THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Interfacial Water Many-Body Effects Drive Structural Dynamics and Allosteric Interactions in SARS-CoV-2 Main Protease Dimerization Interface

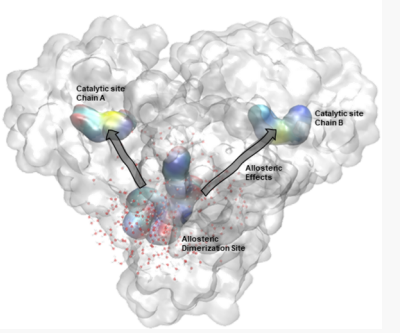
Dina El Ahdab, Louis Lagardère, Théo Jaffrelot Inizan, Fréderic Célerse, Chengwen Liu, Olivier Adjoua, Luc-Henri Jolly, Nohad Gresh, Zeina Hobaika, Pengyu Ren, Richard G. Maroun, and Jean-Philip Piquemal*

Cite This: J. Phys. Chem. Lett. 2021, 12, 6218−6226

Read Online

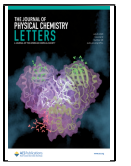
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Following our previous work (*Chem. Sci.* 2021, *12*, 4889−4907), we study the structural dynamics of the SARS-CoV-2 Main Protease dimerization interface (apo dimer) by means of microsecond adaptive sampling molecular dynamics simulations (50 μs) using the AMOEBA polarizable force field (PFF). This interface is structured by a complex H-bond network that is stable only at physiological pH. Structural correlations analysis between its residues and the catalytic site confirms the presence of a buried allosteric site. However, noticeable differences in allosteric connectivity are observed between PFFs and non-PFFs. Interfacial polarizable water molecules are shown to appear at the heart of this discrepancy because they are connected to the global interface H-bond network and able to adapt their dipole moment (and dynamics) to their diverse local physicochemical microenvironments. The water−interface many-body interactions appear to drive the interface volume fluctuations and to therefore mediate the allosteric interactions with the catalytic cavity.



In the context of COVID-19 drug discovery, both structural and nonstructural proteins are considered as promising targets for the development of antiviral agents against the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).[1] Specifically, SARS-CoV-2 M^{pro} plays a pivotal role in controlling viral replication and transcription through proteolytic processing of viral poly proteins.[2] Many studies on inhibitor ligands are based on active site pocket targeting. However, advancing a drug toward clinical trials remains a daunting task[3] (as was the case for SARS-Cov1[4,5]). In practice, because of the dimeric nature of M^{pro}, another strategy can be employed to inhibit its activity through the development of dimerization inhibitors.[2,6] Indeed, dimerization inhibitor design was previously reported for many viral enzymes such as the HIV reverse transcriptase, integrase, herpes simplex virus ribonucleotide reductase, and DNA polymerase.[6,7] In fact, targeting dimerization could potentially affect the substrate pocket and thus inhibit the M^{pro} activity because of allosteric connectivity between the dimerization site and the catalytic site.[2,8] Recently, we provided extensive simulations on M^{pro}[9] using the AMOEBA polarizable force field (PFF)[10−12] and a new highly parallel GPUs-accelerated[13,14] unsupervised adaptive sampling strategy.[9] These multimicrosecond simulations and their associated conformational spaces were compared to available non-PFF long-time scale simulation data from D. E. Shaw Research (DESRES)[15] and RIKEN Center for Biosystems Dynamics Research.[16] It was found[9] that AMOEBA results were closely correlated with experimental data, highlighting the observed strong flexibility of M^{pro}.[17] However, important differences in structural dynamics were observed compared to non-PFFs in key areas of the protease. For example, the overall richer conformational space led to enhanced volume cavities and to different solvation patterns within the active site. In order to drive further our high-resolution M^{pro} analysis, we present here a study of the factors structuring the dimerization interface as a function of different pH and solvation patterns. We particularly focus on the study of the role of many-body effects in the modeling of interfacial water and on their impact in allosteric interactions of the dimerization interface with other cavities/sites. To do so, we analyze more than 50 μs (including more than 12 μs of new simulations produced for the study) of AMOEBA molecular dynamics simulations and more than 110 μs of additional non-PFF simulations from other available data sets. All simulation details can be found in Theoretical Methods at the end of this Letter.

Received: May 6, 2021
Accepted: June 10, 2021
Published: July 1, 2021



https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226

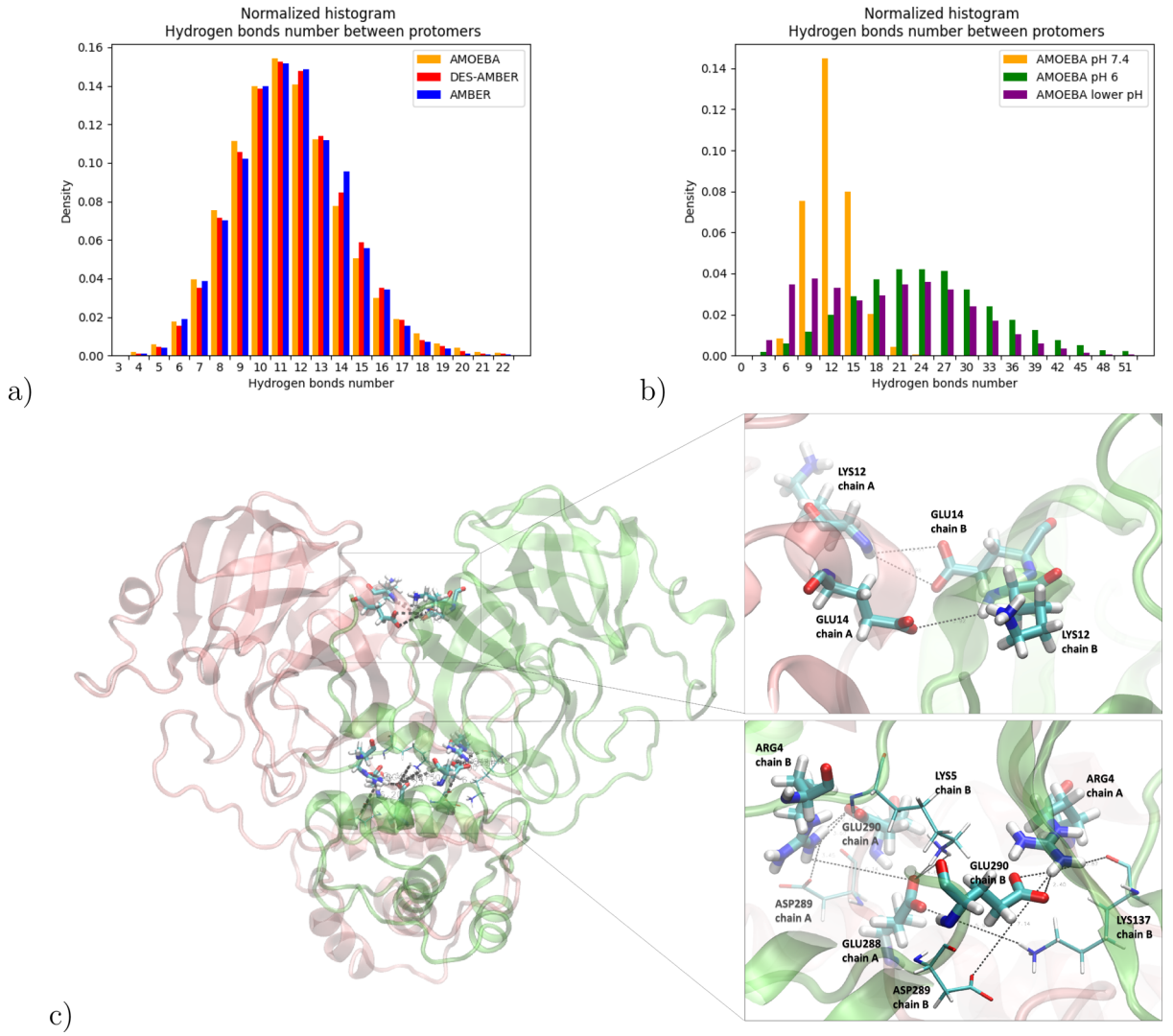


Figure 1. Histogram representation of H-bond probability density for (a) DES-AMBER, AMBER, and AMOEBA force fields at pH 7.4 and for AMOEBA trajectories at pH 7.4, 6, and lower. (b) Representation of the most frequent H-Bond interactions at the dimerization interface. Chains **A** and **B** are presented in pink and lime, respectively, (c).

To start our analysis of the M^{pro} structural dynamics at the dimerization interface, we determined the number of hydrogen bond (H-bond) interactions in order to evaluate the robustness of noncovalent interactions between the two protomers. Starting at physiological pH, we analyzed the DES-AMBER (DESRES), AMBER (RIKEN), and AMOEBA (Tinker-HP) trajectories (see Theoretical Methods for details) provided within the available conformation ensembles. We found relatively similar H-bond interaction probability density functions between the three profiles (see Figure 1a) that all present strong stability of the dimerization interface. Comparing the physiological H-bond distribution to lower pH AMOEBA simulations (see Figure 1b), we found a transition from a sharp Gaussian distribution centered at 14 H-bonds (pH 7.4) to a more diffuse one at pH 6 and below, exhibiting the involvements of weaker, disorganized, interactions. Clearly, our results show a collapse of the dimer interface at pH values lower than physiological as a consequence of the successive protonations of histidine residues (**His172** then **His163**).[9,18,19] Among the observed interactions (see Table 1 in the Supporting Information), **Arg4−Glu290** and **Gly11−Glu14** H-bond interactions have the highest probability density of all over DES-AMBER, AMBER, and AMOEBA trajectories at physiological pH. However, these interactions are not detected at lower pH, which is consistent with experimental studies reporting that low pH is responsible for the loss of the dimer interface.[20,21] It is important to note here that protonation of **His172** at lower pH has recently been shown[9,17,19] to be the source of a partial collapse in the catalytic site as well. Because the dimer interface is known to be fully functional at physiological pH, our multi-pH results reinforce the critical role of the **His172** protonation state and are consistent with Verma et al. findings[19] of a nonprotonated **His172** at physiological pH. A detailed look at the H-bond interaction profile in Table 1 of the Supporting Information highlights the key role of **Arg4** in maintaining the dimerization through several interactions, mainly with **Glu290** but also with **Lys137**, **Ser139**, **Glu288**, and **Asp289** at physiological pH. This is consistent with the description of key residues for the maintenance of SARS-CoV-2 M^{pro} dimerization in the experimental literature:[22] **Arg4**, **Ser10**, **Gly11**, **Glu14**, **Asn28**, **Ser139**, **Phe140**, **Ser147**, **Glu166**, **Glu290**, and **Arg298**. These residues all appear along our analysis, except for **Ser147**. Nevertheless, we were capable here of expanding

https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226

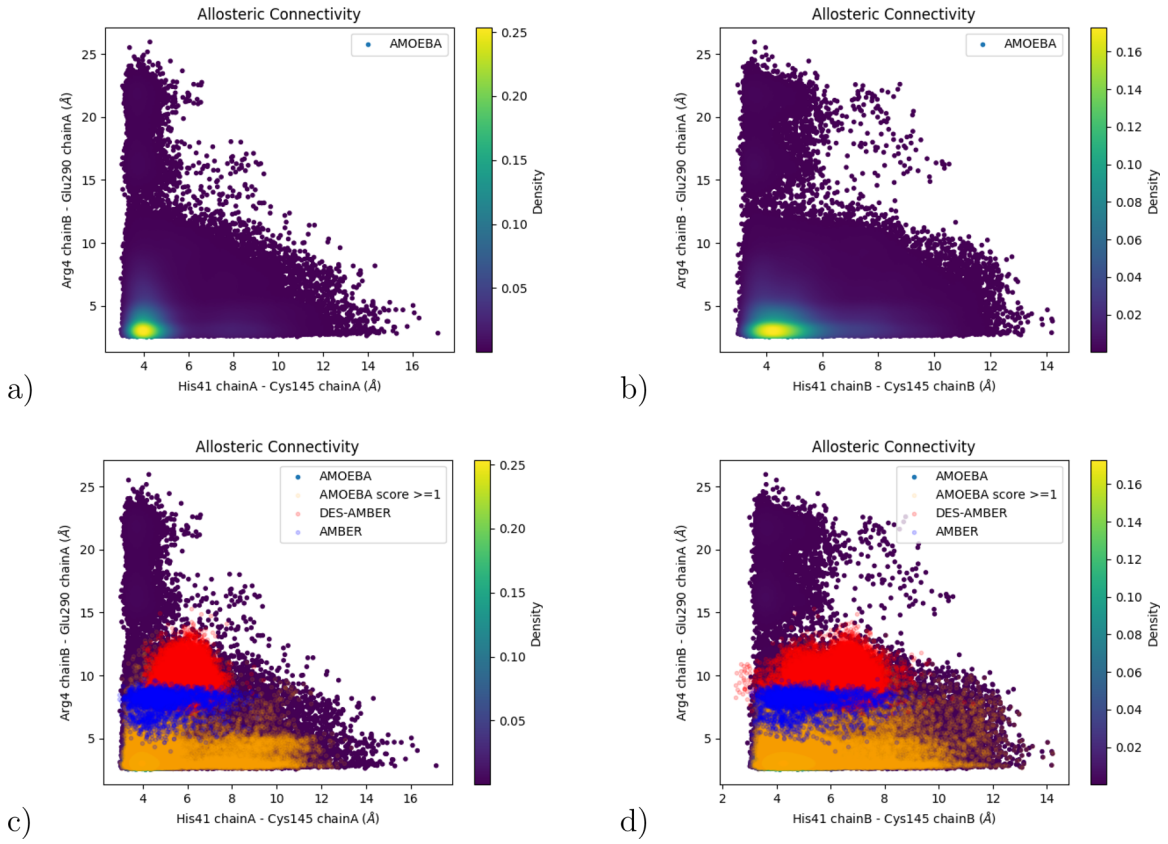


Figure 2. 2D plot representation of ***Arg4*** *chain* ***B−Glu290*** *chain* ***A*** distances vs ***His41*** *chain* ***A−Cys145*** *chain* ***A*** distances (a and c) and vs (b and d) His41 chain **B−Cys145** chain **B**. In panels c and d we have projected on the AMOEBA 15.14 μs, DES-AMBER 100 μs, AMBER 10 μs, and AMOEBA frames with a reweighting score greater than 1.

the list of these residues after a detailed analysis of DES-AMBER, AMBER, and AMOEBA simulations. As shown in Table 1 (Supporting Information), AMOEBA predicts a richer, more exhaustive, list of dimerization-implied residues compared to AMBER and DES-AMBER. The detected special forms of H-bond and other interactions, at physiological pH, are highlighted in Figure 1c. It is important to note that when successive histidine protonations occur, **His172** and **His163** switch from neutral histidines at pH 7.4 to positively charged at pH 6 and below, changing the nature of some of their interactions with other residues and water (for example, moving from H-bonds to salt-bridges in some cases[9,23]). Although pH lowering will affect also other residues that are not all considered in our computations,[19] this physicochemical change in the nature of the histidines interactions is central to the weakening of the interface stability, forcing it to redistribute its H-bond network into a different and less structured configuration. Finally, Table 1 (Supporting Information) also reveals that the Arg4−Glu290 and Gly11−Glu14 interactions are the most important H-bonds responsible for the stabilization of the dimerization interface because they exhibit the highest densities at physiological pH and are absent in the lower pH simulations. Overall, these results highlight the fact that the complex H-bond network is the one driving force stabilizing the interface.

To probe deeper into the complexity of the dimerization interface, we decided to look at its potential allosteric interactions within M^{pro}. Allostery occurs when conformational changes happening at one site of a protein and causing structural or dynamical changes at a topologically independent distant site. Such changes lead to a reduction or an increase in catalytic activity among other structural rearrangements. Structure-based prediction of allosteric sites, modulators, and communication pathway is important for a basic understanding of proteins and can lead drug discovery in order to regulate protein function.[24,25] Because H-bonds play a very important role in the dimerization region, they may be able to influence its volume, which could also have structural effects on other protein surface pockets via allosteric correlations.[24] The druggability of the dimerization interface has been discussed in the literature,[9,20] but fewer contributions looked at the potential allosteric interactions. Indeed, the importance of allosteric connectivity between allosteric and functional sites has been increasingly witnessed during recent years.[26,27] Several potential allosteric sites were recently discussed in order to offer allosteric drug target strategies[28−30] inside SARS-CoV-2 M^{pro}. For example, Stromich et al.[29] studied the scoring of putative allosteric sites and underlined a zone located in the dimerization site showing a high connectivity toward the catalytic active site. They proposed the definition of a potential allosteric dimerization site formed by the six following residues of the interface: **Arg131**, **Asp197**, **Thr199**, **Asp289**, and **Glu290** from chain **A** and **Arg4** from chain **B**. Because several of these residues were shown by our simulations to be instrumental to the interface stabilization (see Table 1, Supporting Information and previous discussion), we decided to study this site. In order to assess for a potential allosteric connectivity of the allosteric dimerization site toward both chains of the catalytic active site and to analyze its structural dynamics, we resorted to extensive bond-to-bond propensity

https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226

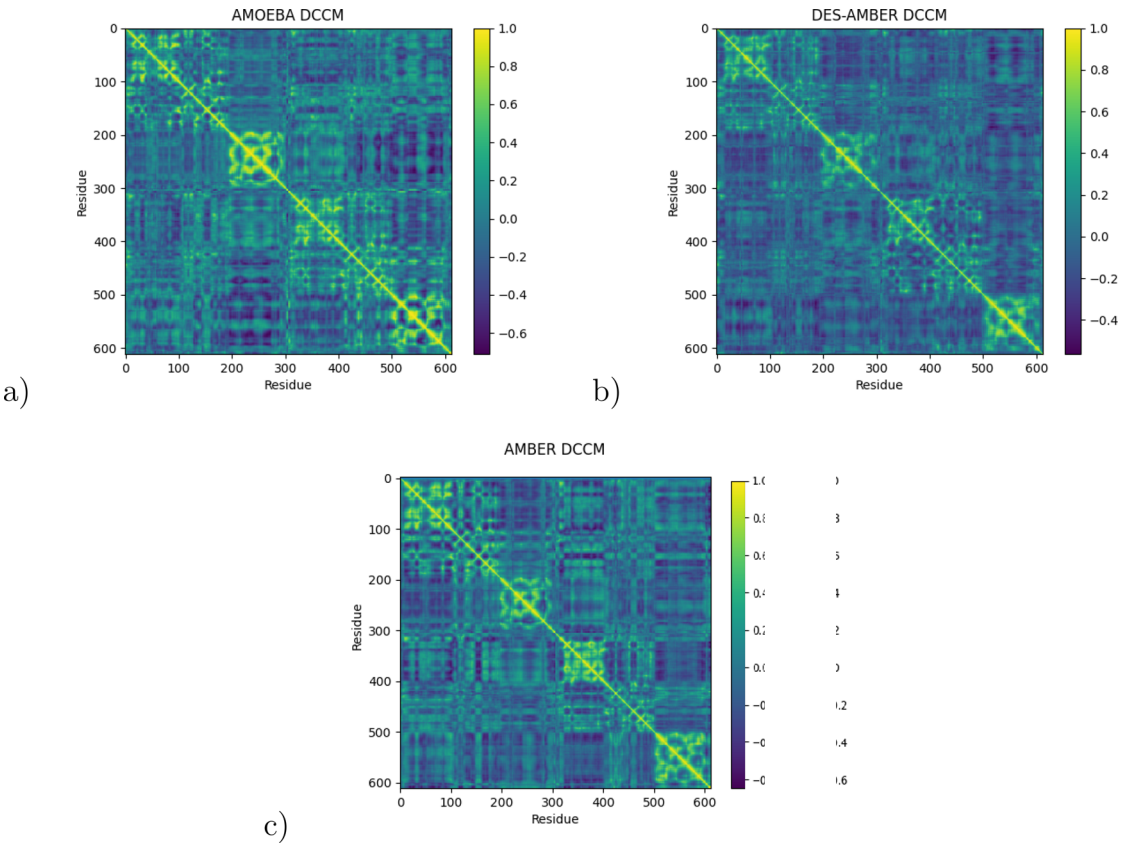


Figure 3. Dynamic cross-correlation maps using the C_α atom of each residue for (a) AMOEBA, (b) DES-AMBER, and (c) AMBER trajectories.

analysis.[31] Using this approach, we measure the fluctuations of given sets of atom−atom interactions and analyze how they affect any other set of interactions located elsewhere within the protein, allowing therefore to measure their instantaneous connectivity at each moment of the dynamics. We calculated first the evolution of distances located inside the allosteric dimerization site with other characteristic distances implicated in the residues forming the catalytic dyad. That way, thanks to well-chosen reference atoms or residues, this study informs us indirectly of the coevolution of the two cavity volumes. Indeed, comparing their volume fluctuations along trajectories can tell us about a possible allosteric connectivity between them.[29,32] We show in Figure 2a,b a 2D plot graphic of the distances separating the residues of the catalytic dyad for both chains **A** and **B** versus the distances between residues from the allosteric dimerization site: **Arg4** chain **B** and **Glu290** chain **A** because they present a robust interaction. AMOEBA trajectories show a high density of structures having both narrow catalytic and allosteric dimerization sites, respectively, around 4 and 3 Å, as shown in Figure 2a,b. However, we are also able to detect a different organization of the structures that are characterized by a narrow allosteric dimerization site and a relaxed catalytic site and, conversely, proposing possible allosteric connectivity between the sizes of the catalytic and allosteric dimerization sites. This additional connectivity found in the AMOEBA simulations is not observed in DES-AMBER nor in AMBER simulations (Figure 2c,d). Within our adaptive sampling scheme, the score is defined as the ratio between the probabilities to obtain the structure q_i in the biased simulation and in an unbiased simulation. Here, we limit ourselves to structures with a reweighting score greater than 1 as they are more likely to be visited during a conventional MD simulation. In contrast, frames with scores less than 1 have been favored by the adaptive algorithm to maximize exploration and are thus less physically relevant to the system statistic (more information can be found in ref 9). Thus, structures presented in orange in Figure 2 are more representative of the true AMOEBA statistics. In this case, we detect mostly structures having a relaxed catalytic site and a narrow allosteric dimerization site. This suggests that this specific dependency is detected thanks to the use of the polarizable AMOEBA FF, whereas the adaptive algorithm sampling is the one responsible for detecting structures associated with both a narrow catalytic site and a relaxed allosteric dimerization site. Similar conclusions can be reached upon considering **Arg131**, **Asp197**, and **Thr199** instead of **Glu290**, as shown in Figure 1 of the Supporting Information. These observations demonstrate the importance of the coupling of the adaptive sampling algorithm to the AMOEBA PFF for bringing out conformations that have escaped nonpolarizable standard MD simulations.

Because some allosteric connection was found between the dimerization and the active sites, we decided to provide another view of the simulation differences observed with the different force fields. To do so, we performed dynamic cross-correlation map (DCCM) analysis[33,34] for the three trajectories. DCCM allows us to investigate the dynamical changes of the system over time and to quantify the correlation coefficients of motions between atoms. The first result to point out is that as seen previously, AMOEBA data differ from the AMBER/DES-AMBER data. DCCM shows more positive/negative values than those obtained from non-PFFs, indicating a stronger correlated/anticorrelated atom motion in PFF simulations (see Figure 3). It is worth mentioning that strong

https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226

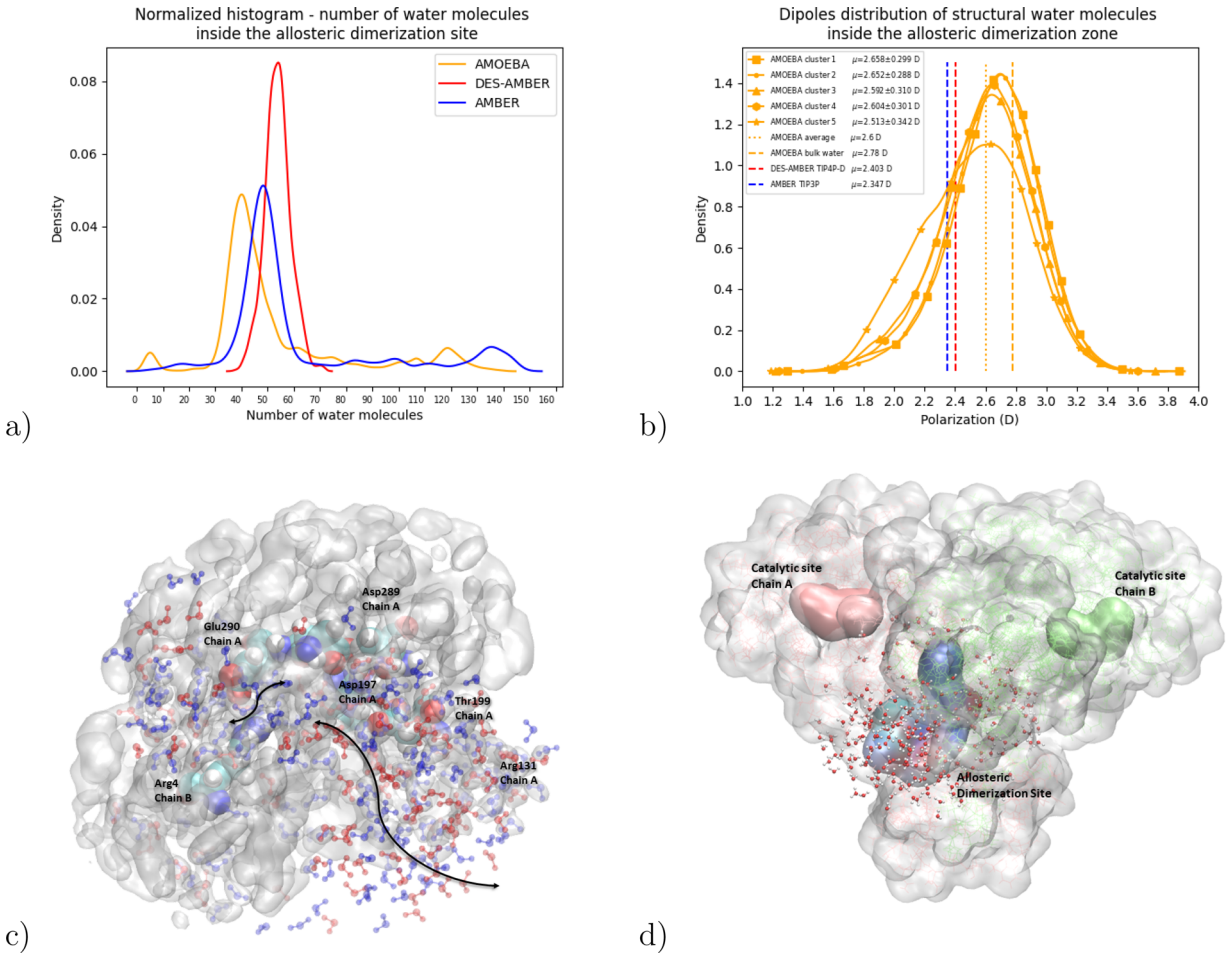


Figure 4. Representation of (a) the probability of structural water molecules number inside the allosteric dimerization site and (b) their dipoles distribution. (c) Representation of the water dipole distribution inside the allosteric dimerization site. Water molecules layered with red have dipole moment ≤2.78 D; those layered with blue have dipole moment ≥2.78 D. Asp and Glu have electrically charged side chains (acidic). Arg have electrically charged side chains (basic). Thr has polar side chain. The distance between **Arg4** and **Glu290** is 5.29 Å. Residues within 10 Å of the allosteric dimerization site are presented in quicksurf mode in white. Black arrows show the flow of water molecules in this site. (d) Global view of the M^{pro}, showing the catalytic site of both chain **A** and **B** and the allosteric dimerization site. Water molecules within 10 Å of the allosteric dimerization site are presented in cpk mode.

anticorrelation motions are observed between the α-helical region of each protomer of M^{pro} (a region strongly participating to the dimerization, i.e., residue range of 220−280 and 470−570) in AMOEBA trajectories. By contrast, the corresponding regions have much weaker (anti)correlation in both DES-AMBER and AMBER trajectories. Figure 2 in the Supporting Information proposes a closer analysis of the regions of interest for the allosteric interactions (i.e., the allosteric dimerization site) and reveals a more global anticorrelated motion between the residues of the allosteric dimerization site and the catalytic dyad of chain **A** than in AMBER/DES-AMBER. For chain **B**, this anti-correlation of the dimerization site with the catalytic dyad residues is also found. In all cases, the stronger correlation DDCM values are found within the AMOEBA simulation. The most positive correlation is found for Cys145 (chain B) and Arg4 (chain B) as the most negative correlation is found for Cy145 (chain A) and Glu290 (chain A). This further confirms the presence of an allosteric correlation between the sites and also supports the hypothesis of a strong asymmetry between protomers.[9]

As our previous analysis confirmed the differences between FF simulations, resulting in different predictions of allosteric connections and correlated motions between sites, we attempted to trace back the discrepancies studying the overall structural dynamics of the interface. As we explained in the first section, the dimerization interface overall stability is linked to a complex H-bond network that is exposed to the water solvent. Within M^{pro}, cavities and pocket volume fluctuations lead to water molecule traffic which is essential to maintain the protein structure. In a sense, the allosteric connection is performed "through water" and the resulting analysis of its presence is therefore impacted by the quality of water modeling. In practice, water molecules are commonly found within enzymatic sites, can form water bridges between the residues, and thus maintain protein secondary structures via H-bond interactions (see ref 35 and references therein). Using polarizable force fields, it has been demonstrated that some structural water molecules exhibit enhanced dipole moments, in kinase active sites for example.[36] Our previous work on M^{pro} clearly also demonstrated a very different behavior of water molecules when they are modeled with the AMOEBA PFF, which takes into account many-body effects.[9] Because water plays an important role in structural and functional activities, we looked for the water molecules present around some key

https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226

interface residues at physiological pH. To do so, we considered a 3.5 Å radius sphere centered at the atom capable of being engaged in hydrogen bonds with water for the most important residues involved in noncovalent interactions between protomers, namely: **Arg4**, **Glu290**, **Gly11**, and **Glu14**. The number of detected water molecules (see Figure 3 in the Supporting Information), presents notably different distribution profiles depending on the simulations: AMOEBA polarizable water, DES-AMBER(TIP4D), and AMBER (TIP3P). In fact, the number of water molecules detected strongly depends on the type of residue, on the considered M^{pro} chain, and on the force field itself. **Arg4** of chain **A**, for example, is found to be mostly interacting with one water molecule for AMBER, 1−2 molecules for DES-AMBER, and 2−3 molecules for AMOEBA. However, **Arg4** of chain **B** is found to interact mostly with 3 water molecules for AMBER and DES-AMBER and with 2 molecules for AMOEBA in line with the predicted asymmetry between protomers found in M^{pro}.[9] Although water traffic is detected for all force fields, the solvation patterns and differences between force fields appear to be residue-dependent. Water molecules extracted from AMOEBA trajectories around the concerned residues are polarizable (and the water model is flexible[10]), and therefore, their distribution is mainly controlled by the physicochemical nature of the residues (polar, apolar, positively/negatively charged, etc.) generating specific polarizing fields. In practice, the AMOEBA bulk water average dipole moment amounts to 2.78 D, in nice agreement with experiment, whereas non-PFF models exhibit smaller fixed dipole moments of 2.40 and 2.35 D for TIP4P-D and TIP3P, respectively. Figure 4 in the Supporting Information shows the average dipole values for the water molecules in the vicinity of the targeted residues. Their mean values (around 2.6 D on average) is below the bulk AMOEBA reference value. This result is consistent with the idea that the dense interface environment generates a global many-body depolarizing effect (compared to bulk water) influencing the water molecule-induced dipoles. Overall, the interface H-bond network connects to the solvent's own H-bond pattern forming a higher level of complexity. Clearly, the water molecule behavior is strongly influenced by the nature of the interface residues through many-body effects, generating various microsolvation patterns according to the local environment. These patterns are themselves affected by their interactions with the solvent in a self-consistent fashion.

In order to further evaluate the difference in solvation patterns, we focused on the previously introduced allosteric dimerization site, a specific location within the interface that allows for water molecules to circulate between the interface residues. To get a better understanding of what is happening, we have to evaluate the number of water molecules present and their lifetimes within this site. It is important to mention here that the six residues forming the allosteric site at the dimerization interface are either ionic or polar. Asp and Glu are negatively charged, whereas His is positively charged. Side-chains such as Thr can retain water molecules inside the cavity. Black arrows in Figure 4 display the flow of water molecules in the buried site. Because the greatest distance separating **Arg4** chain **B** and **Glu290** chain **A** is around 24 Å, we defined a sphere with a (cutoff) radius of 10 Å, centered at the geometrical center of the six residues forming the pocket at the allosteric dimerization site, and calculated the number of water molecules present within this sphere. Figure 4a shows a striking difference between AMOEBA and non-PFF simulations. PFF simulations give far fewer water molecules inside the allosteric dimerization site and a highest probability density of presence centered at 40, to be compared with 50 for AMBER and 55 for DES-AMBER.

We then measured the water lifetimes in the 10 Å sphere using the 400 ns CMD simulations produced with both the AMBER and AMOEBA force fields. We observed an average water lifetime of 0.171 ns for AMBER and a longer lifetime of 0.516 ns for AMOEBA. This clearly shows that many-body polarization effects tend to act as glue between the dimerization interface and the water molecules, specifically at the allosteric dimerization site, retaining them longer at the surface of the residues of the dimerization site (Figure 5 in the Supporting Information). Putting these two findings together allows us to better understand why the water dynamics outside the interface is so different from the (slower) dynamics found in the most confined part of the dimerization allosteric site. The smaller number of water molecules inside the allosteric dimerization site reflects therefore a slower water traffic, because these polarized water molecules tend to move slowly, being engaged into many more H-bonds. Indeed, the AMOEBA diffusion constant is more in line with experiment than the TIP3P and TIP4-D models. However, as we discussed, the AMOEBA water dipole moment values can present strong local variations because of the local microsolvation patterns that cannot be captured by the mean-field approximation, which is the basis of classical non-PFFs.[35] As for the previous situation, Figure 4 displays a rather underpolarized global situation for water that exhibits an average dipole moment lower than that of the bulk. Nevertheless, Figure 4 also highlights the collection of multiple different situations where the microsolvation patterns tend to generate simultaneously partial distributions of highly polarized and underpolarized water molecules in the allosteric dimerization site because this distribution is mainly controlled by the physicochemical nature of the residues. As shown in Figure 4c and in Figure 6 in the Supporting Information, mostly underpolarized water molecules are found in the most buried section of the allosteric dimerization site where confinement generates more depolarizing effects. These are well-known to decrease the average dipole moment values of confined waters and are observed here. Again, AMOEBA exhibits a higher probability density lower than bulk at 2.6 D, whereas DES-AMBER and AMBER water dipoles remain fixed at 2.403 and 2.347 D, respectively (see Figure 4b). Figures 4b also provides a view of the average dipole moments found after clusterization of the AMOEBA trajectories (see ref 9 for more information about the five different clusters). The site maintains a relatively stable average dipole solvent value because of the fluctuation of both the volumes (i.e., different in the different clusters) and the number of water molecules (see Figure 7 in the Supporting Information), highlighting the interconnection of the interface H-bond network and the solvent. This suggests that there is a complex interplay between the distribution of dipoles of polarizable water molecules and the residues (and associated volumes) of the dimerization allosteric site. This interaction network contributes to regulating the allosteric effects with the catalytic site of both protomers. Modeling such connections between cavities requires capturing the subtle equilibrium between the protein and solvent dynamics. The dipolar fluctuations of the water traffic tend to be extremely complex, leading to dramatically different behavior in different parts of the interface where the

https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226

local water dynamics can be quite different (i.e., for the AMOEBA-predicted dynamic slowdown within the buried allosteric dimerization site, etc.). Such water traffic shapes the interface and participates in modulating the allosteric dimerization site structural "breathing" that is involved in the overall allosteric effects with the main catalytic site. Such critical involvement of the "polarizable" water molecule within recognition or regulatory sites of proteins had been postulated before,[36] and it is clear that the number of water molecules within a binding site matters. Indeed, waters interacting with their close environment via through-water binding modes are common and able to strongly influence local electronic properties.[37] Through-water configurations can mediate interactions between an inhibitor (see for example refs 36 and 38) and indirectly bound residues of the recognition site. In such situations, also considered in the context of pFFs, an accurate count of water molecules can be critical because many-body effects (particularly the polarization energy) could tip the (free) energy balance between competing inhibitors. Missing this aspect within the modeling certainly results in a loss in the prediction of signal in the allosteric communication. It is also important to mention that beyond this energetic view of the phenomenon, the connection between interfacial water molecules and protein dynamics/flexibility has been extensively discussed in the experimental literature (see references 39−41 and references therein): protein dynamics and solvation shell dynamics have been characterized regionally. More precisely, it has been observed that flexible regions of proteins generally encompass fast-moving waters, while stable regions are embedded into slower hydration layer water molecules. This is exactly what we see here, and what is new in our results is that such regional dynamics modeling is shown to be strongly affected by many-body effects. Indeed, they strongly influence the dynamics of interfacial water molecules acting on their local "viscosity" and therefore local dynamics. As binding pockets and allosteric sites require being reasonably stable over time to be targeted by drugs, in some situations, non-PFF simulations may tend to predict solvation patterns associated with an excessive water traffic and to too fast-moving interfacial molecules. This could unfortunately lead to the destabilization of druggable hotspots that therefore would potentially remain unknown to molecular modelers.

To conclude, in order to propose a high-quality model of the dimerization interface of SARS-CoV2 M^{pro} that could be used for further drug design, it is important to understand well and model its complex H-bonds network that is embedded within a dynamic dipolar water solvent network. Water appears to be a key player in the overall structural dynamics of the dimerization interface, being one building block of the global allosteric effects between sites through many-body polarization interactions with the interface residues. As we stressed before,[9] M^{pro} is a difficult and complex molecular system that requires the simultaneous ability to (i) accurately describe all types of noncovalent interactions within the protein and solvent requiring therefore an accurate force field able to describe local many-body polarization effects and (ii) perform extensive sampling going beyond the microsecond time scale. Of course, we analyzed here only one example of allosteric interactions within M^{pro} and many other ones may remain to be discovered; we hope that these analyses and molecular dynamics trajectories (available via the BioExcel/MolSSI repository) will help drug hunters targeting the M^{pro} dimerization interface.

THEORETICAL METHODS

To study the dimerization interface we extensively analyzed the all-atom conformation space produced previously[9] using the AMOEBA polarizable force field (AMOEBA protein force field[11,12] and AMOEBA03 flexible water model[10]) as well as the one provided by the RIKEN[16] (using the AMBER ff14SB force field[42] and the TIP3P water model[43]) and DESRES[15] (using the DES-AMBER[44] and TIP4P-D water model[45]) groups. Following the same simulation protocol (reference PDB structure 6LU7[46]) proposed in our previous work,[9] we performed separate additional runs of adaptive simulations for a total of 12 μs with AMOEBA to simulate low pH values. In this case, additional histidine residue protonation occurs. Therefore, to produce additional data to the pH 7.4 and pH 6 simulations proposed in our previous data set,[9] we also successively protonated (2 × 6 μs runs) the two **His163** residues to simulate further pH lowering (see discussion and Table 2 in ref 18). Further 800 ns AMOEBA and AMBER99SB conventional molecular dynamics simulations (400 ns × 2) were produced at physiological pH and restarting from starting points from our previous data set, taking a snapshot every 10 ps to enable an in-depth analysis of the role of the water solvent. All additional all-atom simulations were performed using the newly developed GPUs module[14] within the Tinker−HP package,[13] which is part of the Tinker 8 platform.[47] This recently developed module is able to efficiently leverage mixed precision,[14] offering a strong acceleration of simulations using GPUs. Periodic boundary conditions using a cubic box of side length 100 Å were used. Langevin molecular dynamics simulations were performed using the BAOAB−RESPA1 integrator[48] using a 10 fs outer time step, a preconditioned conjugate gradient polarization solver (with a 10^{-5} convergence threshold), hydrogen−mass repartitioning (HMR), and random initial velocities. Periodic boundary conditions (PBC) were employed using the smooth particle mesh Ewald (SPME) method with a grid of dimension 128 Å × 128 Å × 128 Å. The Ewald-cutoff was taken to 7 Å, and the van der Waals cutoff was taken to be 9 Å. Post processing analysis was done using the MDTraj,[49] Scikit-Learn,[50] and Scipy packages.[51] Dynamical cross-correlation matrices (DCCMs) were generated based on the C_α atom of each residue by using the functionality provided in the MD-TASK package.[52]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.1c01460.

All the residues implicated in H-bond interactions (Table 1); 2D plot representation of distances **His41**$_X$**−Cys145**$_X$ (X = chain **A** or chain **B**) versus distances **Arg4**$_B$**−Arg131**$_A$, **Arg4**$_B$**−Asp197**$_A$ and **Arg4**$_B$**−Thr199**$_A$ showing that allosteric connectivity is present (Figure 1); extracted values from dynamic cross-correlation maps revealing the cross-correlation between residues implicated in allosteric connectivity (Figure 2); number of water molecules detected in a 3.5 Å radius from **Arg4**$_X$, **Gly11**$_X$, **Glu14**$_X$, or **Glu290**$_X$ (X = chain **A** or chain **B**) (Figure 3); dipole distribution of structural water molecules interacting with **Arg4**$_X$, **Gly11**$_X$, **Glu14**$_X$, or **Glu290**$_X$ (X = chain **A** or chain **B**) (Figure 4); water lifetime distribution inside the allosteric

https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226

dimerization site (Figure 5); representation of the water dipole distribution inside the allosteric dimerization site, for 5.29 and 8.7 Å between **Arg4** and **Glu290** (Figure 6); 2D plot representation of the volume of the dimerization site vs the number of water molecules inside the allosteric dimerization site and schematic representation of the SARS-CoV-2 M^{pro} dimer showing the dimerization site and the allosteric dimerization site residues (Figure 7) (PDF)

AUTHOR INFORMATION

Corresponding Author

Jean-Philip Piquemal − *Sorbonne Université, 75005 Paris, France; Institut Universitaire de France, 75005 Paris, France; Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0001-6615-9426; Email: jean-philip.piquemal@sorbonne-universite.fr

Authors

Dina El Ahdab − *Sorbonne Université, 75005 Paris, France; Université Saint-Joseph de Beyrouth, 1104 2020 Beirut, Lebanon*

Louis Lagardère − *Sorbonne Université, 75005 Paris, France; Sorbonne Université, 75005 Paris, France*

Théo Jaffrelot Inizan − *Sorbonne Université, 75005 Paris, France*

Fréderic Célerse − *Sorbonne Université, 75005 Paris, France; Sorbonne Université, 75005 Paris, France;* Present Address: F.C.: EPFL, LCMD, Lausanne, 1015, Switzerland.; orcid.org/0000-0001-8584-6547

Chengwen Liu − *Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0002-3930-7793

Olivier Adjoua − *Sorbonne Université, 75005 Paris, France*

Luc-Henri Jolly − *Sorbonne Université, 75005 Paris, France*

Nohad Gresh − *Sorbonne Université, 75005 Paris, France;* orcid.org/0000-0001-7174-2907

Zeina Hobaika − *Université Saint-Joseph de Beyrouth, 1104 2020 Beirut, Lebanon*

Pengyu Ren − *Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0002-5613-1910

Richard G. Maroun − *Université Saint-Joseph de Beyrouth, 1104 2020 Beirut, Lebanon*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.1c01460

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No 810367), project EMC2 (J.-P.P.). D.E.A. acknowledges funding from the Lebanese National Council for Scientific Research, CNRS-L. F.C. acknowledges funding from the French state funds managed by the CalSimLab LABEX and the ANR within the Investissements d'Avenir program (reference ANR11-IDEX-0004-02) and support from the Direction Génerale de l'Armement (DGA) Maîtrise NRBC of the French Ministry of Defense. Adaptive sampling computations have been performed at GENCI thanks to a COVID19 emergency allocation on the Jean Zay machine (IDRIS, Orsay, France) on Grant No. A0070707671 and on the Irene Joliot Curie machine thanks to a PRACE COVID-19 emergency grant (Project COVID-HP). Additional conventional AMOEBA and AMBER MD simulations have been performed on the Amazon cloud platform thanks to an AWS COVID-19 special grant. The authors thank the Swiss National Supercomputing Center (CSCS) for hosting our data through the FENIX infrastructure. J.-P.P. acknowledges a special COVID-19 funding from Sorbonne Université. P.R. is grateful for support by National Science Foundation (CHE-1856173) and National Institutes of Health (R01GM106137 and R01GM114237).

REFERENCES

(1) Zumla, A.; Chan, J. F.; Azhar, E. I.; Hui, D. S.; Yuen, K.-Y. Coronaviruses-Drug Discovery and Therapeutic Options. *Nat. Rev. Drug Discovery* **2016**, *15*, 327−347.

(2) Ding, L.; Zhang, X.-X.; Wei, P.; Fan, K.; Lai, L. The interaction between severe acute respiratory syndrome coronavirus 3C-like proteinase and a dimeric inhibitor by capillary electrophoresis. *Anal. Biochem.* **2005**, *343*, 159−165.

(3) Cui, W.; Yang, K.; Yang, H. Recent Progress in the Drug Development Targeting SARS-CoV-2 Main Protease as Treatment for COVID-19. *Front. Biosci.* **2020**, *7*, 398.

(4) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome−Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595−6628.

(5) Kuo, C.; Liang, P. Characterization and Inhibition of the Main Protease of Severe Acute Respiratory Syndrome Coronavirus. *ChemBioEng Rev.* **2015**, *2*, 118−132.

(6) Boggetto, N.; Reboud-Ravaux, M. Dimerization Inhibitors of HIV-1 Protease. *Biol. Chem.* **2002**, *383*, 1321−1324.

(7) Zutshi, R.; Brickner, M.; Chmielewski, J. Inhibiting the assembly of protein-protein interfaces. *Curr. Opin. Chem. Biol.* **1998**, *2*, 62−66.

(8) Wei, P.; Fan, K.; Chen, H.; Ma, L.; Huang, C.; Tan, L.; Xi, D.; Li, C.; Liu, Y.; Cao, A.; Lai, L. The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochem. Biophys. Res. Commun.* **2006**, *339*, 865−872.

(9) Jaffrelot Inizan, T.; Célerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; Monmarché, P.; Piquemal, J.-P. High-resolution mining of the SARS-CoV-2 main protease conformational space: supercomputer-driven unsupervised adaptive sampling. *Chem. Sci.* **2021**, *12*, 4889−4907.

(10) Ren, P. Y.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.

(11) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046−4063.

(12) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14*, 2084−2108.

(13) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956−972.

(14) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole

https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226

Polarizable Force Fields Using GPUs and Multi-GPU Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2034−2053.
(15) D. E. Shaw Research. *Molecular Dynamics Simulations Related to SARS-CoV-2,*. D. E. Shaw Research Technical Data, 2020, http://www.deshawresearch.com/resources_sarscov2.html.
(16) Komatsu, T. S.; Koyama, Y.; Okimoto, N.; Morimoto, G.; Ohno, Y.; Taiji, M. COVID-19 related trajectory data of 10 microseconds all atom molecular dynamics simulation of SARS-CoV-2 dimeric main protease, V2. *Mendeley Data* **2020**, *10*, 17632.
(17) Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.; Langan, P.; Joachimiak, A.; Coates, L.; Kovalevsky, A. Structural plasticity of SARS-CoV-2 3CL M pro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **2020**, *11*, 3202.
(18) Tan, J.; Verschueren, K. H.; Anand, K.; Shen, J.; Yang, M.; Xu, Y.; Rao, Z.; Bigalke, J.; Heisen, B.; Mesters, J. R.; Chen, K.; Shen, X.; Jiang, H.; Hilgenfeld, R. pH-dependent Conformational Flexibility of the SARS-CoV Main Proteinase (Mpro) Dimer: Molecular Dynamics Simulations and Multiple X-ray Structure Analyses. *J. Mol. Biol.* **2005**, *354*, 25−40.
(19) Verma, N.; Henderson, J. A.; Shen, J. Proton-Coupled Conformational Activation of SARS Coronavirus Main Proteases and Opportunity for Designing Small-Molecule Broad-Spectrum Targeted Covalent Inhibitors. *J. Am. Chem. Soc.* **2020**, *142*, 21883−21890.
(20) Goyal, B.; Goyal, B. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb. Sci.* **2020**, *22*, 297−305.
(21) Chou, C.-Y.; Chang, H.-C.; Hsu, W.-C.; Lin, T.-Z.; Lin, C.-H.; Chang, G.-G. Quaternary Structure of the Severe Acute Respiratory Syndrome (SARS) Coronavirus Main Protease. *Biochemistry* **2004**, *43*, 14958−14970.
(22) Liang, J.; Karagiannis, C.; Pitsillou, E.; Darmawan, K. K.; Ng, K.; Hung, A.; Karagiannis, T. C. Site mapping and small molecule blind docking reveal a possible target site on the SARS-CoV-2 main protease dimer interface. *Comput. Biol. Chem.* **2020**, *89*, 107372.
(23) Liao, S.-M.; Du, Q.-S.; Meng, J.-Z.; Pang, Z.-W.; Huang, R.-B. The multiple roles of histidine in protein interactions. *Chem. Cent. J.* **2013**, *7*, 44.
(24) Monod, J.; Wyman, J.; Changeux, J.-P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **1965**, *12*, 88−118.
(25) Greener, J. G.; Sternberg, M. J. Structure-based prediction of protein allostery. *Curr. Opin. Struct. Biol.* **2018**, *50*, 1−8.
(26) Laskowski, R. A.; Gerick, F.; Thornton, J. M. The structural basis of allosteric regulation in proteins. *FEBS Lett.* **2009**, *583*, 1692.
(27) Suplatov, D. A.; Švedas, V. Study of Functional and Allosteric Sites in Protein Superfamilies. *Acta. Naturae* **2015**, *7*, 34−45.
(28) Günther, S.; et al. X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021**, *372*, 642−646.
(29) Strömich, L.; Wu, N.; Barahona, M.; Yaliraki, S. N. Allosteric Hotspots in the Main Protease of SARS-CoV-2. *BioRxiv* 2020, DOI: 10.1101/2020.11.06.369439.
(30) Carli, M.; Sormani, G.; Rodriguez, A.; Laio, A. Candidate Binding Sites for Allosteric Inhibition of the SARS-CoV-2 Main Protease from the Analysis of Large-Scale Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2021**, *12*, 65−72.
(31) Amor, B. R. C.; Schaub, M. T.; Yaliraki, S. N.; Barahona, M. Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nat. Commun.* **2016**, *7*, 12477.
(32) La Sala, G.; Decherchi, S.; De Vivo, M.; Rocchia, W. Allosteric Communication Networks in Proteins Revealed through Pocket Crosstalk Analysis. *ACS Cent. Sci.* **2017**, *3*, 949−960.
(33) McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press, 1988.
(34) Ichiye, T.; Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 205−217.
(35) Melcr, J.; Piquemal, J.-P. Accurate biomolecular simulations account for electronic polarization. *Front. Mol. Biosci.* **2019**, *6*, 143.
(36) de Courcy, B.; Piquemal, J.-P.; Garbay, C.; Gresh, N. Polarizable Water Molecules in Ligand-Macromolecule Recognition. Impact on the Relative Affinities of Competing Pyrrolopyrimidine Inhibitors for FAK Kinase. *J. Am. Chem. Soc.* **2010**, *132*, 3312−3320.
(37) de Courcy, B.; Pedersen, L. G.; Parisel, O.; Gresh, N.; Silvi, B.; Pilme, J.; Piquemal, J.-P. Understanding Selectivity of Hard and Soft Metal Cations within Biological Systems Using the Subvalence Concept. 1. Application to Blood Coagulation:Direct Cation−Protein Electronic Effects versus Indirect Interactions through Water Networks. *J. Chem. Theory Comput.* **2010**, *6*, 1048−1063.
(38) Gresh, N.; de Courcy, B.; Piquemal, J.-P.; Foret, J.; Courtiol-Legourd, S.; Salmon, L. Polarizable Water Networks in Ligand−Metalloprotein Recognition. Impact on the Relative Complexation Energies of Zn-Dependent Phosphomannose Isomerase with d-Mannose 6-Phosphate Surrogates. *J. Phys. Chem. B* **2011**, *115*, 8304−8316.
(39) Dahanayake, J. N.; Mitchell-Koch, K. R. How Does Solvation Layer Mobility Affect Protein Structural Dynamics? *Front. Biosci.* **2018**, *5*, 65.
(40) Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016**, *116*, 7673−7697.
(41) Combet, S.; Zanotti, J.-M. Further evidence that interfacial water is the main "driving force" of protein dynamics: a neutron scattering study on perdeuterated C-phycocyanin. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4927−4934.
(42) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(44) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a Force Field for the Simulation of Single-Chain Proteins and Protein−Protein Complexes. *J. Chem. Theory Comput.* **2020**, *16*, 2494−2507.
(45) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(46) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C. Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289.
(47) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273−5289.
(48) Lagardère, L.; Aviat, F.; Piquemal, J.-P. Pushing the Limits of Multiple-Time-Step Strategies for Polarizable Point Dipole Molecular Dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 2593−2599.
(49) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(50) Pedregosa, F.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(51) Virtanen, P.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261−272.
(52) Brown, D. K.; Penkler, D. L.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A. R.; Atilgan, C.; Tastan Bishop, O. MD−TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **2017**, *33*, 2768−2771.

https://doi.org/10.1021/acs.jpclett.1c01460
J. Phys. Chem. Lett. 2021, 12, 6218−6226