Cooperative multivalent receptor binding promotes exposure of the SARS-CoV-2 fusion machinery core

Alexander J. Pak[1,⊥], Alvin Yu[1], Zunlong Ke[2], John A. G. Briggs[2], and Gregory A. Voth[1,#]

[1] Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, IL, USA

[2] Structural Studies Division, Medical Research Council Laboratory of Molecular Biology, Cambridge, UK

[⊥] Present address: Department of Chemical and Biological Engineering, Colorado School of Mines, Golden, CO 80401

[#] Corresponding author: gavoth@uchicago.edu

Abstract

The molecular events that permit the spike glycoprotein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to bind, fuse, and enter cells are important to understand for both fundamental and therapeutic reasons. Spike proteins consist of S1 and S2 domains, which recognize angiotensin-converting enzyme 2 (ACE2) receptors and contain the viral fusion machinery, respectively. Ostensibly, the binding of spike trimers to ACE2 receptors promotes the preparation of the fusion machinery by dissociation of the S1 domains. We report the development and use of coarse-grained models and simulations to investigate the dynamical mechanisms involved in viral binding and exposure of the S2 trimeric core. We show that spike trimers cooperatively bind to multiple ACE2 dimers. The multivalent interaction cyclically and processively induces S1 dissociation, thereby exposing the S2 core containing the fusion machinery. Our simulations thus reveal an important concerted interaction between spike trimers and ACE2 dimers that primes the virus for membrane fusion and entry.

Significance Statement

Viral particles of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the COVID-19 pandemic, are decorated by spike trimer proteins that mediate entry into cells. Spike trimers first bind to angiotensin-converting enzyme 2 (ACE2) receptors, and subsequently shed their receptor-binding components to expose and prepare the viral fusion machinery. Here, we present systematically developed coarse-grained models and molecular dynamics simulations of viral binding and shedding prior to membrane fusion. Our analysis shows that spike trimers are capable of binding to multiple ACE2 dimers and that this multivalent interaction facilitates exposure of the viral fusion core machinery. Our findings demonstrate the necessity of concerted, cooperative interactions between viral proteins and host cell receptors, and advance fundamental understanding and therapeutic efforts on coronaviruses, aspects of which are at present difficult if not impossible to obtain by experimental means alone.

Introduction

Infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) begins with binding of viral particles to angiotensin-converting enzyme 2 (ACE2) cell-surface receptors, followed by viral-host cell membrane fusion, and entry of the viral genetic cargo (1). These events are orchestrated by spike glycoproteins – trimeric class I fusion proteins – that decorate the exterior of SARS-CoV-2 particles. Each spike protein consists of S1 and S2 domains, which are post-translationally cleaved by furin and remain non-covalently associated in a metastable prefusion state (2, 3). The S1 domain contains the receptor binding domain (RBD) that is responsible for ACE2 recognition, while the S2 domain contains the fusion machinery. The S2 domains interact as a trimeric core with a characteristic helical stalk that is partially covered by the three S1 domains; the RBDs interact as a trimeric cap around the S2 trimeric core (4, 5). A hinge-like conformational change exposes the RBD interfaces that are amenable to ACE2 association. Receptor binding is believed to result in S1 shedding, thereby exposing the S2 core. For fusion to occur, the second proteolytic site S2' is cleaved, thereby releasing the fusion peptide, after which the S2 core undergoes a dramatic conformational change into an extended helical stalk, i.e., the postfusion state, in a process similar to prior coronaviruses (6).

As spike proteins are indispensable to SARS-CoV-2 infectivity, inhibiting their function is a natural target for therapeutic design. Understanding how spike proteins bind to ACE2 and how binding primes exposure of the S2 trimeric core is therefore essential to viral activity. Structures resolved using cryo-electron microscopy (cryo-EM) combined with docking have shown that ACE2-RBD binding requires spike trimers to be open, and that two spike trimers can be accommodated on one ACE2 dimer (7). Additional cryo-EM structures have shown that up to three soluble ACE2 monomers can bind to the same spike trimer, which may result in exposure of the S2 trimeric core (8); indeed, soluble ACE2 trimers were recently demonstrated as an effective inhibitor of SARS-CoV-2 (9). Both modalities are potentially important for viral avidity, and furthermore, likely dependent upon the flexibility exhibited by both the spike trimer stalk (10) and ACE2 domain (11), as well as the stochastic nature of RBD opening observed throughout the spike trimer (4, 5). It has also been suggested that ACE2 binding to spike trimers induces allosteric effects that heighten (dampen) conformational motions near proteolytic cleavage sites

and the fusion peptide (near the central helix and stalk) (12) or conformational motions that promote subsequent RBD opening (13). Characterization of three circulating variant strains (B.1.1.7, B.1.351, and P.1) has shown that increased infectivity is, in part, due to mutations in the RBD that increase ACE2 binding affinity (14, 15). Clearly, viral binding, fusion, and entry rely on a complex dynamical process that likely leverages the concerted interactions between spike trimers and ACE2 receptors.

To elucidate the molecular events that lead to binding and priming of the spike protein for membrane fusion, we use here molecular dynamics simulations. We first use atomistic molecular dynamics (MD) simulations to systematically derive from "bottom-up" effective coarse-grained (CG) models of spike trimers and ACE2 dimers, which are then validated against conformational state statistics extracted from experimental cryo-EM data. We then use CG MD simulations to investigate the binding between spike trimers and ACE2 receptors, and associated shedding of the S1 domains that expose the S2 trimeric cores. Our simulations indicate that spike trimers cooperatively bind to multiple ACE2 dimers, and show that this multivalent interaction is necessary to productively prime the virus for membrane fusion. At present, such insights are difficult to obtain experimentally.

Results

Our CG MD simulations consist of three constituents, spike trimers, ACE2 dimers, and lipid membranes, each depicted in **Fig. 1A**. We extended our previously reported implicit-solvent, CG spike protein model derived from all-atom molecular dynamics trajectories (16), which were based on the open and closed state atomic models from cryo-EM (Protein Data Bank, PDB: 6VYB, 6VXX) (3). Each spike protomer is modeled as non-covalently interacting S1 and S2 domains with a CG resolution of around 10-15 amino acids per CG "bead" or site, and initially prepared as trimers. Briefly, intra-domain interactions are represented by a heterogeneous effective harmonic bond network while inter-domain interactions are represented by a combination of excluded volume repulsion, short-range attraction, and screened electrostatic terms (see *Methods*). A similar resolution and interaction model are used to represent each ACE2 monomer, initially prepared as dimers. The lipids are represented by a phenomenological CG model with a

bending rigidity of 30 k_BT to emulate the mechanical properties of typical membranes (17). All CG model details and parametrization procedures are described in the *Methods* and *Supplementary Information (SI)* sections.

It is well-known that the RBDs of spike trimers are capable of hinge-like movements, resulting in combinations of open and closed configurational states throughout the trimer (2-5). Therefore, to validate our updated CG spike model, we quantified the configurational sampling of the RBD. We first simulated a single CG spike trimer in membrane (see *Methods*) for 50×10^6 CG timesteps (τ_{CG} = 50 fs). To characterize the motions of the RBD, we quantified the distribution of reciprocal interatomic distances (18) (DRID) for each configuration of the three RBDs; the DRID method computes the first, second, and third moments of the reciprocal distances between each CG site and every other CG site, yielding $3N_{CG}$ features. We next performed dimensional reduction using time-lagged independent component analysis (tICA), a linear projection technique that aims to maximize representation of auto-correlations (19). The first 10 tIC dimensions were clustered using *k*-means clustering (20), revealing three primary conformational states. These three states can be seen in **Fig. 1B**, which depicts the probability distribution of each of the three labeled clusters along the first tIC (tIC_1) dimension. Assessing the root-mean-squared deviation of the configurations from each of the three states (see **Fig. 1D** and **Fig. S2**) reveals that the green-labeled distribution represents "closed" RBDs, the red-labeled distribution represents "open" RBDs, and the blue-labeled distribution represents an intermediate state we denote as "partially closed," which we discuss later. The separability of the open- and closed-state distributions (we include the partially-closed state in the latter) depicted in **Fig. 1B** suggests that tIC_1 can be used as a classifying metric for the open state ($tIC_1 > 0.4$) and closed or partially-closed states ($tIC_1 < 0.4$).

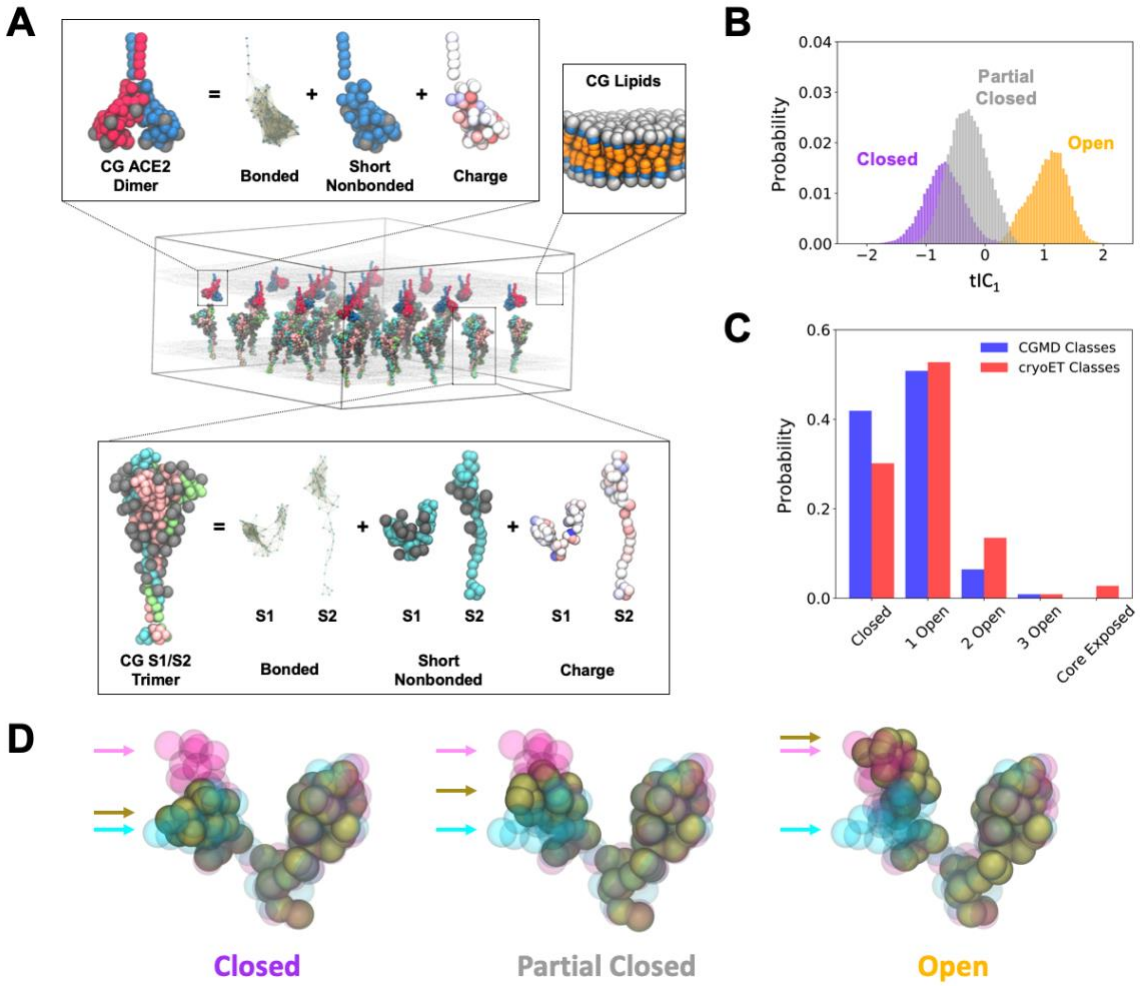

Fig. 1. **Schematic overview and characterization of our coarse-grained (CG) models.** (**A**) Representative depiction of a CG simulation of spike trimers in membrane interacting with an adjacent membrane with ACE2 dimers. The insets depict the CG model components for the spike trimer (bottom), ACE2 dimer (upper left), and lipid membrane (upper right). Note that the protein CG sites are colored by monomer while glycans are represented by grey balls. (**B**) Probability distributions for RBD conformations projected onto the first time-lagged independent component (tIC_1). The purple, grey, and orange distributions denote the three clusters identified by k-means clustering with labels indicating their representative conformational state. (**C**) Comparison of spike trimer conformational state populations (ranging from 0 to 3 open RBDs or shed RBDs, i.e., exposure of the S2 trimeric core) from CG simulations and cryo-ET classification; the core exposed state is in the prefusion form in the CG simulations while it is in the postfusion form in the cryo-ET dataset. (**D**) Representative depiction of the average configuration of CG S1 (brown beads) within the identified k-means clusters. For reference, CG configurations of experimentally-resolved open and closed states of S1 are shown as magenta and cyan beads, respectively. Arrows indicate the positions of each respective RBDs. In all cases, the N-terminal domains of S1 are aligned.

In our prior cryo-ET classification of RBDs within spike trimers along the surface of intact virions, we found that RBD conformations are stochastic and may range from three closed RBDs to three open RBDs, with three closed RBDs and one open (two closed) RBDs representing the two most predominant states (5). Given these observations, we assessed the distribution of open

RBDs within spike trimers as predicted by our CG model. We prepared 25 spike trimers in a flat membrane (around 170×170 nm^2) with periodic boundary conditions and simulated three replicas of the system for 172×10^6 τ_{CG}. The simulated surface density, which is around 1150 nm^2 per spike trimer, was chosen to reflect virions observed using cryo-ET, which had 24 ± 9 spike trimers and an average diameter of 91 nm (5). Using the final 12.5×10^6 τ_{CG}, we characterized the RBDs throughout each of the trimers and computed the population of trimers with 0, 1, 2, and 3 open RBDs. In **Fig. 1C**, we compare the distribution of trimer states from our CG simulations to that of a previously reported cryo-ET dataset (5). Our CG spike trimers undergo a slow yet reversible transition between open and closed states, with around 50% of the total population having 1 open RBD, around 40% having 0 open RBDs (i.e., 3 closed), and diminishing populations of 2 and 3 open RBD trimers. The relative populations of these states are consistent with the cryo-ET classification of RBDs in virions. Note that complete S2 trimeric core exposure (i.e., the precursor to the postfusion state) was not observed throughout these simulations. However, we observed partial exposure in rare cases, i.e., single S1 domains from spike trimers spontaneously dissociated due to thermal fluctuations, which we discuss further later. In contrast, a minority population of postfusion trimers was observed by cryo-ET. At this point, it is unclear whether this discrepancy is due to limitations of the CG model, due to fixation by formaldehyde which may shift the open/closed equilibrium, or both. Nonetheless, our analysis provides qualitative validation that the conformational variability of the RBD in our CG spike trimer is consistent with experimental observations.

Next, to investigate the spike-ACE2 binding process, we simulated spike trimers interfacing with ACE2 dimers. We prepared triplicate simulations of spike trimers in one membrane adjacent to another containing ACE2 dimers (see **Fig. 1A**). The density of spike trimers and ACE2 dimers was fixed such that the relative ratio of ACE2 dimer to spike trimer ($[ACE2_{dimer}]/[S_{trimer}]$) was 2.56, i.e., 64 ACE2 dimers and 25 spike trimers, sufficient to allow multivalent ACE2 binding to each spike trimer. The two membranes (and their constituents) were initially shifted along their normal directions until the closest distance between spike trimers and ACE2 dimers was around 2 nm. The simulations were run for 172×10^6 τ_{CG}.

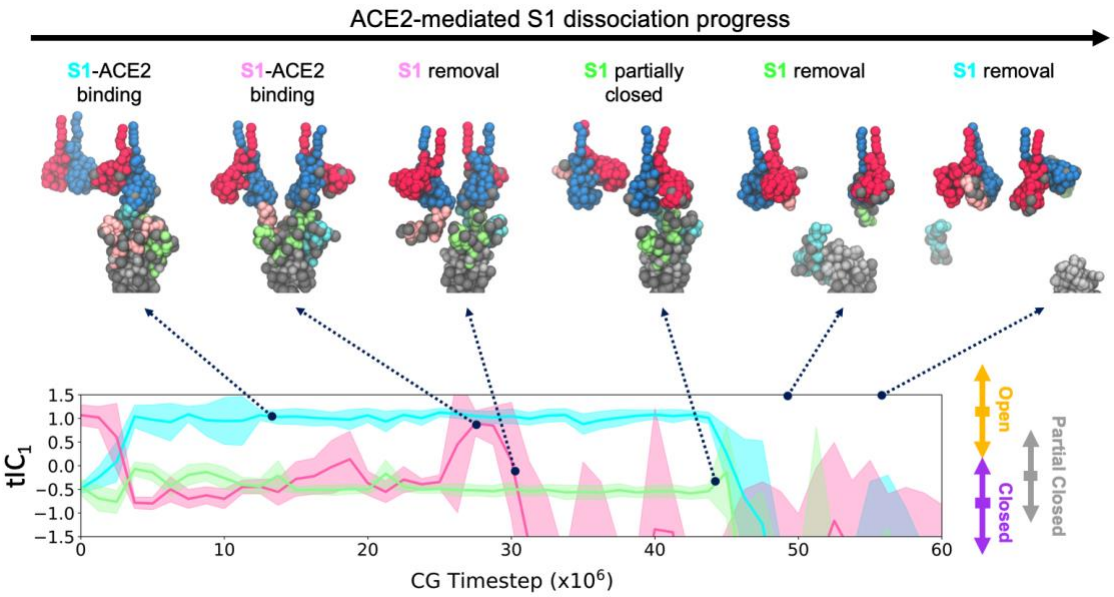

Fig. 2. **Dissociation of S1 facilitated by multivalent ACE2 binding.** (**Top**) Snapshots of one representative S1 dissociation process upon binding of two ACE2 dimers. Each monomer in the ACE2 dimer is represented by red and blue beads, respectively, while each S1 protomer in the spike trimer is represented by cyan, pink, and green beads, respectively. The grey and silver beads denote glycans and S2 protomers, respectively. The depicted process is also shown in **Movie S1**. (**Bottom**) Time series profile depicting tIC_1 evaluated for the RBD of each protomer (sharing the same color) for the spike trimer depicted above, i.e., spike trimer protomers are labeled cyan to pink to green (back to cyan) in counter-clockwise order when viewing from the top-down. To the right of the panel, the double-ended arrows represent the extent of the tIC_1 distributions for each state from **Fig. 1B**, while the rectangle shows the peak of each distribution.

Throughout our simulations, we observed spike trimers fluctuating between the open and closed state, with the former preceding ACE2 binding, as expected. Interestingly, we observed spike trimers that were capable of binding to multiple ACE2 dimers, which required additional opening of RBDs. However, we did not see multiple spike trimers binding to the same ACE2 dimer, suggesting that any multivalent interactions between spike and ACE2 are due to an excess of accessible ACE2 receptors.

Our simulations predict two modalities for ACE2-facilitated dissociation of S1. First, S1 dissociation appears to occur due to thermal fluctuations, which may happen both in the absence and presence of binding between an open RBD and a single ACE2 dimer. The remaining two S1 monomers persist on the spike trimer, although in some cases, complete S1 shedding occurs after secondary RBD opening and binding to a new ACE2 receptor. The second modality was observed more frequently and involves the sequential binding of two ACE2 receptors to the same spike

trimer. Consequently, dissociation of all three S1 domains is accelerated and results in the exposure of the S2 trimeric core (**Movie S1**).

We depict the representative elements of the multivalent ACE2-binding process in **Fig. 2**, in addition to tIC_1 time-series profiles for each RBD (additional examples are shown in **Fig. S3**). Here, the observed effect on the bound spike trimer is one of processive release of S1 domains in a cyclical fashion. The process generally adheres to the following sequential steps. First, RBDs conformationally vary between the open and closed state, as seen for $\tau < 5\times10^6\ \tau_{CG}$ (note the variation of the cyan and pink lines/protomers in **Fig. 2**). One ACE2 dimer successfully binds to the open RBD, pulling the RBD and enforcing the open state, as seen for $\tau = 12\times10^6$ CG (the cyan protomer adopts $tIC_1 = 1.0$ with reduced variance). The cleft formed between the upraised RBD and the N-terminal arm of S1 provides a stabilizing pocket for the clockwise RBD (from a top-down view), or the green protomer as depicted in **Fig. 2** ($tIC_1 = -0.6$), thereby "tightening" the adhesion of this S1 domain. Conversely, the counter-clockwise RBD, or the pink protomer as depicted in **Fig. 2**, appears to "loosen" its adhesion to the same cleft provided by the green S1 domain. Due to the weakened interactions, the pink protomer dynamically adopts the partially-closed state, as seen between $15\times10^6 < \tau < 20\times10^6\ \tau_{CG}$ (pink line/protomer in **Fig. 2** with $tIC_1 = -0.25$). Thermal fluctuations allow the pink protomer to spontaneously adopt the open state ($\tau = 26\times10^6\ \tau_{CG}$). A second ACE2 dimer binds to the newly exposed RBD and "pulls" this S1 domain away from the spike, thereby promoting dissociation ($\tau = 30\times10^6\ \tau_{CG}$). The remaining S1 dimer complex, which remains bound to the first ACE2 dimer, subsequently "loosens" and eventually dissociates due to thermal fluctuations ($\tau = 45\times10^6\ \tau_{CG}$). Finally, the S1-bound ACE2 dimers diffuse away from the spike trimer, which now consists of the exposed S2 trimeric core ($\tau > 50\times10^6\ \tau_{CG}$).

Prior experimental binding assays using soluble ACE2 have demonstrated positive cooperativity when binding to SARS-CoV-2 spike trimers (21). Our simulations show that multivalent ACE2 interactions induce productive S1 dissociation, and further suggest that exposure of the postfusion machinery may be cooperatively correlated with ACE2 expression. To test this hypothesis, we prepared additional triplicate simulations with the same spike trimer density but varied the ACE2 density. Here, we tested stoichiometric ratios of $[ACE2_{dimer}]/[S_{trimer}]$

between 0 and 4, the latter representing 100 ACE2 dimers and 25 spike trimers, and performed simulations for 172×10^6 τ_{CG}. The final 12.5×10^6 τ_{CG} were used for analysis.

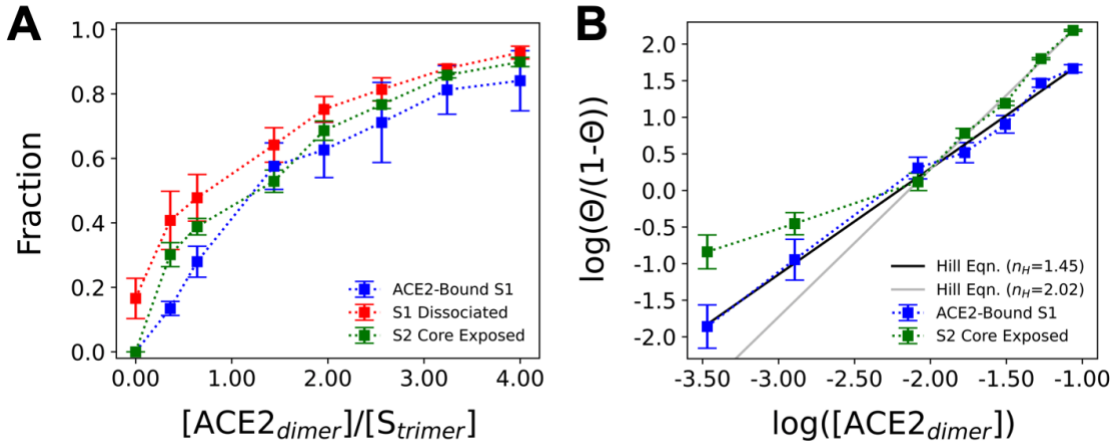

Fig. 3. **Cooperative enhancement of S2 trimeric core exposure by ACE2.** (**A**) Summary statistics for the fraction of S1 monomers bound to ACE2 (blue), S1 monomers dissociated from the spike trimer (red), and spike trimers with complete exposure of the S2 trimeric core (green) as the stoichiometric ratio of ACE2 dimers to spike trimers, $[ACE2_{dimer}]/[S_{trimer}]$, is varied. (**B**) Hill plot comparing the natural log of the surface density of ACE2 (in #/nm^2) to the natural log of the fraction of S1 monomers bound to ACE2 or the fraction of completely exposed S2 trimeric cores (Θ). The black and grey lines are fits to the Hill equation with Hill coefficients (n_H) equal to 1.45 and 2.02. All data points report the mean and standard error from triplicate simulations.

We summarize statistics on the extent of ACE2 binding, S1 dissociation, and S2 trimeric core exposure as $[ACE2_{dimer}]/[S_{trimer}]$ is varied in **Fig. 3A**. We used a distance-based metric (described in *Methods*) to determine both binding and dissociation events. First, we find that S1 monomers are capable of spontaneous dissociation due to thermal fluctuations in the absence of ACE2, as noted above. Within our simulated timescales, however, none of these dissociation events resulted in complete S2 trimeric core exposure. As S1 dissociation was not observed in our atomistic MD simulations, nor during training of the CG model, we speculate that the collective membrane fluctuations, further induced by the spike trimer inclusions, promote S1 dissociation as bound S1 should be considered a metastable state.

The presence of ACE2 has intriguing effects on binding, dissociation, and exposure statistics. To measure the degree of cooperativity, we fit our binding data to the Hill equation:

$$\ln\frac{\Theta}{1-\Theta} = n_H \ln[A] + c$$

where Θ is the fraction of S1 monomers bound to ACE2, $[A]$ is the surface density of ACE2, n_H is the Hill coefficient, and c is a constant. As seen in **Fig. 3B**, we find that $n_H > 1$ (= 1.45) for the

binding curve, indicating that S1 binding to ACE2 is positively cooperative. In **Fig. 3A**, both the extent of S1 dissociation and S2 trimeric core exposure appear to monotonically increase with $[ACE2_{dimer}]/[S_{trimer}]$. Note that the relative ratio between S1 dissociation and S2 core exposure significantly decreases when $[ACE2_{dimer}]/[S_{trimer}]$ ratios exceed 2. In fact, as seen in **Fig. 3B**, the propensity for S2 core exposure when $[ACE2_{dimer}]/[S_{trimer}] > 2$ also exhibits positive cooperativity with $n_H = 2.02$. Together, these results suggest that multivalent ACE2 binding promotes S1 dissociation events that productively expose the S2 trimeric core, and eventually, the postfusion machinery.

Discussion

We developed and applied CG simulation models of SARS-CoV-2 spike trimers and receptor ACE2 dimers to study the dynamical mechanisms involved in viral avidity and dissociation, the latter being a necessary step prior to membrane fusion. Our simulations recapitulate the metastability of the S1 domains, which remain bound to spike trimers after furin cleavage. Although rare, we observed spontaneous dissociation of S1 even in the absence of ACE2 binding, consistent with cryo-EM characterization of postfusion trimers on mature virions (4, 5). It has been shown that the D614G mutation is more infectious and stabilizes interactions between S1 and S2 that prevent dissociation (22), promotes RBD opening (23, 24), and increases ACE2 affinity (25). Ostensibly, the D614G mutation prevents premature S1 dissociation and facilitates productive RBD-ACE2 binding, while ACE2-mediated S1 dissociation remains unimpeded. It would be valuable for future atomistic and CG simulations to ascertain how D614G and related mutations (e.g. the B.1.1.7, B.1.351, B.1.617.2, and P.1 variants) affect spike trimer stability before, during, and after ACE2 binding.

One of our key findings is that the primary binding modality between spike trimers and ACE2 dimers involves multiple ACE2 dimers binding to single spike trimers, rather than multiple spike trimers binding to the same ACE2 dimer, an observation that is consistent with split reporter assays of soluble spike trimers incubated with ACE2 dimers (26). We found that spike trimer binding exhibits positive cooperativity with ACE2 dimer density due to the multivalent nature of ACE2 binding. Soluble ACE2 has also been observed to cooperatively bind to SARS-CoV-

2 spike trimers, yet interestingly, not for SARS-CoV-1 spike trimers (21). Finally, our simulations suggest that the multivalency of the ACE2 to spike trimer interaction is necessary to facilitate complete dissociation of the S1 monomers that occlude the S2 trimeric core, which contains the viral fusion machinery. Binding to ACE2 induces a cyclical tightening and loosening process that processively weakens S1 interactions, both to adjacent S1 monomers and to the underlying S2 core, which leads to S1 dissociation. On the basis of cryo-EM maps of soluble spike trimers bound to soluble ACE2, it was recently proposed that the three RBDs throughout the spike trimer open and bind to ACE2 in a processive fashion, eventually resulting in simultaneous dissociation of the three S1 monomers due to reduced contacts between neighboring monomers (8). The process we observe in our simulations is conceptually similar, although we propose that only two ACE2 dimers are minimally necessary for productive uncoating of S1, and the counter-clockwise S1 (from a top-down view) from the initial ACE2-bound S1 tends to open, bind to a second ACE2, and dissociate first.

One notable difference between the conditions of our simulations and the experiments discussed above is that both our spike trimers and ACE2 dimers are tethered to their respective membranes in a more realistic system than solubilized proteins. The tethers inherently restrict the diffusion and conformational motions available to both spike and ACE2 proteins. One potential consequence is that the relative populations of different spike/ACE2 complexes for solubilized proteins may not reflect populations at the interface between virions and host cells. Likewise, these populations are likely dependent upon the relative stoichiometry of spike trimers and ACE2 dimers, as our simulations indicate. We suggest that the use of fluid membrane-bound spike trimers and ACE2 receptors in biochemical and structural characterization of viral avidity, e.g., as seen in the work of Yang et al. and Lu et al. (27, 28), compared to assays using soluble components, may prove to be characteristically different. For example, single-molecule Förster resonance energy transfer analysis suggests that spike trimers bound to one or two ACE2 receptors exhibit asymmetry throughout the three spike protomers (28), which is consistent with the dynamical behavior observed in our CG simulations.

Our analysis also motivates several directions for future simulations. While researchers have explored the free energy surface of single RBD open-close transitions (29, 30), the

thermodynamics of secondary and tertiary RBD opening, as well as the role of multivalent ACE2 binding, should be thoroughly investigated. We note that our current CG model will also likely require additional features to holistically probe SARS-CoV-2 virion binding, fusion, and entry. Here, we investigated spike trimer and ACE2 dimer association at the interface of two planar membranes. However, it will be important to elucidate the effects of virion curvature (between 70-120 nm diameters (5)), which likely impose additional restrictions to spike trimer conformational sampling while seeking ACE2 receptors. Sophisticated modeling approaches to incorporate large-scale conformational changes, such as the extension of the heptad repeat coiled coil in the postfusion state (6), will also be necessary. In this work, we linearly mixed two model Hamiltonians that represent different configurational states (open and closed RBDs). Several additional approaches, including plastic network models (31) and multi-configurational coarse-graining (32), may prove useful to construct CG models capable of state transitions with increased accuracy for state populations and transition rates. These efforts will expand the capabilities of CG modeling, which in turn will provide valuable insight into collective molecular phenomena during SARS-CoV-2 infection, such as the role of cooperative multivalent ACE2 binding to induce spike trimer uncoating reported herein.

Methods

Below, we describe the computational methods used throughout this work. Additional details on the CG model and parameterization procedure are provided in the *Supplementary Information*.

Atomistic MD simulations. Trajectories from prior atomistic MD simulations of the spike trimer in membrane were used to derive effective CG Hamiltonians (16, 33). Statistics for the open and closed state of the spike trimer were supplemented with additional triplicate simulations of the spike ectodomain using PDB 6VYB and 6VXX as initial structures, respectively (3). Triplicate simulations of the ACE2 dimer interacting with bound RBDs were also prepared using an initial structure from PDB 6M17 (7). All simulations were prepared using CHARMM-GUI (34) and the CHARMM36m (35) force field. Each system was solvated with 2 nm of TIP3P water from the

protein edge to the simulation domain boundary and with 150 mM NaCl. All simulations were performed using GROMACS 2020 (36). Minimization was performed using steepest descent until the maximum force was reduced to 1,000 kJ/mol/nm^2. Then, equilibration was performed in several phases. First, 10 ns were integrated in the constant *NVT* ensemble using the stochastic velocity rescaling thermostat (37) with a damping time of 0.2 ps and a timestep of 1 fs. During this phase, the Cα backbone of the protein was harmonically restrained with a force constant of 1,000 kJ/mol/nm^2. An additional 40 ns were integrated in the constant *NPT* ensemble using the stochastic velocity rescaling thermostat (37) (2 ps damping time) and the Parrinello-Rahman barostat (38) (10 ps damping time) and a timestep of 2 fs. Finally, an additional 1650 ns were integrated in the constant *NVT* ensemble using the Nose-Hoover chain thermostat (39) (2 ps damping time) and a timestep of 2 fs. Throughout this procedure, H-containing bonds were constrained using the LINCS algorithm (40). Statistics were gathered every 100 ps and the final 1,200 ns of each trajectory was used for CG model generation.

CG modeling. CG model generation for the protein constituents proceeded in three phases. The first phase determined the mapping between atomistic and CG sites. Mapping was performed using essential dynamics coarse-graining (EDCG) (41) with resolutions chosen to balance between computational efficiency and least-squared error in the represented principal component subspace. Monomers of S1, S2, and ACE2 were mapped to 60, 50, and 70 CG sites, respectively; each N-linked glycan (42) was mapped to a single site, totaling 13, 9, and 7 sites, respectively.

The second phase determined the effective CG Hamiltonian for each protein. We systematically determined parameters for S1/S2 in the open state, S1/S2 in the closed state, and ACE2 in the bound state, as well as their inter-domain interactions. The Hamiltonian (E_{tot}) was decomposed into four terms: intra-domain fluctuations (E_{intra}), inter-domain exclusion (E_{excl}), inter-domain attraction (E_{attr}), and screened electrostatics (E_{yukawa}):

$$E_{tot} = E_{intra} + E_{excl} + E_{attr} + E_{yukawa}$$

where E_{intra} was represented as a hetero-elastic network model (hENM) (43) with bond energies $k\ (r - r_0)^2$ where k is the spring constant of a particular bond and r_0 is the equilibrium bond length. These parameters were optimized using the hENM method with a cutoff distance of 3, 4,

and 3 nm for S1, S2, and ACE2, respectively. For E_{excl}, a soft cosine potential, $A\left[1+\cos\frac{\pi r}{r_c}\right]$, was used, where A = 25 kcal/mol and r_c is the onset for excluded volume determined from pair correlation functions. For E_{attr}, the sum of two Gaussian potentials, $A_1 \exp\left[-\frac{(r_{ij}-r_1)^2}{2\sigma_1^2}\right] + A_2 \exp\left[-\frac{(r_{ij}-r_2)^2}{2\sigma_2^2}\right]$, was used, where $r_{1/2}$ and $\sigma_{1/2}$ are the mean and standard deviation determined by a fit to the pair correlation between CG sites i and j through least-squares regression. The constants A_1 and A_2 were optimized using relative-entropy minimization (REM) (44). For E_{yukawa}, a Yukawa potential, $\frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}}\exp(-\kappa\, r_{ij})$, was used, where q_i is the aggregate charge of CG site i, κ = 1.274 nm^{-1} is the inverse Debye length for 150 mM NaCl, and ε_r is the effective dielectric constant of the protein environment, approximated as 17.5 (45).

The third phase determined an effective Hamiltonian for the spike trimer that was capable of continuously transitioning between the open and closed state. Here, we modified E_{intra}, E_{excl}, and E_{attr} as follows. For E_{intra}, we kept bonds that appeared in both open and closed states and used the mean k and r_0 constants. For E_{excl}, we used the minimum r_c from both open and closed states. For E_{attr}, we tested different linear combinations of the Gaussians parametrized from the open and closed state. We found that an equal mix recapitulated open/closed statistics the best, i.e., $E_{attr} = 0.5E_{attr,open} + 0.5E_{attr,closed}$. No modification of E_{yukawa} was needed since this term is not state dependent.

The lipid model used the functional form from Ref. (17) with tuned parameters listed in the Supplementary Information. For the purposes of modeling the interaction between the transmembrane domains of the spike trimer/ACE2 dimer and lipids, we effectively considered the transmembrane domain as having the same interaction as the lipids. Complete details on protein and lipid model parametrization are provided in Supplementary Information.

Model parameters and input files are available on Github: https://github.com/uchicago-voth/MSCG-models/tree/master/Spike_ACE2

CG simulations. All simulations were prepared using PACKMOL and Moltemplate (46, 47). Briefly, CG spike trimers and ACE2 dimers were arranged in a grid using each density described in the

main text. Two CG lipid membranes were randomly packed around the transmembrane domains of each grid of proteins in a 200×200 nm^2 lateral (*xy*) domain; the *z* dimension of the box was fixed to 160 nm. The two membranes were then translated along the *z* direction until the minimum *z* distance between spike trimers and ACE2 dimers was 2 nm. All simulations were performed using LAMMPS (48). Conjugate gradient energy minimization was performed on each system until the change in force was less than 10^{-6}. Then, proteins were held fixed while lipids were integrated using a Langevin thermostat at 310 K (0.5 ps damping constant) and a Berendsen barostat at 1 atm (0.5 ps damping constant, applied over the lateral *xy* dimension) over 2×10^6 τ_{CG} (with τ_{CG} = 50 fs). During this process, the *xy* simulation domain compressed to around 170×170 nm^2. We then integrated both proteins and lipids using a Langevin thermostat at 310 K (5 ps damping constant) over 170×10^6 τ_{CG}. Trajectory snapshots were saved every 25×10^3 τ_{CG}.

Analysis. All analysis was performed using a combination of custom Tcl scripts using VMD (49) and Python scripts using the MSMBuilder (50) and MDTraj (51) libraries. The configurational state of each RBD was featurized using the DRID method (18). Dimensional reduction (via tICA (19)) of this feature set into 10 tICs was performed using a lag time of 40×10^3 τ_{CG}; the values of the first tIC eigenvector are shown in **Fig. S1**. The data were then clustered into 3 sets using k-means clustering (20). We featurized and projected the configuration of each RBD extracted from our CG simulations onto the first tIC to classify its state. We also used the following distance metric to classify the bound state of the S1 domain to both its partner S2 domain and ACE2:

$$\left| \frac{1}{N_{ii}} \sum_{ii}^{N_{ii}} r_{S1,ii} - \frac{1}{N_{jj}} \sum_{jj}^{N_{jj}} r_{S2(ACE2),jj} \right| = d_{S1-S2(ACE2)}$$

where *ii* and *jj* refer to CG site indices out of a set with N_{ii} and N_{jj} total sites, *r* is the position, *d* is the distance, and | | is the norm of the vector. In other words, we compute the distance between the centers of geometry of sets of CG sites. We chose CG site clusters that were observed to have close association distances with minimal fluctuation during CG model training. For S1/S2 binding, these included site indices 31-32 and 54-57 for S1 and 77, 87-89, 98-99, and 102 for S2. For S1/ACE2 binding, these included site indices 43-49 for S1 and 134-137, 158, and 162 for ACE2 (residue mapping for each site is shown in Supplementary Tables 1-3). Using this metric, we

respectively classified S1/S2 and S1/ACE2 as bound states when $d_{S1\text{-}S2}$ and $d_{S1\text{-}ACE2}$ were less than 4 nm, respectively.

Acknowledgements

This work was supported in part by the National Science Foundation through NSF RAPID grant CHE-2029092 (A.J.P and G.A.V). This study was also supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-CoG-648432 MEMBRANEFUSION to J.A.G.B.) and the Medical Research Council as part of UK Research and Innovation (MC_UP_1201/16 to J.A.G.B.). Alexander J. Pak gratefully acknowledges support from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under grant F32 AI150477. Alvin Yu gratefully acknowledges support from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under grant F32 AI150208. This work used resources, services, and support provided via the COVID-19 HPC Consortium (https://covid19-hpc-consortium.org/), which is a unique private-public effort to bring together government, industry, and academic leaders who are volunteering free compute time and resources in support of COVID-19 research. The authors acknowledge the use of the Witherspoon cluster at IBM Research that has contributed to this work. This research also used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The authors also acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC and visualization resources that have contributed to the research results reported within this paper.

References

1. J. Shang *et al.*, Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences* **117**, 11727 (2020).
2. D. Wrapp *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260 (2020).
3. A. C. Walls *et al.*, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **183**, 281-292 (2020).
4. Y. Cai *et al.*, Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586 (2020).
5. Z. Ke *et al.*, Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498-502 (2020).
6. A. C. Walls *et al.*, Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proceedings of the National Academy of Sciences* **114**, 11157 (2017).
7. R. Yan *et al.*, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444 (2020).
8. D. J. Benton *et al.*, Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, 327-330 (2020).
9. T. Xiao *et al.*, A trimeric human angiotensin-converting enzyme 2 as an anti-SARS-CoV-2 agent. *Nature Structural & Molecular Biology* **28**, 202-209 (2021).
10. B. Turoňová *et al.*, In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203 (2020).
11. E. P. Barros *et al.*, The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection. *Biophysical Journal* **120**, 1072-1084.
12. P. V. Raghuvamsi *et al.*, SARS-CoV-2 S protein:ACE2 interaction reveals novel allosteric targets. *eLife* **10**, e63646 (2021).
13. C. Xu *et al.*, Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science Advances* **7**, eabe5575 (2021).
14. D. Zhou *et al.*, Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348-2361.e2346 (2021).
15. P. Supasa *et al.*, Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* **184**, 2201-2211.e2207 (2021).
16. A. Yu *et al.*, A Multiscale Coarse-Grained Model of the SARS-CoV-2 Virion. *Biophysical Journal* **120**, 1097-1104 (2020).
17. J. M. A. Grime, J. J. Madsen, Efficient Simulation of Tunable Lipid Assemblies Across Scales and Resolutions. *arXiv:1910.05362v1* (2019).
18. T. Zhou, A. Caflisch, Distribution of Reciprocal of Interatomic Distances: A Fast Structural Metric. *Journal of Chemical Theory and Computation* **8**, 2930-2937 (2012).
19. Y. Naritomi, S. Fuchigami, Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. *The Journal of Chemical Physics* **134**, 065101 (2011).
20. S. Lloyd, Least squares quantization in PCM. *IEEE Transactions on Information Theory* **28**, 129-137 (1982).

21. S. P. Anand *et al.*, Interaction of Human ACE2 to Membrane-Bound SARS-CoV-1 and SARS-CoV-2 S Glycoproteins. *Viruses* **12** (2020).
22. L. Zhang *et al.*, SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature Communications* **11**, 6013 (2020).
23. R. A. Mansbach *et al.*, The SARS-CoV-2 Spike variant D614G favors an open conformational state. *Science Advances* **7**, eabf3671 (2021).
24. D. J. Benton *et al.*, The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. *Proceedings of the National Academy of Sciences* **118**, e2022586118 (2021).
25. S. Ozono *et al.*, SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nature Communications* **12**, 848 (2021).
26. I. Lui *et al.*, Trimeric SARS-CoV-2 Spike interacts with dimeric ACE2 with limited intra-Spike avidity. *bioRxiv* 10.1101/2020.05.21.109157 (2020).
27. J. Yang *et al.*, Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor. *Nature Communications* **11**, 4541 (2020).
28. M. Lu *et al.*, Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles. *Cell Host & Microbe* **28**, 880-891.e888 (2020).
29. M. Gur *et al.*, Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *The Journal of Chemical Physics* **153**, 075101 (2020).
30. L. Casalino *et al.*, AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. *The International Journal of High Performance Computing Applications* 10.1177/10943420211006452, 10943420211006452 (2021).
31. P. Maragakis, M. Karplus, Large Amplitude Conformational Change in Proteins Explored with a Plastic Network Model: Adenylate Kinase. *Journal of Molecular Biology* **352**, 807-822 (2005).
32. M. E. Sharp, F. X. Vázquez, J. W. Wagner, T. Dannenhoffer-Lafage, G. A. Voth, Multiconfigurational Coarse-Grained Molecular Dynamics. *Journal of Chemical Theory and Computation* **15**, 3306-3315 (2019).
33. L. Casalino *et al.*, Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* **6**, 1722-1734 (2020).
34. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859-1865 (2008).
35. J. Huang *et al.*, CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71 (2016).
36. M. J. Abraham *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
37. G. Bussi, T. Zykova-Timan, M. Parrinello, Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **130**, 074101 (2009).
38. M. Parrinello, A. Rahman, Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **45**, 1196-1199 (1980).
39. G. J. Martyna, M. L. Klein, M. Tuckerman, Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635-2643 (1992).
40. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463-1472 (1997).

41. Z. Zhang *et al.*, A Systematic Methodology for Defining Coarse-Grained Sites in Large Biomolecules. *Biophysical Journal* **95**, 5073-5083 (2008).
42. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330-333 (2020).
43. E. Lyman, J. Pfaendtner, G. A. Voth, Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophysical Journal* **95**, 4183-4192 (2008).
44. M. S. Shell, The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics* **129**, 144108 (2008).
45. L. Li, C. Li, Z. Zhang, E. Alexov, On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *Journal of Chemical Theory and Computation* **9**, 2126-2136 (2013).
46. L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **30**, 2157-2164 (2009).
47. A. I. Jewett, Z. Zhuang, J.-E. Shea, Moltemplate a Coarse-Grained Model Assembly Tool. *Biophysical Journal* **104**, 169a (2013).
48. S. Plimpton, Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **117**, 1-19 (1995).
49. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graphics* **14**, 33-38 (1996).
50. K. A. Beauchamp *et al.*, MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J Chem Theory Comput* **7**, 3412-3419 (2011).
51. Robert T. McGibbon *et al.*, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528-1532 (2015).