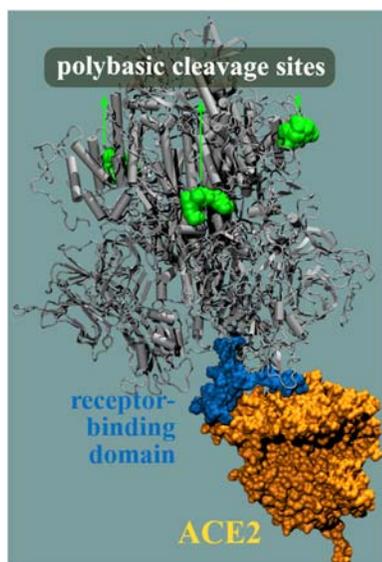# The Distal Polybasic Cleavage Sites of SARS-CoV-2 Spike Protein Enhance Spike Protein-ACE2 Binding

Baofu Qiao and Monica Olvera de la Cruz [*]

The receptor-binding domain (RBD) of the SARS-CoV-2 spike protein plays a crucial role in binding the human cell receptor ACE2 that is required for viral entry. Many studies have been conducted to target the structures of RBD-ACE2 binding and to design RBD-targeting vaccines and drugs. Nevertheless, mutations distal from the SARS-CoV-2 RBD also impact its transmissibility and antibody can target non-RBD regions, suggesting the incomplete role of the RBD region in the spike protein-ACE2 binding. Here, in order to elucidate distant binding mechanisms, we analyze complexes of ACE2 with the wild type spike protein and with key mutants via large-scale all-atom explicit solvent molecular dynamics simulations. We find that though distributed approximately 10 nm away from the RBD, the SARS-CoV-2 polybasic cleavage sites enhance, via electrostatic interactions and hydration, the RBD-ACE2 binding affinity. A negatively charged tetrapeptide (GluGluLeuGlu) is then designed to neutralize the positively charged arginine on the polybasic cleavage sites. We find that the tetrapeptide GluGluLeuGlu binds to one of the three polybasic cleavage sites of the SARS-CoV-2 spike protein lessening by 34% the RBD-ACE2 binding strength. This significant binding energy reduction demonstrates the feasibility to neutralize RBD-ACE2 binding by targeting this specific polybasic cleavage site. Our work enhances understanding of the binding mechanism of SARS-CoV-2 to ACE2, which may aid the design of therapeutics for COVID-19 infection.

TOC: The SARS-CoV-2 spike protein-ACE2 complex showing the polybasic cleavage sites

Keywords: COVID-19, SARS-CoV-2, polybasic cleavage sites, peptide inhibitor, molecular dynamics simulations

The outbreak of the SARS-CoV-2 related disease COVID-19 has caused more than 6 million infections and over 0.3 million deaths globally. The SARS-CoV-2 spike protein exhibits two notable features; its receptor-binding domain (RBD) is optimized to bind the human cell receptor angiotensin-converting enzyme 2 (ACE2) and each subunit of the spike protein trimer has a polybasic cleavage site. [1] Given the structural similarity between the spike proteins of the SARS-CoV-2 and its close relative SARS-CoV (2003), the RBD of the SARS-CoV-2 has been quickly recognized in targeting the ACE2 receptor, [2-4] which is required for subsequent viral entry. The atomic structures of the contact region between the SARS-CoV-2 RBD and ACE2 have been experimentally obtained by different groups. [5-7] Owing to the crucial role of the SARS-CoV-2 RBD, numerous vaccine and drug candidates have been proposed to target the RBD to inhibit the RBD-ACE2 binding. [8-10]

However, more research is beginning to show that mutations distal from the RBD (non-RBD mutations) of SARS-CoV-2 also influence its transmissibility. For instance, Walls et al. [2] found that by mutating the $T_{678}NSPRRAR_{685}$ residues, which are distributed around 10 nm from the SARS-CoV-2 RBD, to a variant with T_{678}--IL--R_{685}, the transduction efficiency decreased in human ACE2-expressing Baby Hamster Kidney cells. The mutant D614G (a single change in the genetic code; D = aspartic acid, G = glycine) of the SARS-CoV-2 spike protein, which began spreading in Europe in February, displayed stronger transmissibility. [11] The D_{614} and G_{614} residues are located about 7 - 10 nm from the SARS-CoV-2 RBD. Surprisingly, Yuan et al. [12] found that the antibody CR3022 isolated from a recovered SARS-CoV patient targeted a highly conserved epitope of SARS-CoV-2 and SARS-CoV, which is also distal from their RBDs. That is, the antibody CR3022 does not directly bind the spike protein RBD. All these works support the significance of non-RBD mutations and non-RBD-targeting in vaccine and drug design, which remains elusive thus far.

In the present work, large scale all-atom explicit solvent molecular dynamics (MD) simulations reveal that the spike protein polybasic cleavage sites, which are distributed approximately 10 nm away from the RBD, can enhance the binding affinity between the SARS-CoV-2 RBD and ACE2. This information is used to design a negatively charged tetrapeptide, GluGluLeuGlu, that targets the polybasic cleavage sites. The negatively charged glutamic acids (Glu) are introduced to charge neutralize the polybasic cleavage sites and the nonpolar leucine (Leu) is included to increase the hydrophobicity of the tetrapeptide, which favors binding to the polybasic cleavage site. The tetrapeptide GluGluLeuGlu is found to bind to one specific polybasic cleavage site, resulting in a remarkable weakening (34%) of the spike protein-ACE2 binding energy. Our work provides guidelines to design therapeutic peptides to inhibit SARS-CoV-2 RBD-ACE2 binding.
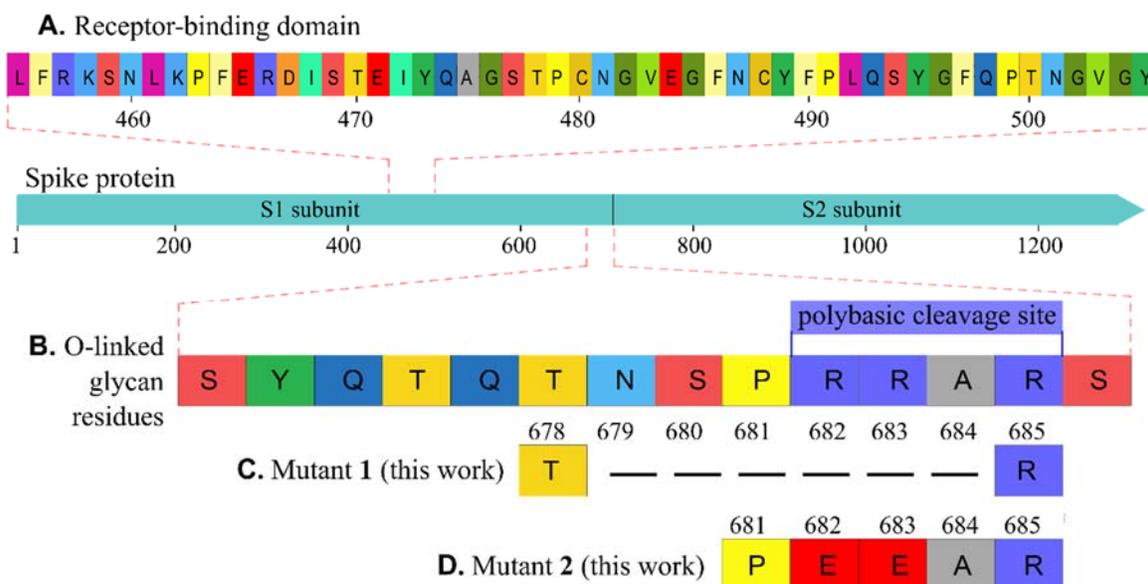
A. Receptor-binding domain

L F R K S N L K P F E R D I S T E I Y Q A G S T P C N G V E G F N C Y F P L Q S Y G F Q P T N G V G Y

460   470   480   490   500

Spike protein

S1 subunit   S2 subunit

1   200   400   600   800   1000   1200

polybasic cleavage site

B. O-linked glycan residues

S	Y	Q	T	Q	T	N	S	P	R	R	A	R	S
						678	679	680	681	682	683	684	685

C. Mutant 1 (this work)   T   — — — — — —   R

681   682   683   684   685

D. Mutant 2 (this work)   P   E   E   A   R

Figure 1. Wild type SARS-CoV-2 spike protein and the two mutants investigated here. (**A**) Receptor binding domain from L_{456} to Y_{505}. [1] (**B**) O-linked glycan residues from S_{673} to S_{686}, among which $P_{681}RRA_{684}$ is unique to SARS-CoV-2 compared to the other lineage B coronavirus. [2, 13, 14] $R_{682}RAR_{685}$ form a polybasic cleavage site. (**C**) Mutant **1,** where $N_{679}SPRRA_{684}$ residues were removed, is similar to the experimental variant [2] and (**D**) mutant **2** ($E_{682}E_{683}$), both investigated in this work.

RESULTS AND DISCUSSION

Impact of the Polybasic Cleavage Site on RBD-ACE2 Binding Affinity.

Each subunit of the SARS-CoV-2 trimeric spike protein has one polybasic cleavage site ($R_{682}RAR_{685}$). [1] The polybasic cleavage sites ($R_{682}RAR_{685}$) have not been reported for SARS-CoV or any other lineage B coronaviruses, and therefore, they are thought to be unique to SARS-CoV-2. [2, 13, 14] In comparison to the extensive studies carried out on the SARS-CoV-2 RBD, the polybasic cleavage sites have thus far not been widely investigated, and their function remains elusive. They are believed to be related to the viral transmissibility of SARS-CoV-2. [2, 14] Moreover, they have been found to be essential for spike protein-driven viral entry into lung cells [15].

To elucidate the influence of the polybasic cleavage sites, the wild type (WT) spike protein-ACE2 complex was investigated along with two mutants (**Figure 1**). The SARS-CoV-2 spike protein-ACE2 complex was downloaded from the Zhang-Server, [16] which contains the full length of the wild type SARS-CoV-2 spike protein bound to the ACE2 receptor. The spike protein-ACE2 complex was solvated in a water box with a length of $16×18×24$ nm^3, where 0.15 M NaCl was included. All-atom MD

simulation was conducted to equilibrate the system using the CHARMM 36m force field. [17] To preserve the spike protein-ACE2 binding structure the non-hydrogen atoms of the spike protein trimer and ACE2 were restrained, which were removed in the following simulations.

The WT spike protein was subsequently mutated. For mutant **1**, the $N_{679}SPRRA_{684}$ residues were removed for all the three subunits of the spike protein trimer. This mutant is similar to the variant experimentally investigated (T_{678}--IL--R_{685}) [2]. In the case of mutant **2**, the $R_{682}R_{683}$ residues were changed to $E_{682}E_{683}$ for all three subunits of the trimeric spike protein, where E stands for negatively charged glutamic acid. For each of the three systems (WT and mutants **1** and **2**) all-atom explicit solvent simulations were performed for a duration of 100 ns. And five parallel runs were carried out based on the simulation configurations at 20, 40, 60, 80, and 100 ns. See the Methods section. The five parallel simulations were employed to calculate the binding energy (Table S2 in the Supporting Information) and intermolecular hydrogen bonds between the SARS-CoV-2 RBD and ACE2.
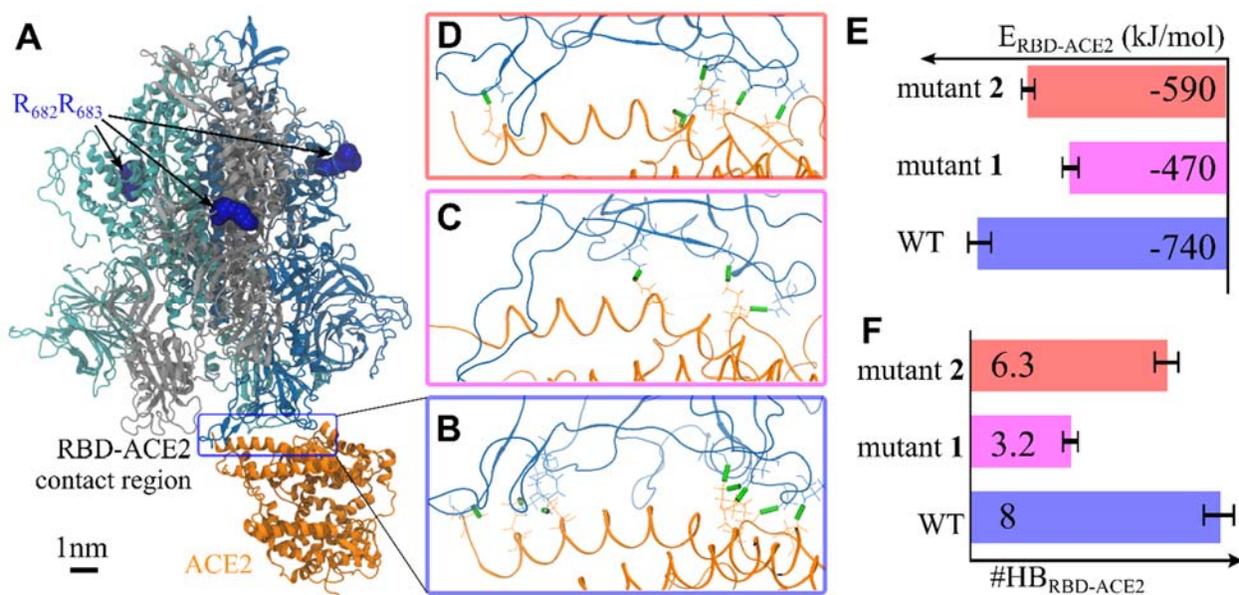

Figure 2. Distal mutants 1 and 2 both weaken the binding affinity between the RBD and ACE2. (A) Final simulation snapshot of the WT spike protein (in tan/gray/dark blue) and ACE2 (in orange). The $R_{682}R_{683}$ residues (blue beads) on the polybasic cleavage sites are approximately 10 - 13 nm away from the RBD. Water and salt ions are omitted for display. **(B)** The contact region between the WT spike protein (in dark blue) and ACE2 (in orange), where the RBD-ACE2 intermolecular hydrogen bonds are highlighted by green sticks. **(C, D)** are similar to **(B)**, but for the mutants **1** and **2**, respectively. **(E)** The potential energy between the SARS-CoV-2 RBD and ACE2. See Table S2 for detailed values. **(F)** The number of intermolecular hydrogen bond (HB) between the RBD and ACE2. In (E, F), the averages and standard deviations are from five parallel runs (**Figure 4**).

Demonstrated in **Figure 2A** is the final simulation snapshot of the wild type spike protein-ACE2 complex. The polybasic cleavage sites are located 10 - 13 nm away from the RBD. In all three systems, the SARS-CoV-2 RBD stayed bound to ACE2 (**Figure 2B-D**). Nevertheless, the potential energy between the RBD and ACE2, which was the summation of the intermolecular short-range Coulomb and Lennard-Jones interactions, exhibited remarkable differences (**Figure 2E**). The RBD-ACE2 binding energy for the WT spike protein is -740 ± 70 kJ/mol. The removal of $N_{679}SPRRA_{684}$ compromises the RBD-ACE2 interaction by 36% to -470 ± 50 kJ/mol, supporting that mutant **1** heavily depresses the binding affinity between the RBD and ACE2. This is in line with the experimental finding that the mutation of $T_{678}NSPRRAR_{685}$ to a variant with T_{678}--IL--R_{685} of a murine leukemia virus decreased the transduction efficiency of human ACE2-expressing Baby Hamster Kidney cells. [2] Meanwhile, the mutation from $R_{682}R_{683}$ to $E_{682}E_{683}$ decreased the RBD-ACE2 binding energy by 20% to -590 ± 40 kJ/mol. Therefore, both mutants greatly destabilized the RBD-ACE2 binding in comparison to the WT spike protein.

Note that ACE2 is highly negatively charged (-28 e), which leads to a highly negatively charged extracellular membrane surface. The Coulomb interactions favor the adsorption of positively charged species, for instance, arginine residues. Also of note is that due to the positively charged nature, arginine has been found to enhance the cellular uptake of peptides [18] and promote the therapeutic delivery of peptides across the blood-brain barrier for Alzheimer's disease [19]. Therefore, the observed drops in the RBD-ACE2 binding energy for mutants **1** and **2** could be primarily ascribed to a change in (long-range) Coulomb interactions between ACE2 and the spike protein (Table S2), even though the polybasic cleavage sites are around 10 nm away from the RBD and ACE2. Moreover, the substitution of positively charged arginine with negatively charged glutamic acid increases protein hydration (Table S2), which is consistent with the fact that negatively charged amino acids exhibit stronger hydration than positively charged ones. [20-22]

Further analysis of the intermolecular hydrogen bonds between the RBD and ACE2 demonstrated a similar influence of the mutants (**Figure 2F**). For the WT spike protein, there exist 8 ± 1 hydrogen bonds between the RBD and ACE2. This is in good agreement with the experimental finding [5] that eight RBD-ACE2 intermolecular polar interactions were suggested based on the cryo-EM structure. For mutant **1** the number of RBD-ACE2 intermolecular hydrogen bonds decreased by 60% to 3.2 ± 0.5, and by 21% (to 6.3 ± 0.8) for mutant **2**. These results further support that both mutants weaken the RBD-ACE2 binding affinity, in line with the calculated RBD-ACE2 intermolecular interaction energy.

Visualization of the distribution of the RBD-ACE2 intermolecular hydrogen bonds (**Figure 2B-D**) suggested that balanced hydrogen bonds at both ends of the RBD-ACE2 contact region play a crucial

role in stabilizing the RBD-ACE2 binding. Specifically, the number of the intermolecular hydrogen bonds decreases at both termini from the WT spike protein to the mutant **2** and to the mutant **1**.

Design of Oligopeptide Inhibitor.

Interestingly, when we were preparing this work, an antibody CR3022 [12] was reported to bind a non-RBD region of the spike protein. Here, we explore the possibility to inhibit SARS-CoV-2 RBD-ACE2 binding by neutralizing the polybasic cleavage sites. To this end, we designed an oligopeptide: a tetrapeptide GluGluLeuGlu (EELE). Negatively charged glutamic acids (Glu) were included to neutralize the positively charged arginine residues at the polybasic cleavage sites. The hydrophobic leucine (Leu) residue was introduced to decrease the hydration, as a test simulation on a tetrapeptide EEEE showed that EEEE became quickly dissolved (within 10 ns) owing to the high hydration of the negatively charged amino acids. [20-22]
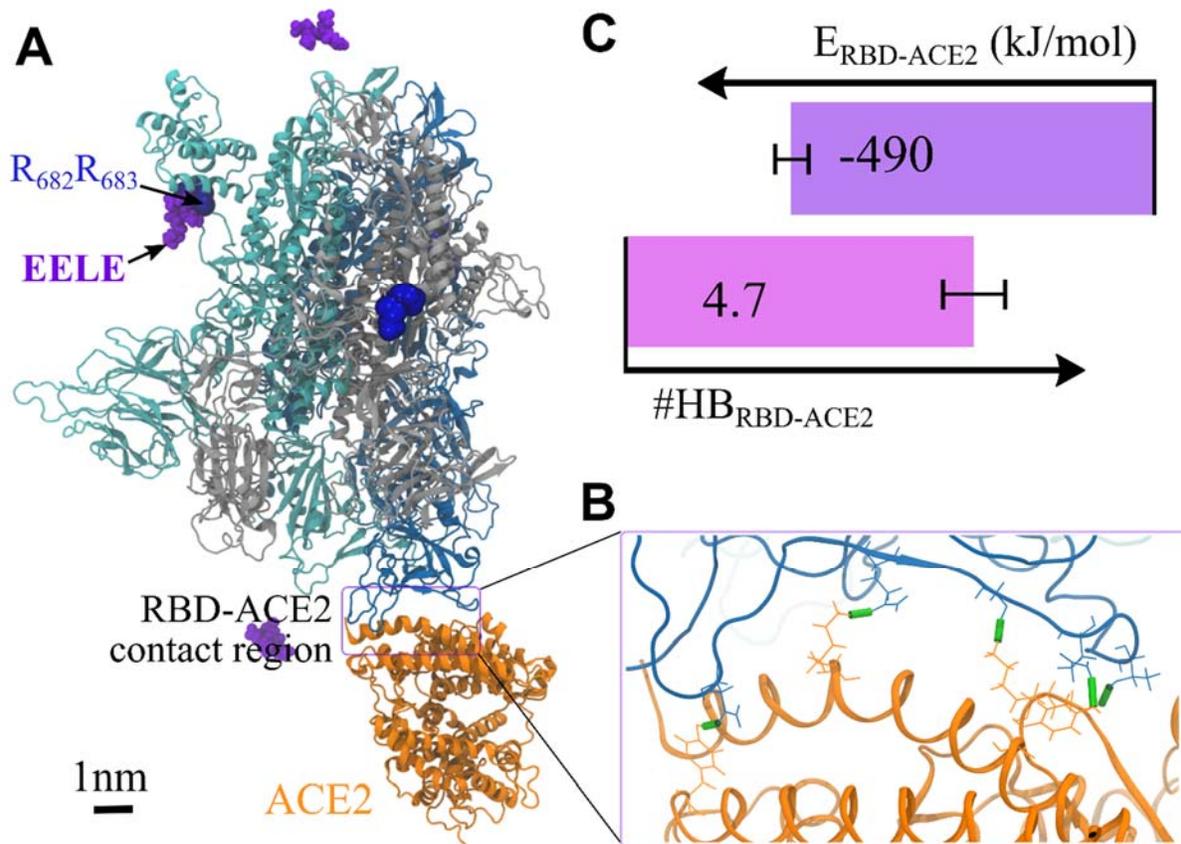

Figure 3 The presence of tetrapeptide EELE lessens the RBD-ACE2 binding. (**A**) Simulation snapshot of a spike-ACE2 complex in the presence of EELE (in violent). (**B**) The RBD-ACE2 contact region, where the RBD-ACE2 intermolecular hydrogen bonds are highlighted via green sticks. (**C**) Potential energy and number of hydrogen bonds between the RBD and ACE2. In (C), the averages and standard deviations are from five parallel runs (**Figure 4**).

Three EELE molecules were initially put next to the three polybasic sites of the trimeric spike protein. The system was equilibrated using the same process as that in the simulations in the absence of the tetrapeptide. One typical configuration is presented in **Figure 3A**. The inclusion of the tetrapeptide EELE can still preserve the spike protein-ACE2 complex (**Figure 3A, B**). Nevertheless, the RBD-ACE2 intermolecular potential energy (**Figure 3C**) remarkably dropped by 34% from -740 ± 70 kJ/mol without EELE to -490 ± 50 kJ/mol in the presence of EELE. Similarly, the number of RBD-ACE2 intermolecular hydrogen bonds dropped by 41%. Therefore, the presence of EELE greatly lessened RBD-ACE2 binding.

Of note is that only one subunit of the trimeric spike protein directly binds to the ACE2 receptor. Consequently, the three polybasic cleavage sites have different distances to the ACE2 receptor and distinct local environment. Furthermore, the ACE2 receptor and the tetrapeptide EELE are both negatively charged ($-28e$ for ACE2 and $-3e$ for EELE). These effects collectively lead to the observation that the three tetrapeptides exhibited different binding behavior to the neighboring polybasic cleavage sites. Specifically, the polybasic cleavage site distributed the farthest from ACE2 stably binds to the tetrapeptide EELE for the whole simulation duration of 100 ns (**Figure 3A**); in contrast, the two polybasic cleavage sites closer to ACE2 form weaker interactions with their neighboring tetrapeptides EELE, which became unbound at around 40 ns and 74 ns.

To experimentally design polybasic cleavage site-targeting clinical therapeutic peptides, a further increase in the oligopeptide hydrophobicity is required to elevate the stability in the SARS-CoV-2 spike protein-oligopeptide binding to determine if binding to all of the polybasic cleavage sites reduces the RBD-ACE2 overall binding energy even further. Meanwhile, the delivery of therapeutic peptides is known to be challenging concerning their short half-life due to the rapid proteolytic degradation and short circulation time due to the low molecular weight. [23, 24] These could be potentially overcome by covalently grafting oligopeptides at the side chains of peptide brush polymers, [25] or integrating them onto peptide amphiphiles [26] or PEGylation [27].

CONCLUSIONS

Using large-scale all-atom explicit solvent simulations, we investigated the impact of the SARS-CoV-2 polybasic cleavage sites, which are distributed approximately 10 nm away from the RBD, on the binding affinity of the RBD and ACE2. It is found that in comparison to the wild type SARS-CoV-2 spike protein, a mutant with the deletion of $N_{679}SPRRA_{684}$ and a mutant with the substitution of $R_{682}R_{683}$ to $E_{682}E_{683}$ can both lessen the binding strength of the SARS-CoV-2 RBD and ACE2. The mutation-driven difference is ascribed to the electrostatic interactions between the spike proteins (wild type and mutants)

and ACE2 and their hydration. In line with recent experimental findings, [2, 11] this work supports that distal mutations can impact the SARS-CoV-2 RBD-ACE2 binding affinity. Our design of a tetrapeptide, GluGluLeuGlu, that binds to the polybasic cleavage site of SARS-CoV-2, demonstrates that the polybasic cleavage site is a target for neutralizing SAR-CoV-2 RBD-ACE2 binding. This supports a recent finding that the non-RBD region of SARS-CoV-2 can be targetted by an antibody. [12] This work besides shedding light on the mechanism by which the SARS-CoV-2 spike protein binds to human cells, suggests therapeutic peptides design to target the polybasic cleavage sites that inhibit SARS-CoV-2 RBD binding to ACE2.

METHODS

All-atom Simulations on the Wild Type SARS-CoV-2 Spike Protein–ACE2 Complex.

All-atom explicit solvent molecular dynamics (MD) simulations were performed using the package GROMACS (version 2016.4). [28] The CHARMM 36m potential [17] was employed for the proteins, Na^+ and Cl^- ions. The recommended CHARMM TIP3P water model [29] was employed with the structures constrained using the SETTLE algorithm [30].

The SARS-CoV-2 spike protein-ACE2 binding structure was downloaded from the Zhang-Server. [16] The spike protein-ACE2 complex was reconstructed using the C-I-TASSER model [31] based on the protein identification number QHD43416 [32] for the spike protein. Each subunit of the trimeric spike protein included the residues from M_1 to T_{1273} with the net charge of $-7e$, and the ACE2 has the residues from S_{19} to D_{615} with the net charge of $-28e$, with e being the elementary charge.

The spike protein-ACE2 complex was first solvated in a simulation box with an edge length of $16\times18\times24$ nm^3 (**Figure 4**). 0.15 M NaCl was added along with 49 Na^+ counterions of spike protein and ACE2. The system composition is listed in Table S1 in the Supporting Information. The energy of the simulation box was first minimized, which was followed by a simulation of 1 ps using the canonical ensemble (constant number of particles, volume, and temperature, NVT). The integration time step of 2 fs was employed with all the hydrogen-involved covalent bonds constrained using the LINCS algorithm [33, 34]. Another simulation of 1 ps was subsequently conducted using the isothermal-isobaric ensemble (constant number of particles, pressure, and temperature, NPT). The velocity rescale temperature coupling was employed along with the Berendsen pressure coupling. In the following equilibration simulation of 10 ns, the Nosé-Hoover temperature coupling was applied along with the Parrinello-Rahman pressure coupling.[35] In all the equilibration simulations above, the coordinates of the non-hydrogen atoms of both

the spike protein trimer and ACE2 were restrained using a force constant of 1000 kJ/mol/nm^2 to preserve the spike protein-ACE2 binding structure. The other parameters were the same as those in the production simulation below.
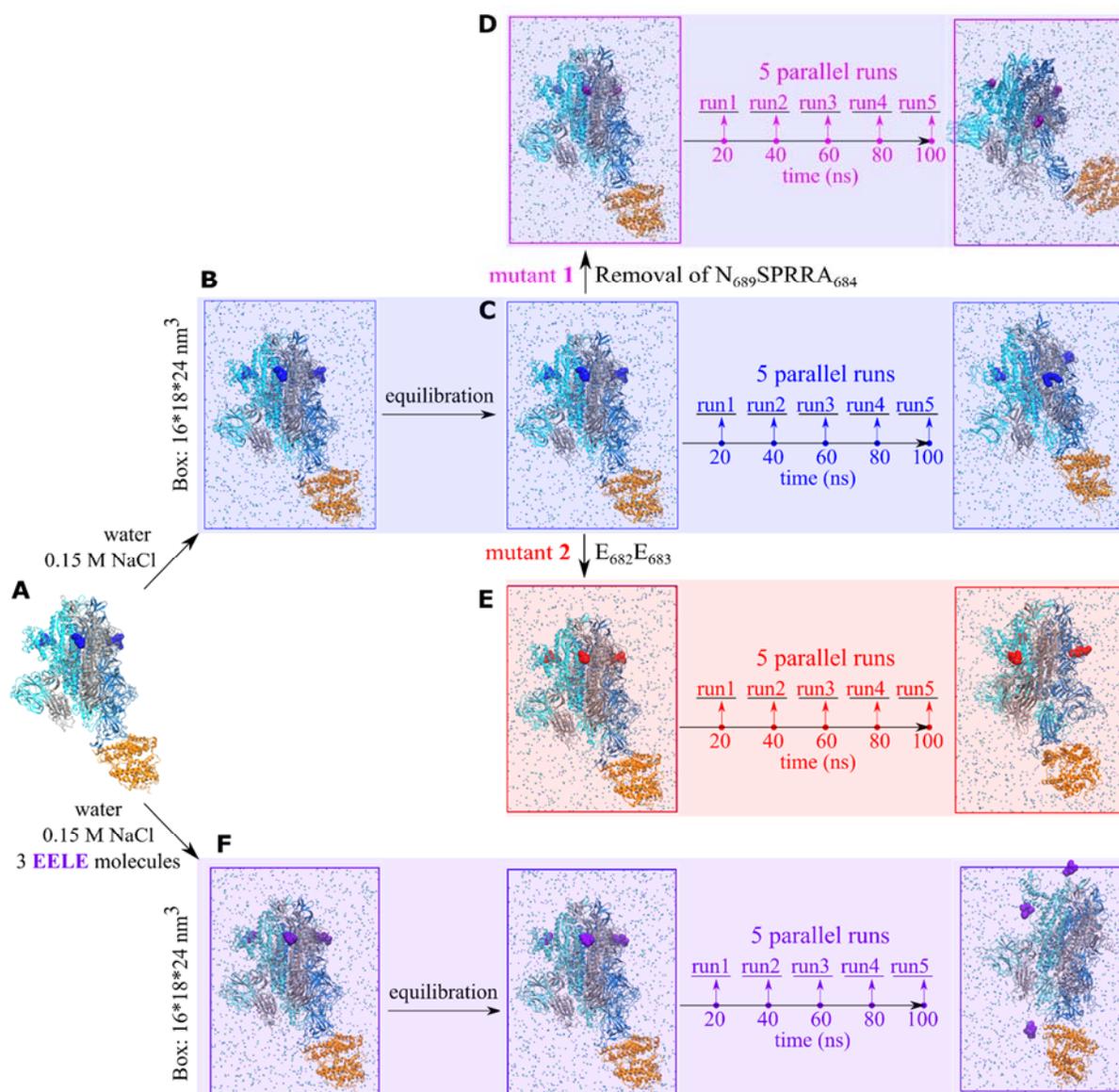

Figure 4. Simulation method in the absence (B-E) and presence (F) of the tetrapeptides EELE. The spike protein-ACE2 complex (**A**) was solvated in a simulation box with 0.15 M NaCl (**B**). The spike protein trimer is colored cyan, silver, and blue, respectively, and ACE2 in orange. Na$^+$ and Cl$^-$ ions are displayed using small blue and cyan beads, respectively, with water omitted for display. The polybasic cleavage sites (R$_{682}$RAR$_{685}$) are highlighted with bigger blue beads. The system was equilibrated (**C**). The WT spike protein was mutated by removing N$_{679}$SPRRA$_{684}$ (mutant **1** with R$_{685}$ in purple; **D**) and changing R$_{682}$R$_{683}$ to E$_{682}$E$_{683}$ (mutant **2** with E$_{682}$EAR$_{685}$ in red; **E**). Each simulation lasted 100 ns (**C**, **D**, **E**). Five parallel runs (10 ns each) were performed based on the configurations at 20, 40, 60, 80, and 100 ns. Same process was applied to the system with EELE tetrapeptides (**F**).

In the production simulation (**Figure 4C**), the periodic boundary conditions were employed in all three dimensions. The neighbor searching was conducted up to 12 Å using the Verlet particle-based method and was updated every 20-time steps. The Lennard-Jones (LJ) 12-6 interactions were switched off from 10 to 12 Å via the potential-switch method in GROMACS. The short-range Coulomb interactions were truncated at the cut-off distance of 12 Å, and the long-range interactions were calculated using the Smooth Particle Mesh Ewald (PME) algorithm [36, 37]. The temperatures of proteins, ions, and water were separately coupled using the Nosé-Hover algorithm (reference temperature 298K, characteristic time 1 ps). The isotropic Parrinello-Rahman barostat was employed with the reference pressure of 1 bar, the characteristic time 4 ps, and the compressibility 4.5×10^{-5} bar^{-1}. All the covalent bonds were constrained, which supported an integration time step of 2.5 fs. These parameters were recommended for the accurate reproduction of the original CHARMM simulation on lipid membrane system [38] and were verified in simulations on proteins [20, 39, 40] and lipid membranes [41]. The production simulation lasted 100 ns. Calculation of the root-mean-square deviations of the RBD and ACE2 backbone atoms (**Fig S1** in the Supporting Information) supported that the structures of the RBD and ACE2 were both preserved except that mutant **1** displayed a relatively higher fluctuation. Based on the configurations at 20, 40, 60, 80, and 100 ns, 5 parallel simulations were performed, each of which run 10 ns.

The system had around 700,000 atoms and was computationally expensive. The performance was around 14 ns/day with 8 high-performance computing nodes, each with 24 CPU cores (Intel Haswell E5-2680, 2.5 GHz, 2 × 9.6 GT/s Intel® QPI, 2500 MHz).

The short-range Coulomb and LJ interaction energies (**Table S2** in the Supporting Information) were calculated using the GROMACS program *gmx energy*. The intermolecular hydrogen bonds were calculated using the GROMACS program *gmx hbond*. The structural criteria of hydrogen bond were applied that the donor (D) – acceptor (A) distance $r_{DA} \leq 3.5$ Å and the hydrogen−donor−acceptor angle $\theta_{HDA} \leq 30°$.[42, 43]

Mutants 1 and 2 of SARS-CoV-2 Spike Protein.

Based on the equilibrated structure, the wild type spike protein was mutated. For Mutant **1**, the $N_{679}SPRRA_{684}$ residues were deleted from all the three subunits of the spike protein trimer. This was to mimic the experimentally reported mutant by Walls et al. [2] that the residues of $T_{678}NSPRRAR_{685}$ in the wild type spike protein of SARS-CoV-2 were mutated to T_{678}--IL--R_{685}. Accordingly, six Cl$^-$ ions, which were distributed next to the six arginine residues of $R_{682}R_{683}$ were removed to neutralize the system. The system was further equilibrated to relax the mutated region, which was followed by the production

simulations. The simulation parameters were the same as those in the simulation on the wild type spike protein. Similarly, five parallel simulations were carried out based on the production simulation of 100 ns (**Figure 4D**).

For Mutant **2**, the $R_{682}R_{683}$ residues were changed to $E_{682}E_{683}$ for all the three subunits of the spike protein trimer. Accordingly, the six Cl^- ions, which were distributed next to the six arginine residues ($R_{682}R_{683}$) were replaced with Na^+ ions to neutralize the system. Again, the simulation parameters were the same as those in the simulation on the wild type spike protein. The production simulation lasted 100 ns, where five parallel simulations were performed (**Figure 4E**).

SARS-CoV-2 Spike Protein-ACE2 Complex in the Presence of Tetrapeptide EELE.

Three tetrapeptides GluGluLeuGlu (EELE) were included, which were initially put next to the three polybasic cleavage sites ($R_{682}RAR_{685}$) of the trimeric wild type spike protein (**Figure 4F**). The spike protein-ACE2 complex and the tetrapeptide EELE were then solved in a simulation box with an edge length of $16 \times 18 \times 24$ nm^3. 0.15 M NaCl was added along with 9 Na^+ counterions of EELE and 49 Na^+ counterions of spike protein and ACE2 (**Table S1** in the Supporting Information). The equilibration and productions were the same as whose for the wild type spike protein in the absence of the tetrapeptide. The calculated RBD-ACE2 intermolecular Coulomb and Lennard-Jones interactions are presented in **Table S2** in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

The following files are available free of charge at http://pubs.acs.org/doi/XXX.

> Supporting figure of root-mean-square deviations of the RBD and ACE2 backbone atoms. Supporting tables of system compositions and detailed interaction energies.

AUTHOR INFORMATION

Corresponding Author

Monica Olvera de la Cruz - Department of Materials Science and Engineering, Department of Chemical & Biological Engineering, Department of Chemistry, and Department of Physics and

Astronomy, Northwestern University, Evanston, IL 60208, USA. Orcid.gov/0000-0002-9802-3627; E-mail: m-olver@northwestern.edu.

Authors

Baofu Qiao - Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA. Orcid.gov/0000-0001-8870-5985.

Author Contributions

B.Q. developed, performed, and analyzed the simulations. B.Q. and M.O.d.l.C. designed the research, analyzed data, and wrote the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

References

1. Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F., The Proximal Origin of SARS-CoV-2. *Nat. Med.* **2020,** *26*, 450-452.
2. Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020,** *181*, 281-292.e6.
3. Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S., Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science* **2020,** *367*, 1260-1263.
4. Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F., Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **2020,** *94*, e00127-20.
5. Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q., Structural Basis for the Recognition of SARS-CoV-2 by Full-length Human ACE2. *Science* **2020,** *367*, 1444-1448.
6. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F., Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature* **2020,** *581*, 221-224.
7. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X., Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020,** *581*, 215-220.

8.   Pinto, D.; Park, Y.-J.; Beltramello, M.; Walls, A. C.; Tortorici, M. A.; Bianchi, S.; Jaconi, S.; Culap, K.; Zatta, F.; De Marco, A.; Peter, A.; Guarino, B.; Spreafico, R.; Cameroni, E.; Case, J. B.; Chen, R. E.; Havenar-Daughton, C.; Snell, G.; Telenti, A.; Virgin, H. W.; Lanzavecchia, A.; Diamond, M. S.; Fink, K.; Veesler, D.; Corti, D., Cross-neutralization of SARS-CoV-2 by a Human Monoclonal SARS-CoV Antibody. *Nature* **2020**.

9.   Zhang, G.; Pomplun, S.; Loftis, A. R.; Loas, A.; Pentelute, B. L., The First-in-class Peptide Binder to the SARS-CoV-2 Spike Protein. *bioRxiv* **2020**, 2020.03.19.999318.

10.   Han, Y.; Král, P., Computational Design of ACE2-Based Peptide Inhibitors of SARS-CoV-2. *ACS Nano* **2020**, *14*, 5143-5147.

11.   Korber, B.; Fischer, W.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Foley, B.; Giorgi, E.; Bhattacharya, T.; Parker, M.; Partridge, D.; Evans, C.; Freeman, T.; de Silva, T.; LaBranche, C.; Montefiori, D., Spike Mutation Pipeline Reveals the Emergence of a more Transmissible Form of SARS-CoV-2. *bioRxiv* **2020**, 2020.04.29.069054.

12.   Yuan, M.; Wu, N. C.; Zhu, X.; Lee, C.-C. D.; So, R. T. Y.; Lv, H.; Mok, C. K. P.; Wilson, I. A., A Highly Conserved Cryptic Epitope in the Receptor Binding Domains of SARS-CoV-2 and SARS-CoV. *Science* **2020**, *368*, 630-633.

13.   Coutard, B.; Valle, C.; de Lamballerie, X.; Canard, B.; Seidah, N. G.; Decroly, E., The Spike Glycoprotein of the new Coronavirus 2019-nCoV Contains a Furin-like Cleavage Site Absent in CoV of the Same Clade. *Antivir. Res.* **2020**, *176*, 104742.

14.   Wang, Q.; Qiu, Y.; Li, J.-Y.; Zhou, Z.-J.; Liao, C.-H.; Ge, X.-Y., A Unique Protease Cleavage Site Predicted in the Spike Protein of the Novel Pneumonia Coronavirus (2019-nCoV) Potentially Related to Viral Transmissibility. *Virol. Sin.* **2020**, 1-3.

15.   Hoffmann, M.; Kleine-Weber, H.; Pöhlmann, S., A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell* **2020**, *78*, 779-784.e5.

16.   Zhang, C.; Zheng, W.; Huang, X.; Bell, E. W.; Zhou, X.; Zhang, Y., Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique Similarity between Its Spike Protein Insertions and HIV-1. *J. Proteome Res.* **2020**, *19*, 1351-1360.

17.   Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell Jr, A. D., CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Meth.* **2017**, *14*, 71-73.

18.   Blum, A. P.; Kammeyer, J. K.; Gianneschi, N. C., Activating Peptides for Cellular Uptake via Polymerization into High Density Brushes. *Chem. Sci.* **2016**, *7*, 989-994.

19.   Wong, K. H.; Riaz, M. K.; Xie, Y.; Zhang, X.; Liu, Q.; Chen, H.; Bian, Z.; Chen, X.; Lu, A.; Yang, Z., Review of Current Strategies for Delivering Alzheimer's Disease Drugs across the Blood-Brain Barrier. *Int. J. Mol. Sci.* **2019**, *20*, 381.

20.   Qiao, B.; Jiménez-Ángeles, F.; Nguyen, T. D.; Olvera de la Cruz, M., Water Follows Polar and Nonpolar Protein Surface Domains. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 19274-19281.

21.   Kramer, Ryan M.; Shende, Varad R.; Motl, N.; Pace, C. N.; Scholtz, J. M., Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. *Biophys. J.* **2012**, *102*, 1907-1915.

22.   Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L.-Y., A Simple Method for Improving Protein Solubility and Long-Term Stability. *J. Am. Chem. Soc.* **2004**, *126*, 8933-8939.

23.   Lau, J. L.; Dunn, M. K., Therapeutic Peptides: Historical Perspectives, Current Development Trends, and Future Directions. *Biorg. Med. Chem.* **2018**, *26*, 2700-2707.

24.   Fosgerau, K.; Hoffmann, T., Peptide Therapeutics: Current Status and Future Directions. *Drug Discovery Today* **2015**, *20*, 122-128.

25.   Blum, A. P.; Kammeyer, J. K.; Yin, J.; Crystal, D. T.; Rush, A. M.; Gilson, M. K.; Gianneschi, N. C., Peptides Displayed as High Density Brush Polymers Resist Proteolysis and Retain Bioactivity. *J. Am. Chem. Soc.* **2014**, *136*, 15422-15437.

26.   Sato, K.; Hendricks, M. P.; Palmer, L. C.; Stupp, S. I., Peptide Supramolecular Materials for Therapeutics. *Chem. Soc. Rev.* **2018,** *47*, 7539-7551.

27.   Harris, J. M.; Chess, R. B., Effect of PEGylation on Pharmaceuticals. *Nature Reviews Drug Discovery* **2003,** *2*, 214-221.

28.   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008,** *4*, 435-447.

29.   MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998,** *102*, 3586-3616.

30.   Miyamoto, S.; Kollman, P. A., SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992,** *13*, 952-962.

31.   Zheng, W.; Li, Y.; Zhang, C.; Pearce, R.; Mortuza, S. M.; Zhang, Y., Deep-learning contact-map guided protein structure prediction in CASP13. *Proteins* **2019,** *87*, 1149-1164.

32.   Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M.-L.; Zhang, Y.-L.; Dai, F.-H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E. C.; Zhang, Y.-Z., A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *579*, 265-269.

33.   Hess, B., P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008,** *4*, 116-122.

34.   Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997,** *18*, 1463-1472.

35.   Parrinello, M.; Rahman, A., Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981,** *52*, 7182-7190.

36.   Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

37.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995,** *103*, 8577-93.

38.   Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016,** *12*, 405-413.

39.   Qiao, B.; Lopez, L.; Olvera de la Cruz, M., "Mirror"-like Protein Dimers Stabilized by Local Heterogeneity at Protein Surfaces. *J. Phys. Chem. B* **2019,** *123*, 3907-3915.

40.   Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera de la Cruz, M.; Xu, T., Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018,** *359*, 1239-1243.

41.   Jiang, T.; Hall, A.; Eres, M.; Hemmatian, Z.; Qiao, B.; Zhou, Y.; Ruan, Z.; Couse, A. D.; Heller, W. T.; Huang, H.; Olvera de la Cruz, M.; Rolandi, M.; Xu, T., Single Chain Heteropolymers Transport Proton Selectively and Rapidly. *Nature* **2020,** *577*, 216-220.

42.   Luzar, A.; Chandler, D., Structure and Hydrogen Bond Dynamics of Water-Dimethyl Sulfoxide Mixtures by Computer Simulations. *J. Chem. Phys.* **1993,** *98*, 8160-8173.

43.   Qiao, B.; Demars, T.; Olvera de la Cruz, M.; Ellis, R. J., How Hydrogen Bonds Affect the Growth of Reverse Micelles around Coordinating Metal Ions. *J. Phys. Chem. Lett.* **2014,** *5*, 1440-1444.