Biophysical Journal
Article

Biophysical Society

Dynamic allostery highlights the evolutionary differences between the CoV-1 and CoV-2 main proteases

Paul Campitelli,[1] Jin Lu,[1] and S. Banu Ozkan[1,*]
[1]Center for Biological Physics, Arizona State University, Physical Sciences F-Wing, Tempe, Arizona

ABSTRACT The SARS-CoV-2 coronavirus has become one of the most immediate and widely studied systems since its identification and subsequent global outbreak from 2019 to 2022. In an effort to understand the biophysical changes as a result of mutations, the mechanistic details of multiple different proteins within the SARS-CoV-2 virus have been studied and compared with SARS-CoV-1. Focusing on the main protease (mPro), we explored the long-range dynamics using the Dynamic Coupling Index (DCI) to investigate the dynamic coupling between the catalytic site residues and the rest of the protein, both inter- and intrachain, for the CoV-1 and CoV-2 mPro. We found that there is significant cross-chain coupling between these active sites and specific distal residues in the CoV-2 mPro not present in CoV-1. The enhanced long-distance interactions, particularly between the two chains, suggest subsequently enhanced cooperativity for CoV-2. A further comparative analysis of the dynamic flexibility using the dynamic flexibility index (DFI) between the CoV-1 and CoV-2 mPros shows that the inhibitor binding near active sites induces change in flexibility to a distal region of the protein, opposite in behavior between the two systems; this region becomes more flexible upon inhibitor binding in CoV-1, while it becomes less flexible in the CoV-2 mPro. Upon inspection, we show that, on average, the dynamic flexibility of the sites substituted from CoV-1 to CoV-2 changes significantly less than the average calculated across all residues within the structure, indicating that the differences in behaviors between the two systems is likely the result of allosteric influence, in which the new substitutions in CoV-2 induce flexibility and dynamic changes elsewhere in the structure.

SIGNIFICANCE Here, we have conducted a comparative analysis between the SARS-CoV-1 and SARS-CoV-2 mPro systems to shed mechanistic insight on the biophysical changes associated with the mutations between these two enzymes. Our work shows that the CoV-2 mPro system exhibits enhanced cross-chain communication between catalytic site residues and the rest of the structure. Furthermore, both dynamic coupling and dynamic flexibility analyses indicates that, largely, the dynamic changes as evaluated by DCI and DFI occur at sites other than the mutation sites themselves, indicating that the functional differences between these two proteins are a result of dynamic allostery.

INTRODUCTION

Since its initial onset in late 2019, severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) has spread rapidly through more than 200 countries, resulting in millions of deaths worldwide and becoming the cause of a global pandemic that is unprecedented in the modern era. In a massive effort to combat the viral spread and subsequent toll on human health and life, large swathes of the scientific community, including disciplines ranging from genetics to evolutionary biology, biological physics, data science, immunology, and other disciplines have taken part in a focused effort to stymie the contagion outbreak, resulting in an unparalleled development and production of vaccinations within 1 year of discovery of the virus.

Given the high rate of infection, profound global impact of the CoV-2 disease and the implication that the virus will continue to go through mutations that may result in more transmissible strains that may prove resistant to currently approved immunization procedures, the continued investigation into additional vaccinations or treatment methods remains critical. At the heart of further drug discovery, an understanding of the biophysical behavior of the virus is required. In particular, it is important to obtain mechanistic insights into critical proteins of the virus that regulate its ability to interact with host cells and

Submitted August 31, 2021, and accepted for publication March 8, 2022.
*Correspondence: banu.ozkan@asu.edu
Editor: Margaret Shun Cheung.
https://doi.org/10.1016/j.bpj.2022.03.012

successfully self-replicate for two major reasons: First, to find or design novel drugs or possible allosteric inhibitors, the dynamic behavior of viral proteins must be understood; and second, the mechanistic details of these proteins and, subsequently, how the shape of the mutational landscape regulates protein dynamics is necessary to understand whether observed mutations will confer resistance to developed drugs.

To that end, we focus on the SARS-CoV-2 main protease (mPro), an enzyme critical for the successful reproduction of the virus upon host cell infection. The mPro processes two major polyproteins into several nonstructural proteins, which, in turn, are ultimately responsible for the production of structural proteins comprising the envelope, membrane, spike, and nucleocapsid structural proteins (1,2). Thus, the mPro has undergone significant investigation as a potential drug target (3–6). When compared to CoV-1, the CoV-2 mPro contains 12 amino acid variations (henceforth, "mutations" refers to these variations between the two systems), none of which have been implicated to play important roles in the enzymatic activity of the protein (7). With a 96% sequence identity between the two structures and a root-mean-square deviation (RMSD) of ~0.73 Å, logic dictates that inhibitors developed for the CoV-1 mPro should exhibit some level of effectiveness in the CoV-2 system; however, in both *in vitro* and cell-based assay experiments, most SARS-CoV-1 mPro inhibitors that showed nanomolar-level activity were relatively impotent against the SARS-CoV-2 mPro in enzymatic assays (8–10). Furthermore, current studies often report different enzymatic turnover rates for each system; reports range from the CoV-2 mPro exhibiting approximately similar catalytical rate constant/Michaelis-Menten constant (k_{cat}/K_M) values (10) to a twofold increase in k_{cat}/K_M as compared to CoV-1 (0.21 μM^{-1} s^{-1} and 0.11 μM^{-1} s^{-1}, respectively) (11,12).

Importantly, both the CoV-1 and CoV-2 mPro systems are only biologically active as a homodimer, remaining inert in their monomeric forms (10,13). The dimerization dissociation constant has also been reported to range widely, from similar K_D values (~2.5 μM) (10) to large differences and ranges (0.14 ± 0.03 μM for CoV-2 and ranges of 230 ± 30 μM down to 0.19 ± 0.03 μM for CoV-1) (14–16). However, it has been shown that small-molecule inhibitors can affect the CoV-2 monomer-dimer equilibrium (7,16).

A previous study conducted by McLeish and Dubanevics (17) used the elastic network model (ENM) using low-frequency modes to study allosteric interactions and dynamics of the CoV-2 mPro. This work showed that several regions of the protein exhibited strong cross-chain dynamic coupling as measured by a residue-residue dynamic cross-correlation map. Furthermore, their work suggested that there are several residues located on the dimeric interface critical to allosteric interactions with the CoV-2 mPro catalytic sites and that there are additional allosteric sites that can significantly change the active site coupling by the slight stiffening of local harmonic restraints used within the ENM model.

In an effort to understand the specific residues that may have an effect on the differences in communication between chains of the CoV-1 and CoV-2 mPros, particularly to elucidate how evolutionary changes give rise to differences in dynamic behavior between the two, we used two metrics, the dynamic flexibility index (DFI) and the dynamic coupling index (DCI), to evaluate the dynamics and site-specific interactions between different regions of the two proteins using coordinate information from molecular dynamics (MD) simulation data of two PDB molecules from each system (CoV-1: PDB: 3TNT (18) and PDB: 1UK3 (19); CoV-2: PDB: 5R7Y (20) and PDB: 6Y84 (21)). Additional ENM analysis was performed using SARS-CoV-2 structures PDB: 6M03 (22) (unbound) and PDB: 7BUY (5) (bound) and SARS-CoV-1 structures PDB: 1UK3 (19) (unbound) and PDB: 3TIU (18) (bound). The DFI parameter measures the sensitivity of each position to perturbations within a network of interactions and represents the ability of a given amino acid position to explore its local conformational space (flexibility). DCI measures the displacement response of an individual position to the perturbation of a second position or group of positions, relative to the average response to any perturbation of all of the possible positions and can capture the dynamic coupling between amino acid pairs. In this mechanistic study, we use DFI and DCI to further understand the differences in cross-chain interactions, site-specific changes in flexibility, and the behavior the systems exhibit when bound to modeled inhibitors. Specifically, we investigate how the sequence variations between two mPros modify the dynamics and its consequence on observed biophysical properties.

MATERIALS AND METHODS

MD modeling

Topology files for all of the structures were prepared using the AMBER LEaP program with the ff14SB force field (23). Hydrogen atoms were added and each structure was surrounded by a 16.0-Å cubic box of water molecules using the TIP3P (24) water model. Na^+ and Cl^- atoms were added for neutralization. Each system was energy minimized using the AMBER SANDER package (23) to remove any unfavorable torsional angles or steric clashes and ensure that the system reached a local energetic minimum. In a first minimization step, the protein was kept fixed with harmonic restraints to allow surrounding water molecules and ions to relax, followed by a second minimization step in which the restraints were removed and the protein solution was further minimized. Both minimization steps used the method of steepest descent followed by conjugate gradient. The systems were then heated from 0 K to 300 K over 250 ps, at which point, long-range electrostatic interactions were calculated using the particle mesh Ewald method (25). Direct-sum, non-bonded interactions were cut off at distances of 9.0 Å or greater and bond lengths of all covalent and hydrogen bonds were constrained using the SHAKE algorithm (26). During production and heat-up, we used a Langevin thermostat to control the temperature at 300 K and a Berendsen barostat to adjust the pressure at 1 bar. A time step of 2 fs for the integrator was used for both the heat-up

and production runs. To ensure the robustness of our analysis, we used two apo PDB structures separately for both CoV-1 (PDB: 3TNT (18) and PDB: 1UK3 (19)) and CoV-2 (PDB: 6Y84 (21) and PDB: 5R7Y (20)) and conducted the simulations above. All four simulations were run for a total of 1,000 ns each, generating 2 μs of total simulation time for each system. The comparisons of each system at multiple time windows through the simulations as well their associated standard error is presented in Figs. S2 and S3. RMSD profiles also show simulation convergence, found in Fig. S4.

A protocol for the convergence of MD simulation dynamics has been established previously (27). To calculate DFI and DCI, covariance matrix data were calculated over 200 ns slices of the trajectory of each simulation, using 50-ns moving windows that overlap by 25 ns. Fundamentally, the use of the Hessian by default indicates that we are restricting ourselves to a harmonic potential, and, as such, we assume that the data are sampled from a Gaussian distribution. Appropriate sampling is met assuming that ergodicity is fulfilled in both simulation time as well as the time windows in which the covariance matrices are used for analysis, resulting in two of the basic conditions: (1) All conformations sampled must belong to the same distribution to ensure the consistency of the potential energy of the systems underlying the equilibrium distributions, and (2) the time windows and subsequent covariance matrices obtained should be independent of the initial atomic coordinates to eliminate global motions and accurately capture equilibrium coordinate information. As such, the final average DFI profiles will be independent of the window size; that is, the averaging of DFI profiles from different time window sizes (i.e., 50 ns vs. 75 ns) will give similar results and the calculated covariance matrices extracted from different times of trajectories should also result in similar DFI profiles.

Dynamic flexibility index

The dynamic flexibility index uses a perturbation response scanning (PRS) technique that combines ENMs, linear response theory (LRT), and PRS, in which mutations or amino acid interactions are modeled as fluctuation responses to force perturbations (28,29). The fluctuation response can be determined as follows:

$$[\Delta \boldsymbol{R}]_{3N\times 1} = [\boldsymbol{H}]^{-1}_{3N\times 3N}[\boldsymbol{F}]_{3N\times 1} \tag{1}$$

Here, ΔR is calculated as the fluctuation response vector of residue j as a result of F perturbation on residue i of the unit force, averaged over multiple unit force directions to simulate an isotropic perturbation. H is the Hessian, a 3 N × 3 N matrix that can be constructed from three-dimensional atomic coordinate information and is composed of the second derivatives of the harmonic potential with respect to the components of the position's vectors of length N. For this work, the Hessian matrix was extracted directly from MD simulations as the inverse of the covariance matrix. This method allows one to implicitly capture specific physiochemical properties and more accurate residue-residue interactions via atomistic force fields and subsequent all-atom simulation data.

Each position in the structure was perturbed sequentially to generate a perturbation response matrix A, as follows:

$$\boldsymbol{A}_{N\times N} = \begin{bmatrix} \left|\Delta R^1\right|_1 & \cdots & \left|\Delta R^N\right|_1 \\ \vdots & \ddots & \vdots \\ \left|\Delta R^1\right|_N & \cdots & \left|\Delta R^N\right|_N \end{bmatrix} \tag{2}$$

where $\left|\Delta R^j\right|_i = \sqrt{(\Delta R)^2}$ is the magnitude of fluctuation response at position i due to perturbations at position j. The DFI value of position i is then treated as the displacement response of position i relative to the net displacement response of the entire protein, which is calculated by sequentially perturbing each position in the structure as in Eq. 3:

$$DFI_i = \frac{\sum_{j=1}^{N}\left|\Delta R^j\right|_i}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\Delta R^j\right|_i} \tag{3}$$

It is also often useful to quantify position flexibility relative to the flexibility ranges unique to individual structures. To that end, DFI can be presented as a percentile rank, as in the following:

$$\%DFI_i = \frac{n_{\le i}}{N} \tag{4}$$

where $n_{\le i}$ is the number of positions with a DFI value less than or equal to DFI_i. The denominator is the total displacement of all of the residues, used as a normalizing factor. All %DFI calculations present in this work used the DFI value of every residue of the full structures for ranking. The DFI parameter can be considered a measure of the ability of a given amino acid position to explore its local conformational space.

DCI

Similar to DFI, the DCI also uses PRS with ENM and LRT. The DCI captures the strength of the displacement response of a given position i upon perturbation to a single functionally important position (or subset of positions) j, relative to the average fluctuation response of position i when all of the positions within a structure are perturbed, as seen in Eq. 5:

$$DCI_i = \frac{\sum_{j}^{N_{functional}}\left|\Delta R^j\right|_i / N_{functional}}{\sum_{j=1}^{N}\left|\Delta R^j\right|_i / N} \tag{5}$$

As above, DCI can also be presented as a percentile rank, as follows:

$$\%DCI_i = \frac{m_{\le i}}{\mathrm{N}} \tag{6}$$

where $m_{\le i}$ is the number of positions with a DCI value less than or equal to DCI_i. As such, this parameter represents a measure of the dynamic coupling between i upon a perturbation to j.

One of the most important aspects of both DFI and DCI is that the entire network of interactions is explicitly included in subsequent calculations without the need for dimensionality reduction techniques such as normal mode analysis through principal-component analysis. If one considers interactions such as allostery as an emergent property of an anisotropic interaction network, then it is critical to include the interactions of the entire network to accurately model the effect one residue can have on another. The DFI and DCI metrics include off-diagonal elements as well, utilizing fully the information contained within covariance matrices and subsequently takes into account the full network of interactions within any given protein system as compared to RMSF, which explicitly only contains information regarding the diagonal elements from a given covariance matrix (see Fig. S5).

RESULTS / DISCUSSION

Cross-chain dynamic coupling is enhanced for the CoV-2 mPro catalytic site residues

In the work presented here, we wanted to investigate the ability for the catalytic sites to allosterically interact with the rest of the structure and potentially identify additional putative allosteric residues. Here, we implement one of our tools, which can determine the strength of dynamic coupling

between two residues, *i* and *j*, the DCI. DCI has been used previously in many different systems to identify important sites of regulation, particularly positions distal to active site residues called dynamic allosteric residue coupling (DARC) sites. DARC sites control the dynamics of the active site through dynamic allostery, and previous work has shown that the mutation of DARC sites could alter function (28–34).

Here, we analyzed the dynamic coupling between the catalytic residues H41 and C145 of chain A and the rest of the structure by applying force perturbations to H41 and C145 simultaneously. The obtained DCI profiles are then rescaled using percentile rankings (i.e., %DCI). Fig. 1 is a graphic depiction of this analysis.

Within a given chain, the %DCI distribution is, as expected, based on proximity; residues close to the catalytic sites are strongly coupled, with those further away exhibiting weak coupling. However, for chain A-to-B interactions, this behavior is nearly the opposite of what one would expect if coupling was a measure of proximity alone. Here, regions of chain B close in space to the chain A catalytic sites are weakly coupled to these sites, whereas some distal portions are very strongly coupled. Chain B sites that exhibited particularly strong coupling to the chain A catalytic residues are circled in red, comprising residues E55, I59, and R60 and residues N277, R279, and L286. Interestingly, position 286 falls into a triad of residues indicated to be of particular importance in the work by McLeish and Dubanevics, in which it was shown that this position also exhibited a positive cross-chain dynamic coupling of motion. In addition, position 277, immediately adjacent to position 276, exhibited the maximum absolute response upon the relaxation of spring constants as determined by the largest negative free energy change (17). Furthermore, McLeish and Dubanevics show that position 285 exhibits large changes in free energy in both the positive and negative directions upon spring relaxation and spring stiffening, respectively. The importance of position E285 in the CoV-1 mPro has also been highlighted in another previous study in which it was shown that this position is critical for communication between several domains and the active site and that this position was critical for successful viral replication (36). Taken together, this work implies that this region of the structure may play a major role in allosteric regulation of the enzyme.

These coupling profiles further indicate that cross-chain communication is likely an important mechanistic regulator for the proper functioning of the SARS-CoV-2 mPro. In fact, a direct comparison of the same analysis to the CoV-1 mPro shows that this strong interchain coupling between the catalytic sites and these residue groups is nearly lost completely (Fig. 1).

Given the noticeable differences in coupling between the two systems, we next wanted to compare the changes in dynamic coupling to the active site residues of the mutation sites themselves. Fig. 2 shows this direct comparison. In Fig. 2 *A*, we have taken the %DCI values from the chain A active site perturbations shown in Fig. 1 and subtracted the CoV-2 %DCI profile from the CoV-1 profile. Here, all of the sites with a difference in %DCI within 1 standard deviation of the mean have been set to zero. This analysis

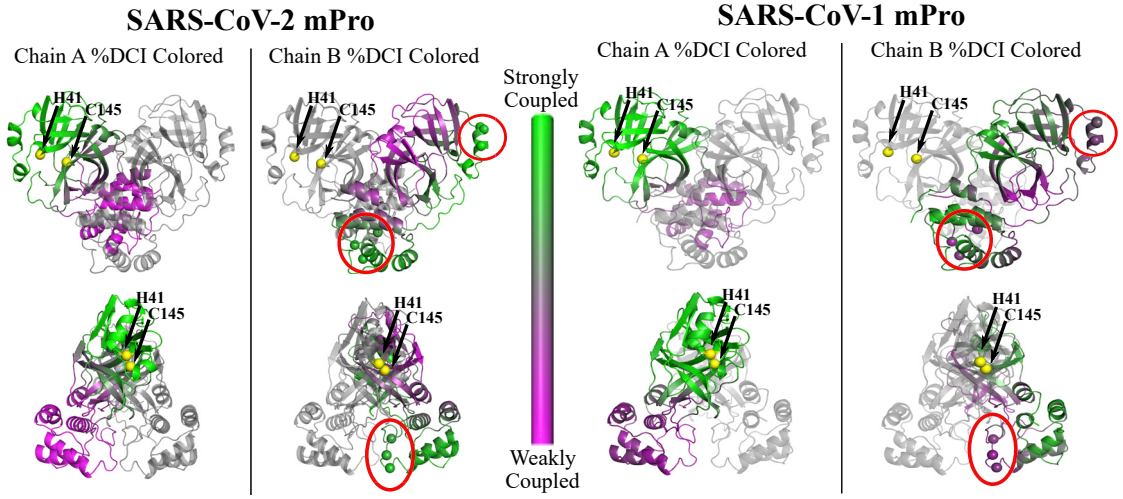

FIGURE 1 Dynamic coupling analysis (measured by %DCI) between catalytic site residues of H41 and C145 of chain A (yellow spheres) and the other residues of chain A and chain B mapped onto the SARS-CoV-2 mPro (left, PDB: 5R7Y (20)) and the SARS-CoV-1 mPro (PDB: 2GZ9 (35)). The green regions of each structure indicate stronger dynamic coupling, while purple represents weak dynamic coupling. Within the chain, the %DCI distribution is as expected based on proximity; residues close to the catalytic sites are strongly coupled, with those further away exhibiting weak coupling, and this behavior is consistent across both systems. However, for A-to-B interactions, this behavior is nearly the opposite of what one would expect if coupling was a measure of proximity alone. Here, regions of chain B close in space to the chain A catalytic sites are weakly coupled to these sites, whereas some distal portions are very strongly coupled. Chain B sites that exhibited particularly strong coupling to the chain A catalytic residues are circled in red and represented as spheres, comprising residues E55, I59, and R60 (top right circles) and residues N277, R279, and L286 (bottom circles). To see this figure in color, go online.

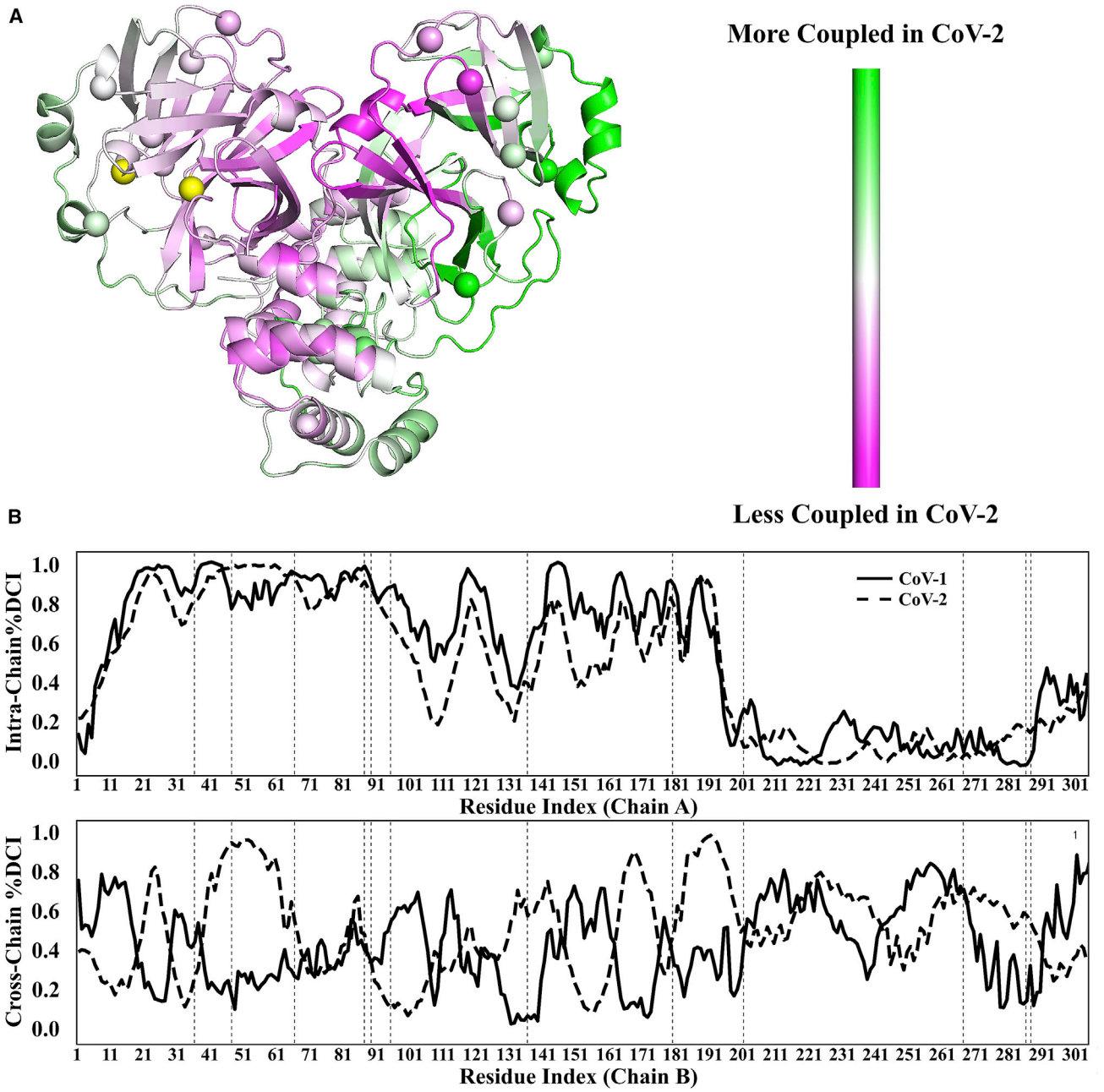

FIGURE 2 (*A*) %DCI difference between the two systems (CoV-2 %DCI and CoV-1 %DCI) mapped onto the CoV-2 (PDB: 5R7Y (20)) main protease using MD covariance matrix data to generate the profiles. The active sites perturbed to generate these values are marked in yellow spheres, as in Fig. 1. Here, all of the sites with a difference in %DCI within 1 standard deviation of the mean have been set to zero. Thus, purple indicates a significant loss in dynamic coupling, whereas green indicates a significant gain in dynamic coupling in CoV-2. Sites of mutation (positions 35, 46, 65, 86, 88, 94, 134, 180, 202, 267, 285, and 286) are also shown as spheres. (*B*) %DCI values used for the subtraction in (*A*), where dashed lines mark mutation sites. In most cases, the mutation sites themselves are not those exhibiting the greatest change in coupling to the active sites. To see this figure in color, go online.

immediately shows that, indeed, sites that exhibit large changes in dynamic coupling are often not at the sites of mutation themselves. The %DCI values used for this comparison are plotted in Fig. 2 *B*, with dashed lines indicating the sites of mutation. Interestingly, we see notable changes in coupling in areas surrounding the active sites of both chains, even though only the active site residues from chain A were perturbed. In particular, we see enhanced coupling to the loop residues 185–201 (marked in black ovals), an area previously reported to assist in stabilizing the active sites (37). In addition, recent work has been performed to identify positions within the SARS-CoV-2 mPro resistant to mutations in human populations. From a dataset containing 19,154 mutations, it was shown that 282 out of 306 residue

positions of SARS-CoV-2 have exhibited at least 1 mutation (38); excluding the active sites 41 and 145, the 24 remaining positions that had not experienced any mutation events were deemed "coldspots" and subjected to further study (39). Interestingly, a comparative analysis of the change in flexibility of these coldspots (of both chains A and B) between CoV-2 and CoV-1 shows that while evolutionarily conserved, these sites experienced a significantly greater change in dynamic flexibility (as measured by %DFI) when compared to all of the residues in the structure (see Fig. S1). In addition, performing dynamic coupling analysis of these coldspot residues uncovered a unique relationship for six specific positions. Through crystallographic studies, residues L141, F185, and Q192 of both chains A and B were shown to be structurally important, involved in the formation of substrate-binding sites (20,22,40). When we analyzed the dynamic coupling of these positions to catalytic site residues H41 and C145, we found that while the intrachain coupling remained relatively unchanged (within 1 standard deviation of the mean), the cross-chain dynamic coupling was significantly enhanced in SARS-CoV-2, with all 3 sites showing an increase in %DCI greater than 1 standard deviation from the mean. Furthermore, positions F185 and Q192, which play critical roles in stabilizing the active sites (20,22,40), exhibited %DCI increases greater than 2 standard deviations from the mean. Since many of the large changes in coupling are located at sites distal to the mutation sites, we wanted to next understand the changes between the two systems via site-specific flexibility as these changes in coupling could be due to dynamic differences that arise from changes in amino acid flexibility.

Dynamical differences between the CoV-1 and CoV-2 proteases occur at sites distal to the mutational sites themselves

In an effort to understand the mechanics underlying the behavior of the SARS-CoV-2 mPro, a comparative analysis of the dynamic differences between SARS-CoV-1 and SARS-CoV-2 was performed to capture any major differences in flexibility. Particularly, we wanted to investigate the dynamic flexibility differences as related to inhibitory binding events. To this end we studied the mPro of each virus using the DFI. Similar to DCI, DFI combines PRS and LRT (28,29) to evaluate the displacement response to random force perturbations of each position at other locations in the protein (see Materials and methods for additional details). This parameter has been used to identify important structural elements of a protein, such as hinges, and to show the flexibility of functionally critical areas, such as binding domains and catalytic regions. Surprisingly, when the flexibility profiles of the two structures were compared, we found that, on average, the larger changes in dynamic flexibility were occurring at positions other than the mutation sites themselves (only ~20% of mutated sites experienced a %DFI difference outside of 1 standard deviation, whereas 31% of all of the other residues fell into this category; Fig. 3). Also, interestingly, the loop regions 185–201, mentioned above as mechanistically important for stabilizing the active sites, do not exhibit a large change in flexibility in the CoV-2 system, which is also in line with previous work (37).

Of specific interest is the identification of sites exhibiting extremely low flexibility (marked by %DFI values $\leq$ 0.2). While positions with low flexibility are often evolutionarily conserved (41–44), our %DFI analysis shows the formation of large regions of low flexibility spanning residues 51–63 of both chains A and B in the SARS-CoV-2, absent in the CoV-1 system. These low flexibility sites called hinges have previously been shown to be mechanistically critical in determining or regulating protein function (33,45,46). Hinges play a pivotal role in the transfer of force through external perturbations throughout the chain in a cascading fashion and often control and mediate protein motion, similar to joints in a skeleton. Furthermore, the formation of hinges, or hinge-shift mechanisms, has been linked to gain-of-function in multiple different enzyme families via protein evolution (30,47–49). Interestingly the same region also has prominent differences in DFI, and a similar region exists between positions 231 and 261. These results indicate that the differences in their biophysical properties such as dimerization (10,14–16) between the two systems may be explained, mechanistically, through allosteric modulation as the formation of the hinge can allow for a greater ability for these regions to communicate across chains as a result of the higher propensity to transfer force elsewhere throughout the structure. Here, the mutations cause changes in the flexibility of amino acid positions located elsewhere in the structure and, subsequently, result in differing dynamics between the two structures.

Inhibitor binding allosterically induces changes in flexibility to distal sites differently between CoV-1 and CoV-2

As comparative flexibility analysis above suggests that allosteric regulation may be a key component in capturing the changes in dynamics between the CoV-1 and CoV-2 mPro systems, we next wanted to determine how the allosteric response to ligand binding events (particularly inhibitor binding) differed between the two. In fact, recent work using Gaussian accelerated MD indicates that there are potential cryptic binding pockets located far from the active site, which may act to inhibit the active sites allosterically when bound to inhibitory drugs (50). To further analyze the allosteric impact of inhibitor binding, we used the ENM formalism for DFI (see Materials and methods), which allows for rapid modeling of ligand binding events by adding additional α carbon atoms at the inhibitor binding sites, thus extending the residue-residue anisotropic network

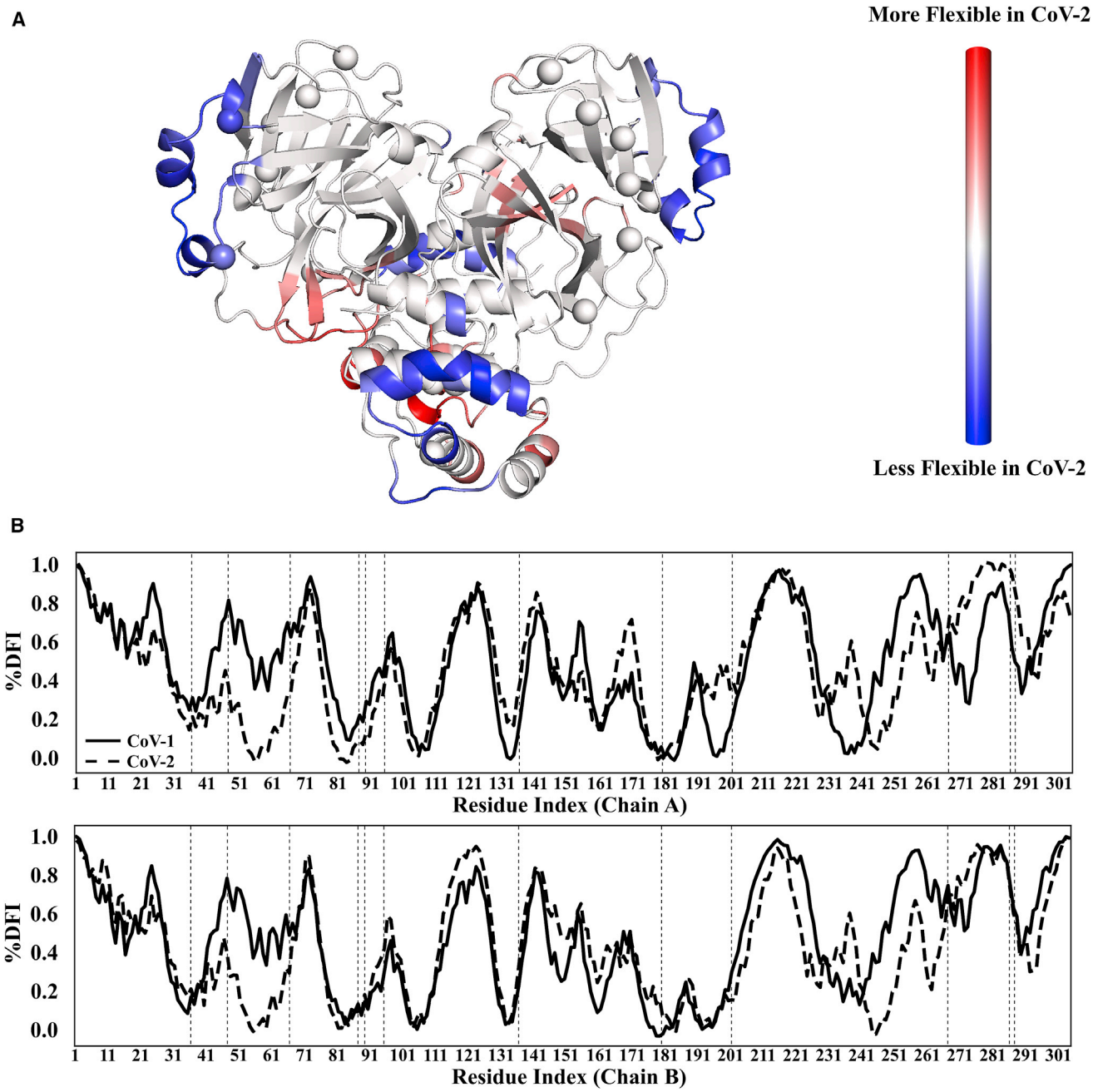

FIGURE 3 (*A*) Chain A flexibility comparison between SARS-CoV-1 CoV-2 measured by the difference in %DFI (CoV-2 %DFI and CoV-1 %DFI) mapped onto the CoV-2 mPro (PDB: 5R7Y (20)) using MD covariance matrices, where red indicates a gain in flexibility, while blue indicates a loss in flexibility. Here, all of the sites with a difference in %DFI within 1 standard deviation of the mean have been set to zero to identify the significant flexibility changes in COV-2. The sites of mutation between the two structures (positions 35, 46, 65, 86, 88, 94, 134, 180, 202, 267, 285, and 286) are marked in spheres. (*B*) The % DFI values used to generate the color coding in (*A*), where mutation sites are marked by dashed lines. Of note, many of the sites that exhibit large differences in %DFI occur at positions other than the mutation sites themselves. To see this figure in color, go online.

of each protein to include modeled inhibitors. For CoV-1, the mPro complexed with an α,β-unsaturated ethyl ester inhibitor SG82 (PDB: 3TIU) was chosen to be the inhibitor bound form. In addition, we used mPro in complex with carmofur (PDB: 7BUY) as the inhibitor bound form for CoV-2. Both inhibitors are bound at the catalytic site, and four carbon atoms among each inhibitor were picked as α carbon atoms, which contribute as nodes in ENM calculations. While less robust, this method is much less computationally expensive and avoids the additional complications of accurately including various molecular compounds in MD simulations. When using different inhibitors, we ultimately have different positions in which the heavy atoms associated with these molecules are used as additional nodes in the

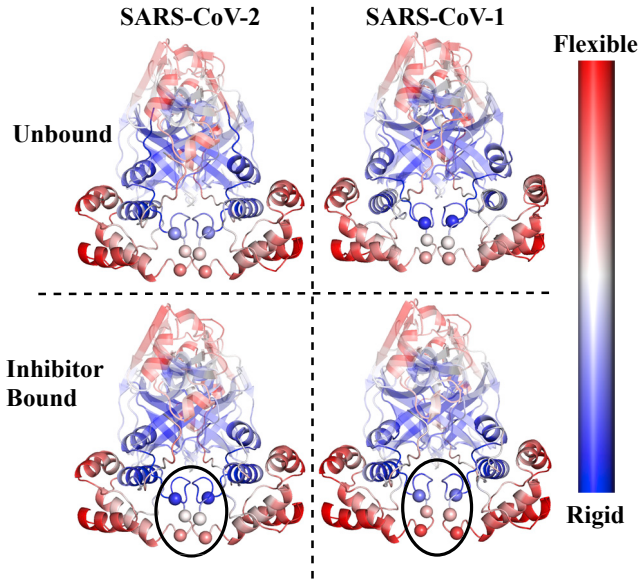

FIGURE 4 Comparison of dynamic flexibility (measured by %DFI) mapped onto the inhibitor bound (top) and inhibitor unbound (bottom) forms of the SARS-CoV-2 (left) PDB: 6M03 (22) (unbound) and PDB: 7BUY (5) (bound) and SARS-CoV-1 (right) PDB: 1UK3 (19) (unbound) and PDB: 3TIU (18) (bound). In general, the dynamic flexibility profiles between structures were similar in both inhibitor-unbound and inhibitor-bound (bottom). A specific difference in behavior was observed for a set of residues (N277, R279, and L286, black circle) located at the dimeric interface of both chains A and B. Notably, when bound to an inhibitor, these residues of the CoV-1 mPro became more flexible, whereas the same residues in the inhibitor-bound CoV-2 mPro became less flexible. These figures were rendered using PyMOL (55). To see this figure in color, go online.

interaction networks of each system. The node-to-node distances scale the spring constant strength (as r^{-6}) and, through DFI, the dynamics of the system will subsequently be different as a result. As such, the changes in the dynamics of the two systems, when bound to similarly modeled inhibitors, should tell us more about the actual mechanical differences associated with the inhibition properties themselves.

Fig. 4 shows the %DFI values mapped onto the structures of each mPro. While overall the dynamic flexibility profiles between structures were similar in both inhibitor unbound (top) and inhibitor bound (bottom), a specific difference in behavior was observed for a set of residues (N277, R279, and L286, black circle) located at the dimeric interface of both chains A and B. This group of residues in particular was identified in previous sections to exhibit enhanced coupling in the CoV-2 mPro as compared to CoV-1, and these positions and immediate neighbors have been shown to play critical roles in both CoV-1 (36) and CoV-2 (17). Here, we see that, when bound to an inhibitor, these residues of the CoV-1 mPro became more flexible, whereas the same residues in the inhibitor-bound CoV-2 mPro became less flexible. Interestingly, normal node analysis using ENM suggests that the first mode of the dimeric protease involves the opening/closing of this region of the protein (51), which may be altered significantly upon a ligand binding event, particularly if bound in the dimeric form. These results indicate that not only does an inhibitor binding event at the catalytic site residues induce flexibility changes at regions distal to the binding site, but does so in a manner opposite between the CoV-1 and CoV-2 mPro, with the residues N277, R279, and L286 becoming more and less flexible, respectively. While additional experimental characterization and mutational studies would be needed to fully understand the importance of this behavior, it follows that a binding event at these positions or near-neighbors could allow for allosteric regulation of the protein by rigidifying this region of the protein in CoV-2, and thus that there is a potential for allosteric regulation between these positions and the important catalytic site residues H41 and C145.

CONCLUSIONS

Here, we have conducted a comparative analysis between the SARS-CoV-1 and SARS-CoV-2 mPro systems to shed mechanistic insight on the biophysical changes associated with the mutations between these two enzymes. When the DCI metric was applied to these two systems, we found that the cross-chain dynamic coupling is enhanced for the CoV-2 mPro catalytic site residues as compared to the CoV-1 mPro system. The DCI profiles indicated that this type of cross-chain communication is likely an important mechanistic regulator and may be a critical functional difference between these two systems. To further understand the mechanistic changes associated with the functional evolution of the virus, we used DFI to analyze the flexibility differences and found that, surprisingly, most of the large changes in amino acid flexibility occurred in CoV-2 at sites other than the sites that are substituted when CoV-1 and CoV-2 sequences are compared. That is, the mutations brought about dynamical changes in the SARS-CoV-2 mPro at locations distal to the sites of mutation, suggesting that allosteric regulation may be a key component in capturing the changes in dynamics between CoV-1 and CoV-2.

To determine how the allosteric response to ligand binding events (particularly inhibitor binding) differed between the two by incorporating the inhibitor interactions at the active site using the inhibitor bound structures. Here, the CoV-1, mPro (PDB: 3TIU) and CoV-2 mPro (PDB: 7BUY) structures were complexed with inhibitors bound to their respective catalytic sites, where four α carbons among each inhibitor were used as additional nodes in the ENM network. From these models, we analyzed the structural dynamics associated with ligand-bound mPro forms. Our results showed that dynamic flexibility differences were observed for a set of residues located at the dimeric interface of both chains, while the DFI profiles of the rest were largely unchanged between CoV-1 and CoV-2 mPros. Notably, when bound to an inhibitor, these residues of the

CoV-1 mPro became more flexible, whereas the same residues in the inhibitor-bound CoV-2 mPro became less flexible, suggesting that interdomain interactions critical for mPro activity (52–54) are much less affected in CoV-2 when the inhibitor is bound to active site.

Overall, our work shows that the CoV-2 mPro system exhibits enhanced cross-chain communication between catalytic site residues and the rest of the structure. Furthermore, both dynamic coupling and dynamic flexibility analyses indicate that, largely, the dynamic changes as evaluated by DCI and DFI occur not at the sites of mutation, but at other, distal regions of the protein. This indicates that the functional differences between these two proteins are a result of dynamic allostery induced elsewhere in the structure upon these mutations. Finally, we show that specific regions of the CoV-2 mPro exhibit markedly different flexibility behavior when bound to an inhibitor as compared to the CoV-1 mPro. Upon comparison between multiple metrics presented here as well as previous experimental and computational work, we highlight a region in which mutations or binding events may be able to inhibit proper functioning of the CoV-2 mPro. Future work will focus on the analysis of these signature regions with the greatest change in coupling and flexibility in an effort to identify putative allosteric binding sites for potential inhibitory drug discovery.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2022.03.012.

AUTHOR CONTRIBUTIONS

P. C. and J. L. each generated MD data for the CoV-1 and CoV-2 systems. P. C. conducted all of the analyses presented here, with the exception of the ENM modeled, ligand-bound work, which was performed by J. L. P. C. wrote the manuscript, with additional writing contributions and editing from both J. L. and S. B. O. S. B. O. was responsible for the oversight and narrative of the analysis found in this manuscript.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation Division of Molecular and Cellular Biosciences (award 1715591) and the Gordon and Betty Moore Foundation Award #8415. The authors disclose no conflicts.

REFERENCES

1. Ren, Z., L. Yan, …, Z. Rao. 2013. The newly emerged SARS-like coronavirus HCOV-EMC also has an "achilles' heel": current effective inhibitor targeting a 3C-like protease. *Protein & cell.* 4:248–250.
2. Ramajayam, R., K.-P. Tan, and P.-H. Liang. 2011. Recent development of 3C and 3CL protease inhibitors for anti-coronavirus and anti-picornavirus drug discovery. *Biochem. Soc. Trans.* 39:1371–1375.
3. Dai, W., B. Zhang, …, H. Liu. 2020. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science.* 368:1331–1335.
4. Das, S., S. Sarmah, …, A. Singha Roy. 2020. An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. *J. Biomol. Struct. Dyn.* 39:3347–3357.
5. Jin, Z., Y. Zhao, …, Z. Rao. 2020. Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat. Struct. Mol. Biol.* 27:529–532.
6. Ullrich, S., and C. Nitsche. 2020. The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* 30:127377.
7. Goyal, B., and D. Goyal. 2020. Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy. *ACS Comb. Sci.* 22:297–305.
8. Gossen, J., S. Albani, …, G. Rossetti. 2021. A blueprint for high affinity SARS-CoV-2 Mpro inhibitors from activity-based compound library screening guided by analysis of protein dynamics. *ACS Pharmacol. Translational Sci.* 4:1079–1095.
9. Zhang, L., D. Lin, …, R. Hilgenfeld. 2020. α-Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. *J. Med. Chem.* 63:4562–4578.
10. Zhang, L., D. Lin, …, R. Hilgenfeld. 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science.* 368:409–412.
11. Kuo, C.-J., T.-L. Chao, …, P. H. Liang. 2021. Kinetic characterization and inhibitor screening for the proteases leading to identification of drugs against SARS-CoV-2. *Antimicrob. Agents Chemother.* 65:e02577–20.
12. Kuo, C.-J., Y.-H. Chi, …, P.-H. Liang. 2004. Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem. biophysical Res. Commun.* 318:862–867.
13. Anand, K., J. Ziebuhr, …, R. Hilgenfeld. 2003. Coronavirus main proteinase ($3CL^{pro}$) structure: basis for design of anti-SARS drugs. *Science.* 300:1763–1767.
14. Chen, S., L. Chen, …, X. Shen. 2005. Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *J. Biol. Chem.* 280:164–173.
15. Chou, C.-Y., H.-C. Chang, …, G. G. Chang. 2004. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry.* 43:14958–14970.
16. Silvestrini, L., N. Belhaj, …, F. Spinozzi. 2021. The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Scientific Rep.* 11:9283.
17. McLeish, T. C., and I. Dubanevics. 2021. Computational analysis of dynamic allostery and control in the SARS-CoV-2 main protease. *J. R. Soc. Interf.* 18:20200591.
18. Zhu, L., and R. Hilgenfeld. 2012. Crystal Structure of SARS Coronavirus Main Protease Complexed With An Alpha, Beta-Unsaturated Ethyl Ester Inhibitor SG82. https://doi.org/10.2210/pdb3tiu/pdb.
19. Yang, H., M. Yang, …, Z. Rao. 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U S A.* 100:13190–13195.
20. Douangamath, A., D. Fearon, …, M. A. Walsh. 2020. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* 11:5047.
21. Owen, C. D., P. Lukacik, …, M. A. Walsh. 2020. SARS-CoV-2 Main Protease With Unliganded Active Site (2019-nCoV, Coronavirus Disease 2019, COVID-19). https://doi.org/10.2210/pdb6Y84/pdb.
22. Jin, Z., X. Du, …, H. Yang. 2020. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature.* 582:289–293.
23. Case, D. A., T. E. Cheatham, …, R. J. Woods. 2005. The AMBER biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
24. MacKerell, A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal Physical Chemistry. B.* 102:3586–3616.

25. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N •Log(N) method for Ewald sums in large systems. *EMBO J.* 98:10089–10092.

26. Pearlman, D. A., D. A. Case, …, P. Kollman. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91:1–41.

27. Modi, T., V. A. Risso, …, S. Banu Ozkan. 2021. Hinge-shift mechanism as a protein design principle for the evolution of β-lactamases from substrate promiscuity to specificity. *Nat. Commun.* 12:1852.

28. Gerek, Z. N., and S. B. Ozkan. 2011. Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLoS Comput. Biol.* 7:e1002154.

29. Nevin Gerek, Z., S. Kumar, and S. Banu Ozkan. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evol. Appl.* 6:423–433.

30. Modi, T., J. Huihui, …, S. B. Ozkan. 2018. Ancient thioredoxins evolved to modern-day stability-function requirement by altering native state ensemble. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373:20170184.

31. Modi, T., and S. B. Ozkan. 2018. Mutations utilize dynamic allostery to confer resistance in TEM-1 β-lactamase. *Int. J. Mol. Sci.* 19:3808.

32. Campitelli, P., L. Swint-Kruse, and S. B. Ozkan. 2020. Substitutions at non-conserved rheostat positions modulate function by re-wiring long-range, dynamic interactions. *Mol. Biol. Evol.* 38:201–214.

33. Campitelli, P., J. Guo, …, S. B. Ozkan. 2018. Hinge-shift mechanism modulates allosteric regulations in human pin1. *Journal Physical Chemistry. B.* 122:5623–5629.

34. Kumar, A., B. M. Butler, …, S. B. Ozkan. 2015. Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. *Curr. Opin. Struct. Biol.* 35:135–142.

35. Lu, I.-L., N. Mahindroo, …, S. Y. Wu. 2006. Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J. Med. Chem.* 49:5154–5161.

36. Stobart, C. C., A. S. Lee, …, M. R. Denison. 2012. Temperature-sensitive mutants and revertants in the coronavirus nonstructural protein 5 protease (3CLpro) define residues involved in long-distance communication and regulation of protease activity. *J. Virol.* 86:4801–4810.

37. Weng, Y. L., S. R. Naik, …, A. Ganesan. 2021. Molecular dynamics and in silico mutagenesis on the reversible inhibitor-bound SARS-CoV-2 main protease complexes reveal the role of lateral pocket in enhancing the ligand affinity. *Scientific Rep.* 11:7429.

38. Badua, C. L. D. C., K. A. T. Baldo, and P. M. B. Medina. 2021. Genomic and proteomic mutation landscapes of SARS-CoV-2. *J. Med. Virol.* 93:1702–1721.

39. Krishnamoorthy, N., and K. Fakhro. 2021. Identification of mutation resistance coldspots for targeting the SARS-CoV2 main protease. *IUBMB life.* 73:670–675.

40. Kneller, D. W., G. Phillips, …, A. Kovalevsky. 2020. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J. Biol. Chem.* 295:17365–17373.

41. Liu, Y., and I. Bahar. 2012. Sequence evolution correlates with structural dynamics. *Mol. Biol. Evol.* 29:2253–2263.

42. Maguid, S., S. Fernandez-Alberti, and J. Echave. 2008. Evolutionary conservation of protein vibrational dynamics. *Gene.* 422:7–13.

43. Maguid, S., S. Fernández-Alberti, …, J. Echave. 2006. Evolutionary conservation of protein backbone flexibility. *J. Mol. Evol.* 63:448–457.

44. Miller, M., Y. Bromberg, and L. Swint-Kruse. 2017. Computational predictors fail to identify amino acid substitution effects at rheostat positions. *Scientific Rep.* 7:41329.

45. Kumar, A., T. J. Glembo, and S. B. Ozkan. 2015. The role of conformational dynamics and allostery in the disease development of human ferritin. *Biophysical J.* 109:1273–1281.

46. Li, Z., A. Bolia, …, C. J. Margulis. 2015. A rigid hinge region is necessary for high-affinity binding of dimannose to cyanovirin and associated constructs. *Biochemistry.* 54:6951–6960.

47. Kim, H., T. Zou, C. Modi, K. Dörner, T. J. Grunkemeyer, and R. M. Wachter. 2015. A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins. *Structure.* 23:34–43.

48. Zou, T., V. A. Risso, J. A. Gavira, J. M. Sanchez-Ruiz, and S. B. Ozkan. 2015. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol. Biol. Evol.* 32:132–143.

49. Zheng, W., B. R. Brooks, S. Doniach, and d. Thirumalai. 2005. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. *Structure.* 13:565–577.

50. Sztain, T., R. Amaro, and J. A. McCammon. 2021. Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 main protease. *J. Chem. Inf. Model.* 61:3495–3501.

51. Chen, Y. W., C.-P. B. Yiu, and K.-Y. Wong. 2020. Prediction of the SARS-CoV-2 (2019-NCoV) 3C-like protease (3CL ^{pro}) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research.* 9:129.

52. Tekpinar, M., and A. Yildirim. 2021. Impact of dimerization and N3 binding on molecular dynamics of SARS-CoV and SARS-CoV-2 main proteases. *J. Biomol. Struct. Dyn.* 1–12. https://doi.org/10.1080/07391102.2021.1880481.

53. Sheik Amamuddy, O., G. M. Verkhivker, and Ö. Tastan Bishop. 2020. Impact of early pandemic stage mutations on molecular dynamics of SARS-CoV-2 Mpro. *J. Chem. Inf. Model.* 60:5080–5102.

54. Suárez, D., and N. Díaz. 2020. SARS-COV-2 main protease: a molecular dynamics study. *J. Chem. Inf. Model.* 60:5815–5831.

55. The Pymol Molecular Graphics System (Schrodinger LLC).