Protein Surface Printer for Exploring Protein Domains

Yang Li,[†] Baofu Qiao,^{*,‡} and Monica Olvera de la Cruz^{*,‡,¶,†}

[†]*Department of Chemical Engineering, Northwestern University, Evanston, Illinois 60201, United States*

[‡]*Department of Material Science and Engineering, Northwestern University, Evanston, Illinois 60201, United States*

[¶]*Department of Chemistry, Northwestern University, Evanston, Illinois 60201, United States*

E-mail: qiaobf@northwestern.edu; m-olvera@northwestern.edu

Phone: +1-847-491-7801

Abstract

The surface of proteins is vital in determining protein functions. Herein, a program, Protein Surface Printer(PSP), is built that performs multiple functions in quantifying protein surface domains. Two proteins, PETase and cytochrome P450, are used to validate that the program supports atomistic simulations with different combinations of programs and force fields. A case study is conducted on the structural analysis of the spike proteins of SARS-CoV-2 and SARS-CoV, and the human cell receptor ACE2. Although the surface domains of both spike proteins are highly similar, their receptor binding domains(RBDs) and the O-linked glycan domains are structurally different. Statistically, the outer surface of ACE2 displays less correlation with the RBD of SARS-CoV-2 than that of SARS-CoV. The O-linked glycan domain of SARS-CoV-2 is highly

positively charged, which may promote binding to negatively charged human cells. Our program paves the way for an accurate understanding of protein binding for aggregation and ligand recognition.

1. Introduction

Surface domains of a protein are defined as the areas where amino acids with similar polarity are distributed close to each other, which collectively impact protein interactions including protein-protein interactions and ligand binding by forming a domain-like structure on protein surfaces. Though many previous studies have unveiled the importance of protein surfaces in functions such as protein adsorption¹ and molecule recognition,² little research has been done to demonstrate the role of surface domains in those processes. Recently, protein surface domains were found to be able to selectively bind amphiphilic random copolymers in foreign environments.^{3,4} Specifically, the polar protein surface domains were found to selectively bind the polar components of the random copolymers, with the nonpolar components of the random copolymers extending into the organic environment. The formed protein-polymer complexes thus display a core (protein) - shell (polymer) reverse micellar structure, which consequently preserved the enzymatic activity of proteins. Furthermore, protein surface domains can distinctly orient water neighbors at the protein surface.⁵ For instance, negatively charged surface amino acids can orient water dipoles towards the proteins, whereas positively charged surface amino acids can orient water dipoles against the proteins. Moreover, different types of surface amino acids demonstrate different capabilities in orienting water neighbors and in influencing water dynamics. Consequently, the water shell near the protein surface forms an alternative pattern according to the protein surface domains.⁵ The protein local hydration behavior is expected to impact protein solubility and assembly,⁶⁻⁸ enzymatic substrate hydrolysis,^{9,10} and ligand recognition.^{6,11}

Given the crucial role of protein surface domains and lack of related studies, it is of paramount importance to quantitatively characterize the protein surface domains and consequently their impacts on protein functions. For instance, by integrating machine learning techniques¹² with protein surface domain information, it may be feasible to predict protein solubility, protein-protein interactions and protein-ligand binding affinity. Currently, there is well-developed software such as VMD¹³ that can visualize proteins and simulation tra-

jectories. However, functions to quantity protein surface domains are limited. It is hard to obtain protein surface information by simple observations of the proteins alone. Furthermore, multiple molecular dynamic (MD) software is prevailing but they produce different output formats, most of which are not processable by other software. The obstacle limits the analysis of results from the MD simulation results if researchers are not familiar with the software.

Inspired by the previous works, we have developed a program, known as Protein Surface Printer (PSP), to analyze protein surface domains and print surface domain information. The program was developed with a Python package MDAnalysis^{14,15} to perform multiple functions 1) analyze MD simulation trajectories in the formats of packages GROMACS,¹⁶ NAMD,¹⁷ AMBER¹⁸ and OpenMM¹⁹ using different atomistic force fields of AMBER,²⁰ CHARMM²¹ and AMOEBA,^{22,23} 2) process a single frame or multiple frames, 3) read and clip trajectories, 4) align proteins for multiframe trajectories, 5) collect and print surface domain information, 6) generate files containing protein surface domain information, 7) create files that can be visualized by software (such as VMD).

Two proteins, PETase and cytochrome P450, were used as examples to test the validity of the new program PSP. PETase is an efficient poly(ethylene terephthalate) plastic-degrading enzyme.⁹ Cytochrome P450 is one of the most versatile biocatalysts in nature.^{24,25} Both proteins are crucial in a broad variety of chemical reactions particularly important for pharmaceuticals, terpenes, gaseous alkanes, etc.²⁴ Around 75% of drugs are metabolized by P450 proteins.²⁶ After validation, a special case study was conducted on the spike proteins of the recent SARS-CoV-2 virus and its close relative SARS-CoV.

2. Methodology

2.1 Protein Surface Printer (PSP)

Surface residues are defined as residues that are in contact with solvent atoms. Some protein molecules are not solid inside and small solvent molecules (such as water molecules) can diffuse inside the protein. Thus, firstly, PSP excludes solvent atoms that are inside the proteins by finding the largest solvent cluster around the protein. Secondly, PSP calculates the distance between an amino acid and its nearest solvent atom. If the distance is less than 3.5 Å, which is the hydrogen donor-acceptor length,²⁷ the amino acid will be labeled as a surface residue. If the simulation trajectory is provided as the input to the program PSP, it will analyze the whole trajectory, which is termed as multi-frame analysis here. Otherwise, the program will only analyze one individual configuration, termed as single-frame analysis here. Lastly, PSP will print out surface domain information depending on user-defined options. The running logic is shown in Scheme 1.

PSP categorizes surface residues into 4 kinds based on their charges and polarities as shown in Table 1.^{5,13}

Table 1: Surface domain categories

Surface domain	Residues
Positively charged	ARG; HSP (charged HIS); LYS
Negatively charged	ASP; GLU
Polar neutral	ASN; CYS; GLN; SER; THR; TYR; GLY; neutral HIS
Nonpolar	ALA; LEU; VAL; ILE; PRO; PHE; MET; TRP;

2.2 PETase and P450

PETase was used to test the multi-frame function. In a previous work,⁵ we equilibrated PETase (+6e), Cl⁻ counterions, and 0.1 M salt of Na⁺ and Cl⁻ in a water box for 300 ns. Here we extended the simulations by another 2 ns using a variety of MD simulation programs (GROMACS,¹⁶ NAMD,¹⁷ AMBER¹⁸ and OpenMM¹⁹) and atomistic simulation force fields

Scheme 1: Flow chart of Protein Surface Printer running progress.

(CHARMM36m,²¹ CHARMM36,²⁸ AMBER14SB²⁰ and AMOEBA 13²³). The isothermal-isobaric ensemble (constant number of particles, temperature and pressure, NTP) was applied to all the simulations. The other parameters are listed in Table 2 and the details can be found in the references therein. The structure of cytochrome P450 was downloaded from the Protein Database with the PDB ID 1BU7²⁹ and was dissolved in water with Gromacs. The resulting file was used for a single frame analysis.

Table 2: Parameters of the MD simulations on PETase

	Software/ Force field	Water Model	Integration	Time Step	Temperature Coupling	Pressure Coupling
Run 1 ^{5,30}	GROMACS CHARMM36m	CHARMM TIP3P	Leap-frog	2fs	Nose-Hoover	Parrinello-Rahman
Run 2 ^{5,31}	NAMD CHARMM36	CHARMM TIP3P	Langevin	2fs	Langevin	Langevin
Run 3 ³⁰	AMBER AMBER14SB	TIP3P	Langevin	2fs	Langevin	Monte Carlo
Run 4 ⁵	OpenMM AMOEBA 13	AMOEBA 03 water	Langevin	1fs	Langevin	Monte Carlo

2.3 SARS-CoV-2 and SARS-CoV

Here the NCBI GenBank access code MN908947 (protein identification number QHD43416.1) is employed for SARS-CoV-2³² and the NCBI GenBank access code AY278741 (protein identification number AAP13441.1) is employed for SARS-CoV.³³ The sequences are provided in the Supporting Information. The sequences were first uploaded to the SwissModel server.³⁴ The experimental structure (PDB ID: 6VSB³⁵) was found to display 99.26% identity to the spike protein of SARS-CoV-2, and the experimental structure (PDB ID 6ACD³⁶) demonstrates 100% identity to the spike protein of SARS-CoV. These two protein structures were then employed to build the homology structure model of SARS-CoV-2 and SARS-CoV. The models are provided in the Supporting Information as 6VSB_model.pdb and

6ACD_model.pdb, respectively.

3. Results and discussion

3.1 Multiple Frames Analysis with Trajectory Alignment

PETase was used as an example to test the multi-frame function and the trajectory alignment function. PETase is positively charged with an overall charge of +6e with e being the elementary charge. PETase is roughly of spherical shape, which is ideal for testing the program.

We carried out MD simulations using different combinations of MD simulation packages and force fields: Gromacs/CHARMM36m, AMBER/AMBER14SB, NAMD/CHARMM36 and OpenMM/AMOEBA 13. The protein surface domain distributions are provided in Fig. 1(a-d), respectively. Besides the legends shown in Fig. 1, the extent of opacity indicates the distance between the residues and the center of mass or center of geometry of the protein depending on the input setting. The more transparent a dot is, the closer the residue represented by the dot locates to the center of the protein. The obtained distributions are highly similar to each other for all four types of protein surface domains. When superimposed, the patterns of surface domains are almost the same. We also calculated the average probability and standard deviation for each type of protein surface domain, which are presented in Fig. 1e. Again a high similarity is observed. These calculations thus demonstrate that the PSP program supports a variety of MD simulation packages and force fields, which are extensively employed for protein simulations. The program is expected to support other types of MD simulation trajectories and force fields that are compatible with MDanalysis.

3.2 Single Frame Analysis

P450 is a negatively-charged protein. Each monomer of the P450 dimer (PDB ID: 1BU7) carries a net charge of -15e (including one heme cofactor).³⁷ In Fig. 2, the protein surface

Fig. 1: Protein surface domains obtained from (a) Gromacs/CHARMM36m (b) NAMD/CHARMM36 (c) AMBER/AMBER14SB (d) OpenMM/AMOEBA 13. The spherical coordinates θ and ϕ were employed. (e) The average probability for all types of protein surface domains. Detailed numbers can be found in Supporting Information table S.1.

pattern has two parts, which are roughly symmetric. This is ascribed to the mirror-like orientation of the two subunits of the P450 dimer. Note that in our previous work,^{5,37} only the surface of one of the two subunits was analyzed. In comparison with the corresponding results for a P450 monomer, the probability of nonpolar surface amino acids increases for the P450 dimer, whereas it drops for the other types of surface amino acids. This suggests that the contact region of the two P450 subunits is predominantly composed of polar amino acids, which has been previously observed.³⁷

Fig. 2: Single frame surface analysis of P450 with pdb ID 1UB7. Legends are same as those of Fig. 1. (a) Spatial distributions of domains (b) percentages of all kinds of surface domains. Detailed numbers can be found in Supporting Information table S.2.

3.3 Quantifying the surface of the spike proteins of SARS-CoV-2 and SARS-CoV, and ACE2

The SARS-CoV-2 has spread to hundreds of countries since its outbreak and infected more than 2,920,000 people at the time of this study. The WHO rated SARS-CoV-2 as *Pandemic*. Early gene study indicated that SARS-CoV-2 (also referred to as 2019-nCoV) is very similar to SARS-CoV.³⁸ They both use a densely glycosylated spike protein to gain entry into host cells. The spike protein is a trimeric class I fusion protein that exists in a metastable pre-fusion conformation that undergoes a substantial structural rearrangement to fuse the viral

membrane with the host cell membrane.³⁵ It is generally believed that the binding between the spike proteins and the human receptor angiotensin I converting enzyme 2 (ACE2) plays a crucial role in cell fusion. In the present work, we are analyzing the surface features of the spike proteins of SARS-CoV-2 and SARS-CoV, their receptor binding domains and O-linked glycan residues (Fig. 3),³⁵ as well as the human ACE2, which is expected to aid the molecular understanding of the binding behavior of SARS-CoV-2 and SARS-CoV to the human cell membrane.

Fig. 3: Sequence of (a) receptor binding domains and (b) O-linked residues.

3.3.1 Spike Protein Surface Domains

The homology structures were analyzed using PSP, and the obtained protein surface features are demonstrated in Fig. 4. The compositions of both spike proteins display high similarity. Both proteins are negatively charged: their surface amino acids are 10% - 11% negatively charged and 9% positively charged. The overall negative surface charge has also been experimentally observed for the inner leaflets of plasma membranes³⁹ and a broad variety of cancer cells.⁴⁰ A minor difference lies in the polar charge-neutral amino acids at the surfaces. The SARS-CoV-2 has a slightly higher probability (47%) of polar charge-neutral

surface amino acids than SARS-CoV (45%). Two spike proteins cannot be distinguished solely by the composition.

3.3.2 Receptor Binding Domain (RBD) and ACE2

Early research⁴¹ suggests that SARS-CoV-2 probably uses the SARS-CoV receptor ACE2 as a cellular entry receptor. Subsequently, more and more research is targeting the binding behavior of SARS-CoV-2 and ACE2.^{42,43} The receptor binding domains of both spike proteins (Fig. 5) are analyzed with the PSP program developed here. Surprisingly, a significant difference was observed. The SARS-CoV-2 is remarkably less charged with an overall probability of 18% of charged surface amino acids when compared to 26% for SARS-CoV. The difference is predominantly ascribed to a smaller number of negatively charged amino acids for SARS-CoV-2 (10%) than for SARS-CoV (17%). SARS-CoV-2 has more polar neutral residues on its surface (50%) than SARS-CoV (39%). As shown in Fig. 6, analysis of the outer surface of ACE2 supports that it is oppositely charged: around 20% of ACE2 outer surface is negatively charged, which surpasses the probability of positively charged surface amino acids of 10%. The oppositely charged feature between ACE2 and the receptor binding domains of both spike proteins favor the binding of the virus and human cell membranes, which is required for cell entry.

Note that ACE2 has a concave structure such that it is hard to simply distinguish the surface. Thus, the protein is analyzed separately, with the concave structure as the inner surface and other parts as the outer surface. Remarkably, it is found that the surface domain distributions of the outer surface of ACE2 display strong correlation with the surface domain distributions of SARS-CoV receptor binding domain: 9% positively charged for the former vs 9% negatively charged for the latter; 19% negatively charged for the former vs 17% positively charged for the latter; and similar probabilities of polar charge-neutral and nonpolar surface amino acids for both ACE2 and the receptor binding domain of SARS-CoV. In contrast, the surface domain distributions of SARS-CoV-2 receptor binding domain are less correlated

Fig. 4: Protein surface domains in Cartesian Coordinates. (a) Model SARS-CoV-2 based on PDB ID 6VSB, (b) model SARS-CoV based on PDB ID 6ACD, and (c) probability of protein surface domains for the spike proteins. Detailed numbers can be found in Supporting Information table S.3.

with the surface domain distributions of ACE2 outer surface. The results suggest that based solely on probability, SARS-CoV-2 may not be ideal for binding human ACE2 in contrast to SARS-CoV. These findings are consistent with a prediction based on the structural analysis of the contact region between the two spike proteins and ACE2.⁴⁴ However, experimental data revealed that RBD of SARS-CoV-2 has a higher affinity than that of SARS-CoV.⁴¹ The contradictory result suggests that the high transmissivity of SARS-CoV-2 may not rely solely on the receptor binding domains. For instance, the influence of O-linked glycan residues remains elusive.

3.3.3 O-linked Glycan residues

A unique feature of SARS-CoV-2 is the O-linked glycan residues.³⁸ Compared with SARS-CoV, SARS-CoV-2 has four extra residues PRRA located on sequence 681-684, where P, R and A stands for proline, arginine and alanine, respectively. Residues RRAR (sequence 682-685) form a polybasic cleavage site. The functionality of the cleavage site in SARS-CoV-2 remains unclear but is suspected to increase its transmissibility and pathogenesis since the development of the cleavage site was observed with increasing transmissibility of other coronaviruses.³⁸ As illustrated in Fig. 7, the inserted residues form a continuous positively charged surface domain. Given the fact that arginine is positively charged and most cell membranes possess negative charges, arginine has previously been shown to improve the cellular uptake of peptides, which is promising for therapeutics applications.⁴⁵ And arginine has been investigated to enhance drug delivery across the blood-brain barrier for Alzheimer's disease.⁴⁶ It may be hypothesized that in addition to the spike protein-ACE2 binding, the O-linked glycan residues may promote binding of the SARS-CoV-2 with cell membranes, leading to fast cell entry. Mutating arginine to charge neutral or negatively charged amino acids could further illustrate the role of the O-linked glycan residue in immunoevasion.

Fig. 5: Protein surface domains in Cartesian Coordinates of the receptor binding domains of (a) SARS-CoV-2, (b) SARS-CoV, and (c) probability of each domain. Color legends are the same as those in Figure 1. Detailed numbers can be found in Supporting Information table S.4.

Fig. 6: ACE2 surface domains in spherical coordinates. Color legends are the same as those in Figure 1. (a) Outer surface, (b) Inner surface, and Surface domain probability of (c) outer and (d) inner surfaces. Detailed numbers can be found in Supporting Information table S.5.

Fig. 7: O-like glycan residues. (a) SARS-CoV-2 (b) SARS-CoV. Purple: polar neutral; Blue: positively charged; Red: negatively charged; Black: non-polar

4. Conclusion

A program, PSP, was developed which performs a quantitative investigation of protein surface domains from MD simulation trajectory and single-frame configuration. An analysis of the spike proteins of SARS-CoV-2 and SARS-CoV demonstrates that the two spike proteins are very similar in surface composition in terms of charge and polarity. The largest difference is only 1.5% on polar neutral residues. The receptor binding domain of the new virus SARS-CoV-2 is dominated by polar neutral amino acids (50%), while SARS-CoV consists of comparative amounts of polar neutral residues (39%) and nonpolar residues (35%). An analysis of the outer surface of human cell receptor ACE2 predicts that SARS-CoV-2 may not interact with ACE2 as strongly as SARS-CoV does, which is in agreement with previous structural analysis,⁴⁴ but in contradiction with experimental findings.⁴¹ This suggests that the high affinity of SARS-CoV-2 may not be exclusively caused by the composition of the receptor binding domains. The highly positively charged feature of the O-linked glycan residues on SARS-CoV-2 may promote binding to the negatively charged human cells. These findings on the SARS-CoV-2 spike protein may provide some clues of value to researchers working on drug discovery and viral pathogenesis.

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.

Supporting Information Available

The following files are available free of charge.

Tables with the detailed distribution of surface domains of PETase, P450, the spike proteins of SARS-CoV-2 and SARS-CoV and their receptor binding domains, and the human ACE2. Amino acid sequences of the spike proteins of SARS-CoV-2 and SARS-CoV. 3D print-out of the surface domains. The model protein (6VSB_model.pdb) for SARS-CoV-2 and the model protein (6ACD_model.pdb) for SARS-CoV. The program ProteinSurfacePrinter.py

References

- (1) Hlady, V.; Buijs, J. Protein adsorption on solid surfaces. *Current Opinion in Biotechnology* **1996**, *7*, 72–77, DOI: 10.1016/S0958-1669(96)80098-X.
- (2) Fletcher, S.; Hamilton, A. D. Protein surface recognition and proteomimetics: Mimics of protein surface structure and function. *Current Opinion in Chemical Biology* **2005**, *9*, 632–638, DOI: 10.1016/j.cbpa.2005.10.006.
- (3) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, 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, DOI: 10.1126/science.aao0335.
- (4) Nguyen, T. D.; Qiao, B.; Olvera de la Cruz, M. Efficient encapsulation of proteins with random copolymers. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115*, 6578–6583, DOI: 10.1073/pnas.1806207115.
- (5) Qiao, B.; Jiménez-Ángeles, F.; Nguyen, T. D.; Olvera de la Cruz, M. Water follows polar and nonpolar protein surface domains. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 19274–19281, DOI: 10.1073/pnas.1910225116.
- (6) Chong, S. H.; Ham, S. Interaction with the surrounding water plays a key role in determining the aggregation propensity of proteins. *Angewandte Chemie - International Edition* **2014**, *53*, 3961–3964, DOI: 10.1002/anie.201309317.
- (7) 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. *Biophysical Journal* **2012**, *102*, 1907–1915, DOI: 10.1016/j.bpj.2012.01.060.

- (8) Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L. Y. A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society* **2004**, *126*, 8933–8939, DOI: 10.1021/ja049297h.
- (9) Yoshida, S.; Hiraga, K.; Takehana, T.; Taniguchi, I.; Yamaji, H.; Maeda, Y.; Toyohara, K.; Miyamoto, K.; Kimura, Y.; Oda, K. A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **2016**, *351*, 1196–1199, DOI: 10.1126/science.aad6359.
- (10) Simonian, A. L.; Good, T. A.; Wang, S. S.; Wild, J. R. Nanoparticle-based optical biosensors for the direct detection of organophosphate chemical warfare agents and pesticides. *Analytica Chimica Acta* **2005**, *534*, 69–77, DOI: 10.1016/j.aca.2004.06.056.
- (11) Peri, C.; Morra, G.; Colombo, G. Surface energetics and protein-protein interactions: Analysis and mechanistic implications. *Scientific Reports* **2016**, *6*, 24035, DOI: 10.1038/srep24035.
- (12) Wang, J.; Cao, H.; Zhang, J. Z.; Qi, Y. Computational Protein Design with Deep Learning Neural Networks. *Scientific Reports* **2018**, *8*, 6349, DOI: 10.1038/s41598-018-24760-x.
- (13) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38, DOI: 10.1016/0263-7855(96)00018-5.
- (14) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016; pp 98–105, DOI: 10.25080/majora-629e541a-00e.
- (15) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327, DOI: 10.1002/jcc.21787.

- (16) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447, DOI: 10.1021/ct700301q.
- (17) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802, DOI: 10.1002/jcc.20289.
- (18) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005**, *26*, 1688, DOI: 10.1002/jcc.20290.
- (19) Eastman, P. et al. OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of Chemical Theory and Computation* **2013**, *9*, 461–469, DOI: 10.1021/ct300857j.
- (20) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713, DOI: 10.1021/acs.jctc.5b00255.
- (21) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2016**, *14*, 71–73, DOI: 10.1038/nmeth.4067.
- (22) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *Journal of Physical Chemistry B* **2003**, *107*, 5933–5947, DOI: 10.1021/jp027815+.

- (23) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *Journal of Chemical Theory and Computation* **2013**, *9*, 4046–4063, DOI: 10.1021/ct4003702.
- (24) Ortiz De Montellano, P. R. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chemical Reviews* **2010**, *110*, 932–948, DOI: 10.1021/cr9002193.
- (25) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie - International Edition* **2018**, *2018*, 4143, DOI: 10.1002/anie.201708408.
- (26) Guengerich, F. P. Cytochrome P450 and chemical toxicology. *Chemical Research in Toxicology* **2008**, *21*, 70–83, DOI: 10.1021/tx700079z.
- (27) 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. *Journal of Physical Chemistry Letters* **2014**, *5*, 1440–1444, DOI: 10.1021/jz500495p.
- (28) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 Dihedral Angles. *Journal of Chemical Theory and Computation* **2012**, *8*, 3257–3273, DOI: 10.1021/ct300400x.
- (29) Sevrioukova, I. F.; Li, H.; Zhang, H.; Peterson, J. A.; Poulos, T. L. Structure of a cytochrome P450-redox partner electron-transfer complex. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 1863–1868, DOI: 10.1073/pnas.96.5.1863.
- (30) Lee, J. et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* **2016**, *12*, 405–413, DOI: 10.1021/acs.jctc.5b00935.

- (31) Qiao, B.; Jiang, W. All-Atom Molecular Dynamics Study of Water-Dodecane Interface in the Presence of Octanol. *Journal of Physical Chemistry C* **2018**, *122*, 687–693, DOI: 10.1021/acs.jpcc.7b10997.
- (32) Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269, DOI: 10.1038/s41586-020-2008-3.
- (33) Rota, P. A. et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **2003**, *300*, 1394–1399, DOI: 10.1126/science.1085952.
- (34) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018**, *26*, W296–W303, DOI: 10.1093/nar/gky427.
- (35) 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, DOI: 10.1126/science.aax0902.
- (36) Song, W.; Gui, M.; Wang, X.; Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathogens* **2018**, *14*, DOI: 10.1371/journal.ppat.1007236.
- (37) Qiao, B.; Lopez, L.; Olvera de la Cruz, M. "Mirror"-like Protein Dimers Stabilized by Local Heterogeneity at Protein Surfaces. *Journal of Physical Chemistry B* **2019**, *123*, 3907–3915, DOI: 10.1021/acs.jpcc.9b01394.
- (38) Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F. The proximal origin of SARS-CoV-2. *Nature Medicine* **2020**, DOI: 10.1038/s41591-020-0820-9.

- (39) Yeung, T.; Grinstein, S. Lipid signaling and the modulation of surface charge during phagocytosis. *Immunological Reviews* **2007**, *219*, 17–36, DOI: 10.1111/j.1600-065X.2007.00546.x.
- (40) Le, W.; Chen, B.; Cui, Z.; Liu, Z.; Shi, D. Detection of cancer cells based on glycolytic-regulated surface electrical charges. *Biophysics Reports* **2019**, *5*, 10–18, DOI: 10.1007/s41048-018-0080-0.
- (41) Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273, DOI: 10.1038/s41586-020-2012-7.
- (42) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TM-PRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, DOI: 10.1016/j.cell.2020.02.052.
- (43) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)* **2020**, *367*, 1444–1448, DOI: 10.1126/science.abb2762.
- (44) Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *Journal of Virology* **2020**, *94*, e00127, DOI: 10.1128/jvi.00127-20.
- (45) Blum, A. P.; Kammeyer, J. K.; Gianneschi, N. C. Activating peptides for cellular uptake via polymerization into high density brushes. *Chemical Science* **2016**, *7*, 989–994, DOI: 10.1039/c5sc03417e.
- (46) 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. *International Journal of Molecular Sciences* **2019**, *20*, 381, DOI: 10.3390/ijms20020381.

Graphical TOC Entry

b)

O-linked

SARS-CoV-2

SARS-CoV

a)

RBD

SARS-CoV-2

SARS-CoV

■ SARS-CoV-2

■ SARS-CoV

■ SARS-CoV-2

■ SARS-CoV