

# Enhanced Binding of SARS-CoV-2 Spike Protein to Receptor by Distal Polybasic Cleavage Sites

Baofu Qiao and Monica Olvera de la Cruz\*

Cite This: <https://dx.doi.org/10.1021/acsnano.0c04798>

Read Online

ACCESS |



Metrics &amp; More

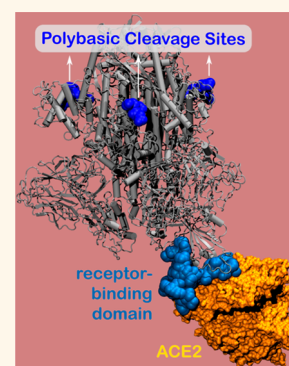


Article Recommendations



Supporting Information

**ABSTRACT:** 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.



**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 11 million infections and over 0.5 million deaths globally as of July 4, 2020. 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.<sup>1</sup> 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,<sup>2–4</sup> 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.<sup>5–7</sup> 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.<sup>8–10</sup>

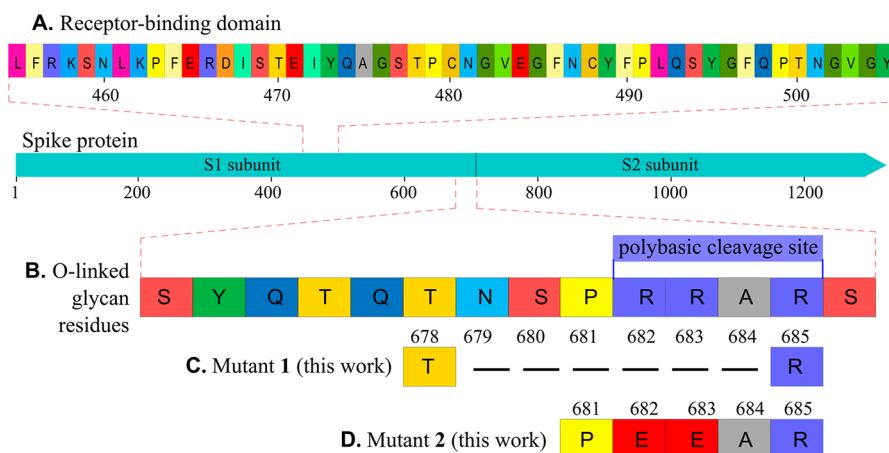
However, more experimental research is beginning to show that mutations distal from the RBD (non-RBD mutations) of SARS-CoV-2 also influence its transmissibility. For instance, Walls *et al.*<sup>2</sup> found that by mutating the T<sub>678</sub>NSPRRAR<sub>685</sub> residues, which are distributed around 10 nm from the SARS-

CoV-2 RBD to a variant with T<sub>678</sub>--IL--R<sub>685</sub>, 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 early February and became the dominant form globally at the end of March, displayed stronger transmissibility.<sup>11</sup> The D<sub>614</sub> and G<sub>614</sub> residues are located about 7–10 nm from the SARS-CoV-2 RBD. Surprisingly, antibodies were also experimentally reported to bind to non-RBD sites of SARS-CoV-2 spike protein. For instance, antibody CR3022<sup>12</sup> 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. Antibody 47D11<sup>13</sup> was also reported to bind

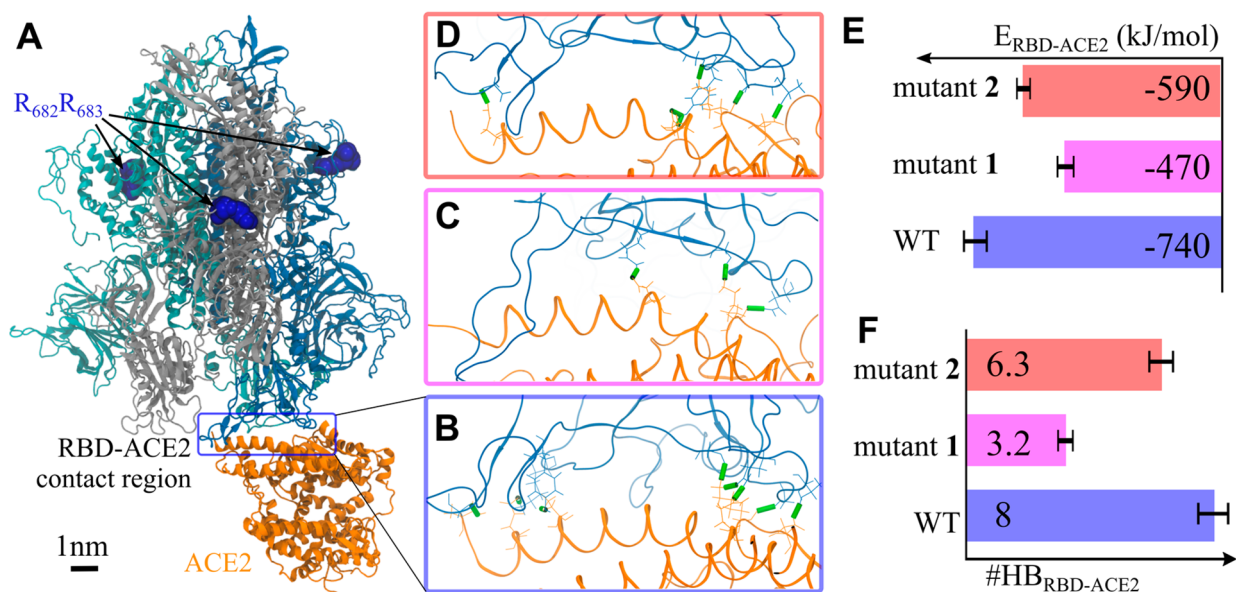
**Received:** June 8, 2020

**Accepted:** August 1, 2020

**Published:** August 3, 2020



**Figure 1.** Wild-type SARS-CoV-2 spike protein and the two mutants investigated here. (A) Receptor binding domain from L<sub>455</sub> to Y<sub>505</sub>.<sup>1</sup> (B) O-linked glycan residues from S<sub>673</sub> to S<sub>686</sub>, among which P<sub>681</sub>RRA<sub>684</sub> is unique to SARS-CoV-2 compared to the other lineage B coronavirus.<sup>2,15,16</sup> R<sub>682</sub>RAR<sub>685</sub> form a polybasic cleavage site. (C) Mutant 1, where N<sub>679</sub>SPRRA<sub>684</sub> residues were removed, is similar to the experimental variant<sup>2</sup> and (D) mutant 2 (E<sub>682</sub>E<sub>683</sub>), both investigated in this work.



**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 trimer (in tan/gray/dark blue) and ACE2 (in orange). The R<sub>682</sub>R<sub>683</sub> 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) Contact region between the WT spike protein (in dark blue) and ACE2 (in orange), where the eight 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) Potential energy between the SARS-CoV-2 RBD and ACE2. See Table S2 for detailed values. (F) Number of intermolecular hydrogen bonds (HB) between the RBD and ACE2. In E and F, the averages and standard deviations are from five parallel runs (Figure 4).

to non-RBD sites of both SARS-CoV-2 and SARS-CoV. The very recently reported antibody 4A8<sup>14</sup> binds to the N-terminal domains of SARS-CoV-2 spike protein, which are located approximately 4–8 nm from the SARS-CoV-2 RBD–ACE2 contact region. 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 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.

## 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<sub>682</sub>RAR<sub>685</sub>).<sup>1</sup>

The polybasic cleavage sites ( $R_{682}R_{685}$ ) 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.<sup>2,15,16</sup> 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.<sup>2,16</sup> Moreover, they have been found to be essential for spike protein-driven viral entry into lung cells.<sup>17</sup>

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,<sup>18</sup> which contains residues 1–1273 of the wild-type SARS-CoV-2 spike protein bound to the ACE2 receptor (residues 19–615). The spike protein–ACE2 complex was solvated in a water box with a length of  $16 \times 18 \times 24 \text{ nm}^3$ , where 0.15 M NaCl was included. All-atom MD simulation was conducted to equilibrate the system using the CHARMM 36m force field.<sup>19</sup> To preserve the spike protein–ACE2 binding structure in the equilibration simulations the non-hydrogen atoms of the spike protein trimer and ACE2 were restrained, which were removed in the following production 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 ( $T_{678}\text{-IL--}R_{685}$ ).<sup>2</sup> 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. 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.

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 observable differences (Figure 2E). The RBD–ACE2 binding energy for the WT spike protein is  $-740 \pm 70 \text{ kJ/mol}$ . The removal of  $N_{679}SPRRA_{684}$  compromises the RBD–ACE2 interaction by 36% to  $-470 \pm 50 \text{ kJ/mol}$ , 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  $T_{678}\text{-IL--}R_{685}$  of a murine leukemia virus decreased the transduction efficiency of human ACE2-expressing Baby Hamster Kidney cells.<sup>2</sup> Meanwhile, the mutation from  $R_{682}R_{683}$  to  $E_{682}E_{683}$  decreased the RBD–ACE2 binding energy by 20% to  $-590 \pm 40 \text{ kJ/mol}$ . Therefore, both mutants greatly destabilized the RBD–ACE2 binding in comparison to the WT spike protein.

Note that a more accurate understanding of the spike protein–ACE2 binding requires binding free energy calcu-

lations using, for instance, free energy perturbation, thermodynamic integration, and/or umbrella sampling approaches.<sup>20</sup> However, these methods are also approximate and highly computationally costly due to the large size of the spike protein–ACE2 complex and the flexibility of their contact region. By assuming that the potential energy plays a dominant role in the free energy difference for the wild-type spike protein–ACE2 complex and the mutant-ACE2 complexes, the difference in the binding potential energies (Figure 2E) can then be used to estimate the differences in the binding free energies. This approximation has also been employed to quantify the binding affinity between SARS-CoV-2 spike protein RBD and a variety of peptides in designing RBD-targeting therapeutic peptides.<sup>10</sup> This approach is useful when comparing systems that do not reach a thermodynamic limit due to limitations on the size of the simulation boxes, since enthalpic differences and entropy differences are generally small and dominated by fluctuations when comparing systems structurally similar, such as the WT with the mutants 1 and 2 (even hydration effect differences have been shown recently to be mainly due to difference in potential energy;<sup>21</sup> in our study, for example, the difference in potential energy between the WT and mutant 1 is  $108 k_B T$ , which is much larger than the thermal energy ( $k_B T$ ) at  $T = 300 \text{ K}$  and mainly coming from hydration effects as can be seen in Table S2 in the Supporting Information). These estimates provide insights on the mechanism responsible for the surprising mutation effect at such a large distance from the binding site, as discussed below.

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<sup>22</sup> and promote the therapeutic delivery of peptides across the blood–brain barrier for Alzheimer’s disease.<sup>23</sup> 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 as evidenced in Table S2 in the Supporting Information, 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.<sup>24–26</sup> Our recent work<sup>24</sup> also demonstrated the existence of long-range hydration<sup>27</sup> at protein surfaces. Specifically, negatively charged surface amino acids orient neighboring water molecules up to around 16 Å from the surface of protein PETase (net charge of  $+6e$ ), and such correlation persists up to around 12 Å for positively charged surface amino acids away from the protein surface. These correlations in water orientation strongly affect Coulomb interactions *via* a decrease in the effective dielectric permittivity ( $\epsilon_r$ ) at protein surfaces (from 80 in bulk water to 2 in confinement by hydrophobic surfaces<sup>28</sup> and to 10 in water confined by polar surfaces<sup>29</sup>) that increase the strength of the Coulomb interactions and reduces the Debye screening length since the salt ions become strongly correlated in low dielectric permittivity media.<sup>30</sup> A simple estimate using a dielectric permittivity reduction from 80 to 8 up to  $\sim 1 \text{ nm}$  from the protein surface changes the Coulomb energy ( $E$ ) over thermal energy ( $k_B T$ ) between two elementary

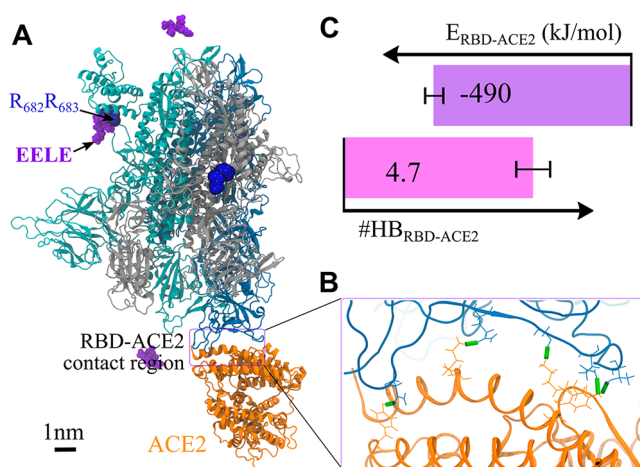
charges ( $e$ ) separated by a distance  $d$  [ $E/k_B T = e^2/(4\pi\epsilon_0\epsilon_r d)/k_B T = l_B/d$ ] from 0.7 to 7 when  $d = 1$  nm, also strongly reducing the Debye length from the value in water ( $\sim 1$  nm in bulk water with 150 mM of NaCl). An increase in protein's absolute net charge, which drives stronger interfacial correlations,<sup>24</sup> therefore, explains the long-range impact of the polybasic cleavage sites on the RBD–ACE2 binding given the fact that the SARS-CoV-2 spike protein trimer and ACE2 are both highly negatively charged with a net charge of  $-21e$  and  $-28e$ , respectively.

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  $8 \pm 1$  hydrogen bonds between the RBD and ACE2. This is in good agreement with the experimental finding<sup>5</sup> 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  $3.2 \pm 0.5$  and by 21% (to  $6.3 \pm 0.8$ ) 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 intermolecular hydrogen bonds decreases at both termini from the WT spike protein to the mutant 2 and to the mutant 1. Furthermore, MARTINI coarse-grained simulations using the MARTINI 2.2 potential<sup>31</sup> and the elastic network model<sup>32</sup> in preserving protein structures also supported that both mutations destabilized the RBD–ACE2 binding. Specifically, in comparison to the wild-type spike protein–ACE2 complex, the RBD–ACE2 binding structure for both mutants became deformed (Figure S2C–E in the Supporting Information), the RBD backbone particles of both mutants displayed larger fluctuations (Figure S2F in the Supporting Information), and the RBD–ACE2 binding interactions were weakened for both mutants (Figure S2G in the Supporting Information).

**Design of Oligopeptide Inhibitor.** Interestingly, while we were preparing this work and the work was under review, some antibodies<sup>12–14</sup> were reported to bind non-RBD regions 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.<sup>24–26</sup>

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) dropped by 34% from  $-740 \pm 70$  kJ/mol without EELE to  $-490 \pm 50$  kJ/mol in the presence of



**Figure 3.** Presence of tetrapeptide EELE lessens the RBD–ACE2 binding. (A) Simulation snapshot of a spike–ACE2 complex in the presence of EELE (in violet). (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).

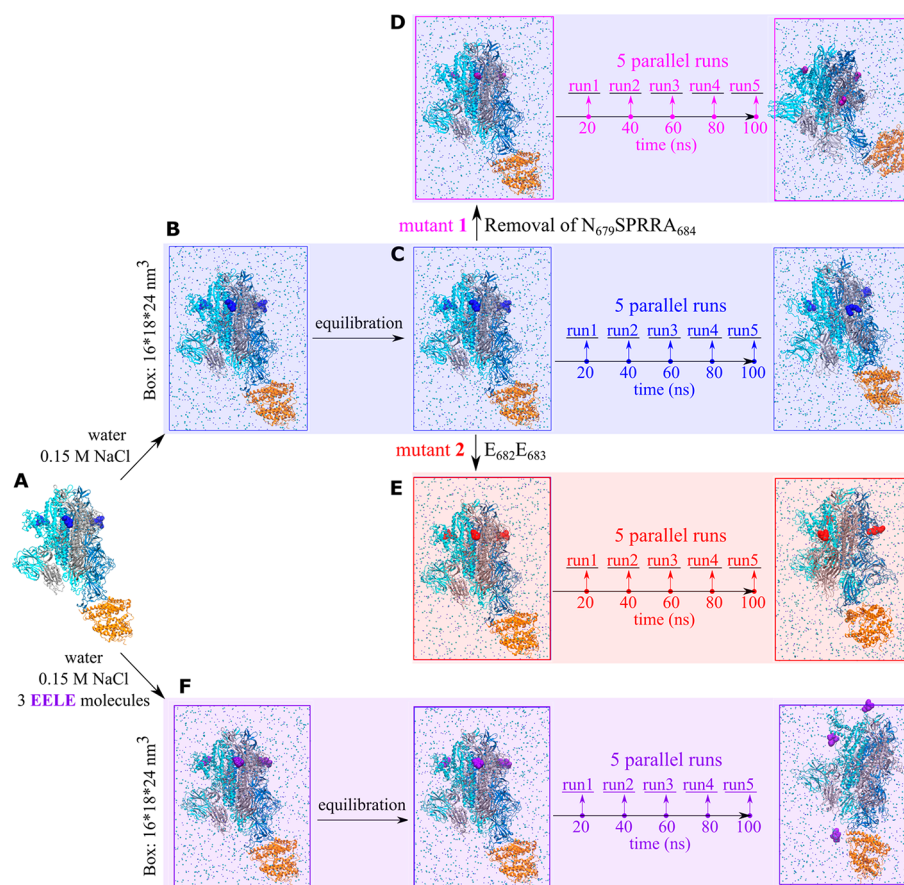
EELE. Similarly, the number of RBD–ACE2 intermolecular hydrogen bonds dropped by 41%. Therefore, the presence of EELE greatly lessened RBD–ACE2 binding.

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 and 74 ns.

The oligopeptide EELE serves as a model peptide, which supports our hypothesis that the polybasic cleavage sites could be targeted to inhibit SARS-CoV-2 spike protein–ACE2 binding. 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.<sup>33,34</sup> These could be potentially overcome by covalently grafting oligopeptides at the side chains of peptide brush polymers,<sup>35</sup> or integrating them onto peptide amphiphiles<sup>36</sup> or PEGylation.<sup>37</sup>

## 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



**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.  $\text{Na}^+$  and  $\text{Cl}^-$  ions are displayed using small blue and cyan beads, respectively, with water omitted for display. The polybasic cleavage sites ( $\text{R}_{682}\text{RAR}_{685}$ ) are highlighted with bigger blue beads. The system was equilibrated (C). The WT spike protein was mutated by removing  $\text{N}_{679}\text{SPRRA}_{684}$  (mutant 1 with  $\text{R}_{685}$  in purple; D) and changing  $\text{R}_{682}\text{R}_{683}$  to  $\text{E}_{682}\text{E}_{683}$  (mutant 2 with  $\text{E}_{682}\text{EAR}_{685}$  in red; E). Each simulation lasted 100 ns (C–E). Five parallel runs (10 ns each) were performed based on the configurations at 20, 40, 60, 80, and 100 ns. The same process was applied to the system with EELE tetrapeptides (F).

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  $\text{N}_{679}\text{SPRRA}_{684}$  and a mutant with the substitution of  $\text{R}_{682}\text{R}_{683}$  to  $\text{E}_{682}\text{E}_{683}$  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,<sup>2,11</sup> 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 targeted by antibodies.<sup>12–14</sup> 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 2016.3).<sup>38</sup> The CHARMM 36m potential<sup>19</sup> was employed for the proteins and  $\text{Na}^+$  and  $\text{Cl}^-$  ions. The recommended CHARMM TIP3P water model<sup>39</sup> was employed with the structures constrained using the SETTLE algorithm.<sup>40</sup>

The SARS-CoV-2 spike protein–ACE2 binding structure was downloaded from the Zhang-Server.<sup>18</sup> The spike protein–ACE2 complex was reconstructed using the C–I-TASSER model<sup>41</sup> based on the protein identification number QHD43416<sup>42</sup> for the spike protein. Each subunit of the trimeric spike protein included the residues from  $\text{M}_1$  to  $\text{T}_{1273}$  with the net charge of  $-7e$ , and the ACE2 has the residues from  $\text{S}_{19}$  to  $\text{D}_{615}$  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  $16 \times 18 \times 24 \text{ nm}^3$  (Figure 4). NaCl (0.15 M) was added along with 49  $\text{Na}^+$  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.<sup>43,44</sup> 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.<sup>45</sup> In all of 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<sup>2</sup> to preserve the spike protein–ACE2 binding structure. These restraints were removed in the following production simulations. The other parameters were the same as those in the production simulation below.

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 cutoff distance of 12 Å, and the long-range interactions were calculated using the Smooth Particle Mesh Ewald (PME) algorithm.<sup>46,47</sup> The temperatures of proteins, ions, and water were separately coupled using the Nosé–Hoover algorithm (reference temperature 298 K, 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  $4.5 \times 10^{-5}$  bar<sup>-1</sup>. 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<sup>48</sup> and were verified in simulations on proteins<sup>24,49,50</sup> and lipid membranes.<sup>51</sup> The production simulation lasted 100 ns. Calculation of the root-mean-square deviations of the RBD and ACE2 backbone atoms (Figure 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, five parallel simulations were performed, each of which run 10 ns.

The system had around 700000 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,  $2 \times 9.6$  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  $\theta_{HDA} \leq 30^\circ$ .<sup>52,53</sup>

**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<sub>679</sub>SPRRAR<sub>684</sub> 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.*<sup>2</sup> that the residues of T<sub>678</sub>NSPRRAR<sub>685</sub> in the wild-type spike protein of SARS-CoV-2 were mutated to T<sub>678</sub>-IL--R<sub>685</sub>. Accordingly, six Cl<sup>-</sup> ions, which were distributed next to the six arginine residues of R<sub>682</sub>R<sub>683</sub> 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<sub>682</sub>R<sub>683</sub> residues were changed to E<sub>682</sub>E<sub>683</sub> for all the three subunits of the spike protein trimer. Accordingly, the six Cl<sup>-</sup> ions, which were distributed next to the six arginine residues (R<sub>682</sub>R<sub>683</sub>) were replaced with Na<sup>+</sup> 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<sub>682</sub>RAR<sub>685</sub>) 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<sup>3</sup>. 0.15 M NaCl was added along with 9 Na<sup>+</sup> counterions of EELE and 49 Na<sup>+</sup> counterions of spike protein and ACE2 (Table S1 in the Supporting Information). The equilibration and productions were the same as those 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 Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsnano.0c04798>.

Supporting figure of root-mean-square deviations of the RBD and ACE2 backbone atoms. Supporting tables of system compositions and detailed interaction energies. Details of MARTINI coarse-grained simulations and results (PDF)

## 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, Illinois 60208, United States; [orcid.org/0000-0002-9802-3627](https://orcid.org/0000-0002-9802-3627); Email: [m-olvera@northwestern.edu](mailto:m-olvera@northwestern.edu)

### Author

Baofu Qiao – Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, United States; [orcid.org/0000-0001-8870-5985](https://orcid.org/0000-0001-8870-5985)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsnano.0c04798>

### 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

This work was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, under Award No. DE-FG02-08ER46539, the Sherman Fairchild Foundation, and the Center for Computation and Theory of Soft Materials at Northwestern University. We thank Yang Li for assistance in generating Figure 1.

## 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.
- (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.; Jacoani, S.; Culap, K.; Zatta, F.; De Marco, A.; Peter, A.; Guarino, B.; Spreafico, R.; Camerani, E.; Case, J. B.; Chen, R. E.; Havenar-Daughton, C.; Snell, G.; Telenti, A.; Virgin, H. W.; et al. Cross-Neutralization of SARS-CoV-2 by a Human Monoclonal SARS-CoV Antibody. *Nature* **2020**, *583*, 290–295.
- (9) Zhang, G.; Pomplun, S.; Loftis, A. R.; Tan, X.; Loas, A.; Pentelute, B. L. Investigation of ACE2 N-Terminal Fragments Binding to SARS-CoV-2 Spike RBD. **2020**, DOI: 10.1101/2020.03.19.999318 (accessed 2020-07-23).
- (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. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; Hastie, K. M.; Parker, M. D.; Partridge, D. G.; Evans, C. M.; Freeman, T. M.; de Silva, T. I.; McDanal, C.; Perez, L. G.; Tang, H.; Moon-Walker, A.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 1–16.
- (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) Wang, C.; Li, W.; Drabek, D.; Okba, N. M. A.; van Haperen, R.; Osterhaus, A. D. M. E.; van Kuppeveld, F. J. M.; Haagmans, B. L.; Grosveld, F.; Bosch, B.-J. A Human Monoclonal Antibody Blocking SARS-CoV-2 Infection. *Nat. Commun.* **2020**, *11*, 2251.
- (14) Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y.; Chen, Z.; Guo, Y.; Zhang, J.; Li, Y.; Song, X.; Chen, Y.; Xia, L.; Fu, L.; Hou, L.; Xu, J.; et al. A Neutralizing Human Antibody Binds to the N-Terminal Domain of the Spike Protein of SARS-CoV-2. *Science* **2020**, No. eabc6952.
- (15) 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. *Antiviral Res.* **2020**, *176*, 104742.
- (16) 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. *Viol. Sin.* **2020**, *35*, 1–3.
- (17) 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.
- (18) 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.
- (19) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (20) Chipot, C. Frontiers in Free-Energy Calculations of Biological Systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 71–89.
- (21) Jiménez-Angeles, F.; Kwon, H.-K.; Sadman, K.; Wu, T.; Shull, K. R.; Olvera de la Cruz, M. Self-Assembly of Charge-Containing Copolymers at the Liquid–Liquid Interface. *ACS Cent. Sci.* **2019**, *5*, 688–699.
- (22) 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.
- (23) 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.
- (24) Qiao, B.; Jiménez-Angeles, 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.
- (25) Kramer, R. M.; Shende, V. 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.
- (26) 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.
- (27) Chong, S.-H.; Ham, S. Interaction with the Surrounding Water Plays a Key Role in Determining the Aggregation Propensity of Proteins. *Angew. Chem., Int. Ed.* **2014**, *53*, 3961–3964.
- (28) Fumagalli, L.; Esfandiari, A.; Fabregas, R.; Hu, S.; Ares, P.; Janardanan, A.; Yang, Q.; Radha, B.; Taniguchi, T.; Watanabe, K.; Gomila, G.; Novoselov, K. S.; Geim, A. K. Anomalous Low Dielectric Constant of Confined Water. *Science* **2018**, *360*, 1339–1342.
- (29) Schlaich, A.; Knapp, E. W.; Netz, R. R. Water Dielectric Effects in Planar Confinement. *Phys. Rev. Lett.* **2016**, *117*, 048001.
- (30) Sun, C.; Shen, M.; Chavez, A. D.; Evans, A. M.; Liu, X.; Harutyunyan, B.; Flanders, N. C.; Hersam, M. C.; Bedzyk, M. J.; Olvera de la Cruz, M.; Dichtel, W. R. High Aspect Ratio Nanotubes Assembled from Macrocyclic Iminium Salts. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 8883–8888.
- (31) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.
- (32) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **2009**, *5*, 2531–2543.
- (33) Lau, J. L.; Dunn, M. K. Therapeutic Peptides: Historical Perspectives, Current Development Trends, and Future Directions. *Bioorg. Med. Chem.* **2018**, *26*, 2700–2707.
- (34) Fosgerau, K.; Hoffmann, T. Peptide Therapeutics: Current Status and Future Directions. *Drug Discovery Today* **2015**, *20*, 122–128.
- (35) 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.
- (36) Sato, K.; Hendricks, M. P.; Palmer, L. C.; Stupp, S. I. Peptide Supramolecular Materials for Therapeutics. *Chem. Soc. Rev.* **2018**, *47*, 7539–7551.
- (37) Harris, J. M.; Chess, R. B. Effect of PEGylation on Pharmaceuticals. *Nat. Rev. Drug Discovery* **2003**, *2*, 214–221.
- (38) 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.
- (39) 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.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

- (40) 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.
- (41) Zheng, W.; Li, Y.; Zhang, C.; Pearce, R.; Mortuza, S. M.; Zhang, Y. Deep-Learning Contact-Map Guided Protein Structure Prediction in CASP13. *Proteins: Struct., Funct., Genet.* **2019**, *87*, 1149–1164.
- (42) 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.
- (43) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- (44) 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.
- (45) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (46) 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.
- (47) 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.
- (48) 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.
- (49) Qiao, B.; Lopez, L.; Olvera de la Cruz, M. Mirror<sup>2</sup>-Like Protein Dimers Stabilized by Local Heterogeneity at Protein Surfaces. *J. Phys. Chem. B* **2019**, *123*, 3907–3915.
- (50) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, L.; 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.
- (51) 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.
- (52) Luzar, A.; Chandler, D. Structure and Hydrogen Bond Dynamics of Water-Dimethyl Sulfoxide Mixtures by Computer Simulations. *J. Chem. Phys.* **1993**, *98*, 8160–8173.
- (53) 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.