence of Nsp10 (18). We provide a structural rationale for this observation by elucidating the structural dynamics of Nsp16 in its monomeric state, which has remained inaccessible to experimental studies, and comparing it to the structural dynamics of the Nsp16/Nsp10 complex. Here, we find that Nsp10 activates Nsp16 by opening its SAM- and RNA-binding loops, allowing them to accommodate their respective ligands. Guided by this activation mechanism, we identify structural states of Nsp16 that are incompatible with substrate binding and also contain potential drug binding sites. Specifically, we find a pocket formed between $\beta3$ and $\beta4$ of Nsp16 that collapses the SAM-binding pocket when open. The region of the pocket has an overlap with where Nsp10 binds to Nsp16, so targeting this cryptic pocket could inhibit both substrate (SAM) and Nsp10 binding. Therefore, this cryptic site is a promising target for small-molecule inhibitor development. Furthermore, we find that this cryptic pocket is conserved in MERS and SARS-CoV-1 Nsp16 but not in the human homolog CMTr1, suggesting its potential for development of a pan-coronavirus, broad-spectrum inhibitor that may be efficacious against COVID-19 and yet unseen coronavirus outbreaks.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.03.024.

AUTHOR CONTRIBUTIONS

N.V., M.D.W., S.S., and G.R.B. designed the research. N.V., M.D.W., M.I.Z., B.N., J.H.B., and S.S. performed research. N.V., M.D.W., J.H.B., and G.R.B. wrote the manuscript.

ACKNOWLEDGMENTS

We are grateful to the citizen scientists who contribute to Folding@home by running simulations on their personal computers. We are grateful to Catherine Knoverek for her thoughtful feedback when editing initial drafts of the writing.

This work was funded by National Science Foundation CAREER Award MCB-1552471, National Science Foundation RAPID 58628, and National Institutes of Health grants R01 GM124007 and RF1AG067194. G.R.B. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and a Packard Fellowship for Science and Engineering from The David & Lucile Packard Foundation.

REFERENCES

1. Zhou, P., X. L. Yang, ..., Z. L. Shi. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 579:270–273.
2. Zimmerman, M. I., J. R. Porter, ..., G. R. Bowman. 2020. SARS-CoV-2 simulations go exascale to capture spike opening and reveal cryptic pockets across the proteome. *bioRxiv* https://doi.org/10.1101/2020.06.27.175430.
3. Wu, A., Y. Peng, ..., T. Jiang. 2020. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe.* 27:325–328.
4. Johns Hopkins Coronavirus Resource Center. 2020. COVID-19 Map. Johns Hopkins University & Medicine, https://coronavirus.jhu.edu/map.html<span class="role">web.
5. Chan-Yeung, M., and R. H. Xu. 2003. SARS: epidemiology. *Respirology.* 8 (*Suppl*):S9–S14.
6. Zaki, A. M., S. van Boheemen, ..., R. A. M. Fouchier. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367:1814–1820.
7. Chan, J. F.-W., K. K.-W. To, ..., K.-Y. Yuen. 2013. Interspecies transmission and emergence of novel viruses: lessons from bats and birds. *Trends Microbiol.* 21:544–555.
8. Belete, T. M. 2020. A review on Promising vaccine development progress for COVID-19 disease. *Vacunas.* 21:121–128.
9. Callaway, E. 2020. COVID vaccine excitement builds as Moderna reports third positive result. *Nature.* 587:337–338.
10. Jackson, L. A., E. J. Anderson, ..., J. H. Beigel; mRNA-1273 Study Group. 2020. An mRNA vaccine against SARS-CoV-2 - preliminary report. *N. Engl. J. Med.* 383:1920–1931.
11. Huang, Y., C. Yang, ..., S.-W. Liu. 2020. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* 41:1141–1149.
12. da Silva, S. J. R., C. T. Alves da Silva, ..., L. Pena. 2020. Role of nonstructural proteins in the pathogenesis of SARS-CoV-2. *J. Med. Virol.* 92:1427–1429.
13. Snijder, E. J., E. Decroly, and J. Ziebuhr. 2016. The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv. Virus Res.* 96:59–126.
14. Ramanathan, A., G. B. Robb, and S. H. Chan. 2016. mRNA capping: biological functions and applications. *Nucleic Acids Res.* 44:7511–7526.

15. Daffis, S., K. J. Szretter, …, M. S. Diamond. 2010. 2′-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature.* 468:452–456.

16. Decroly, E., I. Imbert, …, B. Canard. 2008. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2’O)-methyltransferase activity. *J. Virol.* 82:8071–8084.

17. Sawicki, S. G., D. L. Sawicki, …, S. G. Siddell. 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. *PLoS Pathog.* 1:e39.

18. Chen, Y., C. Su, …, D. Guo. 2011. Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2′-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog.* 7:e1002294.

19. Decroly, E., C. Debarnot, …, B. Canard. 2011. Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2′-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog.* 7:e1002059.

20. Menachery, V. D., K. Debbink, and R. S. Baric. 2014. Coronavirus non-structural protein 16: evasion, attenuation, and possible treatments. *Virus Res.* 194:191–199.

21. Smietanski, M., M. Werner, …, J. M. Bujnicki. 2014. Structural analysis of human 2′-O-ribose methyltransferases involved in mRNA cap structure formation. *Nat. Commun.* 5:3004.

22. Hodel, A. E., P. D. Gershon, and F. A. Quiocho. 1998. Structural basis for sequence-nonspecific recognition of 5′-capped mRNA by a cap-modifying enzyme. *Mol. Cell.* 1:443–447.

23. Bouvet, M., C. Debarnot, …, E. Decroly. 2010. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog.* 6:e1000863.

24. Rosas-Lemus, M., G. Minasov, …, K. Satchell. 2020. The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv* https://doi.org/10.1101/2020.04.17.047498.

25. Viswanathan, T., S. Arya, …, Y. K. Gupta. 2020. Structural basis of RNA cap modification by SARS-CoV-2. *Nat. Commun.* 11:3718.

26. Lin, S., H. Chen, …, G. Lu. 2020. Crystal structure of SARS-CoV-2 nsp10/nsp16 2′-O-methylase and its implication on antiviral drug design. *Signal Transduct. Target. Ther.* 5:131.

27. Debarnot, C., I. Imbert, …, B. Canard. 2011. Crystallization and diffraction analysis of the SARS coronavirus nsp10-nsp16 complex. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 67:404–408.

28. Krafcikova, P., J. Silhan, …, E. Boura. 2020. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat. Commun.* 11:3717.

29. Khan, R. J., R. K. Jha, …, A. K. Singh. 2020. Targeting SARS-CoV-2: a systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2′-O-ribose methyltransferase. *J. Biomol. Struct. Dyn* Published online April 20, 2020. https://doi.org/10.1080/07391102.2020.1753577.

30. Petrossian, T. C., and S. G. Clarke. 2011. Uncovering the human methyltransferasome. *Mol. Cell. Proteomics.* 10:M110.000976.

31. Wang, Y., Y. Sun, …, D. Guo. 2015. Coronavirus nsp10/nsp16 methyltransferase can Be targeted by nsp10-derived peptide in vitro and in vivo to reduce replication and pathogenesis. *J. Virol.* 89:8416–8427.

32. Ke, M., Y. Chen, …, D. Guo. 2012. Short peptides derived from the interaction domain of SARS coronavirus nonstructural protein nsp10 can suppress the 2′-O-methyltransferase activity of nsp10/nsp16 complex. *Virus Res.* 167:322–328.

33. Lee, A. C.-L., J. L. Harris, …, J. H. Hong. 2019. A comprehensive review on current advances in peptide drug development and design. *Int. J. Mol. Sci.* 20:2383.

34. Šali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.

35. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

36. Hess, B. 2008. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4:116–122.

37. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

38. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

39. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.

40. Zimmerman, M. I., and G. R. Bowman. 2015. FAST conformational searches by balancing exploration/exploitation trade-offs. *J. Chem. Theory Comput.* 11:5747–5757.

41. Zimmerman, M. I., and G. R. Bowman. 2016. How to run FAST simulations. *Methods Enzymol.* 578:213–225.

42. Zimmerman, M. I., K. M. Hart, …, G. R. Bowman. 2017. Prediction of new stabilizing mutations based on mechanistic insights from Markov state models. *ACS Cent. Sci.* 3:1311–1321.

43. Hendlich, M., F. Rippmann, and G. Barnickel. 1997. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* 15:359–363, 389.

44. Ward, M. D., M. I. Zimmerman, …, G. R. Bowman. 2020. DiffNets: self-supervised deep learning to identify the mechanistic basis for biochemical differences between protein variants. *bioRxiv* https://doi.org/10.1101/2020.07.01.182725.

45. Bowman, G. R., V. S. Pande, and F. Noé. 2014. An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. Springer, Dordrecht, the Netherlands.

46. Chodera, J. D., and F. Noé. 2014. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* 25:135–144.

47. Schütte, C., and M. Sarich. 2015. A critical appraisal of Markov state models. *Eur. Phys. J. Spec. Top.* 224:2445–2462.

48. Porter, J. R., M. I. Zimmerman, and G. R. Bowman. 2019. Enspara: modeling molecular ensembles with scalable data structures and parallel computing. *J. Chem. Phys.* 150:044108.

49. Shrake, A., and J. A. Rupley. 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79:351–371.

50. McGibbon, R. T., K. A. Beauchamp, …, V. S. Pande. 2015. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109:1528–1532.

51. Gonzalez, T. F. 1985. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.* 38:293–306.

52. Bowman, G. R., X. Huang, and V. S. Pande. 2009. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods.* 49:197–201.

53. Gentle, J. E., L. Kaufman, and P. J. Rousseuw. 1991. Finding groups in data: an introduction to cluster analysis. *Biometrics.* 47:i-iv.

54. Zimmerman, M. I., J. R. Porter, …, G. R. Bowman. 2018. Choice of adaptive sampling strategy impacts state discovery, transition probabilities, and the apparent mechanism of conformational changes. *J. Chem. Theory Comput.* 14:5459–5475.

55. Porter, J. R., K. E. Moeder, …, G. R. Bowman. 2019. Cooperative changes in solvent exposure identify cryptic pockets, switches, and allosteric coupling. *Biophys. J.* 116:818–830.

56. Madeira, F., Y. M. Park, …, R. Lopez. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47:W636–W641.

57. Waterhouse, A. M., J. B. Procter, …, G. J. Barton. 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 25:1189–1191.

58. Pettersen, E. F., T. D. Goddard, …, T. E. Ferrin. 2004. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.

59. Joseph, J. S., K. S. Saikatendu, …, P. Kuhn. 2006. Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *J. Virol.* 80:7894–7901.

Biophysical Journal, Volume *120*

Supplemental information

SARS-CoV-2 Nsp16 activation mechanism and a cryptic pocket with pan-coronavirus antiviral potential

Neha Vithani, Michael D. Ward, Maxwell I. Zimmerman, Borna Novak, Jonathan H. Borowsky, Sukrit Singh, and Gregory R. Bowman

Figure S1.

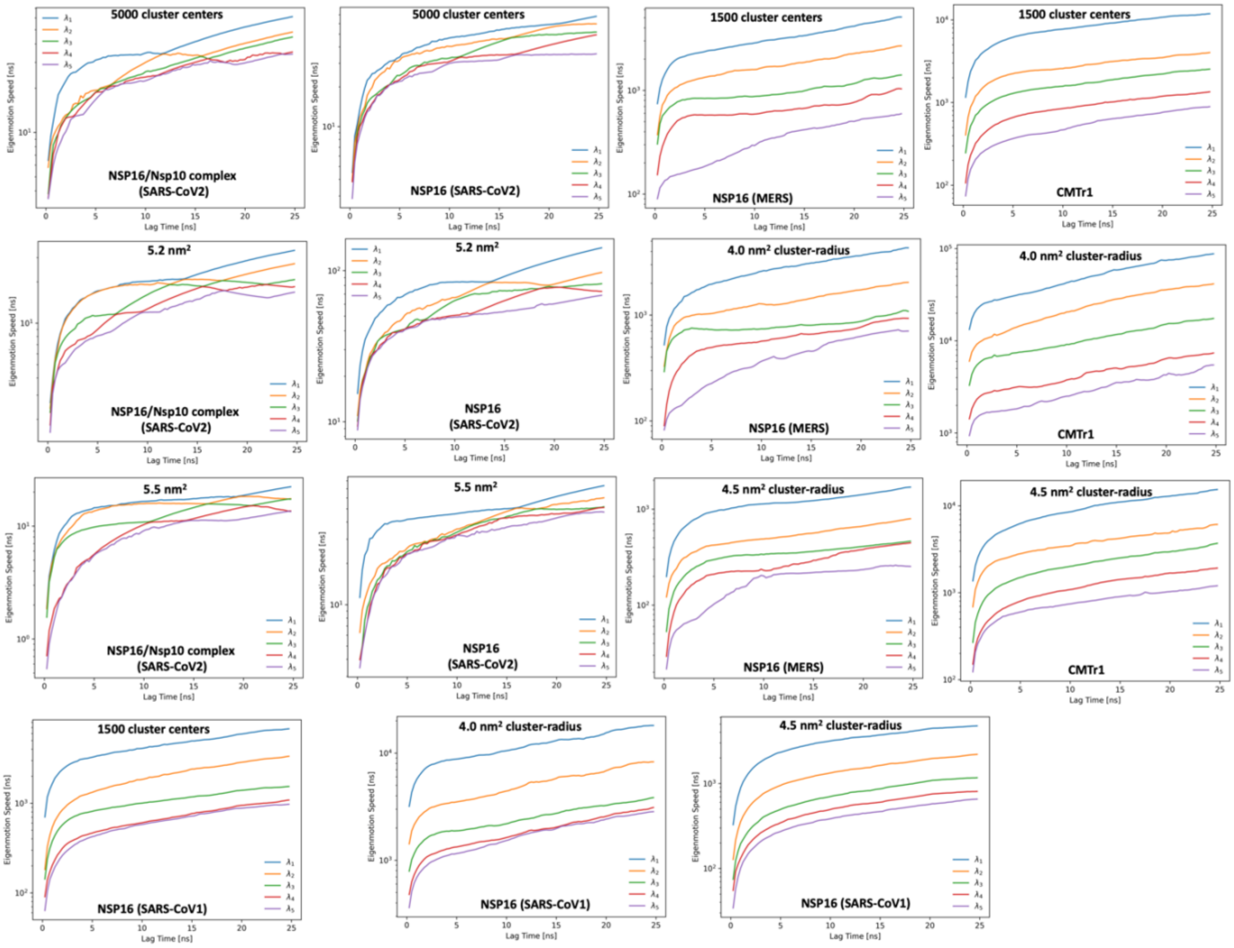


Figure S1. Implied timescales plotted as a function of the lag time for MSMs for Nsp16/Nsp10 complex (SARS-CoV2), Nsp16 homologs (SARS-CoV2, SARS-CoV1, & MERS) and human CMTr1 for different clustering cut-offs. 5 ns lag-time was selected to build the final MSMs used in this study.

Figure S2.

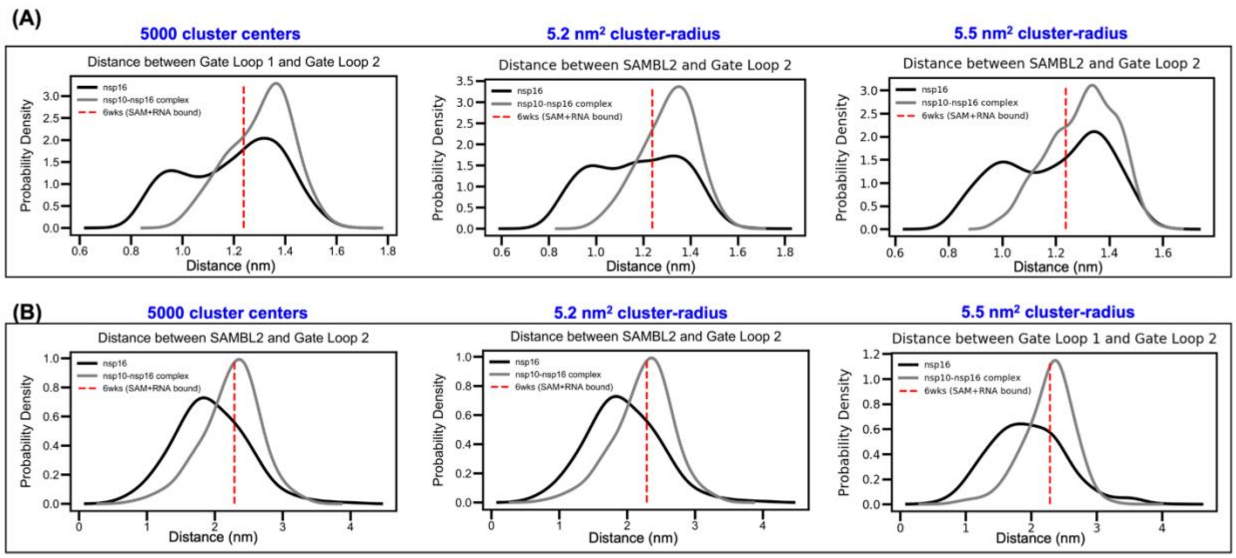


Figure S2. (A) Probability-weighted distance distribution between RNA-binding gate loops 1 and 2 comparing monomeric Nsp16 (black) to the Nsp10-Nsp16 complex (gray) are shown for three different clustering cut-offs. (B) Probability-weighted distance distribution between SAM-binding loop 2 and gate loop 2, comparing monomeric Nsp16 (black) to the Nsp10-Nsp16 complex (gray) are shown for three different clustering cut-offs. The distance for a SAM and RNA bound crystal structure is also plotted (red dotted line) in all figures.

Figure S3.

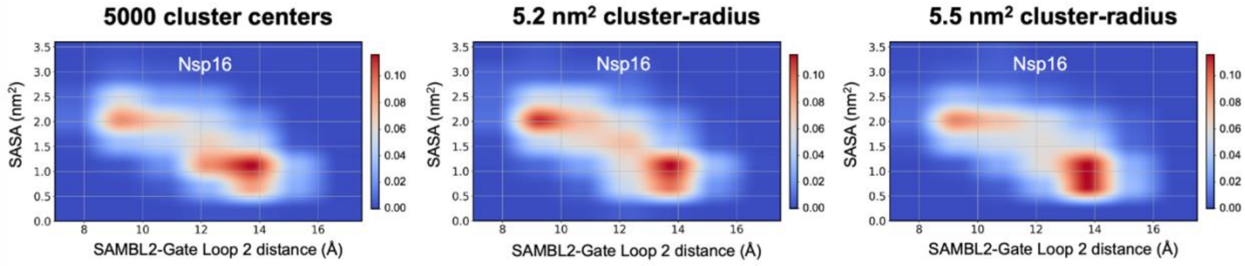


Figure S3. Equilibrium probability weighted 2D histograms of solvent-accessible surface area (SASA) of the cryptic pocket residues and the distance between SAMBL2 and gate loop 2 in Nsp16, derived from MSMs built with three different clustering cut-offs (5000 cluster centers, 5.2 nm^2 cluster radius and 5.5 nm^2 cluster radius)

Figure S4.

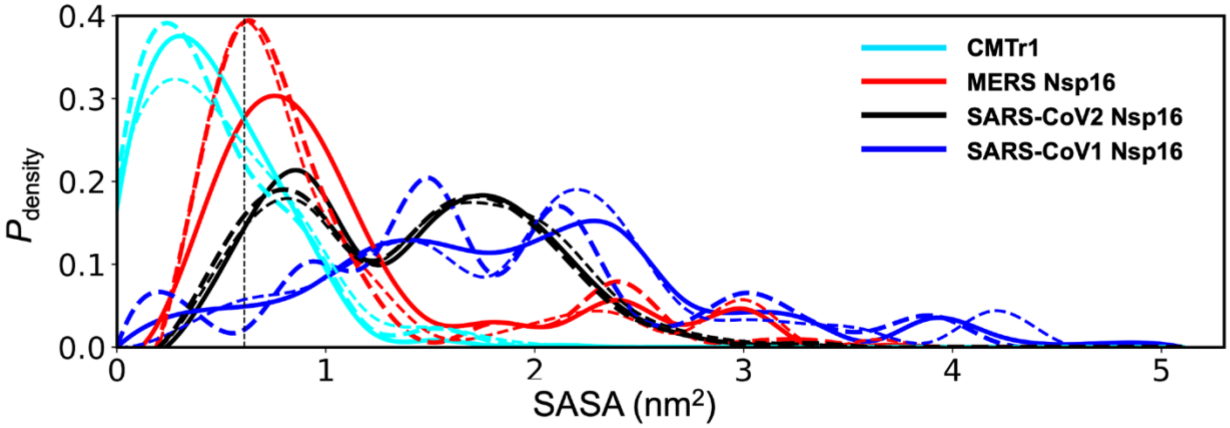


Figure S4. Equilibrium probability-weighted distribution of the solvent exposure of pocket forming residues for SARS-CoV2 (black), SARS-CoV1 (blue), MERS (red) and CMTr1 (cyan). Solid lines show the distributions derived from MSM built on 5000 clusters (for SARS-CoV2 Nsp16) and 1500 clusters (for other homologs). Thick dashed lines show the distributions derived from MSM built on clustering with 5.5 nm^2 (for SARS-CoV2 Nsp16) and 4.5 nm^2 clusters (for other homologs). Thin dashed lines show the distributions derived from MSM built on clustering with 5.2 nm^2 (for SARS-CoV2 Nsp16) and 4.0 nm^2 clusters (for other homologs). Black dotted line depicts SASA of pocket residues in the crystal structure of Nsp16/Nsp10 complex (PDB: 6wks).

Figure S5.

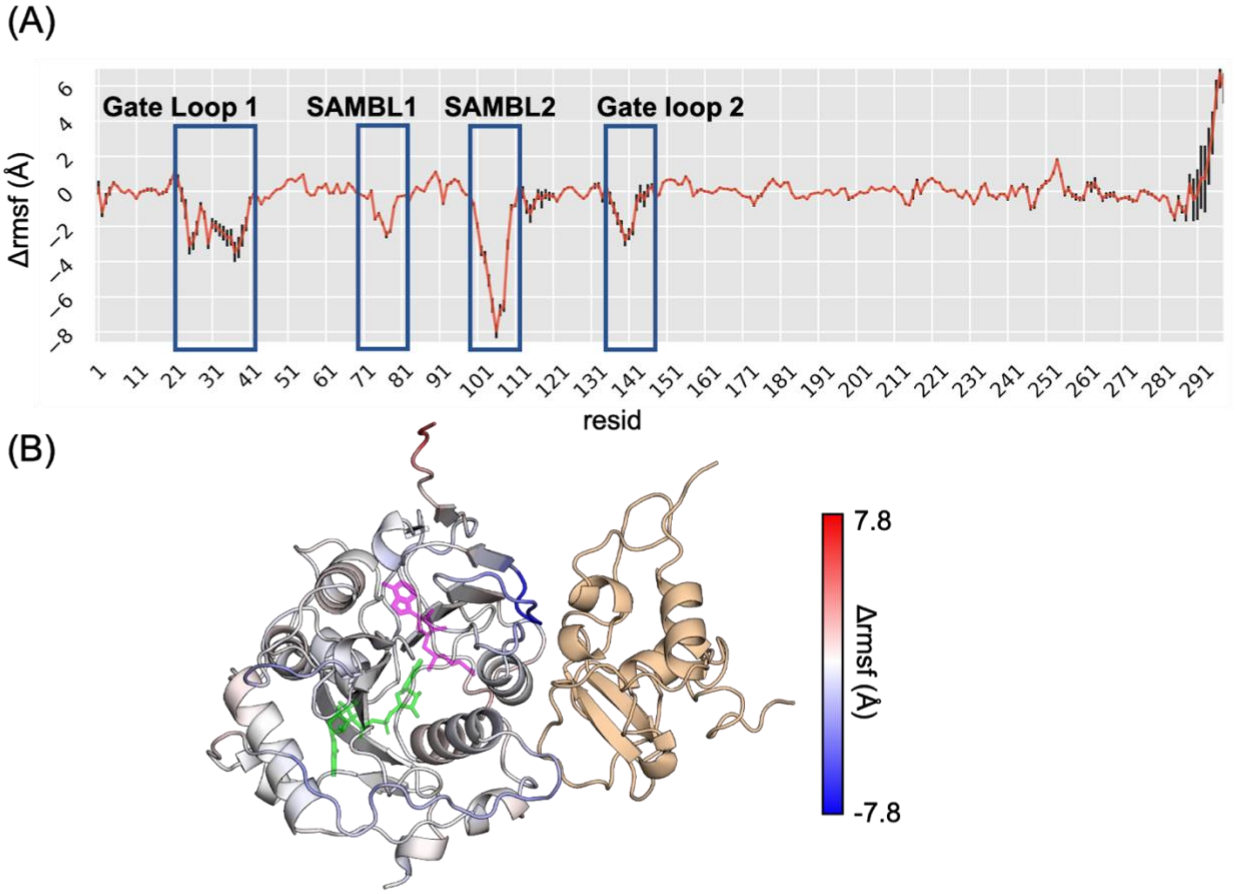


Figure S5. Change in root mean square fluctuation (rmsf) of Nsp16 upon Nsp10 association. (A) Probability weighted Δrmsf of Nsp16' residues upon Nsp10 binding is plotted. Negative values represent a decrease in rmsf upon Nsp10 binding. RNA binding loops (gate loop 1 and 2) and SAM binding loops (SAMBL1 and 2) are highlighted by the blue colored boxes. (B) Probability weighted Δrmsf of Nsp16 is mapped on its structure, with negative values shown in blue and positive values in red.

Figure S6.

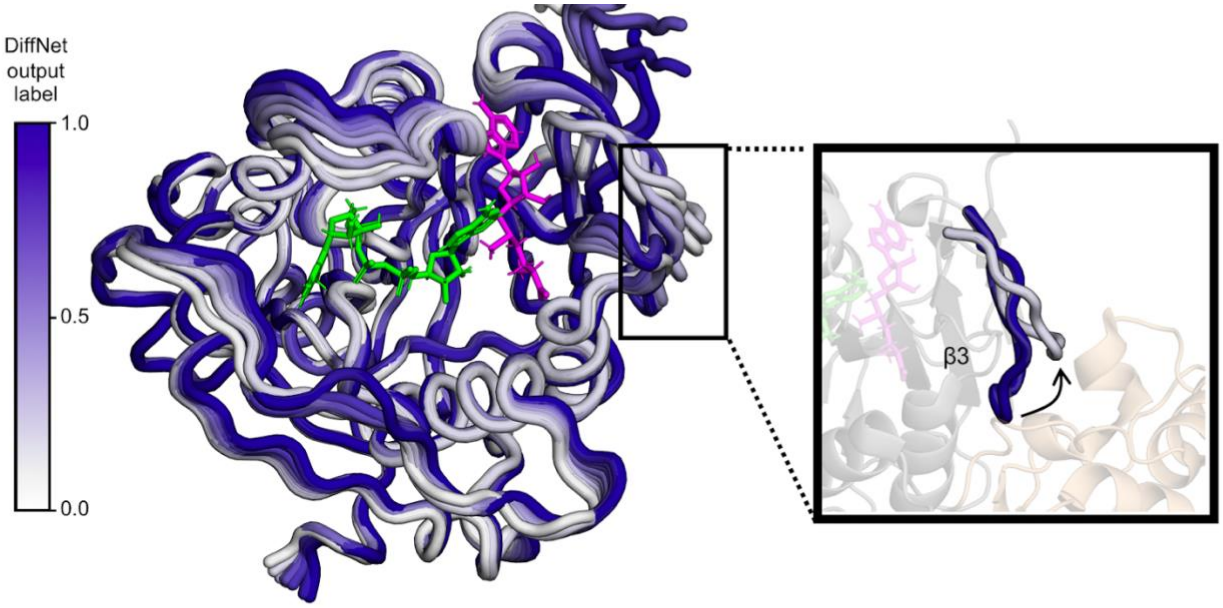


Figure S6. DiffNets predict that β4 peels away from β3 in Nsp16 inactive structural states. (Left) Structural states changing from inactive to active (white to purple) as predicted by the DiffNet. (Right) The loop connecting β3 and β4 peels away from β3 into solution in predicted inactive states.

Figure S7.

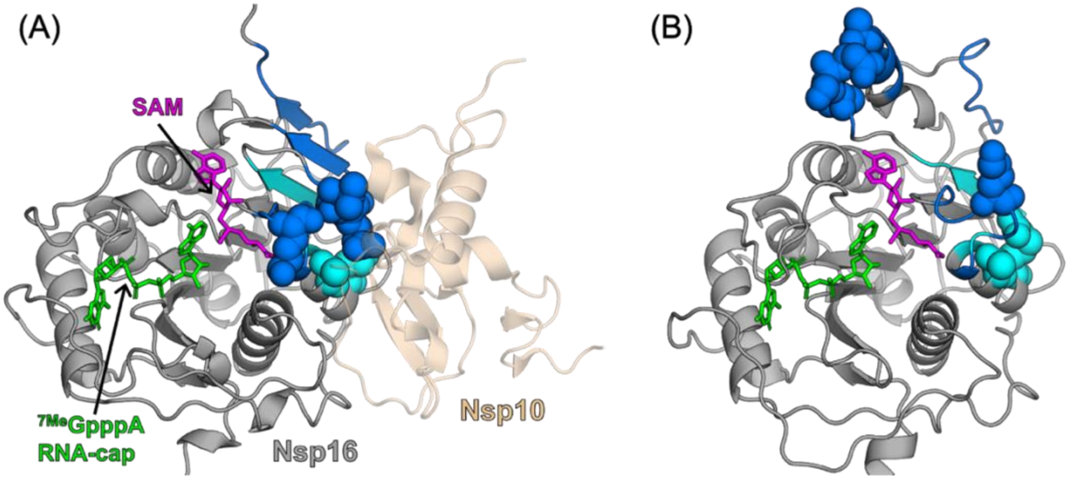


Figure S7. Displacement of Nsp10 binding residues by cryptic pocket opening. (A) Structure of Nsp16 in cryptic pocket closed state is shown in grey. Cryptic pocket forming residues and the residues undergoing opening motion are shown in cyan and blue, respectively. Cryptic pocket residues that contact Nsp10 are depicted in spheres. (B) Opening motion of the cryptic pocket shows the displacement of Nsp10 binding residues.

Figure S8.

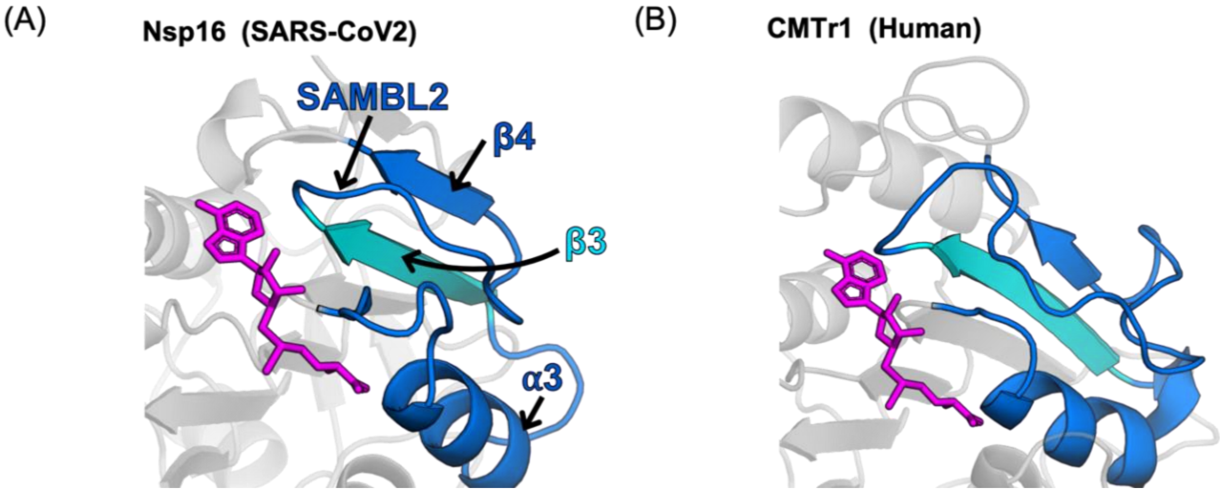
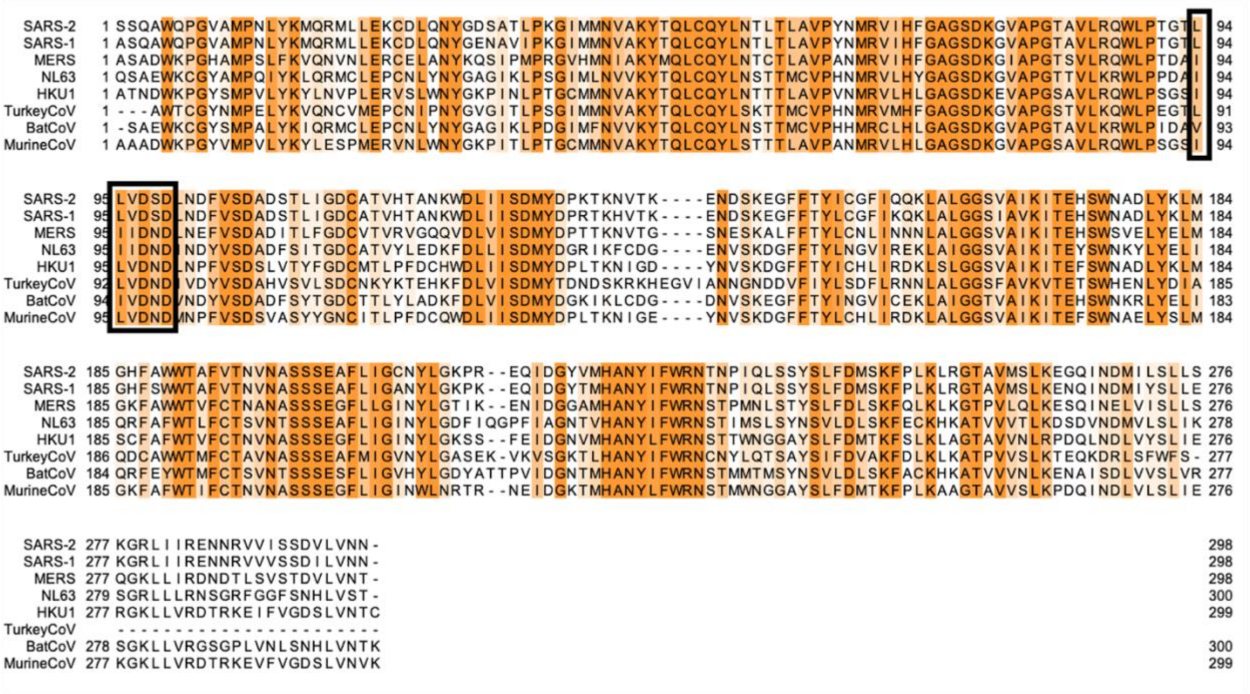


Figure S8. Structural comparison of β3-β4 cryptic pocket in SARS-CoV2 Nsp16 and human CMTr1. (A) β3 and residues lining the cryptic pocket in SARS-CoV2 are shown in cyan and blue, respectively. (B) Regions of human CMTr1, structurally equivalent to β3 and the pocket lining regions are depicted in cyan and blue, respectively.

Table S1. Timescales for transitioning between the pocket closed and open states in Nsp16 homologs.

Nsp16 homolog	Transition time (microseconds) 'Closed' → 'Open'	Transition time (microseconds) 'Open' → 'Closed'
SARS-CoV2	77.0	81.4
SARS-CoV1	26.0	13.5
MERS	19.5	9.9

Figure S9.



```
SARS-2      1 SSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDSATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTL  94
SARS-1      1 ASQAWQPGVAMPNLYKMQRMLLEKCDLQNYGENAVIPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTL  94
MERS        1 ASADWKPGHAMPSLFKVQNVNLERCELANYKQSIPMPRGVHMNIAKYMQLCQYLNTCTLAVPANMRVIHFGAGSDKGIAPGTSVLRQWLPTDAI  94
NL63        1 QSAEWKCGYAMPQIYKLQRMCLEPCNLYNYGAGIKLPSGIMLNVVKYTQLCQYLNSTTMCVPHNMRVLHYGAGSDKGVAPGTTVLKRWLPPDAI  94
HKU1        1 ATNDWKPGYSMPVLYKYLNVPLERVSLWNYGKPINLPTGCMMNVAKYTQLCQYLNTTTLAVPVNMRVLHLGAGSDKEVAPGSAVLRQWLPSGSI  94
TurkeyCoV   1 ---AWTCGYNMPELYKVQNCVMEPCNIPNYGVGITLPSGIMMNVAKYTQLCQYLSKTTMCVPHNMRVMHFGAGSDKGVAPGSTVLKQWLPEGTL  91
BatCoV      1 -SAEWKCGYSMPALYKIQRMCLEPCNLYNYGAGIKLPDGIMFNVVKYTQLCQYLNSTTMCVPHHMRCLHLGAGSDKGVAPGTAVLKRWLPIDAV  93
MurineCoV   1 AAADWKPGYVMPVLYKYLESPMERVNLWNYGKPITLPTGCMMNVAKYTQLCQYLSTTTLAVPANMRVLHLGAGSDKGVAPGSAVLRQWLPSGSI  94

SARS-2     95 LVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTK----ENDSKEGFFTYICGFIQQKLALGGSVAIKITEHSWNADLYKLM 184
SARS-1     95 LVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPRTKHVTK----ENDSKEGFFTYLCGFIKQKLALGGSIAVKITEHSWNADLYKLM 184
MERS       95 IIDNDLNEFVSDADITLFGDCVTVRVGQQVDLVISDMYDPTTKNVTG----SNESKALFFTYLCNLINNNLALGGSVAIKITEHSWSVELYELM 184
NL63       95 IIDNDINDYVSDADFSITGDCATVYLEDKFDLLISDMYDGRIKFCDG----ENVSKDGFFTYLNGVIREKLAIGGSVAIKITEYSWNKYLYELI 184
HKU1       95 LVDNDLNPFVSDSLVTYFGDCMTLPFDCHWDLIISDMYDPLTKNIGD----YNVSKDGFFTYICHLIRDKLSLGGSVAIKITEFSWNADLYKLM 184
TurkeyCoV  92 LVDNDIVDYVSDAHVSVLSDCNKYKTEHKFDLVISDMYTDNDSKRKHEGVIANNGNDDVFIYLSDFLRNNLALGGSFAVKVTETSWHENLYDIA 185
BatCoV     94 IVDNDVNDYVSDADFSYTGDCTTLYLADKFDLVISDMYDGKIKLCDG----DNVSKEGFFTYINGVICEKLAIGGTVAIKITEHSWNKRLYELI 183
MurineCoV  95 LVDNDMNPFVSDSVASYYGNCITLPFDCQWDLIISDMYDPLTKNIGE----YNVSKDGFFTYLCHLIRDKLALGGSVAIKITEFSWNAELYSLM 184

SARS-2    185 GHFAWWTAFVTNVNASSSEAFLIGCNYLGKPR--EQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKEGQINDMILSLLS 276
SARS-1    185 GHFSWWTAFVTNVNASSSEAFLIGANYLGKPK--EQIDGYTMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKENQINDMIYSLLE 276
MERS      185 GKFAWWTVFCTNANASSSEGFLLGINYLGTIK--ENIDGGAMHANYIFWRNSTPMNLSTYSLFDLSKFQLKLKGTPVLQLKESQINELVISLLS 276
NL63      185 QRFAFWTLFCTSVNTSSSEAFLIGINYLGDFIQGPFIAGNTVHANYIFWRNSTIMSLSYNSVLDLSKFECKHKATVVVTLKDSDVNDMVLSLIK 278
HKU1      185 SCFAFWTVFCTNVNASSSEGFLIGINYLGKSS--FEIDGNVMHANYLFWRNSTTWNGGAYSLFDMTKFSLKLAGTAVVNLRPDQLNDLVYSLIE 276
TurkeyCoV 186 QDCAWWTMFCTAVNASSSEAFMIGVNYLGASEK-VKVSGKTLHANYIFWRNCNYLQTSAYSIFDVAKFDLKLKATPVVSLKTEQKDRLSFWFS- 277
BatCoV    184 QRFEYWTMFCTSVNTSSSESFLIGVHYLGDYATTPVIDGNTMHANYIFWRNSTMMTMSYNSVLDLSKFACKHKATVVVNLKENAISDLVVSLVR 277
MurineCoV 185 GKFAFWTIFCTNVNASSSEGFLIGINWLNRTR--NEIDGKTMHANYLFWRNSTMWNGGAYSLFDMTKFPLKAAGTAVVSLKPDQINDLVLSLIE 276

SARS-2    277 KGRLIIRENNRVVISSDVLVNN-                                                                         298
SARS-1    277 KGRLIIRENNRVVVSSDILVNN-                                                                         298
MERS      277 QGKLLIRDNDTLSVSTDVLVNT-                                                                         298
NL63      279 SGRLLLRNSGRFGGFSNHLVST-                                                                         300
HKU1      277 RGKLLVRDTRKEIFVGDSLVNTC                                                                         299
TurkeyCoV     -----------------------
BatCoV    278 SGKLLVRGSGPLVNLSNHLVNTK                                                                         300
MurineCoV 277 KGKLLVRDTRKEVFVGDSLVNVK                                                                         299
```

Figure S9. Multiple sequence alignment of Nsp16 homologs from coronaviruses. The color ranges from white to orange for the sequence conservation score ranging from 0 to 10, where 10 denotes 100% sequence identity. Residues of β3 are enclosed in the black box. Uniprot ids of the sequences used for the alignment are given in the Methods section.