

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Atomistic Modeling and Computational Study of Reactive Systems and Biological Applications

Permalink

<https://escholarship.org/uc/item/27q665r8>

Author

Huang, Yihan

Publication Date

2022

Peer reviewed|Thesis/dissertation

Atomistic Modeling and Computational Study of Reactive Systems and Biological
Applications

By

YIHAN HUANG
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Materials Science and Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Roland Faller, Chair

Ricardo Castro

Jeremy Mason

Committee in Charge

2022

©2022 – Yihan Huang

All rights reserved.

Atomistic Modeling and Computational Study of Reactive Systems and Biological Applications

Abstract

This dissertation covers several atomistic modeling techniques and applications in various fields that were classified into two major categories: quantum chemistry and reactive molecular modeling for materials designs (Chapters 2-4) and protein simulation for biological systems (Chapter 5-7). Despite the differences in modeling techniques and application areas with the two categories, all studies employed atomistic simulations to characterize the structures, thermodynamics, and kinetics of the systems of interest, helped understand the fundamental mechanics and driving forces of materials/proteins behaviors that cannot be solely described by experiments, and predicted optimal processing conditions or molecular structures following the process-structure-property relationship.

The fundamentals of density functional theory (DFT) and molecular dynamics (MD) simulations are introduced in Chapter 1. DFT is a computational modeling method that is based on quantum mechanics and investigate the electronics structure of many-body systems, which allows for the understanding of atomic bonding and reaction mechanisms. However, its limitations in time scale and length scale made it unsuitable for applications where dynamics is important and reactions happen with larger molecules or systems. Therefore, we utilized reactive MD by parametrizing empirical reactive force fields, Tersoff and ReaxFF, to study the structure changes, reaction mechanisms, and thermodynamic and kinetic properties of surface modification, aqueous reaction, and photo-initiated polymerization. The force fields were fitted against DFT-calculated

structures, atomic charges, and energetics, and detailed DFT methods, parametrization algorithms and process, and ReaxFF formalism were described in Chapter 1.

Chapter 2 focuses on applying the above techniques to develop a Tersoff force field for atomic layer etching where incident chlorine adsorbed on germanium (001) surface and modified the surface, followed by energetic Argon bombardment. Reactive MD simulations were then carried out using the developed Tersoff force field to study the correlations between the chlorination energy and penetration depth/ surface coverage, and the Ar bombardment energy and etched product species. Threshold bombardment energy and completion bombardment energy were also determined. This was a collaborative project with Lam Research, and our computational study gave insights with experimental parameters optimization. Chapter 3 focuses on the development of a ReaxFF force field for validating the reaction mechanism of a novel CO₂ absorbent – phosphoenol pyruvate (PEP). MD simulations with this force field described the process where PEP in water transferred bicarbonate, which was converted from CO₂, into carboxyphosphate and enolate, and investigated the correlation between the phosphorous' reactivity and partial charge. Chapter 4 focuses on the development of a ReaxFF force field for photo-initiated acrylate free-radical polymerization. This is a collaborative project with the volumetric additive manufacturing (VAM) team in Lawrence Livermore National Lab, as photopolymer resin design is a crucial part in the VAM process. MD simulations with this force field compared the structures and dynamic properties of three acrylates with different numbers of vinyl groups, sizes and shapes, and depicted their polymerization processes. This study not only provided a reactive force field suitable for acrylate polymerization, but also enabled future benchmark studies for resin materials selection.

Chapter 5-7 are the works that studied the interactions between the glycosylated recombinant ACE2 and the receptor-binding-domain (RBD) of the SARS-CoV-2 spike protein. In Chapter 5 we developed fully glycosylated computational models of ACE2-Fc fusion proteins that were therapeutically designed to neutralize SARS-CoV-2 spike protein, and this study was further extended to investigate the relationship between glycosylation and the binding of SARS-CoV-2 spike to ACE2 in Chapter 6. Two different glycovariants of ACE2 were built, pulling simulation were conducted for the binding strength between SARS-CoV-2 spike and ACE2, and hydrogen bond interactions analysis and spatial analysis of glycan interactions were used to study the binding regimes under the influence of glycans. These computational analyses offer insights for future designs of glycoproteins as therapeutic baits. The final Chapter 7 focuses on a collaborative work where we performed structural stability comparisons of various spike truncations that are experimentally synthesized in Chinese hamster ovary cells for therapeutic and diagnostic applications.

Contents

Abstract	ii
Acknowledgements	x
1. Introduction	1
1.1. Density Functional Theory and Molecular Dynamics Simulation	1
1.2. Reactive Force Fields	10
1.3. SARS-CoV-2 and Glycoproteins	18
2. Computational modelling of atomic layer etching of chlorinated germanium surfaces by argon	21
2.1. Introduction	21
2.2. Tersoff Potential and Simulation Methods	23
2.3. Results and Discussion	26
2.4. Conclusions	35
2.5. Supporting Information	38
2.5.1. Tersoff Force Field Training	38
2.5.2. Ge Surface Relaxation at Room Temperature	44
2.5.3. Other Characterizations in Chlorination and Bombardment Process	46
3. Development of a ReaxFF Force Field for Aqueous Phosphoenolpyruvate as a Novel Biomimetic Carbon Capture Absorbent	50
3.1. Introduction	51
3.2. Computational Methods	54
3.3. ReaxFF Formalism	54
3.3.1. Training Set Generation	57

3.3.2. Charge Parameter Correction	59
3.3.3. MD Simulation	60
3.4. Results and Discussion	61
3.4.1. Force Field Parametrization	61
3.4.2. Solution Validation	64
3.4.3. Analysis and Prediction	66
3.5. Conclusions	68
3.6. Supporting Information 1: Training Set and Additional Training Results	73
3.7. Supporting Information 2: Final Force Field	77
4. Development of a Reactive Force Field for Photo-initiated Acrylate Polymerization in Volumetric Additive Manufacturing	85
4.1. Introduction	86
4.1.1. Acrylates as Photopolymer Resin	86
4.1.2. ReaxFF for Acrylate Free-radical Polymerization	88
4.2. Computational Methods	90
4.2.1. Training Set Generation	90
4.2.2. Parallel Parametrization Scheme	92
4.2.3. MD Simulations	93
4.3. Results and Discussion	94
4.3.1. Force Field Development	95
4.3.2. Liquid Monomers	97
4.3.3. Free-radical Reactions	100
4.4. Conclusions	103
4.5. Supporting Information 1: Training Set and Additional Training Results	107

4.6. Supporting Information 2: Final Force Field	109
5. Development and simulation of fully glycosylated molecular models of ACE2-Fc fusion proteins and their interaction with the SARS-CoV-2 spike protein binding domain ..	118
5.1. Introduction	119
5.2. Materials and Methods	124
5.2.1. Sequences and Initial Structure	124
5.2.2. Simulated Systems	126
5.2.3. Simulation Procedure	127
5.3. Results and Discussion	129
5.4. Conclusions	132
5.5. Supporting Information	137
6. SARS-CoV-2 spike binding to ACE2 is stronger and longer ranged due to glycan interaction	142
6.1. Introduction	143
6.2. Materials and Methods	147
6.2.1. Simulation	147
6.2.2. Experiments	151
6.2.2.1. Protein Deglycosylation	151
6.2.2.2. Biolayer Interferometry	151
6.3. Results	152
6.4. Discussions	166
6.5. Conclusion	168
6.6. Supporting Information	175
6.6.1. Pull Force vs Pull distance	175

6.6.2.	Hydrogen Bonding Script	175
6.6.3.	Hydrogen Bonding Maps	176
6.6.4.	Hydrogen Bond Occupancy	180
6.6.5.	Angle Correlation Functions and Dihedral Correlation Functions	184
6.6.6.	Principal Components	187
7.	Production of Novel SARS-CoV-2 Spike Truncations in Chinese Hamster Ovary Cells Leads to High Expression and Binding to Antibodies	191
7.1.	Introduction	192
7.2.	Methods	194
7.2.1.	Plasmids	194
7.2.2.	Cell Culture and Transfection	194
7.2.3.	Protein Purification and Concentration	195
7.2.4.	SDS-PAGE and Western Blot	196
7.2.5.	Enzyme-linked Immunosorbent Assay (ELISA)	197
7.2.6.	Bradford Assay	198
7.2.7.	Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) and Sequence Alignment	198
7.2.8.	Circular Dichroism (CD)	199
7.2.9.	Simulations	199
7.3.	Results	200
7.3.1.	Expression and Purification of Spike and RBD	200
7.3.2.	Novel Truncations to Improve Protein Titers	204
7.3.3.	Binding Sensitivities against Antibodies	208
7.3.4.	Structural Characterization of Truncations	209

7.4. Discussion	214
7.5. Conclusions	216
7.6. Supplemental Information	221
8. Summary and Outlook	233
9. Bibliography	236

Acknowledgements

First and foremost, I thank my boyfriend, friends and family for the support they gave me during the pursuit of this degree. I thank my lovely dog for giving me endless laughter even though she is far away from me. I thank my advisor Prof. Roland Faller for his mentorship, guidance, patience, and insight. I thank my colleagues at UC Davis Dr. Austen Bernardi, Dr. Vincent Ustach, Dr. Tommy Harrellson, Dr. Bradley Harris, Dr. Haoyan Sha, Dr. Shenli Zhang, Bingxi Li, Shiaki Minami, and Christy Fox for the friendship and fruitful discussions. I thank my collaborators for fundings and a lot of meaningful discussions: Prof. Anthony Wexler, Prof. Keith Bein, and Prof. François Gygi on the biomimetic carbon capture project; Prof. Karen McDonald, Prof. Somen Nandi, and Prof. Priya Shah for all the SARS-CoV-2 projects; Dr. Maxim Shusteff and Dr. John Karnes from Lawrence Livermore National Lab on the volumetric additive manufacturing project; Gulcin Tetiker from Lam Research (now at KLA) on the atomic layer etching project. I also thank current and past staff members in the department of Materials Science and Engineering and HPC support team for their help. Finally, I thank Prof. Ricardo Castro and Prof. Jeremy Mason for their assistance as my dissertation committee members.

Chapter 1

Introduction

1.1 Density Functional Theory and Molecular Dynamics Simulation

Computational modeling is an important tool in materials science and biophysics, as it can be used to advance theories and discover new physical behaviors of materials and proteins, explain experimental findings, and predict optimal conditions in material designs. Quantum and atomistic details are needed in many situations, including studying small-scale materials like small molecules, nanoparticles, thin films, interfaces and proteins, and investigating microscopic structures and mechanisms of materials.¹⁻¹³ The collective behavior of the atoms allows for the understanding of macroscopic phenomena. Density functional theory (DFT) is a computational quantum mechanical simulation method that calculates the electronic structure of many-body systems, from which the properties of atoms, molecules, and solids can be determined by using functional (a function of a function, in the case of DFT the functional is electron density).¹⁻⁴ It belongs to the family of *ab initio* methods, which can predict materials properties for unknown systems without experimental inputs, and has become one of the most popular and versatile methods in the fields of computational physics and computational chemistry in the past few decades.¹⁴⁻¹⁶ Molecular dynamics (MD) simulation is another powerful computational modeling method that analyzes the physical movements of atoms and molecules in a certain period of time and predicts the evolution of the system. The trajectories follow Newton's equation of motion and the potential energy is determined by empirical force fields. If the system meets the ergodic hypothesis, the averages of a MD simulation can represent the system's macroscopic thermodynamic and kinetic properties.^{17,18} Using DFT and MD simulations together can provide

detailed electronic and atomistic information and describe properties and evolutions of systems as large as hundreds of nanometers and in the length scale of hundreds of nanoseconds, as shown in figure 1.1.

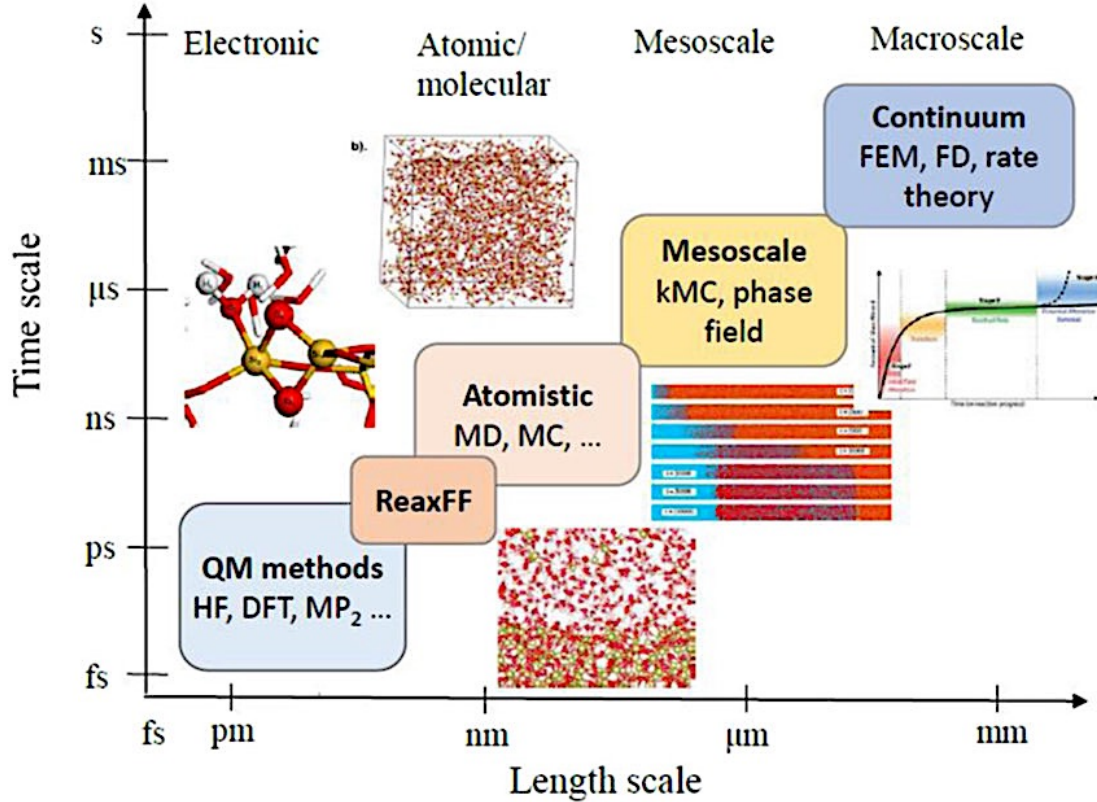


Figure 1.1: Multiscale modeling techniques at different time scale and length scale.¹⁹

DFT²⁰⁻²² is derived from the N-particle Schrödinger equation following the Born–Oppenheimer (BO) approximation,^{23,24} which assumes that the nuclei of the treated molecules or clusters can be treated as fixed due to its relatively larger mass compared to an electron, generating a static external potential in which electrons are moving, as in equation 1.1,

$$\hat{H}\Psi = [\hat{T} + \hat{V} + \hat{U}]\Psi = \left[\sum_{i=1}^N \left(-\frac{\hbar^2}{2m_i} \nabla_i^2 \right) + \sum_{i=1}^N V(\mathbf{r}_i) + \sum_{i<j}^N U(\mathbf{r}_i, \mathbf{r}_j) \right] \Psi = E\Psi \quad (1.1)$$

where \hat{H} is the Hamiltonian, E is the total energy, \hat{T} is the kinetic energy, \hat{V} is the potential energy from the external field due to positively charged nuclei, and \hat{U} is the electron–electron interaction energy. For a normalized Ψ , the functional, electron density $n(\mathbf{r})$, is given by

$$n(\mathbf{r}) = N \int d^3\mathbf{r}_2 \cdots \int d^3\mathbf{r}_N \Psi^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (1.2)$$

Ψ is a unique functional of the ground-state density $n_0(\mathbf{r})$, $\Psi_0 = \Psi[n_0]$, where Ψ_0 is the corresponding ground-state wave function. Therefore, the ground-state expectation value of an observable \hat{O} is also a functional of n_0 :

$$O[n_0] = \langle \Psi[n_0] | \hat{O} | \Psi[n_0] \rangle \quad (1.3)$$

In particular, the ground-state energy is a functional of n_0 :

$$E_0 = E[n_0] = \langle \Psi[n_0] | \hat{T} + \hat{V} + \hat{U} | \Psi[n_0] \rangle \quad (1.4)$$

where the contribution from the external potential can be written as

$$\langle \Psi[n_0] | \hat{V} | \Psi[n_0] \rangle = V[n_0] = \int V(\mathbf{r}) n_0(\mathbf{r}) d^3\mathbf{r} \quad (1.5)$$

and more generally,

$$V[n] = \int V(\mathbf{r}) n(\mathbf{r}) d^3\mathbf{r} \quad (1.6)$$

Substituting back into equation 1.4 can get the general form:

$$E[n] = T[n] + U[n] + \int V(\mathbf{r}) n(\mathbf{r}) d^3\mathbf{r} \quad (1.7)$$

where $T[n]$ and $U[n]$ are universal functionals, and $V[n]$ is the external potential acting on the system. With good approximation of $T[n]$ and $U[n]$, minimizing $E[n]$ with respect to $n(\mathbf{r})$ can yield all other ground-state observables.

In order to approximate the kinetic and electron-electron functionals, Kohn and Sham introduced a fictitious Kohn–Sham system of non-interacting electrons that generate the same density as any given system of interacting electrons.²⁵

And so, the total energy of the system can be written as

$$E[n] = \langle \Psi_s[n] | \hat{T}_s + \hat{V}_s | \Psi_s[n] \rangle = T_s[n] + V_s[n] + \int V(\mathbf{r})n(\mathbf{r})d^3\mathbf{r} \quad (1.8)$$

where V_s is the Kohn–Sham potential in the Kohn–Sham equation

$$\left(-\frac{\hbar^2}{2m} \nabla_i^2 + V_s(\mathbf{r}) \right) \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) \quad (1.9)$$

which yields the orbital energy ε_i and the corresponding orbital $\varphi_i(\mathbf{r})$ that reproduces the electron density $n(\mathbf{r})$ of the original many-body system

$$n(\mathbf{r}) = \sum_{i=1}^N |\varphi_i(\mathbf{r})|^2 \quad (1.10)$$

and T_s is the Kohn–Sham kinetic energy expressed as

$$T_s[n] = \sum_{i=1}^N \int d\mathbf{r} \varphi_i^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m} \nabla_i^2 \right) \varphi_i(\mathbf{r}) \quad (1.11)$$

Equation 1.8 can further be derived as

$$E[n] = T_s[n] + \int V(\mathbf{r})n(\mathbf{r})d^3\mathbf{r} + \frac{e^2}{2} \int d\mathbf{r} \int d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{XC}[n] \quad (1.12)$$

in which the third term is the Hartree (or Coulomb) energy, and $E_{XC}[n]$ is the exchange-correlation energy.

The exact functionals for exchange and correlation for most systems, however, are not known. Therefore, approximations are used and allow for relatively accurate calculations of physical properties. For example, one of the widely-used approximations is the local-density approximation (LDA), which assumes the density is the same everywhere:

$$E_{XC}^{LDA}[n] = \int \varepsilon_{XC}(n) n(\mathbf{r}) d^3\mathbf{r} \quad (1.13)$$

If including electron spin, this expression can be straightforwardly extended to the local spin-density approximation (LSDA):

$$E_{XC}^{LSDA}[n_{\uparrow}, n_{\downarrow}] = \int \varepsilon_{XC}(n_{\uparrow}, n_{\downarrow}) n(\mathbf{r}) d^3\mathbf{r} \quad (1.14)$$

Examples of LDA functionals include VWN²⁶ and VWN5²⁷ functional. Since LDA tends to underestimate the exchange energy and over-estimate the correlation energy,²⁸ an improvement in the accuracy can be obtained by generalized gradient approximations (GGA) functionals:²⁹

$$E_{XC}^{GGA}[n_{\uparrow}, n_{\downarrow}] = \int \varepsilon_{XC}(n_{\uparrow}, n_{\downarrow}, \nabla n_{\uparrow}, \nabla n_{\downarrow}) n(\mathbf{r}) d^3\mathbf{r} \quad (1.15)$$

Examples of GGA functionals include PBE³⁰ and PW91³¹.

There are also other more complicated but also more accurate functionals such as meta-GGA and hybrid functionals. B3LYP^{32,33} is a popular hybrid functional and is frequently used in the work presented in this dissertation.

In a MD simulation, the microscopic time evolution of a many-body system is obtained by numerically solving the classical equations of motion subjecting to periodic boundary conditions. A “real” system might contain about 6.02×10^{23} molecules, but simulations under periodic boundary conditions consider the whole system as a periodic and infinite replication of a small

box, so they are useful for emulating large systems. Figure 1.2 shows the schematic of MD periodic boundary conditions.

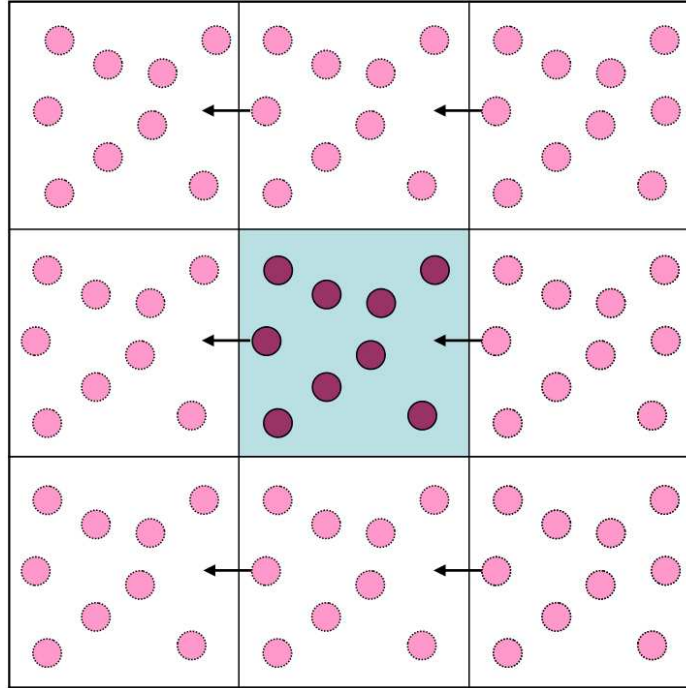


Figure 1.2: A two-dimensional representation of the periodic boundary conditions. A particle that leaves the box on one side is replaced by an image particle that enters from the other side. ³⁴

A MD simulation generates a sequence of points in phase space, which form trajectories of all particles in the system, and time averages can be calculated from the trajectories. Motion of the particles through phase space is governed by Hamiltonian equations of motion:

$$\frac{d\mathbf{q}_j}{dt} = \frac{\partial H}{\partial \mathbf{p}_j} = \frac{\mathbf{p}_j}{m_j}, \quad j = 1, \dots, N \quad (1.16)$$

$$\frac{d\mathbf{p}_j}{dt} = -\frac{\partial H}{\partial \mathbf{q}_j} = \mathbf{F}_j, \quad j = 1, \dots, N \quad (1.17)$$

where N is the total number of particles, \mathbf{p}_j is particle j 's momentum vector, \mathbf{q}_j is the particle's coordinate vector, t is time, m_j is the mass of the particle, \mathbf{F}_j is the force acting on j 's particle, and H is the Hamiltonian that can be expanded into kinetic and potential components:

$$H(\mathbf{p}_1, \dots, \mathbf{p}_N, \mathbf{q}_1, \dots, \mathbf{q}_N) = \frac{1}{2} \sum_{j=1}^N \frac{\|\mathbf{p}_j\|^2}{m_j} + V(\mathbf{q}_1, \dots, \mathbf{q}_N) \quad (1.18)$$

where V is the potential energy. There are three equations per particle, corresponding to the orthogonal Euclidean coordinates x , y , and z , so

$$\mathbf{p}_j = \langle p_{j,x}, p_{j,y}, p_{j,z} \rangle, \quad \mathbf{q}_j = \langle q_{j,x}, q_{j,y}, q_{j,z} \rangle \quad (1.19)$$

Substituting equation 1.18 into 1.17 and using $\mathbf{p} = m(d\mathbf{q}/dt)$, Newton's second law can be derived:

$$m \frac{d^2 \mathbf{q}_j}{dt^2} = - \frac{\partial V}{\partial \mathbf{q}_j} \quad (1.20)$$

This is the equation of motion and can be approximated using Taylor series expansions. There are multiple ways to integrate the equation of motion, popular numerical integration methods include Position Verlet,³⁵ Velocity Verlet,³⁶ Leapfrog Verlet.³⁷ In the work presented in this dissertation, Velocity Verlet is used, which is expressed as

$$\mathbf{q}_j(t + \Delta t) = \mathbf{q}_j(t) + \mathbf{v}_j(t)\Delta t + \frac{\mathbf{F}_j(t)}{2m_j} \Delta t^2, \quad j = 1, \dots, N \quad (1.21)$$

$$\mathbf{v}_j(t + \Delta t) = \mathbf{v}_j(t) + \frac{1}{2m_j} \left(\mathbf{F}_j(t) + \mathbf{F}_j(t + \Delta t) \right) \Delta t, \quad j = 1, \dots, N \quad (1.22)$$

where \mathbf{v}_j is particle j 's velocity vector and Δt is the elapsed time between integration steps, also known as simulation timestep. For numerical stability and accuracy in conserving the energy, a good timestep is usually one order of magnitude lower than the fastest oscillation in the system,

but not too small that exploits too much computer time. In the Velocity Verlet integration the local discretization error in position is $O(\Delta t^4)$ and the local discretization error in velocity is $O(\Delta t^2)$, whereas the global position error is $O(\Delta t^2)$ and the global velocity error is $O(\Delta t^2)$. Since global error is more important than local error, Velocity Verlet is known as a second-order integrator. Higher order expansion will result in higher accuracy but make the calculation more expensive. Velocity Verlet integration generates velocity at the same value of the time variable, and offers promising accuracy at reasonable computational expense. Initial velocities can be assigned per Maxwell-Boltzmann distribution at the desired simulation temperature. The forces acting on each particle are typically the negative of the derivative of the sum of bonded, van der Waals, and Electrostatic energies, generated by force fields.

Equations 1.21 and 1.22 describe the dynamics for a microcanonical (NVE) system.³⁸ However, isothermal (NVT) or isothermal/isobaric (NPT) systems are often used in practice, since temperature and pressure are the relevant state variables in the real world. The NVT and NPT ensembles are approximated by the addition of temperature and pressure controlling terms (thermostats and barostats respectively) on equation 1.20. Several thermostats and barostats have been developed, such as the Nose-Hoover thermostat and barostat,³⁹ the Velocity-rescale thermostat,⁴⁰ the Parinello-Rahman barostat,⁴¹ and the GJF thermostat and barostat.⁴² Figure 1.3 shows a typical MD simulation workflow using Velocity Verlet integration.

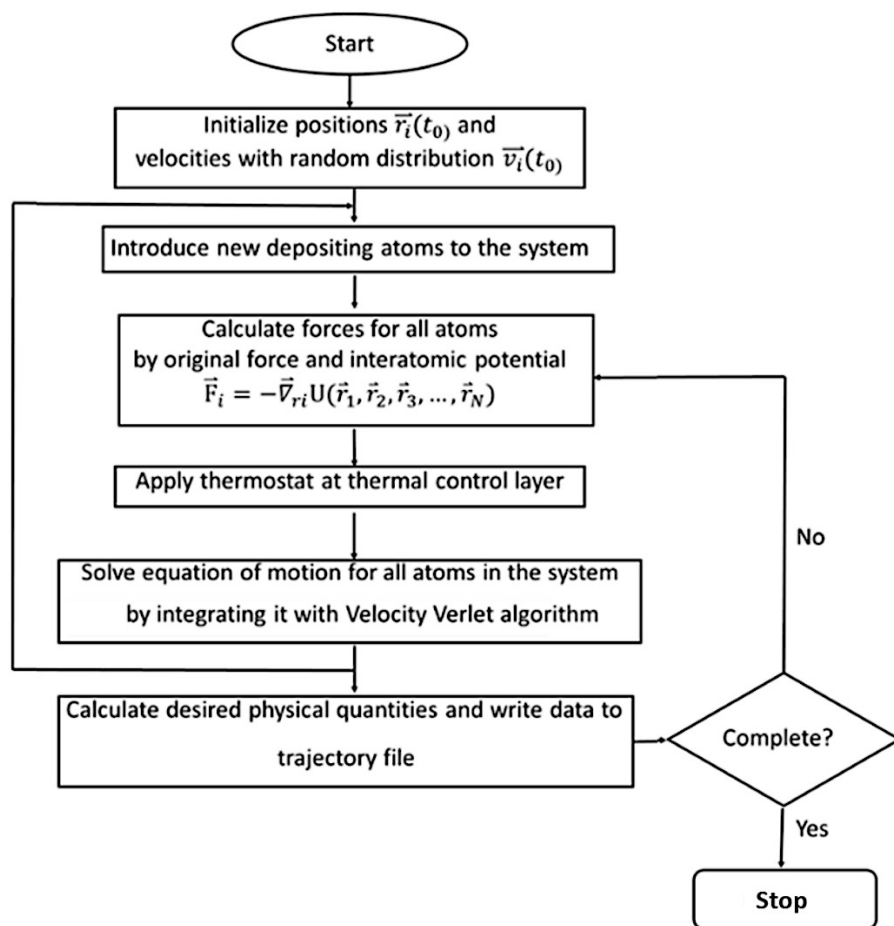


Figure 1.3: Flow chart of a typical MD simulation.⁴³

Although DFT and MD simulations together allow for precise and detailed information on simulated systems, there are still limitations. One of the significant limitations with both techniques is the computation cost. This can be improved by the advent of better computing facilities such as graphical processing units (GPUs) and super computers. Another limitation is the difficulty of simulating reactions during the evolution of the system. Although DFT allows the modeling of bonding, charge distribution, etc., that can describe chemical reactions, the system size using this technique is very limited and the solvent effect is usually not very accurate. Classical MD simulation, on the other hand, predefines atom connectivity so reactions cannot be depicted.

In recent decades, several reactive force fields are developed for MD simulations, many of them are empirical force fields trained from QM calculation results. The next section is dedicated to two of the reactive force fields, Tersoff and ReaxFF, which were used for the applications discussed in this dissertation.

1.2 Reactive Force Fields

Force fields are interatomic potentials that describe the energy landscape, and the negative of the gradient of the potential energy with respect to the particle coordinates is the force acting on each particle. A force field is comprised of the functional forms and the parameter set that are used to calculate the potential energy.⁴⁴ Specifically, reactive empirical bond-order potentials used in MD simulation are designed to describe chemical reactions.

The Tersoff potential was the first reactive empirical bond-order potential that incorporates the structural chemistry of covalently bonded systems. It was first developed for silicon and then extended to a variety of III-V semiconductors including germanium and carbon. The general form of the potential is

$$E = \sum_i E_i = \frac{1}{2} \sum_i \sum_{j \neq i} V_{ij} \quad (1.23)$$

$$V_{ij} = f_C(r_{ij}) [f_R(r_{ij}) + b_{ij} f_A(r_{ij})] \quad (1.24)$$

where E is the total energy of the system, E_i is the energy for atom i , V_{ij} is the interaction energy between atoms i and j , r_{ij} is the distance between two atoms, $f_R(r_{ij})$ and $f_A(r_{ij})$ are pair-additive repulsive and attractive interactions respectively, $f_C(r_{ij})$ is a cutoff function to smoothly transit full interaction to zero, and b_{ij} is a three-body term added to the attraction part, which includes all

neighbors j and k of atom i and describes the local bonding environment or coordination.^{45, 46} For the Tersoff potential, the repulsive and attractive terms are represented by the Morse-type functions

$$f_R(r) = A \exp(-\lambda_1 r) \quad (1.25)$$

$$f_A(r) = -B \exp(-\lambda_2 r) \quad (1.26)$$

where A , B , λ_1 , and λ_2 are all positive constants with $\lambda_1 > \lambda_2$. The Tersoff potential does not assume different forms for the angular functions for different hybridizations. The bond order term b_{ij} has the form

$$b_{ij} = (1 + \beta^n \zeta_{ij}^n)^{-\frac{1}{2n}} \quad (1.27)$$

$$\zeta_{ij} = \sum_{k \neq i, j} f_C(r_{ik}) g(\theta_{ijk}) \exp[\lambda_3^m (r_{ij} - r_{ik})^m] \quad (1.28)$$

$$g(\theta) = \gamma_{ijk} \left(1 + \frac{c^2}{d^2} - \frac{c^2}{[d^2 + (\cos\theta - \cos\theta_0)^2]} \right) \quad (1.29)$$

where θ_{ijk} is the bond angle between bonds ij and ik .

The cutoff function $f_C(r_{ij})$ in equation 1.24 has the form

$$f_C(r) = \left\{ \begin{array}{l} 1: r < R - D \\ \frac{1}{2} - \frac{1}{2} \sin\left(\frac{\pi r - R}{2D}\right): R - D < r < R + D \\ 0: r > R + D \end{array} \right\} \quad (1.30)$$

where $R-D$ and $R+D$ are the inner and outer cutoff distance, respectively.

Another powerful reactive empirical bond order potential is ReaxFF. It allows bond dissociation and formation during dynamic simulations, and enables simulations of reactions across interfaces since the mathematical formalism of each atom is transferable across phases.⁴⁷⁻

⁴⁹ ReaxFF was developed by Adri van Duin and William A. Goddard, III, and it employs bond-order components in conjunction with polarizable charge components to describe both reactive and non-reactive interactions between atoms. Most ReaxFF force fields are developed based on the 2008-C/H/O formalism. ⁴⁸ Equation 1.31 describes the overall system energy in ReaxFF formalism,

$$\begin{aligned}
 E_{system} = & E_{bond} + E_{lp} + E_{over} + E_{under} + E_{val} + E_{pen} + E_{coa} \\
 & + E_{C2} + E_{triple} + E_{tors} + E_{conj} + E_{H-bond} \\
 & + E_{vdWaals} + E_{Coulomb}
 \end{aligned} \tag{1.31}$$

where except $E_{vdWaals}$ and $E_{Coulomb}$, all other energies are bond order dependent and are functions of BO'_{ij} that can be calculated directly from the interatomic distance, so that atom connectivity no longer needs to be pre-defined. BO'_{ij} is composed of the contributions from sigma bonds, pi-bonds and double pi bonds, and so enabling a smooth transition from triple bond (if any) to single bond and to dissociation.

$$\begin{aligned}
 BO'_{ij} = & BO'_{ij}{}^{\sigma} + BO'_{ij}{}^{\pi} + BO'_{ij}{}^{\pi\pi} \\
 = & \exp \left[p_{bo1} \cdot \left(\frac{r_{ij}}{r_0^{\sigma}} \right)^{p_{bo2}} \right] + \exp \left[p_{bo3} \cdot \left(\frac{r_{ij}}{r_0^{\pi}} \right)^{p_{bo4}} \right] \\
 & + \exp \left[p_{bo5} \cdot \left(\frac{r_{ij}}{r_0^{\pi\pi}} \right)^{p_{bo6}} \right]
 \end{aligned} \tag{1.32}$$

An uncorrelated coordination Δ'_i , defined as the difference between the total bond order around the atom and the number of its bonding electrons Val , is used to correct the bond order term:

$$\Delta'_i = -Val_i + \sum_{j=1}^{neighbors(i)} BO'_{ij} \tag{1.33a}$$

and $\Delta'_i{}^{boc}$ is used for atoms having lone electrons:

$$\Delta_i^{boc} = -Val_i^{boc} + \sum_{j=1}^{neighbors(i)} BO'_{ij} \quad (1.33b)$$

The corrected bond orders are:

$$BO_{ij}^{\sigma} = BO'_{ij}{}^{\sigma} \cdot f_1(\Delta'_i, \Delta'_j) \cdot f_4(\Delta'_i, BO'_{ij}) \cdot f_5(\Delta'_j, BO'_{ij})$$

$$BO_{ij}^{\pi} = BO'_{ij}{}^{\pi} \cdot f_1(\Delta'_i, \Delta'_j) \cdot f_1(\Delta'_i, \Delta'_j) \cdot f_4(\Delta'_i, BO'_{ij}) \cdot f_5(\Delta'_j, BO'_{ij})$$

$$BO_{ij}^{\pi\pi} = BO'_{ij}{}^{\pi\pi} \cdot f_1(\Delta'_i, \Delta'_j) \cdot f_1(\Delta'_i, \Delta'_j) \cdot f_4(\Delta'_i, BO'_{ij}) \cdot f_5(\Delta'_j, BO'_{ij})$$

$$BO_{ij} = BO_{ij}^{\sigma} + BO_{ij}^{\pi} + BO_{ij}^{\pi\pi} \quad (1.34a)$$

$$f_1(\Delta'_i, \Delta'_j) = \frac{1}{2} \cdot \left(\frac{Val_i + f_2(\Delta'_i, \Delta'_j)}{Val_i + f_2(\Delta'_i, \Delta'_j) + f_3(\Delta'_i, \Delta'_j)} + \frac{Val_j + f_2(\Delta'_i, \Delta'_j)}{Val_j + f_2(\Delta'_i, \Delta'_j) + f_3(\Delta'_i, \Delta'_j)} \right) \quad (1.34b)$$

$$f_2(\Delta'_i, \Delta'_j) = \exp(-p_{boc1} \cdot \Delta'_i) + \exp(-p_{boc1} \cdot \Delta'_j) \quad (1.34c)$$

$$f_3(\Delta'_i, \Delta'_j) = -\frac{1}{p_{boc2}} \cdot \ln \left\{ \frac{1}{2} \cdot [\exp(-p_{boc2} \cdot \Delta'_i) + \exp(-p_{boc2} \cdot \Delta'_j)] \right\} \quad (1.34d)$$

$$f_4(\Delta'_i, BO'_{ij}) = \frac{1}{1 + \exp(-p_{boc3} \cdot (p_{boc4} \cdot BO'_{ij} \cdot BO'_{ij} - \Delta_i^{boc}) + p_{boc5})} \quad (1.34e)$$

$$f_5(\Delta'_j, BO'_{ij}) = \frac{1}{1 + \exp(-p_{boc3} \cdot (p_{boc4} \cdot BO'_{ij} \cdot BO'_{ij} - \Delta_j^{boc}) + p_{boc5})} \quad (1.34f)$$

Therefore, the bond energy E_{bond} in equation 1.31 as a function of the corrected bond orders BO_{ij} is

$$\mathbf{E}_{bond} = -D_e^{\sigma} \cdot BO_{ij}^{\sigma} \cdot \exp[p_{be1}(1 - (BO_{ij}^{\sigma})^{p_{be2}})] - D_e^{\pi} \cdot BO_{ij}^{\pi} - D_e^{\pi\pi} \cdot BO_{ij}^{\pi\pi} \quad (1.35)$$

As in equation 1.33, Δ_i^e has a similar mathematical form and is the difference between the total number of outer shell electrons and the sum of bond orders around an atomic center.

$$\Delta_i^e = -Val_i^e + \sum_{j=1}^{neighbors(i)} BO_{ij} \quad (1.36)$$

As the total bond order associated with an atom starts to exceed its normal coordination, the lone pair gradually breaks up, causing a deviation Δ_i^{lp} from its optimal number of lone pairs $n_{lp,opt}$, which are defined as

$$\Delta_i^{lp} = n_{lp,opt} - n_{lp,i} \quad (1.37)$$

$$n_{lp,i} = \text{int} \left(\frac{\Delta_i^e}{2} \right) + \exp \left[-p_{lp1} \cdot \left(2 + \Delta_i^e - 2 \cdot \text{int} \left\{ \frac{\Delta_i^e}{2} \right\} \right)^2 \right] \quad (1.38)$$

And so the lone pair energy penalty E_{lp} is defined as

$$E_{lp} = \frac{p_{lp2} \cdot \Delta_i^{lp}}{1 + \exp(-75 \cdot \Delta_i^{lp})} \quad (1.39)$$

For an over-coordinated atom ($\Delta_i > 0$), there's an energy penalty on the system,

$$E_{over} = \frac{\sum_{j=1}^{nbond} p_{ovun1} \cdot D_e^\sigma \cdot BO_{ij}}{\Delta_i^{lpcorr} + Val_i} \cdot \Delta_i^{lpcorr} \cdot \left[\frac{1}{1 + \exp(p_{ovun2} \cdot \Delta_i^{lpcorr})} \right] \quad (1.40a)$$

$$\Delta_i^{lpcorr} = \Delta_i$$

$$- \frac{\Delta_i^{lp}}{1 + p_{ovun3} \cdot \exp \left(p_{ovun4} \cdot \left\{ \sum_{j=1}^{neighbors(i)} (\Delta_j - \Delta_j^{lp}) \cdot (BO_{ij}^\pi + BO_{ij}^{\pi\pi}) \right\} \right)} \quad (1.40b)$$

and for an under-coordinated atom ($\Delta_i < 0$), the energy contribution for the resonance of the π -electron between attached under-coordinated atomic centers should also be considered, and therefore,

$$E_{under} = -p_{ovun5} \cdot \frac{1 - \exp(p_{ovun6} \cdot \Delta_i^{lpcorr})}{1 + \exp(-p_{ovun2} \cdot \Delta_i^{lpcorr})} \cdot \frac{1}{1 + p_{ovun7} \cdot \exp \left[p_{ovun8} \cdot \left\{ \sum_{j=1}^{neighbors(i)} (\Delta_j - \Delta_j^{lp}) \cdot (BO_{ij}^\pi + BO_{ij}^{\pi\pi}) \right\} \right]} \quad (1.41)$$

The valence angle terms include angle energy, penalty energy, and three-body conjugation term. The angle energy is expressed as

$$\begin{aligned}
 E_{val} &= f_7(BO_{ij}) \cdot f_7(BO_{jk}) \cdot f_8(\Delta_j) \\
 &\quad \cdot \left\{ p_{val1} - p_{val1} \cdot \exp \left[-p_{val2} \cdot (\Theta_0(BO) - \Theta_{ijk})^2 \right] \right\} \quad (1.42) \\
 f_7(BO_{ij}) &= 1 - \exp(-p_{val3} \cdot BO_{ij}^{p_{val4}}) \\
 f_8(\Delta_j) &= p_{val5} - (p_{val5} - 1) \cdot \frac{2 + \exp(p_{val6} \cdot \Delta_j^{angle})}{1 + \exp(p_{val6} \cdot \Delta_j^{angle}) + \exp(-p_{val7} \cdot \Delta_j^{angle})} \\
 \Delta_j^{angle} &= -Val_j^{angle} + \sum_{n=1}^{neighbors(j)} BO_{jn} \\
 \Theta_0(BO) &= \pi - \Theta_{0,0} \cdot \{1 - \exp[-p_{val10} \cdot (2 - SBO2)]\} \\
 SBO &= \sum_{n=1}^{neighbors(j)} (BO_{jn}^\pi + BO_{jn}^{\pi\pi}) + \left[1 - \prod_{n=1}^{neighbors(j)} \exp(-BO_{jn}^8) \right] \cdot (-\Delta_j^{angle} - p_{val8} \cdot n_{lp,j}) \\
 SBO2 &= 0 \text{ if } SBO \leq 0 \\
 SBO2 &= SBO^{p_{val9}} \text{ if } 0 < SBO < 1 \\
 SBO2 &= 2 - (2 - SBO)^{p_{val9}} \text{ if } 1 < SBO < 2 \\
 SBO2 &= 2 \text{ if } SBO > 2
 \end{aligned}$$

For systems with two double bonds sharing an atom in a valency angle, an additional energy penalty is introduced,

$$E_{pen} = p_{pen1} \cdot f_9(\Delta_j) \cdot \exp \left[-p_{pen2} \cdot (BO_{ij} - 2)^2 \right] \cdot \exp \left[-p_{pen2} \cdot (BO_{jk} - 2)^2 \right] \quad (1.43a)$$

$$f_9(\Delta_j) = \frac{2 + \exp(-p_{pen3} \cdot \Delta_j)}{1 + \exp(-p_{pen3} \cdot \Delta_j) + \exp(-p_{pen4} \cdot \Delta_j)} \quad (1.43b)$$

and for systems with $-\text{NO}_2-$ group, a three-body conjugation term is included.

$$\begin{aligned}
E_{coa} = p_{coa1} \cdot \frac{1}{1 + \exp(p_{coa2} \cdot \Delta_j^{val})} \\
\cdot \exp \left[-p_{coa3} \cdot \left(-BO_{ij} + \sum_{n=1}^{neighbors(i)} BO_{in} \right)^2 \right] \\
\cdot \exp \left[-p_{coa3} \cdot \left(-BO_{jk} + \sum_{n=1}^{neighbors(k)} BO_{kn} \right)^2 \right] \\
\cdot \exp \left[-p_{coa4} \cdot (BO_{ij} - 1.5)^2 \right] \cdot \exp \left[-p_{coa4} \cdot (BO_{jk} - 1.5)^2 \right]
\end{aligned} \tag{1.44}$$

The torsion angle terms include torsion rotation barriers and four-body conjugation term.

The torsion rotation barriers term ensures that the dependence of the energy of torsion angle ω_{ijkl} accounts properly for BO approaching 0 and for BO greater than 1. It is expressed as

$$\begin{aligned}
E_{tor} = f_{10}(BO_{ij}, BO_{jk}, BO_{kl}) \cdot \sin \Theta_{ijk} \cdot \sin \Theta_{jkl} \\
\cdot \left[\frac{1}{2} V_1 \cdot (1 + \cos \omega_{ijkl}) + \frac{1}{2} V_2 \right. \\
\cdot \exp \left\{ p_{tor1} \cdot (BO_{jk}^\pi - 1 + f_{11}(\Delta_j, \Delta_k))^2 \right\} (1 - \cos 2\omega_{ijkl}) + \frac{1}{2} V_3 \\
\left. \cdot (1 + \cos 3\omega_{ijkl}) \right]
\end{aligned} \tag{1.45a}$$

$$f_{10}(BO_{ij}, BO_{jk}, BO_{kl}) = \tag{1.45b}$$

$$[1 - \exp(-p_{tor2} \cdot BO_{ij})] \cdot [1 - \exp(-p_{tor2} \cdot BO_{jk})] \cdot [1 - \exp(-p_{tor2} \cdot BO_{kl})]$$

$$f_{11}(\Delta_j, \Delta_k) =$$

$$\frac{2 + \exp[-p_{tor3} \cdot (\Delta_j^{angle} + \Delta_k^{angle})]}{1 + \exp[-p_{tor3} \cdot (\Delta_j^{angle} + \Delta_k^{angle})] + \exp[p_{tor4} \cdot (\Delta_j^{angle} + \Delta_k^{angle})]} \tag{1.45c}$$

The four-body conjugation term is defined as

$$E_{conj} = f_{12}(BO_{ij}, BO_{jk}, BO_{kl}) \cdot p_{cot1} \cdot [1 + (\cos^2 \omega_{ijkl} - 1) \cdot \sin \Theta_{ijk} \cdot \sin \Theta_{jkl}] \tag{1.43a}$$

$$\begin{aligned}
f_{12}(BO_{ij}, BO_{jk}, BO_{kl}) \\
= \exp[-p_{cot2} \cdot (BO_{ij} - 1.5)^2] \cdot \exp[-p_{cot2} \cdot (BO_{jk} - 1.5)^2] \\
\cdot \exp[-p_{cot2} \cdot (BO_{kl} - 1.5)^2]
\end{aligned} \tag{1.43b}$$

The hydrogen bond interactions for a X-H-Z system are described by the following term.

$$E_{coa} = p_{hb1} \cdot [1 - \exp(-p_{hb2} \cdot BO_{XH})] \cdot \exp \left[p_{hb3} \left(\frac{r_{hb}^0}{r_{HZ}} + \frac{r_{HZ}}{r_{hb}^0} - 2 \right) \right] \cdot \sin^8 \left(\frac{\Theta_{XHZ}}{2} \right) \quad (1.44)$$

A C₂ correction term is also included so that the triple-bonded of C₂ is not too strong and can be de-stabilized by terminal radical electrons,

$$\begin{aligned} E_{C2} &= k_{C2} \cdot (BO_{ij} - \Delta_i - 0.04 \cdot \Delta_i^4 - 3)^2 & \text{if } BO_{ij} - \Delta_i - 0.04 \cdot \Delta_i^4 > 3 \\ E_{C2} &= 0 & \text{if } BO_{ij} - \Delta_i - 0.04 \cdot \Delta_i^4 \leq 3 \end{aligned} \quad (1.45)$$

and a triple bond correction term is included to stabilize C-O triple bond.

$$E_{trip} = p_{trip1} \cdot \exp \left[-p_{trip2} \cdot (BO_{ij} - 2.5)^2 \right] \cdot \frac{\exp \left[-p_{trip4} \cdot \left(\sum_{k=1}^{neighbors(i)} BO_{ik} - BO_{ij} \right) \right] + \exp \left[-p_{trip4} \cdot \left(\sum_{k=1}^{neighbors(j)} BO_{jk} - BO_{ij} \right) \right]}{1 + 25 \cdot \exp \left[p_{trip3} (\Delta_i + \Delta_j) \right]} \quad (1.46)$$

While the above energy terms are all bond order dependent, van der Waals energy and Coulomb energy are bond order independent. A non-bonded cutoff radius R_{cut} is used so that the non-bonded interactions go to zeros at the cutoff boundary. A Taper term, which is a distance-dependent 7th order polynomial is included in non-bonded interactions expression to avoid discontinuities when charged species move in and out of the non-bonded cutoff radius.

$$E_{vdWaal} = Tap \cdot D_{ij} \cdot \left\{ \exp \left[\alpha_{ij} \cdot \left(1 - \frac{f_{13}(r_{ij})}{r_{vdW}} \right) \right] - 2 \cdot \exp \left[\frac{1}{2} \cdot \alpha_{ij} \cdot \left(1 - \frac{f_{13}(r_{ij})}{r_{vdW}} \right) \right] \right\} \quad (1.47)$$

$$f_{13}(r_{ij}) = \left[r_{ij}^{p_{vdW1}} + \left(\frac{1}{\gamma_w} \right)^{p_{vdW1}} \right]^{\frac{1}{p_{vdW1}}} \quad (1.48)$$

$$E_{Coulomb} = Tap \cdot C \cdot \frac{q_i \cdot q_j}{\left[r_{ij}^3 + \left(\frac{1}{\gamma_{ij}} \right)^3 \right]^{1/3}} \quad (1.49)$$

In ReaxFF, atomic charges are calculated using the Electron Equilibration Method (EEM), which is similar to QEq-scheme except that EEM does not use an iterative scheme for hydrogen charges and uses shielded Coulomb potential instead of rigorous Slater orbital approach to account for charge overlap.

ReaxFF formalism is complicated and usually contains hundreds of parameters in the force field. The new parameter sets are usually trained against experimental data, like crystallography data if interested in solids, and quantum mechanical calculations, like small molecules or cluster information and reaction energetics. The parametrization method used for the systems of interest are explained in Chapter 3 and 4. Similarly, parametrization of a Tersoff potential is explained in Chapter 2.

1.3 SARS-CoV-2 and Glycoproteins

The second half of the dissertation (Chapter 5, 6, and 7) is dedicated to the glycosylated protein simulations of SARS-CoV-2 that caused the COVID-19 pandemic.⁵⁰ Among the seven known coronavirus that infect humans, four of them (229E, NL63, OC43, and HKU1) only cause slight symptoms of the common cold, but the other three (SARS-CoV, MERS-CoV, and SARS-CoV-2) can cause severe symptoms and even death. The surface of SARS-CoV-2 is largely covered by

glycosylated S proteins that can bind to the human host cell receptor angiotensin-converting enzyme 2 (ACE2) and mediate viral cell entry.⁵¹ When the virus enters the cell, the viral RNA is released, and replication and transcription of the viral RNA genome occur. Figure 1.4 shows the schematic structure of the S protein and how it binds to ACE2.

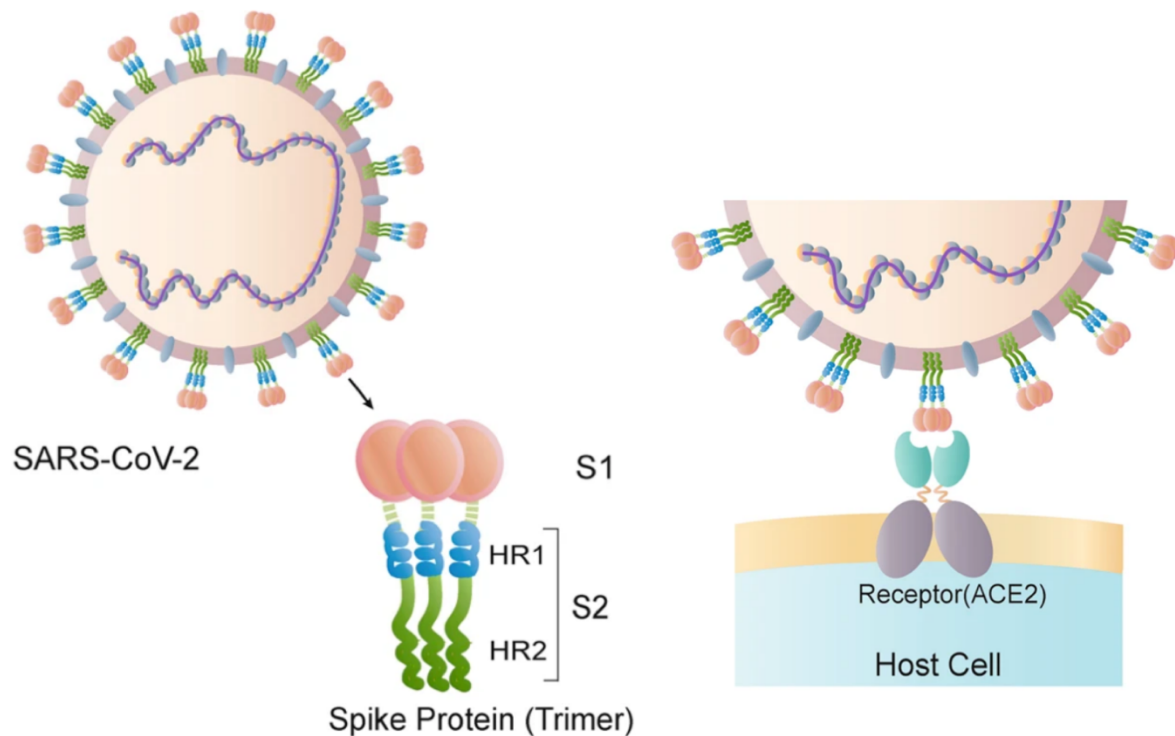


Figure 1.4: Simplified structure of SARS-CoV-2 and S protein, and the schematic of S protein binding to receptor ACE2.⁵²

The S protein consists of an extracellular N-terminus, a transmembrane (TM) domain anchored in the viral membrane, and a short intracellular C-terminal segment.⁵³ As shown in figure 1.4, the S1 receptor binding domain (RBD) in the S protein binds to the human ACE2. S protein are often glycosylated to camouflage them so that they can evade the host immune system during entry, modulate access to protease, and enhance cellular attachment at the viral-host interface.⁵⁴⁻

⁶⁰ The sugar molecules attached to the glycoprotein, like S protein, are called glycans. N-glycosylation and O-glycosylation are the two classes of glycosylation, and for the work in this dissertation, N-glycosylation is studied where the glycans are attached to the side chain nitrogen of asparagine (N) protein residues.⁶¹

Despite the importance of the glycosylation of S protein, there are little attempt in developing the computational model and simulate the interactions between S protein and the ACE2 protein with glycosylation. Chapter 5-7 attempt to better understand the role of glycosylation and provide insights for future therapeutic and diagnostic purpose in this field.

Chapter 2

Computational Modelling of Atomic Layer Etching of Chlorinated Germanium Surfaces by Argon

My contribution to this work included all the DFT calculations and part of the Tersoff potential parametrization.

This chapter originally appeared in the literature as:

Zhang S, Huang Y, Tetiker G, Sriraman S, Paterson A, Faller R. Computational modelling of atomic layer etching of chlorinated germanium surfaces by argon. *Physical Chemistry Chemical Physics*. 2019;21(11):5898-902.

Abstract

The atomic layer etching of chlorinated germanium surfaces under argon bombardment was simulated using molecular dynamics with a newly fitted Tersoff potential. The chlorination energy determines the threshold energy for etching and the number of etched atoms in the bombardment phase. Etch rate is determined by bombardment energy.

2.1 Introduction

With the decrease of electronic device size from the micro, via the nano, down to the atomic scale, the variability of the surface patterning needs to be controlled within dozens of atoms accordingly.^{1,2} Plasma etching is an essential surface patterning technique. Conventionally, it is realized by ionizing non-reactive species such as Ar to transfer energy to the surface, and at the

same time another reactive species such as Cl_2 etches the surface either in its neutral or ionized state.³⁻⁶ However, mixing of energetic and reactive flows leads to coupling of their transport,¹ which increases complexity and makes it hard to achieve highly defined low feature sizes. Moreover, this mixing can lead to the damage of the etching front, as high energy ions distort the surface structure and open a way for reactive species to penetrate deep into the substrate (> 20 nm).¹ Thus, new methods are needed to meet the requirement of precise etching control at the atomic scale, and atomic layer etching (ALE) has been proposed as one solution.^{1, 2, 7, 8} The basic idea is to separate the energetic and reactive flows to avoid the above problems. It removes thin layers of material based on a two-step self-limiting reaction: surface modification and removal. Surface modification makes a surface layer within a well-defined range more easily to be etched afterwards; for example, one can use halogen gas adsorption on the semiconductor surface to weaken the bonds; in the removal step, plasma ions are controlled within a certain range of energy to remove the modified layers while avoiding physical sputtering of untreated materials.

The first report of this technique is by Yoder et al. in 1988 on diamond.⁹ Silicon and GaAs have been the most studied ALE materials since¹. Germanium is another interesting material as a candidate to replace Si for application as transistor channels due to its superior hole mobility.² Several papers have been studied ALE of Ge experimentally,^{8, 10, 11} but the microscopic understanding of the influence of the process variables on etching is still missing. Such a study is crucial to determine appropriate conditions for ALE to achieve optimal accuracy and efficiency. To bridge this gap, we report our results of an ALE simulation using molecular dynamics on a Ge surface. The surface modification step involves chlorine adsorption (chlorination) of the Ge (100) surface. Chlorine has been found to be non-reactive with Ge below $350\text{ }^\circ\text{C}^2$, and thus is suitable

to be used for a self-limiting reaction at room temperature. Removal is then achieved by Ar⁺ bombardment. In our simulation, neutral Ar is used instead of its charged state as the ion will become neutral when approaching the surface.¹²

2.2 Tersoff Potential and Simulation Methods

A new variation of the Tersoff potential based on DFT data was developed for the Ge-Cl interaction, the format used is shown in eq.1-8 (see ESI section 1 for discussion on the choice of potential and its training process).

$$E = \frac{1}{2} \sum_i \sum_{j \neq i} V_{ij} \quad \text{eq.1}$$

$$V_{ij} = f_C(r_{ij})[f_R(r_{ij}) + b_{ij}f_A(r_{ij})] \quad \text{eq.2}$$

$$f_C(r) = \begin{cases} 1: r < R - D \\ \frac{1}{2} - \frac{1}{2} \sin\left(\frac{\pi r - R}{2D}\right): R - D < r < R + D \\ 0: r > R + D \end{cases} \quad \text{eq.3}$$

$$f_R(r) = A \exp(-\lambda_1 r) \quad \text{eq.4}$$

$$f_A(r) = -B \exp(-\lambda_2 r) \quad \text{eq.5}$$

$$b_{ij} = (1 + \beta^n \zeta_{ij}^n)^{-\frac{1}{2n}} \quad \text{eq.6}$$

$$\zeta_{ij} = \sum_{k \neq i, j} f_C(r_{ik}) g(\theta_{ijk}) \exp[\lambda_3^m (r_{ij} - r_{ik})^m] \quad \text{eq.7}$$

$$g(\theta) = \gamma_{ijk} \left(1 + \frac{c^2}{d^2} - \frac{c^2}{[d^2 + (\cos\theta - \cos\theta_0)^2]}\right) \quad \text{eq.8}$$

where f_R is a two-body repulsion term and f_A the attraction term. The bond order term b_{ij} is a three-body term added to the attraction part, which includes all neighbors j and k of atom i . f_C is the

interaction cutoff function to smoothly transit full interaction to zero. R-D and R+D is the inner and outer cutoff distance. Interaction parameters for Ge-Ge and Cl-Cl atom pairs are available from previous literature, so only Ge-Cl interaction terms were trained, as shown in table 2.1 and 2.2.

Fitted Parameters	Ge-Cl
A (eV)	521.8338
B (eV)	62.9673
$\lambda_1(\text{\AA}^{-1})$	2.3248
$\lambda_2(\text{\AA}^{-1})$	0.9857
$r_{min}(\text{\AA})$	2.53
$r_{max}(\text{\AA})$	2.83

Table 2.1 Fitted two-body parameters using annealing algorithm. Rmin=R-D and Rmax=R+D are set to be the average value from Ge-Ge and Cl-Cl cutoff distances.

j-i-k Parameters	Cl-Ge-Cl	Cl-Ge-Ge	Ge-Cl-Ge	Ge-Cl-Cl	Ge-Ge-Cl
m	3.0	3.0	1.0	1.0	3.0
γ_{ijk}	1.0	1.0	4.0	4.0	1.0
λ_3	0.0	0.0	6.0	6.0	0.0
c	1.0643e5	1.0643e5	0.0	0.0	1.0643e5
d	15.652	15.652	1	1	15.652
$\cos\theta_0$	-0.43884	-0.43884	1	1	-0.43884
n	2.23669	0.67543	1	1	0.75627
β	1.234e-4	1.0273e-5	1	1	9.02e-7

Table 2 Fitted parameters in bond order b_{ij} term.

Fig. 2.1 shows the validation of this potential on a small $\text{Ge}_9\text{Cl}_{12}$ cluster, similar to a previous study in Si-F case¹³. The energy difference is obtained by energy minimization of the corresponding structure in both LAMMPS (Tersoff) and Gaussian (DFT). The B3LYP functional with the LanL2DZ basis set is used for DFT. Initial conditions with different Ge-Cl distance and angle were tested for each case, and consistent minimized structure was obtained. The bond length of the surface Ge-Ge in $\text{Ge}_9\text{Cl}_{12}$ was found to be 2.43 Å (DFT: 2.56 Å), and adsorbed Ge-Cl bond length to be 2.22 Å (DFT: 2.20 Å). The energy change calculated by the fitted Tersoff potential predicts exothermic (endothermic) nature for adsorption (desorption), and the desorption energy barrier is well above 350 °C, although the energy difference is 1 eV lower than in DFT.

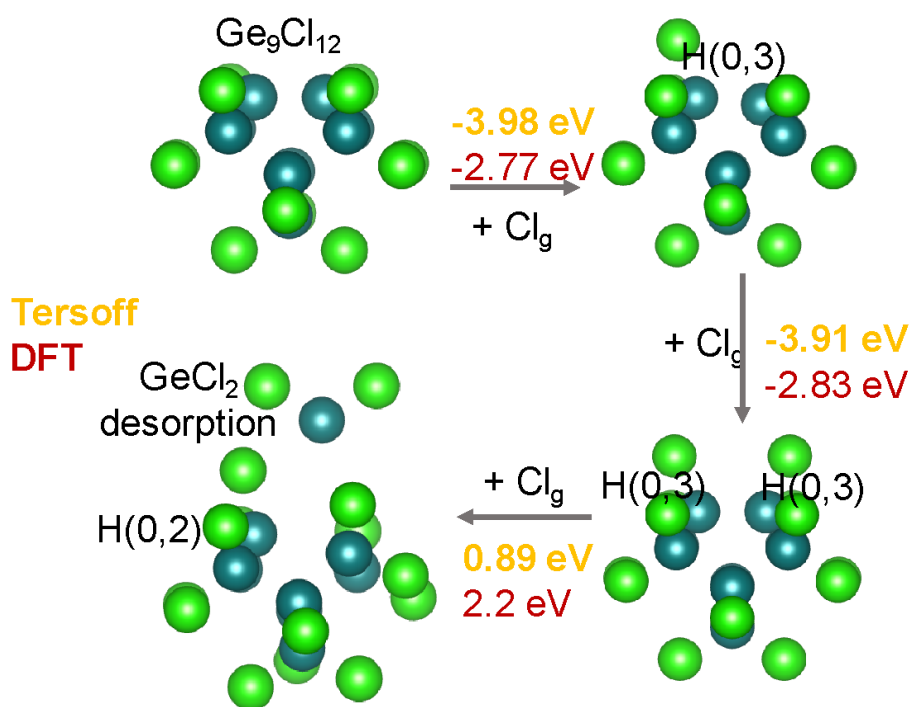


Figure 2.1: DFT and Tersoff calculated energy change during Cl_2 adsorption and desorption from $\text{Ge}_9\text{Cl}_{12}$ cluster. Light green: Cl. Cyan: Ge. H(0,2) and H(0,3) marks the number of neighbour atoms k of Cl and Ge

type for the surface Ge atom, not including its neighbour atom j . For example, H(0,2) means the surface Ge has zero Cl neighbours if not including the adsorbed Cl atom j , and has 2 other Ge atoms as its neighbors.

2.3 Results and Discussions

A Ge slab of surface dimension $64 \text{ \AA} \times 64 \text{ \AA}$ (82×82 in lattice units, about 256 Ge atoms on the top surface) and 113 \AA thick was first equilibrated at room temperature in LAMMPS¹⁴ in the NPT ensemble (anisotropic zero pressure in surface xy plane), with $T_{\text{damp}}=0.1 \text{ ps}$ and $P_{\text{damp}}=10 \text{ ps}$. Dimers were found to form, consistent to both DFT and experimental results (see ESI section 2 for details). 800 Cl atoms were then introduced with a flux normally incident onto the surface with a defined kinetic energy. One Cl atom at a time was introduced to the simulation box every 0.1 ps and the system was then equilibrated for 2 ns. Single Cl atoms instead of Cl_2 molecules were used to shorten the equilibration time for the adsorption process, as with Cl_2 it also involves the Cl-Cl bond breaking step when it is adsorbed. Charged Cl species were not considered similar to the reason that Ar was used instead of Ar^+ . Four different chlorination energies were tested: 5 eV, 10 eV, 25 eV and 50 eV, and simulation was repeated with different initial Cl positions to generate three chlorinated configurations for each condition. Non-adsorbed Cl atoms or Cl_2 molecules (Cl_2 can be formed from two Cl atoms) were removed from the simulation cell. The adsorption criteria are based on the Ge-Cl bond length 2.53 \AA , which means a Cl atom is considered non-adsorbed if it is more than 2.53 \AA away from a surface Ge atom. Only chemisorption is considered. Chlorinated configurations were then subjected to Ar bombardment with energies varying from 20 eV to 100 eV. The interaction between Ar and Ge/Cl is described by Ziegler-Biersack-Littmark (ZBL) potential,¹⁵ which describes high-energy collisions between atoms. The Ge slab was coupled to a

room-temperature thermostat. During a bombardment cycle, one Ar atom is introduced towards the surface and stays in the simulation box for 1.5 ps. Then the Ar atom is deleted and the system equilibrates for 0.5 ps. Etched products were then deleted before the start of a new bombardment cycle. Two hundred bombardment cycles were performed for each test condition, and each condition was repeated three times with different chlorinated configurations. The determination of etching products is again based on bond distance r_{ij} . If $r_{ij} < R_{ij}(1)$ (which is the inner cutoff distance for the i, j interaction in Tersoff potential), then atoms i and j are considered to be in the same cluster.

During chlorination, Cl atoms quickly adsorb and partially penetrate beneath the surface due to the strong incident energy. Few single Ge atoms and GeCl products leave the surface. Fig. 2.2a shows the side view of Ge surface after chlorination, where increasing Cl adsorption and penetration with Cl incident energy can be seen. The adsorbed/absorbed Cl number increases from 3205 atoms (5 eV) to 6164 atoms (50 eV). Defining Cl penetration depth as the Full Width at Half Maximum of Cl density peak, Fig. 2b demonstrates that the penetration depth increases with Cl incident energy. Notice the coverage on surface already reaches 100 % at 5 eV (coverage is defined as the Cl-bonded Ge numbers to the number of top layer Ge). From Ge-Cl coordination number (CN) characterization (Table 3), most Ge are bonded to one Cl, however increasing chlorination energy will increase the bonded Cl number (from the increase of the average CN). Also, more Ge will bond with Cl (from the increase of the coverage). The high population of Ge-1Cl bonding is because the energy decrease from Ge-Ge to Ge-1Cl is much larger than that from Ge-1Cl to Ge-nCl. It is worth noting that the result could be sensitive to force field training, if the bond energy change from GeCl to GeCl₂ is larger than Ge-Ge to Ge-1Cl, then Ge-2Cl can be a more favorable configuration. The decrease of energy with Cl penetration can be also seen in the system energy

(Fig. 2.2b). The system energy is defined as the total energy divided by the number of atoms in the system. It decreases with chlorination energy, or Cl numbers, as bonding between Ge and Cl is stronger than that among Ge. The Ge surface structure becomes highly distorted with increasing chlorination energy, as the radial distribution function in Fig. 2.2c shows the eventual formation of amorphous structures at 25 eV and 50 eV. It is also seen that the Ge-Ge bond length increases from 2.8 Å (in the perfect crystal) to 3.2 Å due to Cl penetration, as the first peak in RDF shifted towards longer bond distance with increasing energy.

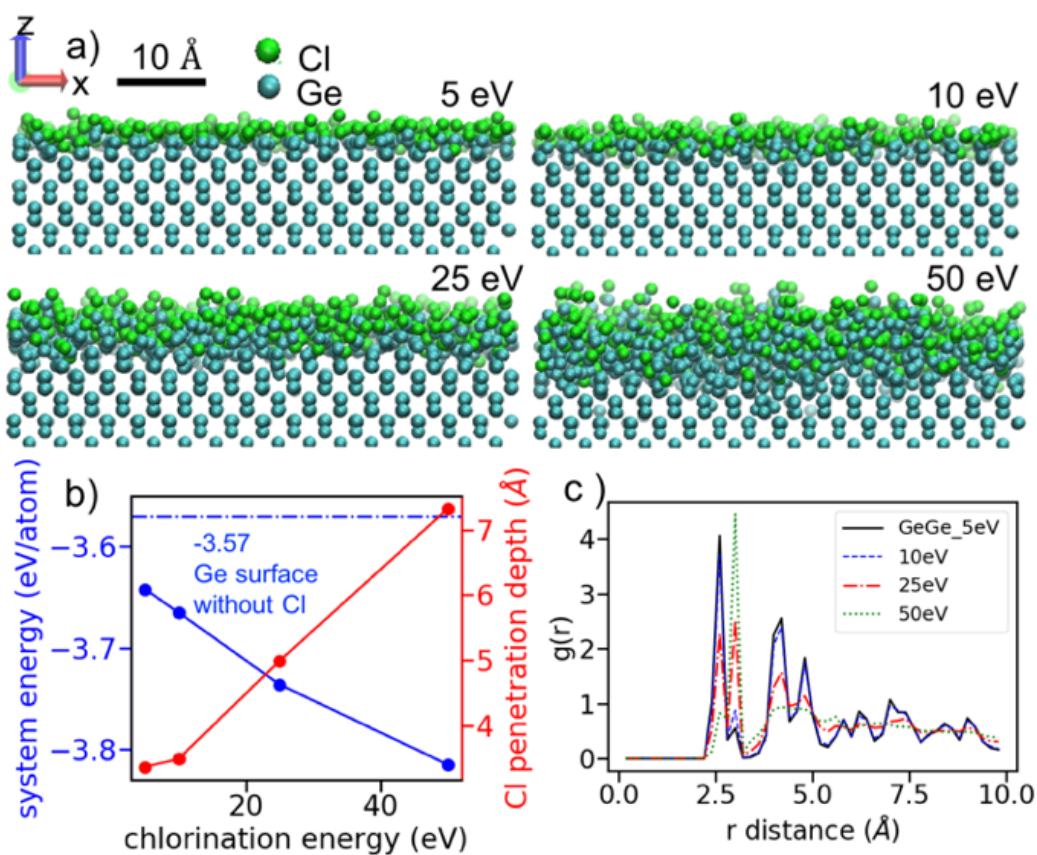


Figure 2.2: a) Side view of Ge surface after chlorination with different energies. The layer thickness showing here is around 16 Å. Non-adsorbed Cl/Cl₂ away from surface have been deleted. b) System energy (including both kinetic and potential energy) after equilibration (error bar smaller than the symbol) and Cl

penetration depth change with chlorination energy. The dashed blue line represents the system energy in pure Ge surface structure, which is -3.57 eV/atom. c) Radial distribution function (RDF) of Ge-Ge atom pair within 10 Å from the chlorinated surface generated by different chlorination energy.

The influence of chlorination energy extends to the Ar bombardment stage. During the bombardment, energy from Ar was transferred to the surface and the weakened Ge-Ge bonds break, allowing Ge-Cl clusters or single atoms to leave as etching products. Few residual Cl atoms remain on the surface (Fig. 2.3, also from the difference between the etched Cl number and total adsorbed Cl number during chlorination).

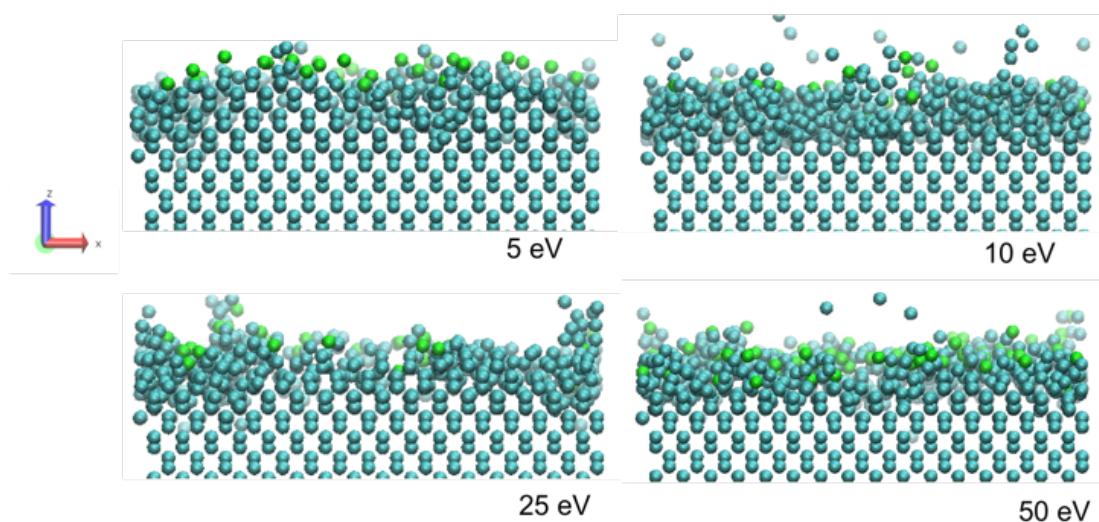


Figure 2.3: Snapshots of Ge surface after 100 eV Ar bombardment (200 times). Light green: Cl. Cyan: Ge.

To compare the surface morphology with different chlorination energies before and after the bombardment, density profiles (Fig. 2.4a and 2.4b) are used to show the etching front evolution. The Z axis is the surface normal pointing from the bulk to the top surface. The Ge density profile starts with a plateau in the bulk phase and then decreases to zero, where the transition region

represents the surface layers. Correspondingly, Cl shows a density peak in this transition region due to surface adsorption and penetration. For initial configurations, the 50 eV chlorinated case shows a wider Cl peak distribution than 5 eV and Cl penetrates deeper as the peak position moves towards bulk, consistent with Fig. 2.2b. After bombardment with 100 eV Ar, slightly more Ge layers have been etched away with the highly chlorinated configuration as the profile recedes more into the bulk region, while a similar amount of Cl residues stays on the surface. This could be understood by the fact that more Cl was adsorbed on the surface originally.

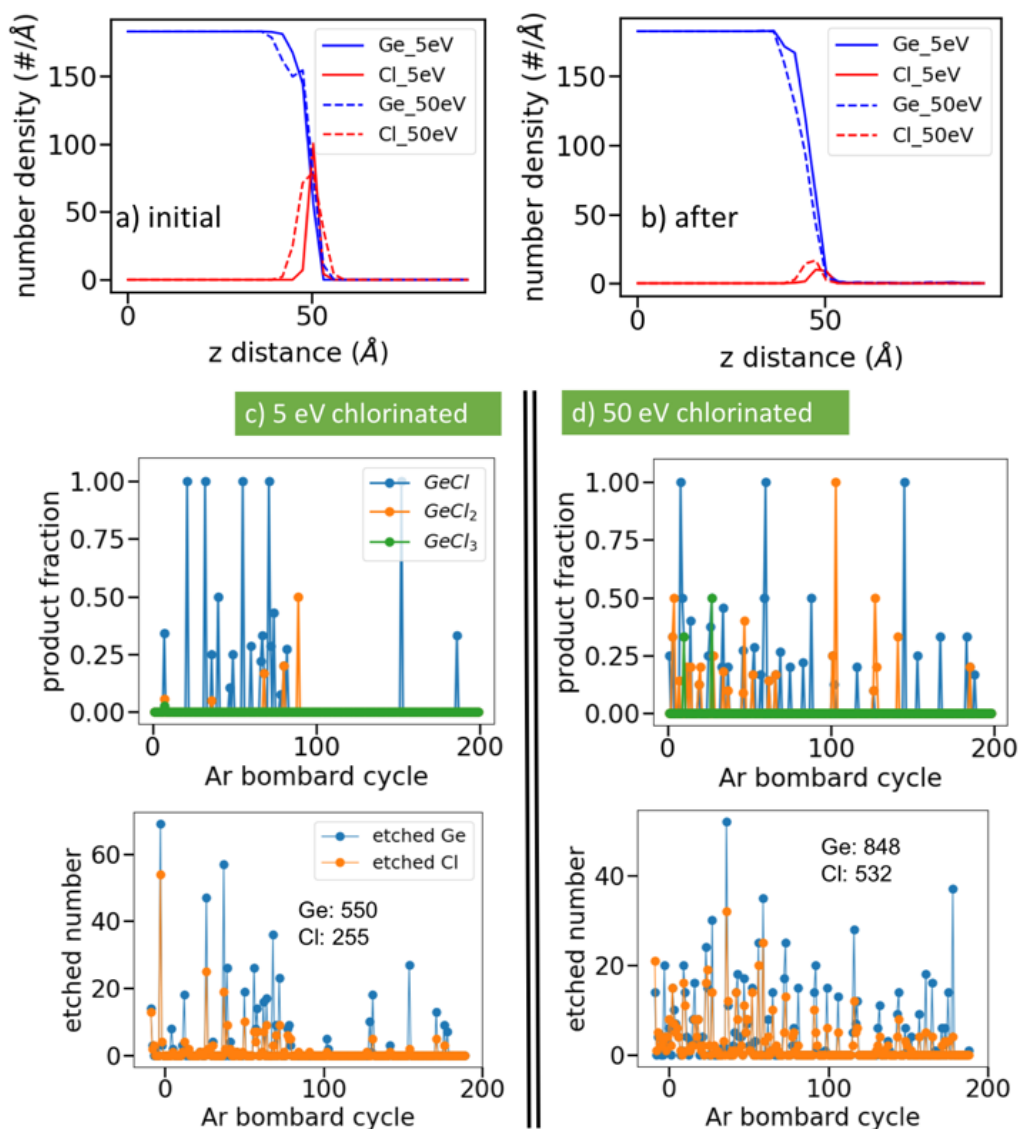


Figure 2.4: Number density change of 5 eV chlorinated and 50 eV chlorinated surface a) before and b) after 100 eV Ar bombardment. Etching products fraction (GeCl, GeCl₂ and GeCl₃) and element etched number change in each Ar bombardment cycle for c) 5 eV and d) 50 eV chlorinated surface. The insets number on etched number graph shows the total etched Ge and Cl number. Sample 1 is used as an example here. Characterizations for 10 eV and 25 eV chlorinated surface are shown in ESI Fig. S7.

Next, to characterize the influence of chlorination energy on etching products, the product fraction, defined as the ratio of the number of a product to the total number of all the products (including physical sputtering products), is plotted for each Ar bombardment cycle, as shown in Fig. 2.4c and 2.4d. Ge_xCl_y (with $x>y$) clusters and single Ge atoms also appear during bombardment cycles (not shown in the above figure, fraction higher than GeCl). They are likely due to physical sputtering of several atoms together. Cl_2 was barely observed as an etching product in our case, almost all Cl were attached to Ge. GeCl was found to be the main chemical etching product, fluctuating around 55%-85% depending on energy conditions (see ESI Fig. S8 and Table S6), if only GeCl_n clusters are considered. GeCl_2 is of between 15 %-35 %, followed by a very small fraction of GeCl_3 (below 10%). The latter may be due to a slight inaccuracy in the forcefield (see ESI). No GeCl_4 is found under any condition. This distribution follows the Ge-Cl CN trend in the chlorination step, as Ge-1Cl has the highest fraction, followed by Ge-2Cl and Ge-3Cl. It's likely the chlorination morphology determines the etching product. Besides, for surface with low chlorination energy such as the 5 eV case (Fig. 2.4c), both Ge and Cl etched number decrease greatly after about hundred bombardment cycle, as the surface becomes deficient in Cl. This indicates two things: first, chlorination is necessary to weaken Ge-Ge bond strength and makes etching possible; second, amorphization of top Ge surface layers also weakens Ge-Ge bond strength such that physical sputtering is only possible within modified layers. With higher chlorination energy, the self-terminated etching process could require more Ar bombardment cycles.

By varying the Ar bombardment energy from 25 eV to 100 eV, we see that the chlorination energy influences the threshold energy for etching. The density profile (Fig. 2.5a) indicates whether etching occurs. Fig. 2.5a shows the evolution of element density profile of 25 eV

chlorinated surface under different bombardment condition (see ESI Fig. S9-S11 for other chlorinated conditions). At 25 eV Ar energy, Cl density peak already starts to decrease, and Ge surface recedes towards the bulk phase. With 50 eV Ar, the etching effect is more obvious, and most Cl has been etched away when Ar energy reaches 75 eV. Thus, the threshold energy should be below 25 eV Ar, matching well with the experimental operation window.² Similarly, the 5 eV chlorinated surface has a threshold energy between 50-55 eV. For the 10 eV chlorinated case, this threshold energy slightly decreases to 25-50 eV, and for 25 eV and 50 eV chlorinated surface, etching starts below 25 eV. Thus, a higher chlorination energy generally decreases the threshold energy for bombardment, possibly because the highly disordered surface structure makes it more vulnerable for detachment.

Meanwhile, we notice the complete etching (defined as when the Cl peak density approaches zero and no longer evolves) occurs between 60-75 eV bombardment energy for 5 eV chlorinated surface, while it increases to the 75-100 eV range for 10 and 25 eV chlorinated surfaces, and higher than 100 eV for the 50 eV case, possibly because the Cl atoms penetrate deeper into the surface for high chlorination energy. Besides, both chlorination energy and bombardment energy have an impact on the etching efficiency. Fig. 2.5b shows the change of the total etched number of Ge and Cl atoms with Ar energy for each chlorinated condition. Several things can be noticed here. First, both Ge and Cl etching increases with Ar energy. Since the etching time and bombardment flux is fixed, this indicates a higher etching rate and efficiency at higher bombardment energy. Second, the etching numbers change of Ge and Cl are coupled. Both etched Cl and Ge number in 5 eV chlorinated case tend to reach a plateau with increasing Ar energy, due to the limited Cl number and modification layer in the system (self-limited etching process), which further proves the bond weakening by chlorination. Higher chlorination energy such as 50 eV postpones the appearance of

the plateau, as more Cl are originally in the system. If the Ar energy is further increased beyond the plateau, the etched Ge number should increase quickly again due to pure physical sputtering.¹ This naturally leads to the third conclusion that this limit of chemically etched number depends on chlorination energy, or the number of Cl available in the system.

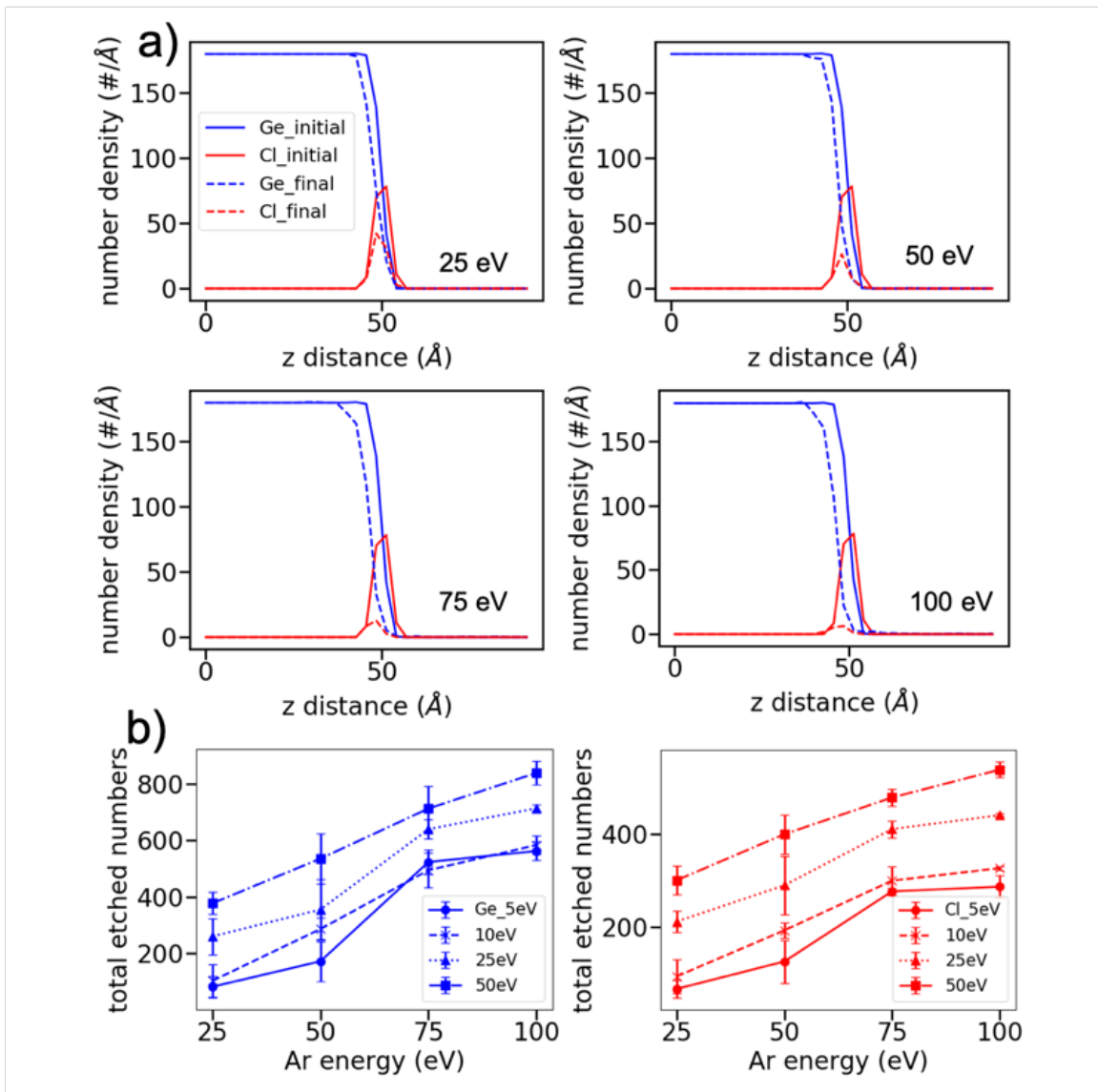


Figure 2.5: a) Density profile of Ge and Cl element change before and after 200 times Ar bombardment with different bombardment energy on 25 eV chlorinated surface. b) Total etched number change (blue: Ge, red: Cl) with Ar bombardment energy for each chlorination condition, averaged from three samples with error bars calculated.

2.4 Conclusion

In conclusion, we used molecular dynamics with a newly improved and validated Ge-Cl Tersoff potential to understand the influence of operation energy on each stage in ALE of Ge and their synergy effect. The chlorination energy is the key factor in surface modification step. Increasing chlorination energy increases Cl adsorption and its penetration depth. It also increases Ge-Cl coordination number from Ge-1Cl to Ge-2Cl and turns the surface into amorphous layer after 25 eV. The configuration and composition of modified surface then dominates the etching results in Ar bombardment step. First, GeCl is the main chemical etching product in our case, as Ge-1Cl is also the main CN during chlorination. Second, the chlorination quantity determines the limit of chemically etched number in bombardment stage. This reflects the self-limiting characteristic in ALE, as the etching termination point is based on the number of Ge-Ge bonds that have been weakened. The more Cl atoms are bonded in the system, the more etched atoms leave. Third, the extent of chlorination determines the threshold for etching. Highly chlorinated surfaces lead to a lower threshold energy, and at the same time a higher energy or more Ar bombardment cycles for complete etching. The lowest threshold energy is below 25 eV for the 25 and 50 eV chlorinated cases. In terms of the influence of bombardment energy, the effective self-terminated etching would require a specific energy range, higher than the threshold energy but lower than

spontaneous physical sputtering. And within this range, the increase of bombardment energy leads to an increase of etching rate.

Our work demonstrates the microscopic link between the surface modification and bombardment step in Ge-Cl reaction system. Future work should focus on the improvement of the Ge-Cl force field, a full characterization of the operation energy window, and experimental validation.

Acknowledgement

We thank the Silicon Valley Community Foundation for a gift which partially supported this research.

References

1. K. J. Kanarik, T. Lill, E. A. Hudson, S. Sriraman, S. Tan, J. Marks, V. Vahedi and R. A. Gottscho, *Journal of Vacuum Science & Technology A*, 2015, 33, 020802.
2. K. J. Kanarik, S. Tan, W. Yang, T. Kim, T. Lill, A. Kabansky, E. A. Hudson, T. Ohba, K. Nojiri, J. Yu, R. Wise, I. L. Berry, Y. Pan, J. Marks and R. A. Gottscho, *Journal of Vacuum Science & Technology A*, 2017, 35, 05C302.
3. J. W. Coburn and H. F. Winters, *Journal of Applied Physics*, 1979, 50, 3189-3196.
4. R. A. Gottscho, C. W. Jurgensen and D. J. Vitkavage, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 1992, 10, 2133-2147.
5. K. Nojiri, *Dry etching technology for semiconductors*, Springer, 2015.
6. V. M. Donnelly and A. Kornblit, *Journal of Vacuum Science & Technology A*, 2013, 31, 050825.
7. G. S. Oehrlein, D. Metzler and C. Li, *ECS Journal of Solid State Science and Technology*, 2015, 4, N5041-N5053.
8. K. J. Kanarik, S. Tan and R. A. Gottscho, *The Journal of Physical Chemistry Letters*, 2018, 9, 4814-4821.
9. Patent 4,756,794, 1988.
10. T. Sugiyama, T. Matsuura and J. Murota, *Applied Surface Science*, 1997, 112, 187-190.
11. K. Ikeda, S. Imai and M. Matsumura, *Applied Surface Science*, 1997, 112, 87-91.
12. H. D. Hagstrum, *Physical Review*, 1954, 96, 325-335.
13. D. Humbird and D. B. Graves, *The Journal of Chemical Physics*, 2004, 120, 2405-2412.
14. S. Plimpton, *Journal of computational physics*, 1995, 117, 1-19.
15. J. F. Ziegler, J. Biersack and U. Littmark, *I*, Pergamon Press, New York, 1985.

2.5 Supporting Information

2.5.1 Tersoff force field training

2.5.1.1 Discussion on the choice of the force field

Tersoff force field has been selected to describe the Ge-Cl interaction, because this bond order potential was originally calibrated for covalent elements including C, Si and Ge,¹⁻³ which have similar structural and chemical properties. Improved versions of Tersoff have been developed and used widely in C and Si reaction studies,⁴⁻⁷ however we stick to the original version here for several reasons. First, some improvements are to achieve a better description for double and conjugation bond, such as the Brenner's correction terms in REBO potential.⁵ This is important for carbon system. However, for either Si or Ge, due to their increased metallicity, double and conjugation bonds rarely form, and such corrections can be unnecessary. Second, previous studies noticed the Si-F bond energy trend has an abnormal value in SiF₃ structure⁴ similar to our GeCl₃ case due to hybridization states (shown in section 1.3). A correction term in bond order b_{ij} can be added for better fitting the bond energies.^{4, 8} However, due to our limited training data, the addition of such correction term leads to an overfitting problem, such that the bond energy change can be fit accurately, while the fitting accuracy of bond length and bond force constant both decrease. Since each property weighs differently based on the purpose of a simulation study, here we choose to lower the total fitting error (eq.1 in Fig. S1) instead of the bond energy term only. Further optimizations can be certainly made with more training data, and a different weighing on the target terms.

2.5.1.2 The format

We used the following format as implemented in LAMMPS:

$$\begin{aligned}
E &= \frac{1}{2} \sum_i \sum_{j \neq i} V_{ij} \\
V_{ij} &= f_C(r_{ij}) [f_R(r_{ij}) + b_{ij} f_A(r_{ij})] \\
f_C(r) &= \left\{ \begin{array}{l} 1: r < R - D \\ \frac{1}{2} - \frac{1}{2} \sin\left(\frac{\pi r - R}{2D}\right): R - D < r < R + D \\ 0: r > R + D \end{array} \right\} \\
f_R(r) &= A \exp(-\lambda_1 r) \\
f_A(r) &= -B \exp(-\lambda_2 r) \\
b_{ij} &= (1 + \beta^n \zeta_{ij}^n)^{-\frac{1}{2n}} \\
\zeta_{ij} &= \sum_{k \neq i, j} f_C(r_{ik}) g(\theta_{ijk}) \exp[\lambda_3^m (r_{ij} - r_{ik})^m] \\
g(\theta) &= \gamma_{ijk} \left(1 + \frac{c^2}{d^2} - \frac{c^2}{[d^2 + (\cos\theta - \cos\theta_0)^2]} \right)
\end{aligned}$$

where f_R is a two-body repulsion term and f_A the attraction term. The bond order term b_{ij} is a three-body term added to the attraction part, which includes all neighbors j and k of atom i . f_C is the interaction cutoff function to smoothly transit full interaction to zero. $R-D$ and $R+D$ is the inner and outer cutoff distance.

2.5.1.2 Training the parameters

Interaction parameters for Ge-Ge and Cl-Cl atom pairs are available from previous literature.^{2,}

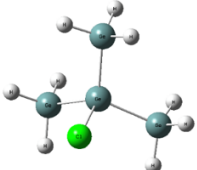
⁴ So only Ge-Cl interaction terms need to be trained. To do this, two-body parameters A , λ_1 , B and λ_2 are first optimized against bond energy, force and force constant of several GeCl_n species (Table S1 and S2) from DFT calculation using the annealing algorithm similar to previous studies, as shown in Fig. S1 and S2.

Table S1 Training data for Cl-Ge-Cl configurations.

Species	Cl-Ge-Cl angle (°)	Ge-Cl bond distance (Å)	Ge-Cl bond energy (eV)	Force constant (N/m)
Ge_2Cl_6	109.47	2.23	4.21	117.72
GeCl_4	109.47	2.22	4.23	163.41
GeCl_3	108.69	2.27	1.47	109.00
GeCl_2	101.09	2.30	4.16	141.26

GeCl	N/A	2.31	3.28	122.60
Ge ₄ H ₉ Cl	N/A	2.32	4.08	131.67

Table S2 Training data for Ge-Ge-Cl configurations.

Species	Ge-Ge-Cl angle (°)	Ge-Ge distance (Å)	bond energy (eV)	bond
Ge ₂ Cl ₆	109.78	2.50	2.91	
Ge ₄ H ₉ Cl 	104.21	2.49	/	

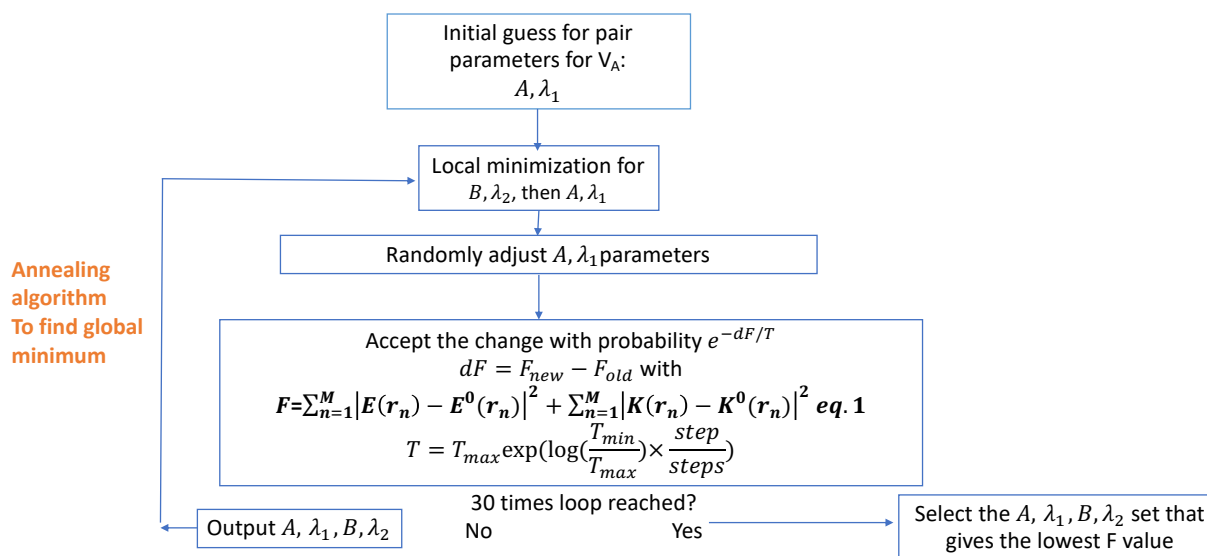


Figure S1 Annealing algorithm training scheme for two-body parameters.

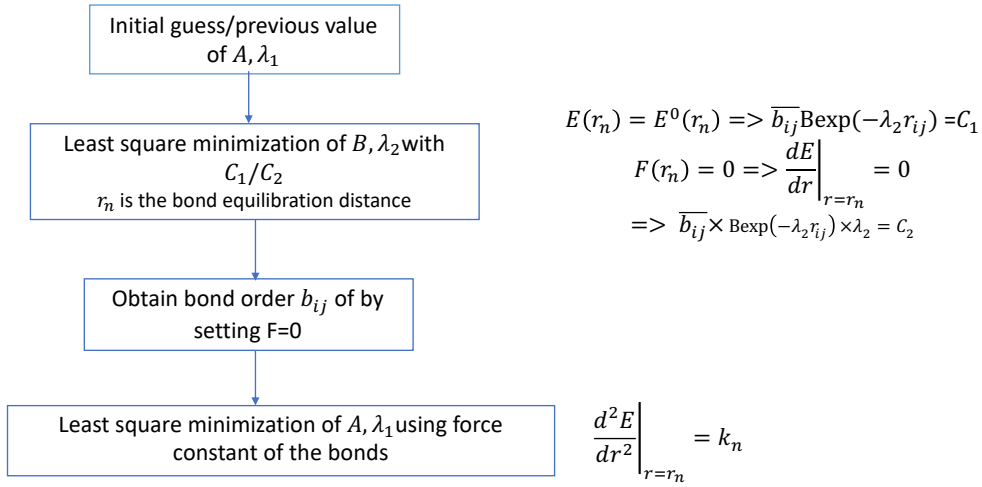


Figure S2 Local minimization scheme as mentioned in the annealing algorithm.

Eq.1 in Fig. S1 is our error function, which tries to minimize the difference in bond energy (E), force (which=0 as we require it to be during local minimization) and force constant (K) in multiple structures at their bond length (DFT value). The fitted results for the two-body parameters are shown in Table S3.

Table S3 Fitted two-body parameters using annealing algorithm. $R_{\min}=R-D$ and $R_{\max}=R+D$ are set to be the average value from Ge-Ge and Cl-Cl cutoff distances.

Fitted Parameters	Ge-Cl
A (eV)	521.8338
B (eV)	62.9673
λ_1 (\AA^{-1})	2.3248
λ_2 (\AA^{-1})	0.9857
r_{\min} (\AA)	2.53
r_{\max} (\AA)	2.83

Bond energy and force constant calculated using the above fitted value are shown in Fig. S3 and Fig. S4. Notice in Fig. S3, GeCl_3 (bond length 2.27 \AA) has an abnormal shallow bond strength comparing to other GeCl_n case, which is hard to capture with the basic Tersoff formalism. Its low bond energy indicates its instability, as this product also seems hard to detect in reality. So one

consequence of our current fitting is we overestimated the bond energy of GeCl_3 and will make it one of the possible etching products.

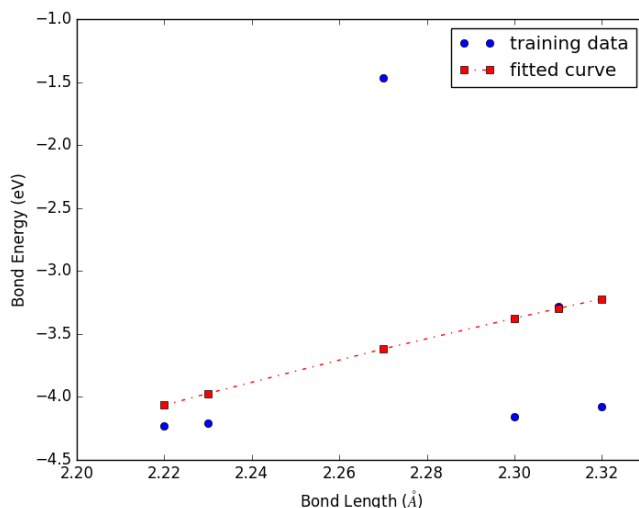


Figure S3 Bond energy fitting curve.

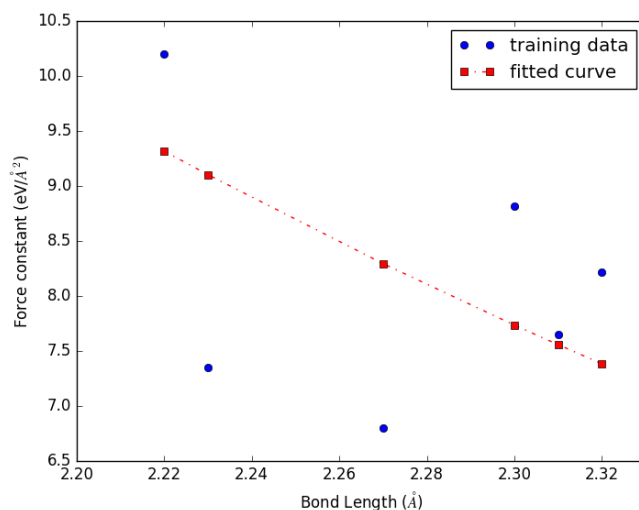


Figure S4 Force constant fitting curve.

Once these two-body term parameters are fixed, the bond order value b_{ij} can be calculated (by setting force $F=0$) and used to train its parameters. Following previous literatures,⁶ i-j-k atom type combinations can be categorized in the way shown in Table S4.

Table S4 Data set for different i-j-k atom type combinations.

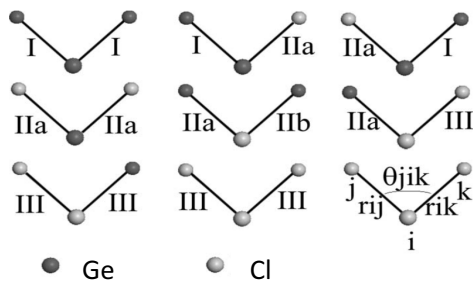
j	i	k	n, β	$g(\theta)$, m, λ_3
----------	----------	----------	------------------------------	---

Ge	Ge	Ge	$^{Ge}Tersoff^a$	$^{Ge}Tersoff^a$
Ge	Ge	Cl	$^{Ge}Tersoff^a$	$^{Ge}Tersoff^a$
Cl	Cl	Cl or Ge	Cl with others & Cl- $^{Cl}David^b$	Cl with others & Cl- $^{Cl}David^b$
Cl	Ge	Cl	Fitting data	$^{Ge}Tersoff^a$
Cl	Ge	Ge	Fitting data	$^{Ge}Tersoff^a$
Ge	Cl	Ge	Fitting data	Cl with others & Cl- $^{Cl}David^b$
Ge	Cl	Cl	Fitting data	Cl with others & Cl- $^{Cl}David^b$

^a Tersoff, PRB, 39.8 (1989):5566-5568

^b David Humbird and David B. Graves, JCP, 120, 2405 (2004)

* The three-body term parameters in $g(\theta)$, along with m and λ_3 , are approximated to the values in either Ge-Ge-Ge or Cl-Cl-Cl case, depending on the central atom type, since their angles are within 10° difference and b_{ij} values within 0.2 difference. Also, the number of training data is far less than the number of parameters, so some of the parameters need be fixed and this is a reasonable approximation.



Least square optimization is used to fit the different parameter sets. Training results are shown in the following.

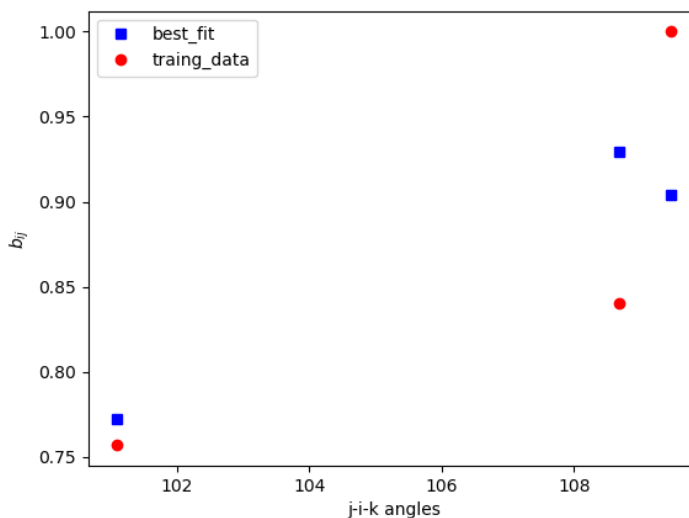


Figure S5 Training results for Cl-Ge-Cl case.

- Training results for Ge-Ge-Cl case

Target b_{ij} : 0.9507, trained b_{ij} : 0.9879

- Training results for Cl-Ge-Ge case

Target b_{ij} : 0.8188, trained b_{ij} : 0.7978

Table S5 Fitted parameters in bond order b_{ij} term.

j-i-k Parameters	Cl-Ge-Cl	Cl-Ge-Ge	Ge-Cl-Ge	Ge-Cl-Cl	Ge-Ge-Cl
m	3.0	3.0	1.0	1.0	3.0
γ_{ijk}	1.0	1.0	4.0	4.0	1.0
λ_3	0.0	0.0	6.0	6.0	0.0
c	1.0643e5	1.0643e5	0.0	0.0	1.0643e5
d	15.652	15.652	1	1	15.652
$\cos\theta_0$	-0.43884	-0.43884	1	1	-0.43884
n	2.23669	0.67543	1	1	0.75627
β	1.234e-4	1.0273e-5	1	1	9.02e-7

2.5.2 Ge surface relaxation at room temperature

The Ge surface was equilibrated under NPT condition ($T=300$ K, anisotropic zero pressure in x and y direction (the surface plane)). As shown in Fig. S6 a) and b), the original minimization

leads to a Ge lattice constant of 5.657 Å (exp: 5.652 Å,⁹ DFT: 5.668 Å). At 300 K, some surface dimers appear to form (2x1) reconstruction (Fig. S6 c) and d)), consistent to experimental observations.¹⁰ The simulated average dimer bond length is 2.53 Å (averaged over 15 dimers, with standard deviation=0.077 Å), fits well to the DFT value of 2.53 Å and experimental value between 2.42 Å¹⁰ and 2.55 Å¹¹

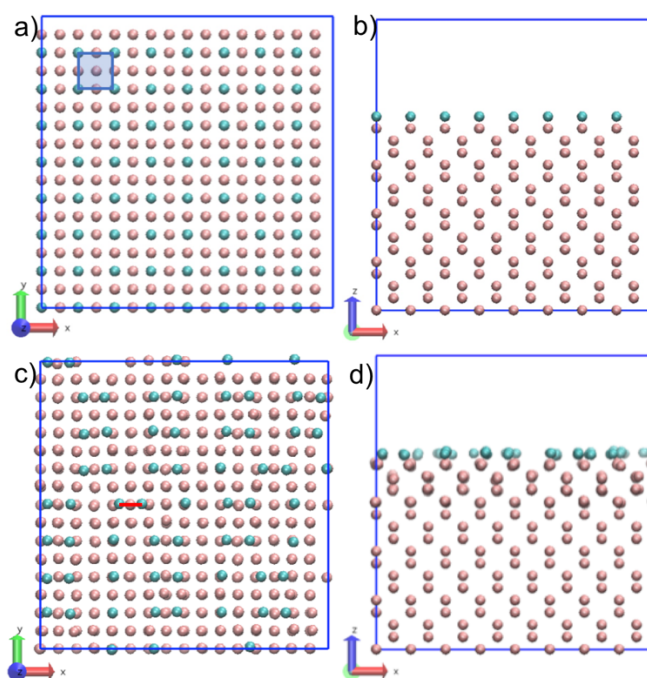


Figure S6 Ge surface minimization a) top view b) side view. Ge surface equilibration at 300 K c) top view d) side view. Green atoms represent top Ge surface layer, red is for other layer Ge. x: [110], y:[-110], z:[001].

2.5.3 Other characterizations in chlorination and bombardment process

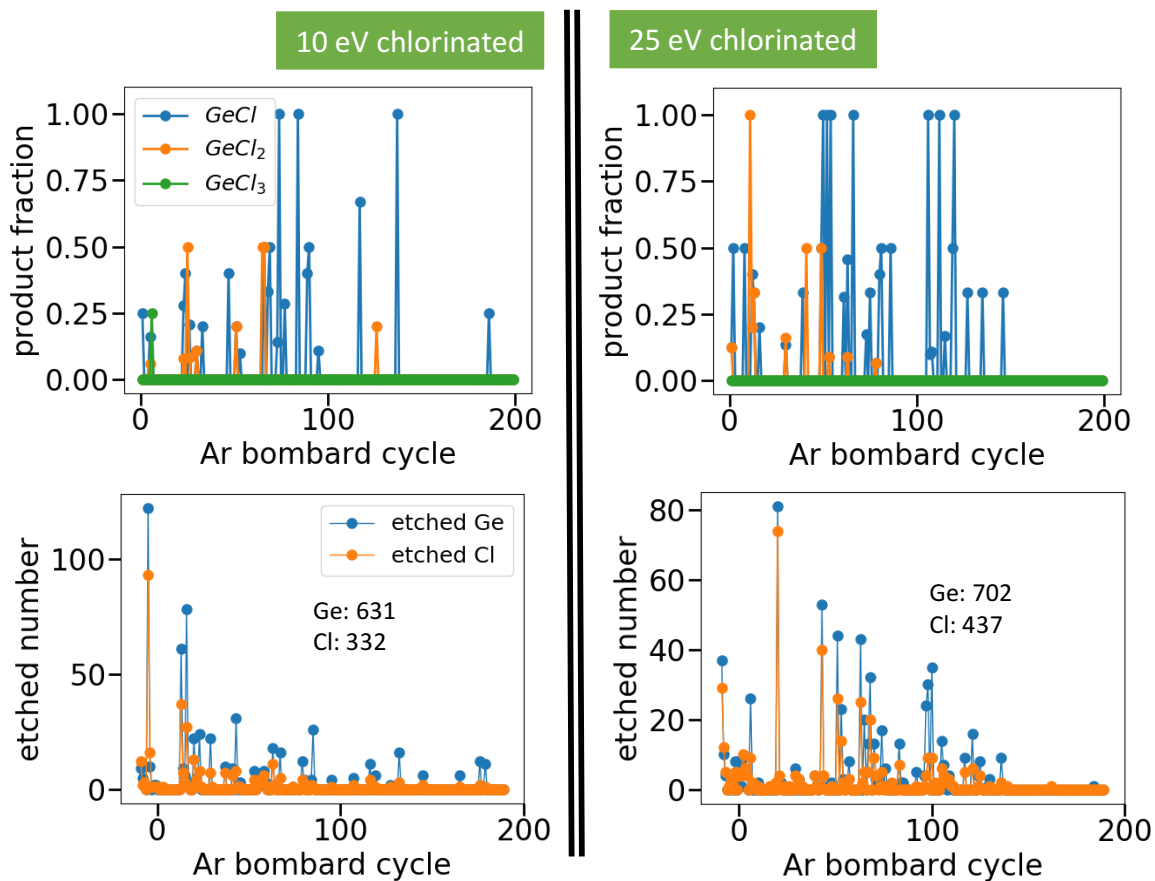


Figure S7 Etched products percentage ($GeCl$, $GeCl_2$ and $GeCl_3$) and element etched number change with Ar bombardment cycle for 10 eV and 25 eV chlorinated surface. The insets number on etched number graph shows the total etched Ge and Cl number. Sample 1 is used as an example here.

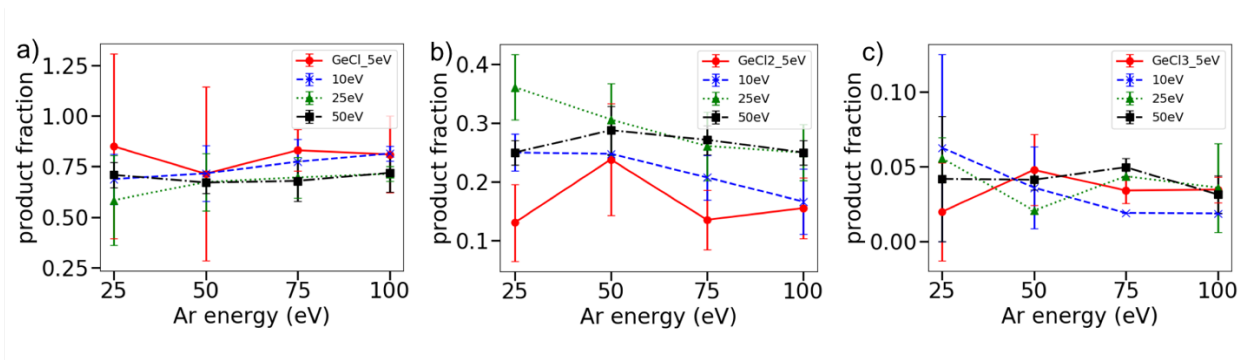


Figure S8 Product fraction (only consider GeCl_n products) of a) GeCl, b) GeCl₂ and c) GeCl₃, averaged from 3 samples with error bars added.

Table S6 Total etched product numbers after 200 Ar bombardment cycles under each condition (averaged from 3 samples with standard deviation).

5 eV chlorination	Ar 25 eV	Ar 50 eV	Ar 75 eV	Ar 100 eV
GeCl	13±7	15±9	49±6	47±11
GeCl ₂	2±1	5±2	8±3	9±3
GeCl ₃	0.3±0.5	1±0.5	2±0.5	2±0.5
10 eV chlorination				
GeCl	11±2	26±5	41±6	44±2
GeCl ₂	4±0.5	9±0	11±2	9±3
GeCl ₃	1±1	1.3±1	1±0	1±0
25 eV chlorination				
GeCl	21±8	33±7	48±7	60±3
GeCl ₂	13±2	15±3	18±4	21±4
GeCl ₃	2±0.5	1±0	3±0	3±2.5
50 eV chlorination				
GeCl	34±3	49±4	55±8	69±9
GeCl ₂	12±1	21±3	22±2	24±2
GeCl ₃	2±1.6	3±0.5	4±0.5	3±1.2

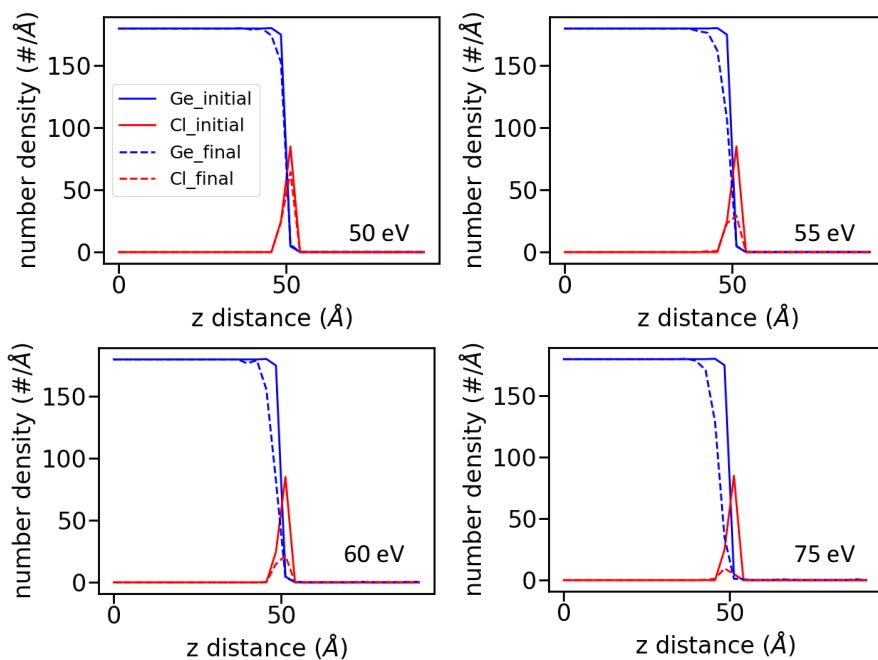


Figure S9 Density profile of Ge and Cl element evolution under different Ar bombardment energy on 5 eV chlorinated surface.

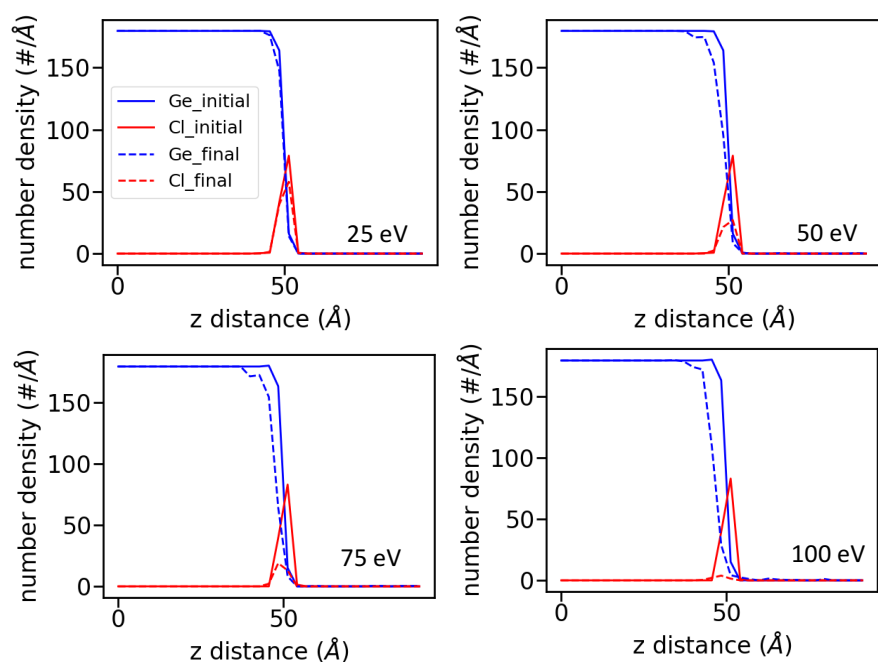


Figure S10 Density profile of Ge and Cl element evolution under different Ar bombardment energy on 10 eV chlorinated surface.

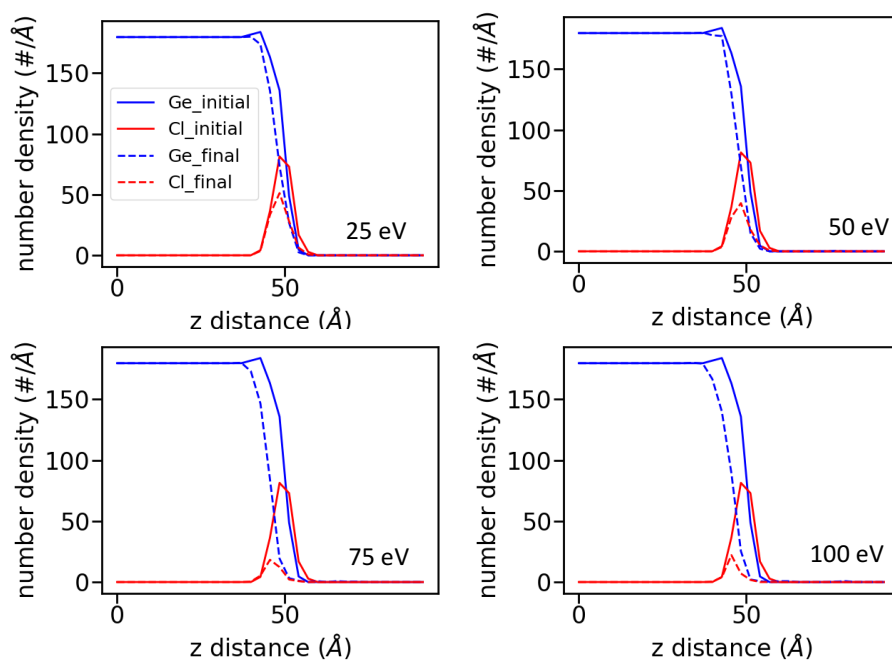


Figure S11 Density profile of Ge and Cl element evolution under different Ar bombardment energy on 50 eV chlorinated surface. Sample 1 is used.

References

1. J. Tersoff, *Physical Review B*, 1988, **37**, 6991.
2. J. Tersoff, *Physical Review B*, 1989, **39**, 5566-5568.
3. P. C. Kelires and J. Tersoff, *Physical Review Letters*, 1989, **63**, 1164-1167.
4. D. Humbird and D. B. Graves, *The Journal of Chemical Physics*, 2004, **120**, 2405-2412.
5. W. B. Donald, A. S. Olga, A. H. Judith, J. S. Steven, N. Boris and B. S. Susan, *Journal of Physics: Condensed Matter*, 2002, **14**, 783.
6. C. F. Abrams and D. B. Graves, *Journal of Applied Physics*, 1999, **86**, 5938-5948.
7. M. V. Ramana Murty and H. A. Atwater, *Physical Review B*, 1995, **51**, 4889-4893.
8. S. P. Walch, *Surface Science*, 2002, **496**, 271-286.
9. M. T. Yin and M. L. Cohen, *Physical Review B*, 1982, **26**, 5668-5687.
10. S. Pflanz, R. Buchtler, W. Moritz and H. Over, *Physical Review B*, 1996, **54**, R8313-R8316.
11. X. Torrelles, H. A. van der Vegt, V. H. Etgens, P. Fajardo, J. Alvarez and S. Ferrer, *Surface Science*, 1996, **364**, 242-252.

Chapter 3

Development of a ReaxFF Force Field for Aqueous Phosphoenolpyruvate as a Novel Biomimetic Carbon Capture Absorbent

My contribution to this work included all the computational work, analysis, and writing.

This chapter originally appeared in the literature as:

Huang Y, Wexler AS, Bein KJ, Faller R. Development of a ReaxFF Force Field for Aqueous Phosphoenolpyruvate as a Novel Biomimetic Carbon Capture Absorbent. *The Journal of Physical Chemistry C*. 2022 May 26;126 (22), 9284-9292

Abstract

Phosphoenolpyruvate (PEP) found in C₄ plants could be a novel green absorbent in biomimetic carbon capture through its Crassulacean Acid Metabolism (CAM) mechanism, and could potentially substitute the most commonly used absorbent Monoethanolamine in future post-combustion capture systems. In this study, a new ReaxFF model has been developed to describe the CAM reactions involving PEP and the atomic interactions in the P/C/O/H system. The ReaxFF force field parameters were fitted against quantum mechanical (QM) training data for partial charges, molecular structures, bond dissociation energies, reaction energies and activation energies. 2nd generation water parameters were combined with P/C parameters for more accurate water description and P's electrostatic parameters were specially treated to correct P/O interactions. The developed P/C/O/H ReaxFF model was able to reproduce the training set for structures and

energetics of the molecules and reactions involved in the CAM process, and accurately describe the aqueous bicarbonate and PEP systems. Molecular dynamics simulation using this ReaxFF model depicts how bicarbonate reacts with PEP and in solution and determines the impact of local structure on reactions necessary to perform carbon capture using PEP, which enables the potential design of PEP variant as the optimal carbon capture absorbent.

3.1 Introduction

Greenhouse gas emissions are the primary cause of climate change, which is the existential challenge of this century. Most human activities consume energy and in the current energy economy most of that energy is generated by combustion of fossil fuels, which releases significant amounts of CO₂, the major greenhouse gas, into the air. Carbon capture and storage (CCS) technologies encompass a suite of approaches that either remove CO₂ from the flue gas or from the atmosphere itself to reduce emissions or reduce atmospheric concentrations, respectively.¹⁻³ This process encompasses capturing CO₂ from gas mixtures and then liquefying it so that it can be transferred and stored underground.⁴⁻⁵ For large-scale applications, the most prevalent form of CSS is post-combustion capture (PCC) wherein CO₂ is removed from the flue gas in an absorber. Currently, this is the preferred form of CCS as it does not require additional steps prior to combustion; therefore, is straightforward to integrate into existing infrastructure.⁶⁻⁷ Monoethanolamine (MEA) is an amine compound which is a commercially available and most frequently used CO₂ solvent in PCC. It typically captures 85% to 90% of the CO₂ produced.⁴ The benchmark solvent for regenerative chemical absorption-based PCC in terms of cost, energy penalty, CO₂ capture efficiency, and physicochemical properties is a 30wt% aqueous solution of MEA.⁸⁻⁹ While typically benign, MEA as a PCC absorbent suffers from thermal and chemical

degradation and high energy production cost.¹⁰⁻¹¹ MEA can break down at high temperatures or when released into the atmosphere forming toxic vapors.¹² In addition, MEA contains nitrogen that is derived by nitrogen fixation from the atmosphere such that producing MEA is energy intensive. These shortcomings have prompted a search for alternative CO₂ solvents that are non-toxic and less energy intensive to produce and use for PCC.

Phosphoenolpyruvate (PEP) is found in C₄ plants that use Crassulacean Acid Metabolism for their photosynthesis and are adapted to arid climates.¹³ At night when it's cooler and more humid, the plants open their stomata to uptake CO₂ that immediately reacts with PEP, to store CO₂ overnight.¹⁴ In the morning when the sun rises and temperatures increase, the stomata close, and the reaction is reversed releasing CO₂ to participate in photosynthesis.¹⁵ To utilize this chemistry in PCC, CO₂ would dissolve in an aqueous solution of PEP, the carbon dioxide would dissolve in this solution forming bicarbonate, which in turn would attack the phosphate group in PEP and split PEP into a carboxyphosphate and the enolate form of pyruvate (Figure 3.1).¹⁶ Based on this concept, PEP could serve as a novel green absorbent in biomimetic CCS and could potentially serve as a template for designing similar molecules to substitute MEA in future PCC systems. In order to further examine the viability of PEP as a CCS solvent requires detailed knowledge of its thermodynamics and kinetics when reacting with bicarbonate.

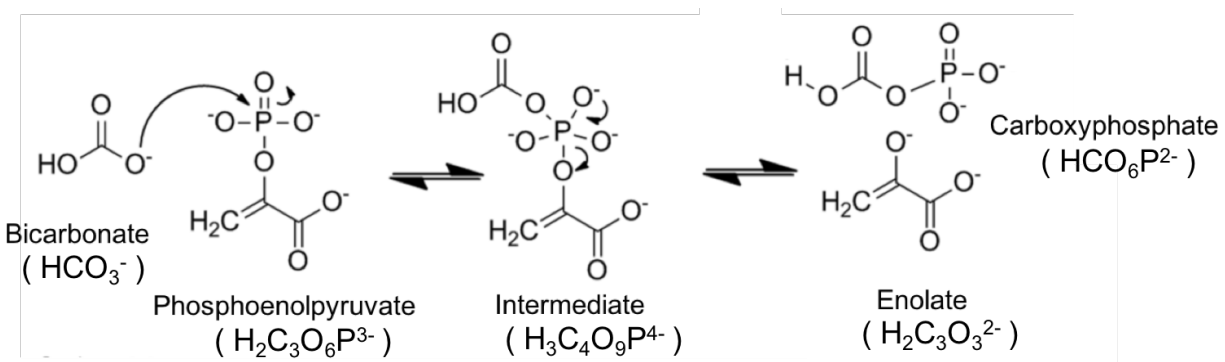


Figure 3.1: Chemistry of bicarbonate reacting with PEP in water.

Computation can be an effective and efficient tool to study the chemistry of this system, and to screen possible PEP variants. In addition, understanding the thermodynamics and kinetics of PEP may assist investigations into the biochemistry and metabolism of C4 plants. Quantum mechanics (QM) -based methods like density functional theory (DFT) ¹⁷⁻¹⁸ are powerful tools to describe chemical reactions on the atomistic scale. However, these calculations are very computationally expensive and thus limit the available time scale and length scales. On the other hand, empirical force field methods like classical molecular dynamics (MD) simulations can study the system's dynamic evolution over nano- to microseconds with tens to hundreds of thousands of atoms but loses the reaction information since the atom connectivity has to be normally predefined. In order to describe the reaction in Figure 1 and evaluate potential PEP variants, we use ReaxFF-based reactive MD that fills in the gap between quantum chemistry and classical empirical force fields. ¹⁹⁻²¹

Little prior work has examined reaction of phosphate-containing molecules. The closest example is that by Zhang and coworkers ²² who developed a ReaxFF model for carbon capture that uses ionic liquids (ILs) to capture CO₂ mixtures. The chosen IL was tetrabutylphosphonium glycinate, [P(C₄)₄]. Although it contains the atoms P, C, O, and H, it was not parametrized for the system and chemistries of our interest, and the phosphorus was bonded to four carbons instead of oxygens like in the PEP system here. Furthermore, phosphorus in [P(C₄)₄] does not participate in the reaction, so this force field does not necessarily describe P-O bond formation and breakage accurately. In addition, the H/O interaction parameters were from the 1st generation water force field from which NPT water density is known to be significantly smaller than the experimental

value, and the H_2O , H_3O^+ , and OH^- diffusion constants are in incorrect order.²³ We deal with an aqueous system so the water behavior needs to be more accurate than the 1st generation water force field. A force field that could accurately describe both the physical and chemical interactions between bicarbonate and PEP in water is needed. In this work, we develop a ReaxFF force field for PEP-based carbon capture, compare against partial charges, molecule structures, and energetics of molecules in vacuum, and then validate and predict molecular behavior in solutions.

3.2 Computational Methods

3.2.1 ReaxFF Formalism

ReaxFF is a reactive force field that allows bond formation and dissociation. The total system energy is comprised of bond order dependent energies as well as non-bonded energies.¹⁹⁻
²⁰ Unlike in classical empirical force fields, each chemical element is represented by only one atom type, and its connectivity is not predefined but derived from bond orders (BO) that are functions of interatomic distances at every MD step. As formulated in the bond order equation in Figure 3.2, the bond order between a pair of atoms consists of three exponential terms, each corresponding to the sigma bond (pbo,1 and pbo,2) which is unity below $\sim 1.5 \text{ \AA}$ but negligible above $\sim 2.5 \text{ \AA}$, a first pi bond (pbo,3 and pbo,4) which is unity below $\sim 1.2 \text{ \AA}$ but negligible above $\sim 1.75 \text{ \AA}$, and a second pi bond (pbo,5 and pbo,6) which is unity below $\sim 1.0 \text{ \AA}$ but negligible above $\sim 1.4 \text{ \AA}$.
¹⁹ So, for example, as two carbon atoms starting from very close contact separate, the bond order gradually drops from three (triple bond) to zero (fully dissociated). Like in a classical force field, the van der Waals and Coulomb interactions are also included for all atom pairs, where a shielding

term is used to avoid excessively close range non-bonded interactions, and polarization effects are calculated through a geometry dependent charge equilibrium scheme.²⁴⁻²⁵

$$E_{\text{system}} = \underbrace{E_{\text{bond}} + E_{\text{over}} + E_{\text{under}} + E_{\text{val}} + E_{\text{tors}} + E_{\text{Hbond}}}_{\text{Bonding}} + \underbrace{E_{\text{vdWaaals}} + E_{\text{Coulomb}}}_{\text{Non-bonding}}$$

(BO'_{ij} - dependent)

E_{bond} = bonded interaction	E_{tors} = torsion interaction
E_{over} = penalty for over coord.	E_{Hbond} = hydrogen bonding
E_{under} = stabilize under coord.	E_{vdWaaals} = van der Waals interaction
E_{val} = valence angle interaction	E_{Coulomb} = Coulomb interaction

$$BO'_{ij} = \exp \left[p_{bo,1} \cdot \left(\frac{r_{ij}}{r_0} \right)^{p_{bo,2}} \right] + \exp \left[p_{bo,3} \cdot \left(\frac{r_{ij}}{r_0} \right)^{p_{bo,4}} \right] + \exp \left[p_{bo,5} \cdot \left(\frac{r_{ij}}{r_0} \right)^{p_{bo,6}} \right]$$

Figure 3.2: The fundamental equations of ReaxFF formulism.

The ReaxFF force field ff.P/N/C/O/H/Na originally developed by Zhang and coworkers²² was used as the starting parameter set in our optimization. All water related parameters were replaced with those from the 2nd generation water force field²³ and kept fixed during parametrization. C and P atomic parameters and C-C, C-O and O-P interaction parameters were re-parametrized to better describe the reaction in Figure1. Table 3.1 shows which parameters are chosen based on their physicochemical importance to our system.

Parameters chosen		Reasons for chosen
Electrostatic	Electronegativity equalization method (EEM) parameters for C, P	Bonding prediction
Valence bond	All bond radii/order/dissociation energy, under/over coordination energy parameters for C, P atoms and C-C, C-O, P-O pairs	Bonding prediction
		Reaction pathway
		Reactant/product/transition state structure
Valence angle	All angle parameters of C, O, P related angles in the molecules	Geometry prediction
Van der Waals	Vdw radii/dissociation energy/shielding for C,P	Bonding prediction

Table 3.1: Choice of optimized force field parameters.

The optimization process works in two stages: training set generation and a parallel search algorithm.²⁶ New parameter values are accepted such that they reduce the error between training set features and ReaxFF fitted features. Figure 3.3 shows the workflow. A detailed description of the algorithm can be found in Deetz' work.²⁶ The scripts we are using are written in MATLAB and Bash, and the parallel search optimization part is run in parallel on a high performance cluster (HPC1/2 at UC Davis). The goal is to minimize the error function P (Figure 3.3), in which the inverse weights i for each feature in the training set are selected so that each section of the training set (e.g., charges, reactions) weighs similarly into the overall error function. Then after the error function converges, weights were readjusted according to the sectional errors or even the individual feature errors to fine-tune the parameters and focus on the ones which most deviate from the target while making sure to not negatively impact the others. This process was repeated until the overall performance of the new force field matches the training set with acceptable errors, and the most important features like the kinetics matches the training data very well.

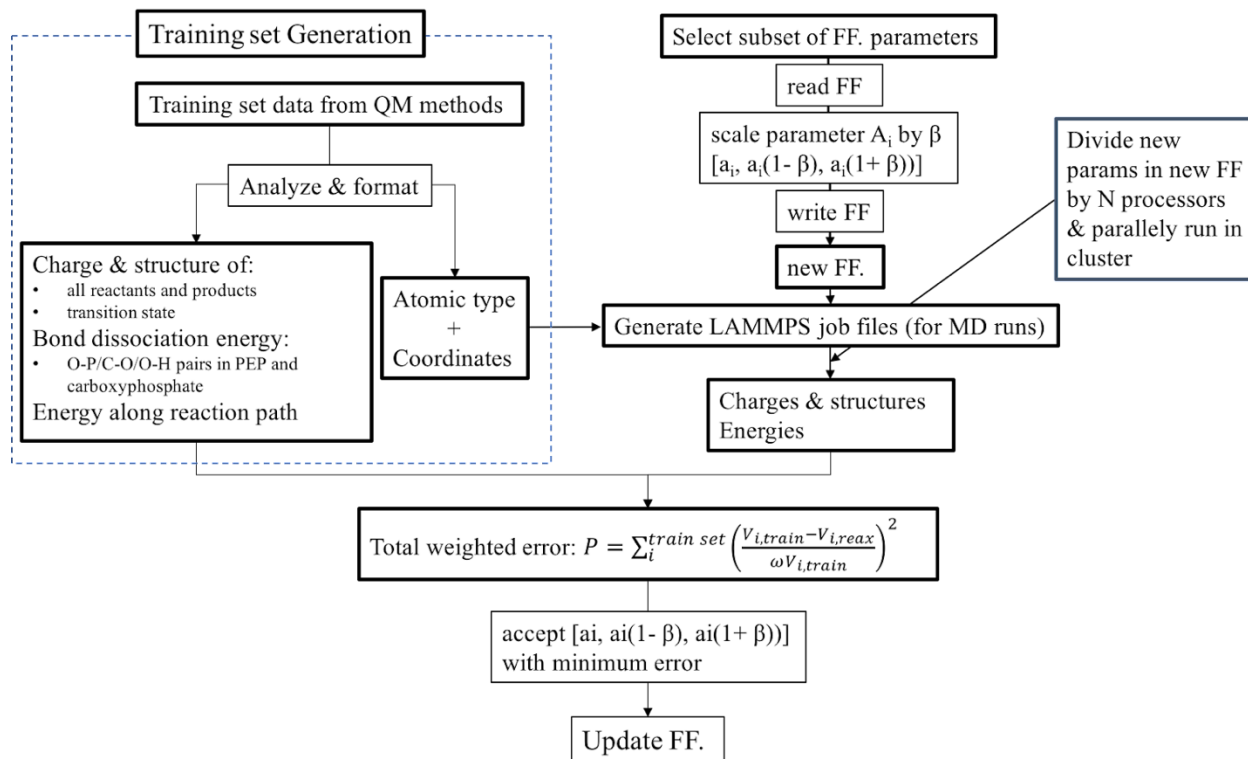


Figure 3.3: Workflow of the whole optimization process.

3.2.2 Training Set Generation

The training set has four sections – charge, geometry, energy, and reaction. DFT calculations were performed to acquire the necessary information using Gaussian 16²⁷, and the relaxed structures were fed to LAMMPS (release version 12 Dec 2018)²⁸⁻²⁹ that runs the MD simulations to calculate the error function. The charge and geometry sections collect the partial charges, bond lengths, and angles of the neutral forms of bicarbonate, PEP, carboxyphosphate, enolate, oxaloacetate, the transition state, as well as their charged states. The transition state refers to the reaction where the PO_3^{2-} group dissociates from the charged PEP^{3-} (Figure 3.4a). Since the molecules were all in vacuum for DFT calculations, and the intermediate species in Figure 1 is not stable in vacuum, it was excluded from the training set. Because LAMMPS requires a neutral

system, counter ions were introduced into the system to balance the anions. The reactions happen in aqueous water, which will self-ionize into H_3O^+ and OH^- , so to keep the nature of environment and prevent any unrelated reactions, H_3O^+ were inserted as counter ions. H_3O^+ were placed at least 4 Å away from the anions to avoid unwanted interactions. Figure 3.4b shows an example of such a neutralized system with PEP^{3-} , other systems are shown in Figure S1. The Becke, 3-parameter, Lee–Yang–Parr functional (B3LYP)³⁰⁻³¹ and the basis set 6311++g(d,p)³²⁻³³ were used in DFT calculations to simulate the geometrically optimized structures,. In order to be consistent with charge equilibrium for ReaxFF,¹⁹ a Mulliken charge calculation population analysis³⁴ with B3LYP/631g(d,p)³⁵ was used to calculate partial charges. The bond section collects the bond dissociation energy scans for P-O and O-H bonds in neutral PEP and carboxyphosphate. The reaction section has the reaction energy of the bicarbonate (HCO_3^-) + PEP ($\text{H}_2\text{C}_3\text{O}_6\text{P}^{3-}$) carboxyphosphate ($\text{HCO}_6\text{P}^{2-}$) + enolate form of pyruvate ($\text{H}_2\text{C}_3\text{O}_3^{2-}$) reaction and the activation energy of the splitting of PEP^{3-} into PO_3^{2-} and the enolate form of pyruvate. The DFT activation energy and the energies along the reaction pathway of the splitting of PEP^{3-} were obtained from intrinsic reaction coordinate (IRC)³⁶⁻³⁷ scans. H_3O^+ were again added to neutralize, and B3LYP/6311++g(d,p) was used to calculate energies.

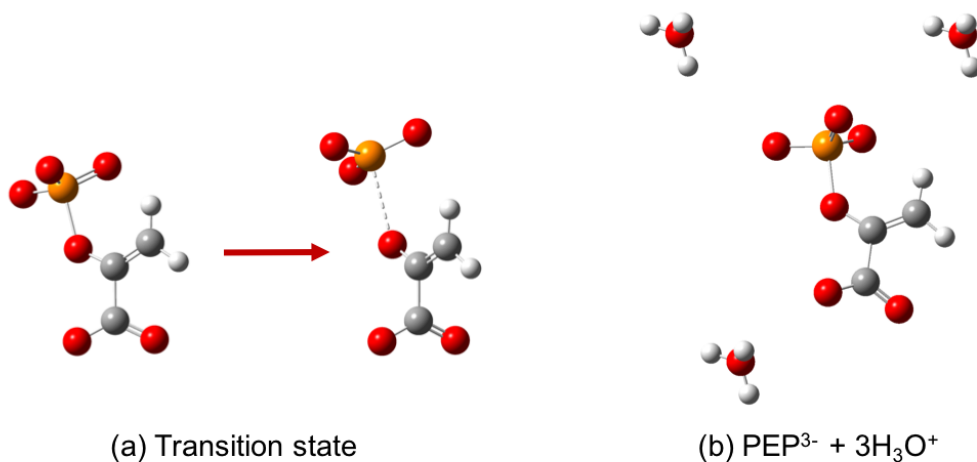


Figure 3.4: (a) transition state of the dissociation of PEP^{3-} , and (b) the neutralized system of charged PEP^{3-} .

3.2.3 Charge Parameter Correction

An intrinsic problem with the force field before re-parametrization was that the atomic hardness parameter (etaEEM in the force field notation) of phosphorous made it too “soft”, so that when the intermediate species formed, P donated too many electrons to the double bonded O, leading to extreme partial charges $+8.6e$ and $-7.2e$, respectively, after a few steps in the MD runs. The theoretical partial charges of the $\text{P}=\text{O}$ pair were $+1.2e$ and $-0.7e$ respectively. This created a huge unphysical local dipole that eventually led to simulation instability and stopped the reaction. This only occurred during the dynamic runs when the reacting PEPs are in solution, so the parametrization process that trained the parameters against static DFT values in vacuum could not address this problem, and a special treatment on the etaEEM parameter of P was needed. The etaEEM parameter of P after standard parametrization was 7.2562, and so a snapshot was taken just before the huge dipole formed and etaEEM of P was set to be 5, 6, 7, 8, and 9 and partial

charges were calculated. It was found that when etaEEM of P was 8, the partial charges of the P=O pair were +0.85e and -0.65e respectively, which were closest to the theoretical values, and the intermediate species did not break. Finally, the force field was fine-tuned with etaEEM of P starting at 8 and was only allowed to take very small scaling steps during the fine-tuning. The final etaEEM of P was 7.9960 and the partial charges of the P=O pair were +1.05e and -0.70e, respectively, and these values kept the simulation stable.

3.2.4 MD Simulation

All MD simulations were performed in LAMMPS. Because the benchmark solvent for chemical absorption-based PCC is 30wt% MEA, 24 PEPs were randomly distributed in 520 waters to create a 30wt% PEP solution. PEPs were kept in their neutral form during equilibration. Counter ions were used to ensure charge neutrality. Similarly, bicarbonates initially were in their neutral form (carbonic acid). Pressure fluctuations converted some carbonic acids to CO₂, so in order to have enough bicarbonates to react with PEP, 36 bicarbonates were randomly distributed. Therefore, the initial box contained 520 H₂O, 24 neutral PEP, 36 carbonic acid, 108 H₃O⁺ and 108 OH⁻, and was first equilibrated in the canonical (NVT) ensemble for 20 ps at 300 K, and then equilibrated in the isothermal-isobaric (NPT) ensemble for another 25 ps at 300 K and 1 atm. A short equilibration time ensured less reactions during equilibration. Temperature was maintained using a Nosé-Hoover thermostat³⁸ with a damping parameter of 25 fs, and pressure was maintained using a Nosé-Hoover barostat with a damping parameter of 250 fs. Then 18 H⁺ were deleted from 6 PEP to create 6 PEP³⁻, 9 H⁺ were deleted from 9 carbonic acid to create 9 bicarbonate, together with 27 OH⁻, and the system was equilibrated in NPT for 10 ps to eliminate potential unphysical dipoles created by the deletions. This deletion and equilibration process was repeated 4 times.

Then the resulting system was simulated as reaction stage for 2 ns in NPT at 300 K and 1 atm. Time steps of 0.25 fs were used for all the NVT and NPT runs.

3.3 Results and Discussions

3.3.1 Force Field Parametrization

The atomic partial charge errors between ReaxFF simulations and QM calculations are negligible, and the bond and angle errors are all less than 6% and 8%, respectively (Figure 3.5). Bond dissociation energy scans in Figure 3.6 (a-c) show good agreement. O-P, C-O, and O-H bonds at the active sites for both neutral PEP and neutral carboxyphosphate were scanned from very short to equilibrium distances and then to very large distances without relaxing the whole structure, and the energies along the scans were acquired. To account for the multiple spin states as the molecules break, both singlet and triplet scans were calculated, and the lower energies were taken. Some energy data points at far distances were removed to reduce computational cost. The structures corresponding to each DFT data point along the scans were then fed to ReaxFF MD simulations to calculate ReaxFF energies. Figure S2 shows additional bond dissociation energy scans. The fitted ReaxFF energies generally match the DFT energies, especially at the regions near the equilibrium distances, and the performance of the force field improved substantially comparing to the original one.

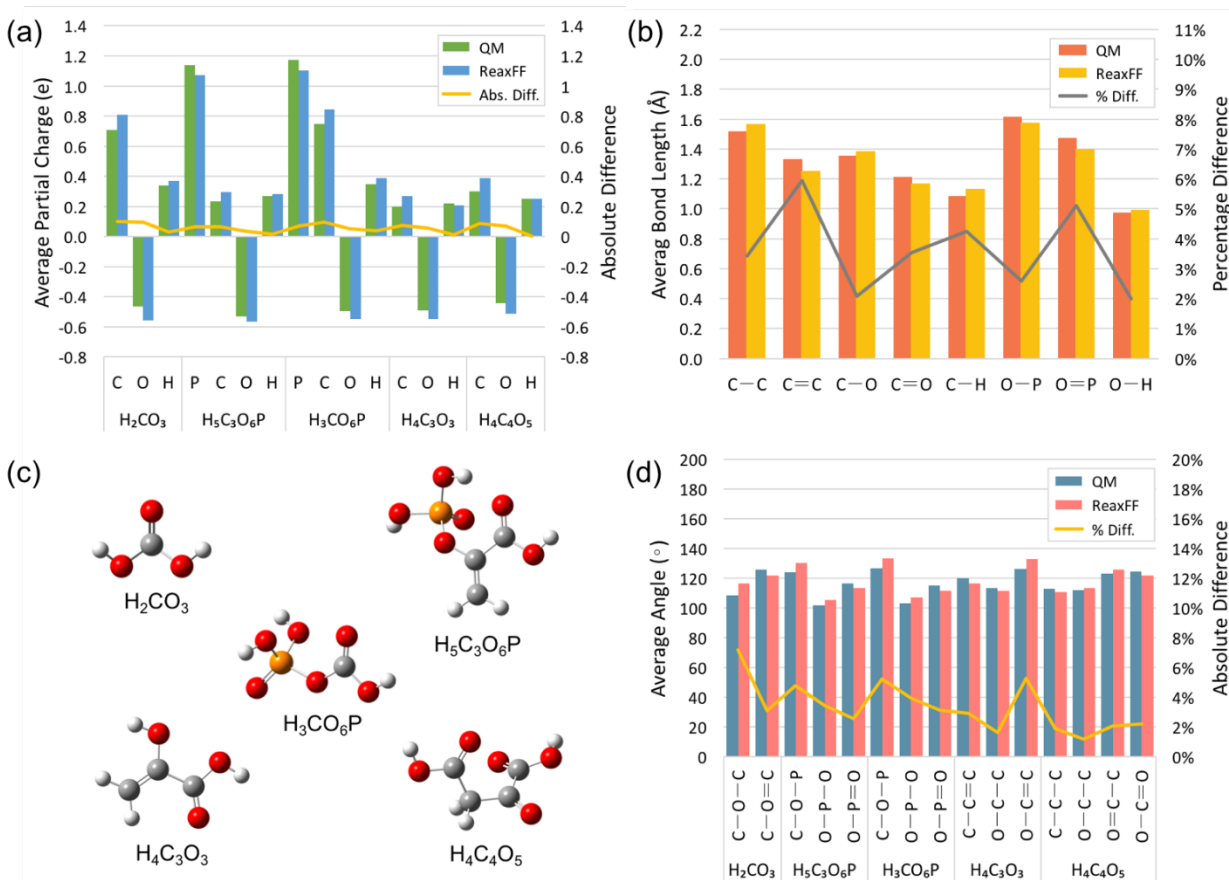


Figure 3.5: Charge and structure validation of the neutral forms of carbonic acid (H_2CO_3), PEP ($H_5C_3O_6P$), carboxyphosphate (H_3CO_6P), enolate form of pyruvate ($H_4C_3O_3$), and oxaloacetate ($H_4C_4O_5$). (a): average partial charge comparison between ReaxFF and QM for all atoms; (b): average bond length comparison between ReaxFF and QM for all type of bonds; (d): average angle comparison between ReaxFF and QM for bonds around active sites in each molecule; (c): representations of the neutral forms of the molecules. Tables with direct number comparisons are in Supplementary Information.

Figure 3.6 (d) shows the comparison of the energies along the reaction pathway and the reaction energy between DFT (B3LYP/6311++g(d,p)) and ReaxFF. The energies along the reaction pathway were obtained from intrinsic reaction coordinate (IRC) calculation using DFT, and the ReaxFF energies were fitted using for the structures along the DFT IRC curve. The barrier

energy from DFT and ReaxFF are 8.87 kcal/mol and 7.08 kcal/mol, respectively, and the reaction energy from DFT and ReaxFF are -16.90 kcal/mol and -12.84 kcal/mol, respectively. Both the barrier energy and the reaction energy prove that the optimized force field can well describe the kinetics and thermodynamics of the reaction.

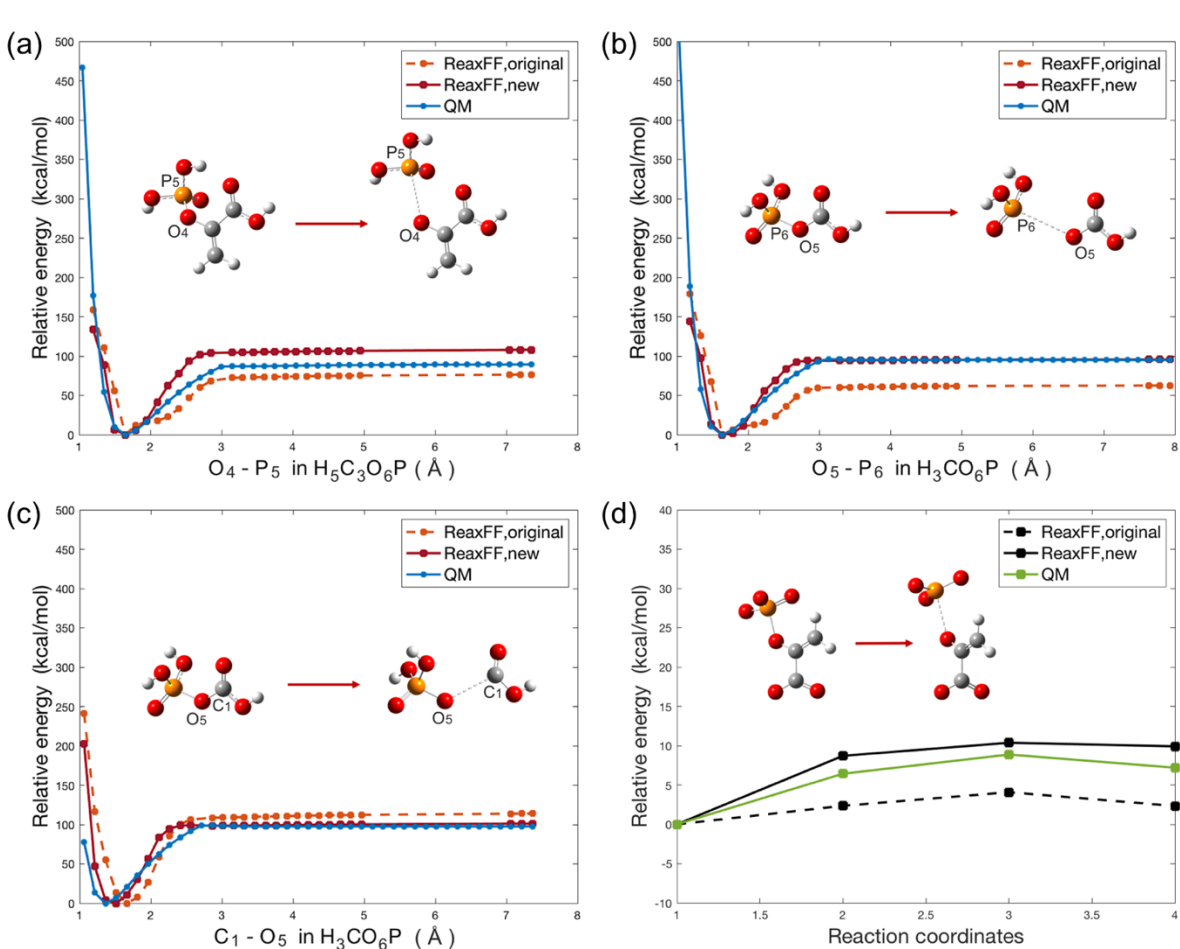


Figure 3.6: (a-c): bond dissociation energy scans comparison between DFT and ReaxFF. Blue represents bond dissociation energy scans from DFT calculations, orange represents the ReaxFF force field before re-parametrization, and red represents ReaxFF after the re-parametrization; (d): intrinsic reaction coordinate (IRC) from DFTvs ReaxFF with barrier and reaction energies. Green represents the IRC from DFT, and black represents ReaxFF fitted values, among which the dash line represents the force field before re-parametrization, and the solid line represents the new ReaxFF after re-parametrization.

3.3.2 Solution Validation

Diffusion coefficients and radial distribution functions (RDFs) were also calculated (Figure 3.7) and compared with literature values³⁹ using non-reactive MD simulations to characterize HCO_3^- , and PEP^{3-} in water. Diffusion coefficients were obtained by fitting to the mean square displacement (MSD) over the time interval where MSD increased linearly. The diffusion coefficient of HCO_3^- at 298 K in Zeebe's work was $1.17 \times 10^{-9} \text{m}^2 \text{s}^{-1}$, and the diffusion coefficient of HCO_3^- at 298 K in our ReaxFF fitted system was $0.67 \times 10^{-9} \text{m}^2 \text{s}^{-1}$. Because both calculations were performed on one HCO_3^- in water, a statistical error was expected, and the discrepancy between the two diffusion coefficients were acceptable particularly as our model was not optimized for dynamics.

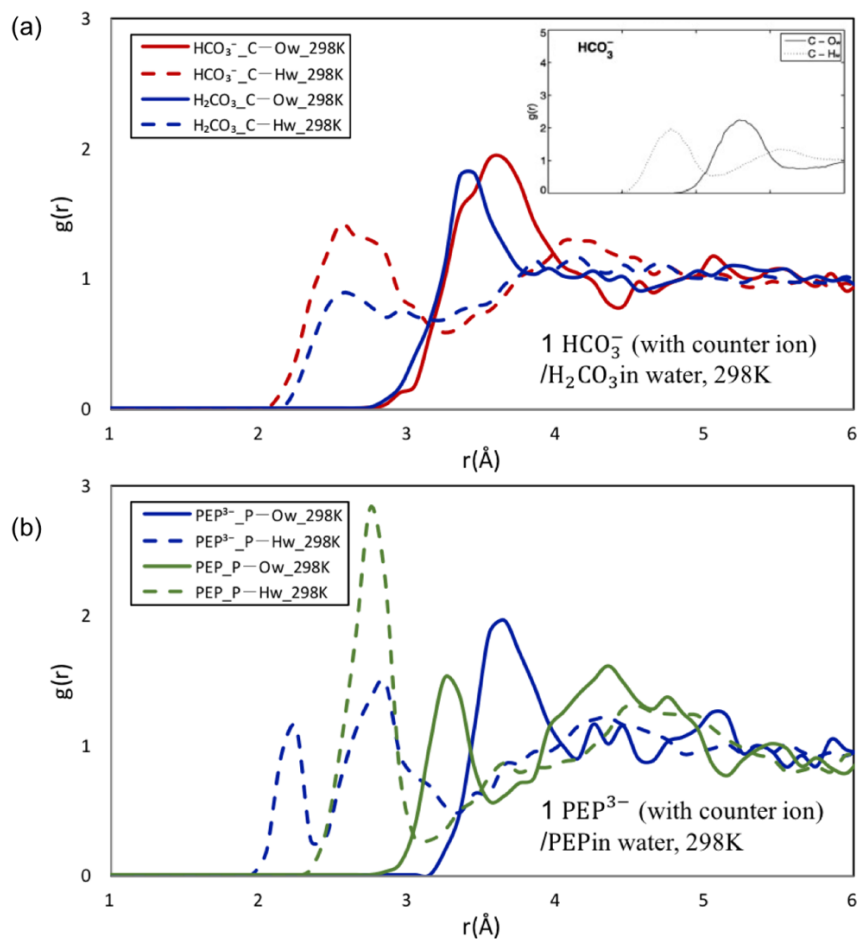


Figure 3.7: RDFs of 1 target molecule in water at 298 K. (a): RDF of HCO_3^- and H_2CO_3 ; (b): RDF of PEP^{3-} and PEP. The inset was reprinted from *Geochimica et Cosmochimica Acta*, volume 75, issue 9, Richard E. Zeebe et al., On the molecular diffusion coefficients of dissolved CO_2 , HCO_3^- , and CO_3^{2-} and their dependence on isotopic mass, Pages No. 16, Copyright (2011), with permission from Elsevier.

The RDF of HCO_3^- was calculated both in Zeebe's work and in our system. The first peak of the RDF between the carbon in HCO_3^- and the oxygen in H_2O was around 3.65 \AA in the ReaxFF system whereas between carbon in HCO_3^- and hydrogen in H_2O around 2.57 \AA in the ReaxFF fitted system. Both matched with Zeebe quantitatively. In addition, the RDFs of H_2CO_3 , PEP^{3-} and

PEP were also calculated. Diffusion coefficients and RDFs compared between the non-reactive MD simulations in Zeebe's work and our ReaxFF model show that the model not only correctly describes vacuum properties but also solution structure and dynamics.

3.3.3 Analysis and Prediction

After optimization and initial validation (see above) long simulations were performed for 30wt% PEP in water with bicarbonate using the newly parametrized ReaxFF model ff.P/C/O/H (see Supplementary Information). Depending on the initial configuration, the reaction sometimes proceeds very easily and sometimes was not observed after the 2 ns runs showing that local structure is important for reaction progress. In the runs where the reactions happened, the species in Figure 3.1 were observed. For example, in one case (Figure 3.8), two intermediate species were formed, weakening the two P-O bonds where the oxygens connect to the carbons. Then one of the intermediate species dissociates into an enolate form of pyruvate and a carboxyphosphate, which then diffuse away from each other, making both the enolate form of pyruvate and the carboxyphosphate stable. These reactions match with the expected reactions (Figure 3.1), showing that the developed ReaxFF model predict the correct reactions without need for biasing.

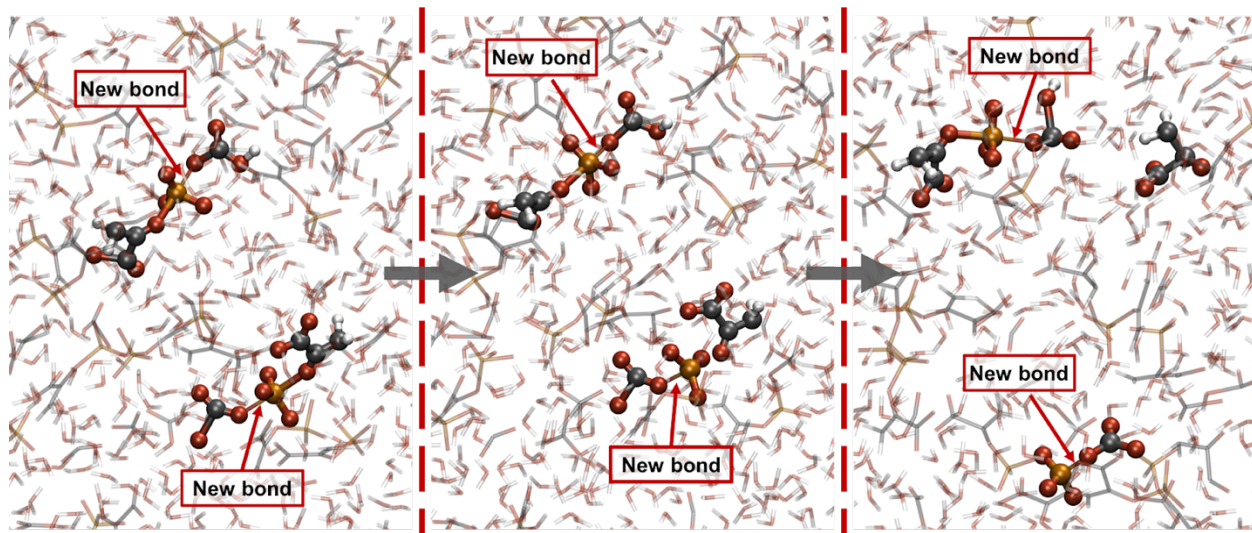


Figure 3.8: snapshots of three different stages of the reactions in one of the runs. The thinner P-O bond indicates the newly formed P-O bond.

The correlation between the partial charges and the reaction was also analyzed. For the system where two PEP-bicarbonate pairs reacted, the partial charge and the coordination number of P in the reacting PEPs were recorded over 25 ps (Figure 3.9 a-b). The coordination number of P in stable PEP was 4, but when the PEP reacted with a bicarbonate to form the intermediate, the P connected with an oxygen from the bicarbonate raising its coordination number to 5. It was observed that the partial charge of the P in the reacting PEP was negatively correlated with the coordination number. The P in the reacting PEP has lower partial charge when it bonded to the O in the bicarbonate.

In another run where only one PEP-bicarbonate pairs reacted, the partial charge and coordinate number of that P were recorded, and those of the P in 10 randomly chosen non-reacting PEPs were also recorded for comparison (Figure 3.9 c-d). We observe that the P in non-reacting PEPs generally had higher partial charges than in the reacting PEP, and in the non-reacting PEPs, the partial charge of the P did not have any correlation with the coordination number of the Ps. Therefore, if the PEP reacts with its neighboring bicarbonate, the partial charge of its P tends to be lowered and is negatively correlated with its coordination number.

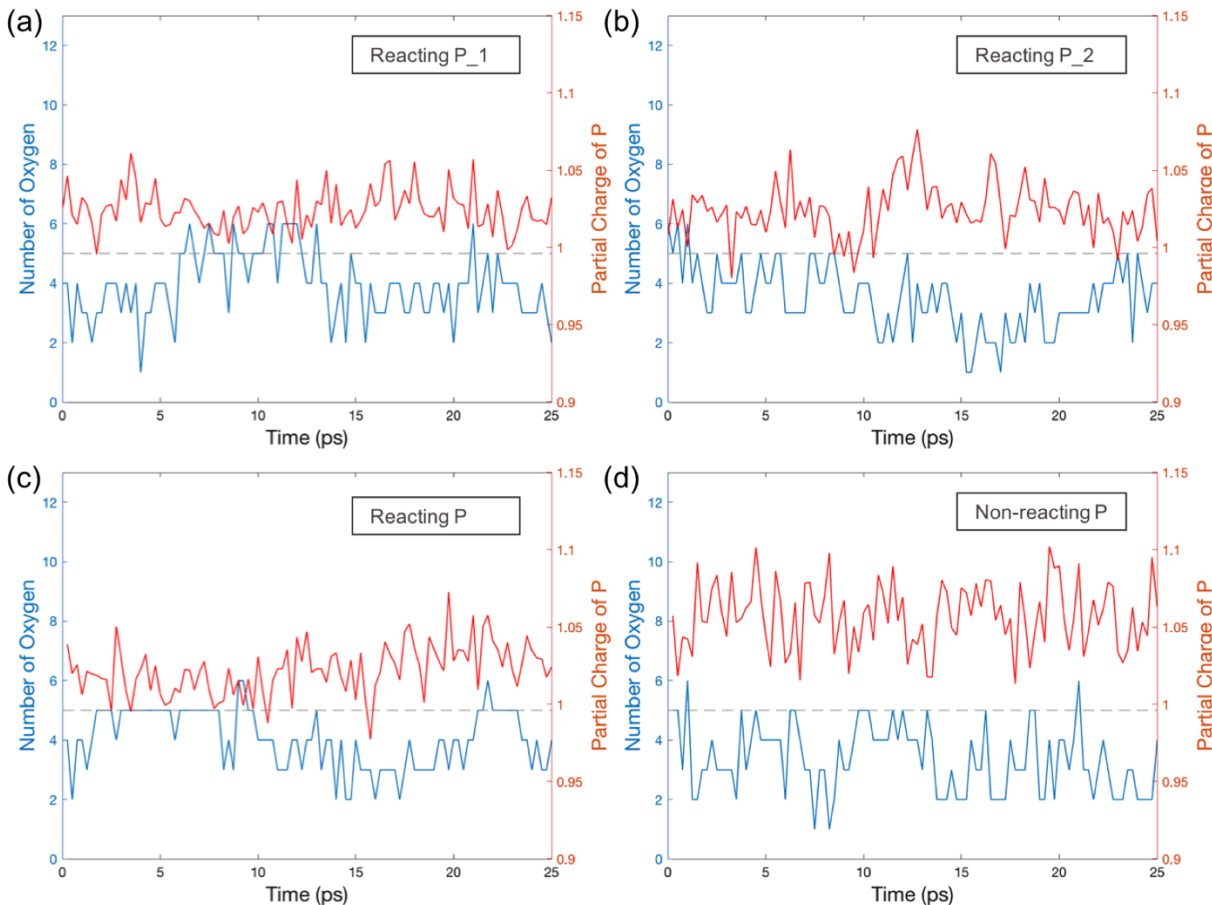


Figure 3.9: partial charges (red) and coordination number of P (blue) in reacting PEPs where there were two reacting PEP-bicarbonate pairs (a-b) and in both reacting and non-reacting PEPs where there were only one reacting PEP-bicarbonate pairs (c-d). The dash line corresponds to coordination number equals 5.

3.4 Conclusions

We developed a ReaxFF model ff.P/C/O/H for biomimetic carbon capture potentially using PEP as a substitute for MEA. During the force field parametrization, we found out that the phosphorous' EEM parameters have some intrinsic problems that needed to be treated. The final ReaxFF model was able to correctly predict vacuum properties of the molecules and reaction related to our target carbon capture chemistries, as well as solution properties. The developed

model also correctly predicted the target reactions during MD simulations, where bicarbonates react with PEPs in solution to form the intermediate species and then dissociate into the carboxyphosphate and enolate form of pyruvate. We also observed that if the PEP reacts with its neighboring bicarbonate, the partial charge of its P tends to be lowered and is negatively correlated with its coordination number. Thus, we have developed a model which can determine the impact of local structure on reactions necessary to perform carbon capture using PEP. This will enable selection of optimal reaction conditions and design of new PEP variants that are more reactive with bicarbonate, more stable after reaction, and possibly more energy-efficient.

Acknowledgement

We thank Francois Gygi for many fruitful discussions. Parts of this work was supported by the National Science Foundation grant CBET 1911267. Additional funding from the California Energy Commission (EPC-2015-072) and the Electric Power Research Institute (EPRI) is acknowledged.

References

1. Bui, M.; Adjiman, C. S.; Bardow, A.; Anthony, E. J.; Boston, A.; Brown, S.; Fennell, P. S.; Fuss, S.; Galindo, A.; Hackett, L. A., Carbon Capture and Storage (CCS): The Way Forward. *Energy Environ. Sci.* 2018, 11, 1062-1176.
2. Raza, A.; Gholami, R.; Rezaee, R.; Rasouli, V.; Rabiei, M., Significant Aspects of Carbon Capture and Storage—a Review. *Petroleum* 2019, 5, 335-340.
3. Tcvetkov, P.; Cherepovitsyn, A.; Fedoseev, S., Public Perception of Carbon Capture and Storage: A State-of-the-Art Overview. *Heliyon* 2019, 5, e02845.
4. Wilberforce, T.; Baroutaji, A.; Soudan, B.; Al-Alami, A. H.; Olabi, A. G., Outlook of Carbon Capture Technology and Challenges. *Sci. Total Environ.* 2019, 657, 56-72.
5. Kelemen, P.; Benson, S. M.; Pilorgé, H.; Psarras, P.; Wilcox, J., An Overview of the Status and Challenges of CO₂ Storage in Minerals and Geological Formations. *Frontiers in Climate* 2019, 1, 9.
6. Koperna Jr, G. J.; Gupta, N.; Godec, M.; Tucker, O.; Riestenberg, D.; Cumming, L., Society of Petroleum Engineers Grand Challenge: Carbon Capture and Sequestration. 2016. https://www.spe.org/en/industry/carbon-capture-sequestration2016/?_ga=2.127508813.1521841451.1651901470-1675539005.1651901470
7. Council, N. P., Meeting the Dual Challenge—a Roadmap to at-Scale Deployment of Carbon Capture, Use, and Storage. National Petroleum Council Washington, DC: 2019.
8. Lepaumier, H.; da Silva, E. F.; Einbu, A.; Grimstvedt, A.; Knudsen, J. N.; Zahlsen, K.; Svendsen, H. F., Comparison of MEA Degradation in Pilot-Scale with Lab-Scale Experiments. *Energy Procedia* 2011, 4, 1652-1659.
9. Feron, P. H.; Cousins, A.; Jiang, K.; Zhai, R.; Garcia, M., An Update of the Benchmark Post-Combustion CO₂-Capture Technology. *Fuel* 2020, 273, 117776.
10. Supap, T.; Idem, R.; Tontiwachwuthikul, P.; Saiwan, C., Kinetics of Sulfur Dioxide-and Oxygen-Induced Degradation of Aqueous Monoethanolamine Solution During CO₂ Absorption from Power Plant Flue Gas Streams. *Int. J. Greenh. Gas Control* 2009, 3, 133-142.
11. Chi, S.; Rochelle, G. T., Oxidative Degradation of Monoethanolamine. *Industrial & engineering chemistry research* 2002, 41, 4178-4186.
12. Lee, D.; Wexler, A. S., Atmospheric Amines – Part III: Photochemistry and Toxicity. *Atmos. Environ.* 2013, 71, 95-103.
13. Sage, R. F.; Sage, T. L.; Kocacinar, F., Photorespiration and the Evolution of C₄ Photosynthesis. *Annu. Rev. Plant Biol.* 2012, 63, 19-47.
14. Sutton, B., The Path of Carbon in CAM Plants at Night. *Funct. Plant Biol.* 1975, 2, 377-387.
15. Nimmo, H. G., The Regulation of Phosphoenolpyruvate Carboxylase in CAM Plants. *Trends Plant Sci.* 2000, 5, 75-80.
16. Vidal, J.; Chollet, R., Regulatory Phosphorylation of C₄ Pep Carboxylase. *Trends Plant Sci.* 1997, 2, 230-237.
17. Parr, R. G.; Yang, W., Density-Functional Theory of the Electronic Structure of Molecules. *Annu. Rev. Phys. Chem.* 1995, 46, 701-728.

18. Parr, R. G.; Yang, W., Density Functional Approach to the Frontier-Electron Theory of Chemical Reactivity. *J. Am. Chem. Soc.* 1984, 106, 4049-4050.
19. Van Duin, A. C.; Dasgupta, S.; Lorant, F.; Goddard, W. A., ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* 2001, 105, 9396-9409.
20. Chenoweth, K.; Van Duin, A. C.; Goddard, W. A., ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation. *J. Phys. Chem. A* 2008, 112, 1040-1053.
21. Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M., The ReaxFF Reactive Force-Field: Development, Applications and Future Directions. *Npj Comput. Mater.* 2016, 2, 1-14.
22. Zhang, B.; van Duin, A. C.; Johnson, J. K., Development of a ReaxFF Reactive Force Field for Tetrabutylphosphonium Glycinate/CO₂ Mixtures. *J. Phys. Chem. B* 2014, 118, 12008-12016.
23. Zhang, W.; Van Duin, A. C., Second-Generation ReaxFF Water Force Field: Improvements in the Description of Water Density and OH-Anion Diffusion. *J. Phys. Chem. B* 2017, 121, 6021-6032.
24. Mortier, W. J.; Ghosh, S. K.; Shankar, S., Electronegativity-Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.* 1986, 108, 4315-4320.
25. Rappe, A. K.; Goddard III, W. A., Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* 1991, 95, 3358-3363.
26. Deetz, J. D.; Faller, R., Parallel Optimization of a Reactive Force Field for Polycondensation of Alkoxysilanes. *J. Phys. Chem. B* 2014, 118, 10966-10978.
27. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16. Gaussian, Inc. Wallingford, CT: 2016.
28. Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D., LAMMPS-A Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* 2022, 271, 108171.
29. Aktulga, H. M.; Fogarty, J. C.; Pandit, S. A.; Grama, A. Y., Parallel Reactive Molecular Dynamics: Numerical Methods and Algorithmic Techniques. *Parallel Comput.* 2012, 38, 245-259.
30. Becke, A. D., Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* 1988, 38, 3098.
31. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* 1988, 37, 785.
32. McLean, A.; Chandler, G., Contracted Gaussian Basis Sets for Molecular Calculations. I. Second Row Atoms, Z= 11–18. *J. Chem. Phys.* 1980, 72, 5639-5648.
33. Krishnan, R.; Binkley, J.; Seeger, R.; Pople, J.A., Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions *J. Chem. Phys.* 1980, 72, 650-654.
34. Mulliken, R. S., Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *J. Chem. Phys.* 1955, 23, 1833-1840.

35. Petersson, G.; Al-Laham, M. A., A Complete Basis Set Model Chemistry. II. Open-Shell Systems and the Total Energies of the First- Row Atoms. *J. Chem. Phys.* 1991, 94, 6081-6090.
36. Deng, L.; Ziegler, T.; Fan, L., A Combined Density Functional and Intrinsic Reaction Coordinate Study on the Ground State Energy Surface of H₂CO. *J. Chem. Phys.* 1993, 99, 3823-3835.
37. Deng, L.; Ziegler, T., The Determination of Intrinsic Reaction Coordinates by Density Functional Theory. *Int. J. Quantum Chem.* 1994, 52, 731-765.
38. Nosé, S., A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* 1984, 81, 511-519.
39. Zeebe, R. E., On the Molecular Diffusion Coefficients of Dissolved CO₂, HCO₃⁻, and CO₃²⁻ and Their Dependence on Isotopic Mass. *Geochim. Cosmochim. Acta* 2011, 75, 2483-2498.

3.5 Supporting Information 1: Training Set and Additional Training Results

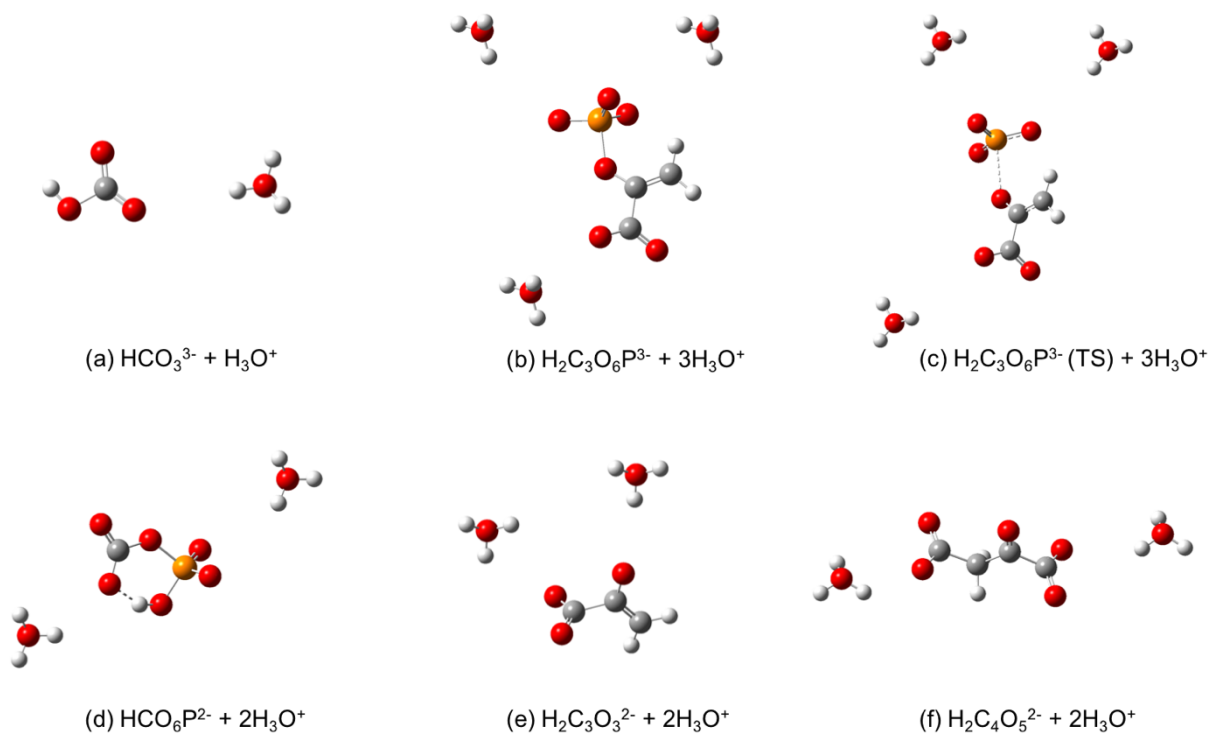


Figure S1. The neutralized systems of all the charged molecules used in the training set.

Species	Atom	Ave. Partial Charge (e)		
		QM	ReaxFF	Abs. Diff.
H ₂ CO ₃	C	0.7082	0.8076	0.0994
	O	-0.4620	-0.5578	0.0958
	H	0.3390	0.3682	0.0293
H ₅ C ₃ O ₆ P	P	1.1386	1.0716	0.0670
	C	0.2332	0.2975	0.0642
	O	-0.5293	-0.5634	0.0341
	H	0.2675	0.2832	0.0157
H ₃ CO ₆ P	P	1.1732	1.1049	0.0683
	C	0.7467	0.8428	0.0961
	O	-0.4950	-0.5468	0.0518
	H	0.3500	0.3876	0.0376
H ₄ C ₃ O ₃	C	0.1974	0.2690	0.0716
	O	-0.4909	-0.5461	0.0552
	H	0.2202	0.2079	0.0123
H ₄ C ₄ O ₅	C	0.3018	0.3873	0.0855
	O	-0.4436	-0.5110	0.0674
	H	0.2528	0.2515	0.0013

Table S1. Averaged partial charges calculated from the parametrized ReaxFF model against quantum mechanical (QM) method (ie. DFT).

Species	Bond	Ave. Bond Length (Å)		
		QM	ReaxFF	% Diff.
H ₂ CO ₃	C - O	1.34	1.42	5.97%
	C = O	1.21	1.18	2.48%
	O - H	0.97	0.98	1.03%
H ₅ C ₃ O ₆ P	C - C	1.49	1.58	5.97%
	C = C	1.33	1.25	5.94%
	C - O	1.36	1.35	0.74%
	C = O	1.22	1.17	4.10%
	C - H	1.08	1.13	4.63%
	O - P	1.61	1.60	1.03%
	O = P	1.47	1.39	5.44%
	O - H	0.97	0.99	2.06%
H ₃ CO ₆ P	C - O	1.35	1.43	5.93%
	C = O	1.20	1.17	2.50%
	O - P	1.61	1.54	4.14%
	O = P	1.47	1.40	4.76%
	O - H	0.97	0.98	1.72%
H ₄ C ₃ O ₃	C - C	1.50	1.59	6.00%
	C = C	1.33	1.25	5.94%
	C - O	1.36	1.35	0.37%
	C = O	1.21	1.16	4.13%
	C - H	1.08	1.13	4.17%
	O - H	0.97	0.99	2.59%
H ₄ C ₄ O ₅	C - C	1.52	1.55	1.75%
	C - O	1.34	1.37	1.87%
	C = O	1.21	1.16	3.87%
	C - H	1.09	1.13	3.67%
	O - H	0.97	1.00	3.09%

Table S2. Averaged bond lengths calculated from the parametrized ReaxFF model against quantum mechanical (QM) method (ie. DFT).

Species	Angle	Ave. Angle (°)		
		QM	ReaxFF	% Diff.
H ₂ CO ₃	C-O-C	108.68	116.50	7.20%
	C-O=C	125.66	121.75	3.11%
H ₅ C ₃ O ₆ P	C-O-P	124.11	130.06	4.79%
	O-P-O	101.83	105.36	3.47%
	O-P=O	116.31	113.31	2.58%
	O-C-C	114.38	114.32	0.05%
	O=C-C	123.91	124.09	0.15%
	O-C=C	121.17	125.98	3.97%
	C-C=C	123.60	118.63	4.02%
H ₃ CO ₆ P	C-O-P	126.62	133.24	5.23%
	O-P-O	103.27	107.34	3.94%
	O-P=O	115.11	111.52	3.12%
	O-C-O	106.89	114.52	7.14%
	O-C=O	126.56	122.74	3.02%
H ₄ C ₃ O ₃	C-C=C	119.93	116.42	2.93%
	O-C-C	113.26	111.40	1.64%
	O-C=C	126.27	132.93	5.27%
	O=C-C	124.02	124.22	0.16%
	O-C=O	123.27	123.63	0.29%
H ₄ C ₄ O ₅	C-C-C	113.09	110.92	1.92%
	O-C-C	112.00	113.33	1.18%
	O=C-C	123.05	125.59	2.07%
	O-C=O	124.43	121.69	2.20%

Table S3. Averaged angle degrees calculated from the parametrized ReaxFF model against quantum mechanical (QM) method (ie. DFT).

3.6 Supporting Information 2: Final Force Field

Reactive MD-force field: PEP carboxylase enzymatic mechanism with 2nd generation of water parameters

39 ! Number of general parameters
50.0000 !Overcoordination parameter
9.5469 !Overcoordination parameter
26.5405 !Valency angle conjugation parameter
1.7224 !Triple bond stabilisation parameter
6.8702 !Triple bond stabilisation parameter
60.4850 !C2-correction
1.0588 !Undercoordination parameter
4.6000 !Triple bond stabilisation parameter
12.1176 !Undercoordination parameter
13.3056 !Undercoordination parameter
-70.5044 !Triple bond stabilization energy
0.0000 !Lower Taper-radius
10.0000 !Upper Taper-radius
2.8793 !Not used
33.8667 !Valency undercoordination
6.0891 !Valency angle/lone pair parameter
1.0563 !Valency angle
2.0384 !Valency angle parameter
6.1431 !Not used
6.9290 !Double bond/angle parameter
0.3989 !Double bond/angle parameter: overcoord
3.9954 !Double bond/angle parameter: overcoord
-2.4837 !Not used
5.7796 !Torsion/BO parameter
10.0000 !Torsion overcoordination
1.9487 !Torsion overcoordination
-1.2327 !Conjugation 0 (not used)
2.1645 !Conjugation
1.5591 !vdWaals shielding
0.1000 !Cutoff for bond order (*100)
2.1365 !Valency angle conjugation parameter
0.6991 !Overcoordination parameter
50.0000 !Overcoordination parameter
1.8512 !Valency/lone pair parameter
0.5000 !Not used
20.0000 !Not used
5.0000 !Molecular energy (not used)

0.0000 !Molecular energy (not used)
2.6962 !Valency angle conjugation parameter
9 !Nr of atoms; cov.r; valency;a.m;Rvdw;Evdw;gammaEEM;cov.r2;#
alfa;gammavdW;valency;Eunder;Eover;chiEEM;etaEEM;n.u.
cov r3;Elp;Heat inc.;n.u.;n.u.;n.u.;n.u.
ov/un;val1;n.u.;val3,vval4

C	1.3350	4.0000	12.0000	1.8999	0.1175	0.6910	1.0742	4.0000
	10.4766	2.1649	4.0000	35.5960	79.5548	4.6160	6.3095	0.0000
	1.2114	0.0000	202.5551	8.7782	35.1696	13.9536	0.8563	0.0000
	-2.9043	2.3031	1.0564	4.0000	3.0494	0.0000	0.0000	0.0000
H	0.8930	1.0000	1.0080	1.3550	0.0930	0.8203	-0.1000	1.0000
	8.2180	33.2894	1.0000	0.0000	121.1250	3.7248	9.6093	1.0000
	-0.1000	0.0000	61.6606	3.0408	2.4197	0.0003	1.0698	0.0000
	-19.4571	4.2733	1.0338	1.0000	2.8793	0.0000	0.0000	0.0000
O	1.2450	2.0000	15.9990	2.3808	0.1038	1.0950	1.0548	6.0000
	9.7942	11.7301	4.0000	37.5000	116.0768	8.5000	8.3134	2.0000
	0.9049	0.1000	59.0626	3.5357	0.6653	0.0021	0.9745	0.0000
	-3.6039	2.7952	1.0493	4.0000	2.9225	0.0000	0.0000	0.0000
N	1.2333	3.0000	14.0000	1.8000	0.0906	0.5584	1.1748	5.0000
	9.9109	12.3604	4.0000	30.0710	100.0000	8.4290	7.0000	2.0000
	1.0433	0.1105	119.9837	0.6571	6.7147	2.8693	0.9745	0.0000
	-2.0000	3.3708	1.0183	4.0000	2.8793	0.0000	0.0000	0.0000
S	1.9405	2.0000	32.0600	2.0677	0.2099	1.0336	1.5479	6.0000
	9.9575	4.9055	4.0000	52.9998	112.1416	6.5000	8.2545	2.0000
	1.4601	9.7177	71.1843	5.7487	23.2859	12.7147	0.9745	0.0000
	-11.0000	2.7466	1.0338	6.2998	2.8793	0.0000	0.0000	0.0000
Si	2.0175	4.0000	28.0600	2.0473	0.1835	0.8925	1.2962	4.0000
	12.3588	1.2523	4.0000	21.7115	139.9309	4.6988	6.0000	0.0000
	-1.0000	0.0000	128.2031	8.7895	23.9298	0.8381	0.8563	0.0000
	-4.7525	2.1607	1.0338	4.0000	2.5791	0.0000	0.0000	0.0000
Na	1.8000	1.0000	22.9898	2.8270	0.1872	0.4000	-1.0000	1.0000
	10.0000	2.5000	1.0000	0.0000	0.0000	-0.9871	6.7728	0.0000
	-1.0000	0.0000	23.0445	100.0000	1.0000	0.0000	0.8563	0.0000
	-2.5000	3.9900	1.0338	8.0000	2.5791	0.0000	0.0000	0.0000
P	1.6337	3.0000	30.9738	1.4512	0.2413	1.0000	1.3000	5.0000
	9.0216	13.7407	5.0000	0.0000	0.0000	1.6505	7.9960	0.0000
	-1.0000	24.8446	125.6300	0.2175	20.7131	14.9053	0.0000	0.0000
	-3.7396	3.4698	1.0338	5.0000	2.9260	0.0000	0.0000	0.0000
X	-0.1000	2.0000	1.0080	2.0000	0.0000	1.0000	-0.1000	6.0000
	10.0000	2.5000	4.0000	0.0000	0.0000	8.5000	1.5000	0.0000
	-0.1000	0.0000	127.6226	8.7410	13.3640	0.6690	0.9745	0.0000
	-11.0000	2.7466	1.0338	6.2998	2.8793	0.0000	0.0000	0.0000

31 !Nr of bonds; Edis1;LPpen;n.u.;pbe1;pbo5;13corr;pbo6
pbe2;pbo3;pbo4;n.u.;pbo1;pbo2;ovcorr
1 1 166.1111 95.8843 78.2202 -0.6988 -0.4764 1.0000 38.7131 0.4147

0.4178 -0.1000 10.0070 1.0000 -0.0872 5.3910 1.0000 0.0000
 1 2 169.4760 0.0000 0.0000 -0.6083 0.0000 1.0000 6.0000 0.7652
 5.2290 1.0000 0.0000 1.0000 -0.0500 6.9136 0.0000 0.0000
 2 2 153.3934 0.0000 0.0000 -0.4600 0.0000 1.0000 6.0000 0.7300
 6.2500 1.0000 0.0000 1.0000 -0.0790 6.0552 0.0000 0.0000
 1 3 126.6982 124.9168 53.1082 0.2896 -0.3319 1.0000 18.7989 0.6624
 0.8382 -0.3047 9.0767 1.0000 -0.1027 5.8495 0.0000 0.0000
 3 3 142.2858 145.0000 50.8293 0.2506 -0.1000 1.0000 29.7503 0.6051
 0.3451 -0.1055 9.0000 1.0000 -0.1225 5.5000 1.0000 0.0000
 1 4 202.6206 117.0591 95.9251 -0.9401 -0.4828 1.0000 7.0000 0.1409
 0.0816 -0.3918 7.2508 1.0000 -0.1691 4.0589 1.0000 0.0000
 3 4 130.8596 169.4551 40.0000 0.3837 -0.1639 1.0000 35.0000 0.2000
 1.0000 -0.3579 7.0004 1.0000 -0.1193 6.8773 1.0000 0.0000
 4 4 157.9384 82.5526 152.5336 0.4010 -0.1034 1.0000 12.4261 0.5828
 0.1578 -0.1509 11.9186 1.0000 -0.0861 5.4271 1.0000 0.0000
 2 3 167.2086 0.0000 0.0000 -0.5770 0.0000 1.0000 6.0000 0.6019
 1.1413 1.0000 0.0000 0.0000 -0.0924 4.2778 0.0000 0.0000
 2 4 235.0480 0.0000 0.0000 -0.4825 0.0000 1.0000 6.0000 0.2330
 7.1149 1.0000 0.0000 1.0000 -0.1183 5.2018 0.0000 0.0000
 1 5 128.9942 74.5848 55.2528 0.1035 -0.5211 1.0000 18.9617 0.6000
 0.2949 -0.2398 8.1175 1.0000 -0.1029 5.6731 1.0000 0.0000
 2 5 151.5159 0.0000 0.0000 -0.4721 0.0000 1.0000 6.0000 0.6000
 9.4366 1.0000 0.0000 1.0000 -0.0290 7.0050 1.0000 0.0000
 3 5 0.0000 0.0000 0.0000 0.5563 -0.4038 1.0000 49.5611 0.6000
 0.4259 -0.4577 12.7569 1.0000 -0.1100 7.1145 1.0000 0.0000
 4 5 0.0000 0.0000 0.0000 0.4438 -0.2034 1.0000 40.3399 0.6000
 0.3296 -0.3153 9.1227 1.0000 -0.1805 5.6864 1.0000 0.0000
 5 5 96.1871 93.7006 68.6860 0.0955 -0.4781 1.0000 17.8574 0.6000
 0.2723 -0.2373 9.7875 1.0000 -0.0950 6.4757 1.0000 0.0000
 1 6 108.3910 95.0233 0.0000 0.1129 -0.5558 1.0000 17.2117 0.4568
 0.2424 -0.2378 10.1163 1.0000 -0.1020 5.7156 1.0000 0.0000
 2 6 250.0000 0.0000 0.0000 -0.7128 0.0000 1.0000 6.0000 0.1186
 18.5790 1.0000 0.0000 1.0000 -0.0731 7.4983 0.0000 0.0000
 3 6 272.8709 18.4462 0.0000 -0.6107 -0.3000 1.0000 36.0000 0.8270
 10.2334 -0.5495 29.9954 1.0000 -0.1277 7.5863 1.0000 0.0000
 4 6 119.7136 41.2405 43.3991 -0.2060 -0.3000 1.0000 36.0000 0.7957
 0.8189 -0.2614 9.4060 1.0000 -0.1245 6.1856 1.0000 0.0000
 6 6 78.0276 54.0531 30.0000 0.5398 -0.3000 1.0000 16.0000 0.0476
 0.2865 -0.8055 7.1248 1.0000 -0.0681 8.6957 0.0000 0.0000
 2 7 0.0000 0.0000 0.0000 -1.0000 -0.3000 1.0000 36.0000 0.7000
 10.1151 -0.3500 25.0000 1.0000 -0.1053 8.2003 1.0000 0.0000
 3 7 45.8933 0.0000 0.0000 -0.1511 -0.3000 1.0000 36.0000 0.3105
 5.8448 -0.3500 25.0000 1.0000 -0.0659 7.9140 1.0000 0.0000
 6 7 0.1000 0.0000 0.0000 0.2500 -0.5000 1.0000 35.0000 0.6000
 0.5000 -0.5000 20.0000 1.0000 -0.2000 10.0000 1.0000 0.0000

```

7 7 60.0000 0.0000 0.0000 -0.3458 0.3000 0.0000 25.0000 0.2477
    2.4578 -0.4000 12.0000 1.0000 -0.0513 4.5180 0.0000 0.0000
1 8 179.9072 92.0000 0.0000 0.1716 -0.1418 1.0000 13.1260 0.4180
    0.2409 -0.1310 10.7257 1.0000 -0.1147 5.4503 1.0000 0.0000
2 8 0.1466 0.0000 0.0000 0.2250 -0.1418 1.0000 13.1260 0.6000
    0.3912 -0.1310 0.0000 1.0000 -0.1029 9.3302 0.0000 0.0000
3 8 195.0941 190.3829 0.0000 1.0000 -0.5000 1.0000 25.0000 0.4873
    0.4733 -0.1642 15.6098 1.0000 -0.2572 7.1078 1.0000 0.0000
4 8 0.0000 0.0000 0.0000 0.2500 -0.5000 1.0000 35.0000 0.6000
    0.5000 -0.5000 20.0000 1.0000 -0.2000 10.0000 1.0000 0.0000
6 8 0.0000 0.0000 0.0000 0.2500 -0.5000 1.0000 35.0000 0.6000
    0.5000 -0.5000 20.0000 1.0000 -0.2000 10.0000 1.0000 0.0000
7 8 0.0000 0.0000 0.0000 0.2171 -0.5000 1.0000 35.0000 0.6000
    0.5000 -0.5000 20.0000 1.0000 -0.2000 10.0000 1.0000 0.0000
8 8 0.0000 0.0000 0.0000 0.2171 -0.5000 1.0000 35.0000 0.6000
    0.5000 -0.5000 20.0000 1.0000 -0.2000 10.0000 1.0000 0.0000
16 ! Nr of off-diagonal terms; Ediss;Ro;gamma;rsigma;rpi;rpi2
1 2 0.1239 1.4004 9.8467 1.1210 -1.0000 -1.0000
2 3 0.0320 1.2896 10.9108 0.9215 -1.0000 -1.0000
2 4 0.1248 1.3000 9.8958 1.0320 -1.0000 -1.0000
1 3 0.0228 1.9391 10.8116 1.2489 1.1490 1.0506
1 4 0.1386 1.8640 9.7844 1.3603 1.2540 1.1315
3 4 0.0951 2.5000 9.0048 1.4348 1.0531 1.2771
1 6 0.0541 2.0811 13.5179 1.7778 1.5840 -1.0000
2 6 0.1659 1.4000 11.7054 1.3437 -1.0000 -1.0000
3 6 0.1330 2.0545 10.8315 1.7043 1.3773 -1.0000
4 6 0.1297 1.9384 10.9856 1.6175 1.4045 -1.0000
1 5 0.1408 1.8161 9.9393 1.7986 1.3021 1.4031
2 5 0.0895 1.6239 10.0104 1.4640 -1.0000 -1.0000
3 7 0.0825 1.5904 11.3396 1.5905 -1.0000 -1.0000
6 7 0.1757 2.0409 13.7267 -1.0000 -1.0000 -1.0000
3 8 0.0520 1.5627 9.1396 1.7764 1.3780 -1.0000
6 8 0.1130 1.7570 9.3000 -0.1000 -1.0000 -1.0000
99 ! Nr of angles;at1;at2;at3;Thetao,o;ka;kb;pv1;pv2;val(bo)
1 1 1 65.0616 33.5007 0.7035 0.0000 0.7272 6.7284 1.0722
1 1 2 65.7758 14.5234 6.2481 0.0000 0.5665 0.0000 1.6255
2 1 2 70.2607 25.2202 3.7312 0.0000 0.0050 0.0000 2.7500
1 2 2 0.0000 0.0000 6.0000 0.0000 0.0000 0.0000 1.0400
1 2 1 0.0000 3.4110 7.7350 0.0000 0.0000 0.0000 1.0400
2 2 2 0.0000 27.9213 5.8635 0.0000 0.0000 0.0000 1.0400
1 1 3 54.4165 19.6475 1.7468 0.0000 2.7430 57.3098 1.0436
3 1 3 90.5094 17.5479 3.1342 -2.4810 1.4258 0.0000 2.9645
1 1 4 65.5461 44.6277 2.0685 0.0000 1.1834 0.0000 2.9572
3 1 4 72.4850 12.4764 0.1000 0.0000 1.1019 0.0000 1.4570
4 1 4 90.0000 43.1792 0.5055 0.0000 1.1155 0.0000 1.0204

```

2	1	3	64.3088	32.5434	2.1997	0.0000	0.1000	0.0000	1.2995
2	1	4	48.1894	40.9792	0.9493	0.0000	0.2000	0.0000	2.8341
1	2	4	0.0000	0.0019	6.3000	0.0000	0.0000	0.0000	1.0400
1	3	1	70.1705	47.2718	1.2877	0.0000	2.9579	0.0000	1.0000
1	3	3	79.7622	43.6373	1.5083	0.0000	1.0356	72.2772	1.4733
1	3	4	69.5983	45.0000	1.4248	0.0000	2.9000	0.0000	2.3286
3	3	3	80.7324	30.4554	0.9953	0.0000	1.6310	50.0000	1.0783
3	3	4	83.5202	33.7933	1.0337	0.0000	2.9000	0.0000	1.3398
4	3	4	67.1317	42.3748	1.7873	0.0000	3.0072	0.0000	1.5832
1	3	2	86.2788	7.8085	6.9850	0.0000	1.2779	0.0000	2.8959
2	3	3	75.6935	50.0000	2.0000	0.0000	1.0000	0.0000	1.1680
2	3	4	72.7348	20.1071	7.5000	0.0000	0.1000	0.0000	1.0746
2	3	2	85.1864	8.5843	2.2985	0.0000	2.9142	0.0000	2.0521
1	4	1	57.1959	24.9572	3.2923	0.0000	2.8702	0.0000	1.2002
1	4	3	73.9745	21.1329	2.3337	0.0000	2.8701	0.0000	1.7170
1	4	4	71.4579	14.0942	2.8540	0.0000	2.8701	0.0000	1.0631
3	4	3	74.2613	20.9008	2.8607	-18.0069	3.0701	0.0000	1.3874
3	4	4	74.2615	27.8669	1.6736	-0.9193	3.0117	0.0000	1.4381
4	4	4	73.3189	24.9685	2.2561	0.0000	2.9983	0.0000	2.1573
1	4	2	69.4924	7.2031	3.5196	0.0000	0.2025	0.0000	1.2709
2	4	3	74.5739	45.0000	1.4078	0.0000	0.3956	0.0000	3.0000
2	4	4	79.7136	45.0000	0.5316	0.0000	0.5437	0.0000	1.0000
2	4	2	81.8118	13.4140	6.8107	0.0000	0.1000	0.0000	1.0000
1	2	3	0.0000	11.1212	0.6706	0.0000	0.0000	0.0000	1.0000
1	2	4	0.0000	0.2694	2.1363	0.0000	0.0000	0.0000	1.8036
1	2	5	0.0000	15.0000	3.0000	0.0000	0.0000	0.0000	1.0400
3	2	3	0.0000	11.8475	2.7571	0.0000	0.0000	0.0000	2.9000
3	2	4	10.0112	0.4699	0.1000	0.0000	0.0000	0.1200	2.7822
4	2	4	7.0790	0.1901	0.4358	0.0000	0.0000	0.1050	2.1684
2	2	3	0.0000	6.4269	2.8500	0.0000	0.0000	0.0000	1.0772
2	2	4	0.0000	0.0019	6.0000	0.0000	0.0000	0.0000	1.0400
1	1	5	74.9397	25.0560	1.8787	0.1463	0.0559	0.0000	1.0400
1	5	1	86.9521	36.9951	2.0903	0.1463	0.0559	0.0000	1.0400
2	1	5	74.9397	25.0560	1.8787	0.0000	0.0000	0.0000	1.0400
1	5	2	86.1791	36.9951	2.0903	0.0000	0.0000	0.0000	1.0400
1	5	5	85.3644	36.9951	2.0903	0.1463	0.0559	0.0000	1.0400
2	5	2	93.1959	36.9951	2.0903	0.0000	0.0000	0.0000	1.0400
2	5	5	84.3331	36.9951	2.0903	0.0000	0.0000	0.0000	1.0400
6	6	6	71.0490	32.4076	1.2648	0.0000	0.0133	0.0000	1.2899
2	6	6	77.2616	5.0190	7.8944	0.0000	4.0000	0.0000	1.0400
2	6	2	75.7983	14.4132	2.8640	0.0000	4.0000	0.0000	1.0400
3	6	6	99.8997	26.6610	2.1237	0.0000	0.0100	0.0000	1.4341
2	6	3	73.6998	40.0000	1.8782	0.0000	4.0000	0.0000	1.1290
3	6	3	98.2184	38.9429	0.7727	0.0000	1.1658	0.0000	2.2641
6	3	6	39.2858	1.3068	5.6478	0.0000	3.8972	0.0000	3.0000

2	3	6	79.2126	4.8973	8.0000	0.0000	1.0859	0.0000	2.1209	
3	3	6	82.7397	32.1198	1.8862	0.0000	0.1058	0.0000	1.5443	
2	2	6	0.0000	47.1300	6.0000	0.0000	1.6371	0.0000	1.0400	
6	2	6	0.0000	27.4206	6.0000	0.0000	1.6371	0.0000	1.0400	
3	2	6	0.0000	7.0550	3.9236	0.0000	1.6371	0.0000	1.0400	
2	2	5	0.0000	0.0019	6.0000	0.0000	0.0000	0.0000	1.0400	
1	1	6	72.5239	22.3583	2.0393	0.0000	1.0031	0.0000	1.0400	
1	6	1	69.1709	18.9268	2.1226	0.0000	1.0031	0.0000	1.0400	
6	1	6	68.6453	18.7377	2.0496	0.0000	1.0031	0.0000	1.0400	
1	6	6	68.9902	19.7021	2.0587	0.0000	1.0031	0.0000	1.0400	
2	1	6	72.6403	13.6964	2.4702	0.0000	1.0000	0.0000	1.0400	
1	6	2	71.8708	14.6864	2.4702	0.0000	1.0000	0.0000	1.0400	
4	6	6	60.6199	17.7559	1.0576	0.0000	2.1459	0.0000	1.0400	
4	6	4	74.1294	20.6494	2.1244	0.0000	0.7689	0.0000	1.0400	
3	6	4	57.0650	9.4985	0.3423	0.0000	0.7689	0.0000	1.0400	
6	4	6	24.1137	1.7457	0.2198	0.0000	4.1125	0.0000	1.0400	
2	6	4	68.7410	15.5851	1.8545	0.0000	0.8613	0.0000	1.0400	
2	4	6	80.9040	4.0560	1.2284	0.0000	1.6982	0.0000	1.0400	
4	4	6	60.0000	10.0000	0.7500	0.0000	1.0000	0.0000	1.0400	
3	4	6	69.8728	32.7155	1.5875	0.0000	2.2466	0.0000	1.0400	
4	3	6	69.8728	27.1273	1.5875	0.0000	2.2466	0.0000	1.0400	
4	2	6	0.0000	31.0427	4.5625	0.0000	1.6371	0.0000	1.0400	
3	11	3	60.4781	11.1095	0.4044	0.0000	1.0000	0.0000	1.5333	
11	3	11	24.3810	6.0806	1.0000	0.0000	1.0251	0.0000	1.1000	
3	3	11	60.2827	26.3290	0.5226	0.0000	1.0000	0.0000	1.7180	
1	3	6	85.8521	12.6881	1.0112	0.0000	1.0000	0.0000	1.3220	
1	6	3	71.7524	35.8987	1.5000	0.0000	1.0000	0.0000	1.0487	
3	1	6	70.0000	5.0250	1.0000	0.0000	1.0000	0.0000	1.2500	
1	2	6	0.0000	2.5000	1.0000	0.0000	1.0000	0.0000	1.2500	
3	8	3	82.6335	17.6654	0.7057	-7.1312	0.1338	0.0000	1.9390	
2	3	8	72.6004	9.6150	0.8905	0.0000	3.5473	0.0000	1.0400	
3	3	8	60.0000	40.0000	4.0000	0.0000	1.0000	0.0000	1.0400	
3	2	8	0.0000	10.0000	1.0000	0.0000	1.0000	0.0000	1.0400	
6	3	8	67.3176	9.1175	2.2677	0.0000	1.0000	0.0000	1.0000	
8	3	8	62.1312	7.5931	0.1000	0.0000	0.5154	0.0000	2.1744	
1	3	8	66.4081	2.1763	2.1783	0.0000	1.0000	0.0000	2.5198	
2	8	3	75.0000	25.0000	2.0000	0.0000	1.0000	0.0000	1.2500	
3	8	8	70.0000	25.0000	2.0000	0.0000	1.0000	0.0000	1.2500	
1	8	3	82.8511	35.1702	2.2155	0.0000	1.0000	0.0000	1.0400	
3	1	8	50.2929	41.6249	2.8868	0.0000	1.0000	0.0000	1.0400	
2	1	8	74.8439	24.3187	3.4300	0.0000	0.0050	0.0000	1.3014	
1	1	8	44.8758	38.4344	2.3166	0.0000	0.7180	6.2933	1.3532	
1	8	1	56.0196	40.1896	1.0567	0.0000	0.7180	6.2933	1.1983	
67	!	Nr of torsions;at1;at2;at3;at4;;V1;V2;V3;V2(BO);vconj;n.u;n								
1	1	1	1	-0.2500	34.7453	0.0288	-6.3507	-1.6000	0.0000	0.0000

1	1	1	2	-0.2500	29.2131	0.2945	-4.9581	-2.1802	0.0000	0.0000
2	1	1	2	-0.2500	31.2081	0.4539	-4.8923	-2.2677	0.0000	0.0000
1	1	1	3	-0.5740	22.4215	0.8787	-2.7603	-1.1000	0.0000	0.0000
2	1	1	3	1.2143	5.0000	0.4930	-8.1288	-1.0978	0.0000	0.0000
3	1	1	3	-2.5000	56.1599	-1.0000	-4.3607	-0.8614	0.0000	0.0000
1	1	3	1	2.5000	14.6490	1.0000	-2.5209	-0.9000	0.0000	0.0000
1	1	3	2	-2.5000	5.0000	-0.5000	-5.7468	-0.9000	0.0000	0.0000
2	1	3	1	-1.0402	26.8401	0.6384	-2.5000	-0.9000	0.0000	0.0000
2	1	3	2	-1.0000	66.0304	0.7580	-5.4593	-1.1000	0.0000	0.0000
1	1	3	3	1.0182	5.3409	0.1292	-4.3356	-2.0544	0.0000	0.0000
2	1	3	3	2.1531	45.9655	1.0000	-2.5000	-2.8274	0.0000	0.0000
3	1	3	1	0.6706	80.0000	-0.2443	-4.7181	-3.0437	0.0000	0.0000
3	1	3	2	-0.6375	25.4544	-0.3400	-5.6706	-3.0476	0.0000	0.0000
3	1	3	3	-1.9346	5.0000	0.6401	-3.3416	-2.7174	0.0000	0.0000
1	3	3	1	1.0469	4.3827	0.8149	-3.4434	-2.7536	0.0000	0.0000
1	3	3	2	-2.5000	-0.5181	0.0268	-5.4085	-2.9498	0.0000	0.0000
2	3	3	2	2.5000	-4.0000	0.9000	-2.5000	-1.0000	0.0000	0.0000
1	3	3	3	2.4118	-24.8219	0.9706	-2.5004	-0.9972	0.0000	0.0000
2	3	3	3	0.8302	-4.0000	-0.7763	-2.5000	-1.0000	0.0000	0.0000
3	3	3	3	-2.5000	-4.0000	1.0000	-2.5000	-1.0000	0.0000	0.0000
1	1	4	2	-0.8437	86.3887	0.0320	-7.6835	-1.9825	0.0000	0.0000
2	1	4	2	-0.9868	96.6106	0.7443	-7.5073	-2.1051	0.0000	0.0000
3	1	4	2	-0.6817	3.5069	0.0849	-6.0257	-2.5261	0.0000	0.0000
3	1	1	4	0.4250	41.9360	0.6083	-3.8680	-0.9511	0.0000	0.0000
4	1	1	4	-1.0000	21.8427	1.0000	-4.0686	-1.7241	0.0000	0.0000
1	1	4	1	1.0000	87.5734	0.6399	-5.9421	-1.6589	0.0000	0.0000
3	1	4	1	-0.5092	24.8952	0.4691	-7.6208	-1.8038	0.0000	0.0000
2	1	1	4	1.0000	98.8297	-0.2745	-4.9954	-1.9000	0.0000	0.0000
4	1	4	2	0.5000	2.8273	-0.1650	-7.9605	-2.0202	0.0000	0.0000
2	1	4	1	-0.0359	88.1512	-0.0959	-7.2511	-1.5996	0.0000	0.0000
0	1	2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	2	2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	2	3	0	0.0000	0.1000	0.0200	-2.5415	0.0000	0.0000	0.0000
0	1	1	0	0.0000	50.0000	0.3000	-4.0000	-2.0000	0.0000	0.0000
0	3	3	0	0.5511	25.4150	1.1330	-5.1903	-1.0000	0.0000	0.0000
0	1	4	0	0.2176	40.4126	0.3535	-3.9875	-2.0051	0.0000	0.0000
0	2	4	0	0.0000	0.1032	0.3000	-5.0965	0.0000	0.0000	0.0000
0	3	4	0	1.1397	61.3225	0.5139	-3.8507	-2.7831	0.0000	0.0000
0	4	4	0	0.7265	44.3155	1.0000	-4.4046	-2.0000	0.0000	0.0000
4	1	4	4	-0.0949	8.7582	0.3310	-7.9430	-2.0000	0.0000	0.0000
0	1	5	0	4.0885	78.7058	0.1174	-2.1639	0.0000	0.0000	0.0000
0	5	5	0	-0.0170	-56.0786	0.6132	-2.2092	0.0000	0.0000	0.0000
0	2	5	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	6	6	0	0.0000	0.0000	0.1200	-2.4426	0.0000	0.0000	0.0000
0	2	6	0	0.0000	0.0000	0.1200	-2.4847	0.0000	0.0000	0.0000

0	3	6	0	0.0000	0.0000	0.1200	-2.4703	0.0000	0.0000	0.0000
1	1	3	3	-0.0002	20.1851	0.1601	-9.0000	-2.0000	0.0000	0.0000
1	3	3	1	0.0002	80.0000	-1.5000	-4.4848	-2.0000	0.0000	0.0000
3	1	3	3	-0.1583	20.0000	1.5000	-9.0000	-2.0000	0.0000	0.0000
1	1	1	8	1.0000	28.0452	0.0382	-8.0000	-1.7255	0.0000	0.0000
8	1	1	8	0.0000	80.5586	0.1104	-8.0928	-1.7255	0.0000	0.0000
0	1	8	0	4.0000	45.8264	0.9000	-4.0000	0.0000	0.0000	0.0000
0	8	8	0	4.0000	45.8264	0.9000	-4.0000	0.0000	0.0000	0.0000
2	1	3	8	-1.5000	18.9285	0.3649	-6.1208	0.0000	0.0000	0.0000
2	3	8	3	1.5000	-1.0000	0.2575	-6.2100	0.0000	0.0000	0.0000
1	3	8	3	-1.3246	-0.8513	1.0985	-2.4363	0.0000	0.0000	0.0000
8	3	8	3	-1.5000	21.5086	-1.0000	-4.8869	0.0000	0.0000	0.0000
1	4	2	3	-2.5000	10.8893	-0.7189	-2.7243	-0.9000	0.0000	0.0000
1	3	2	4	-0.2868	15.6462	-1.0200	-2.6045	-0.9000	0.0000	0.0000
1	3	2	3	-1.7443	13.8935	-1.3212	-2.5000	-0.9000	0.0000	0.0000
3	2	4	2	-1.7508	9.6877	-1.9412	-2.7882	-0.9000	0.0000	0.0000
4	1	3	2	-2.4642	11.8541	-1.8692	-3.1855	-0.9000	0.0000	0.0000
1	1	8	1	0.5864	32.4198	-0.2514	-8.4627	-1.6000	0.0000	0.0000
2	1	1	8	0.2704	32.0532	0.2173	-5.3228	-2.1802	0.0000	0.0000
2	1	8	1	0.5984	7.2907	0.1161	-5.1321	-2.1802	0.0000	0.0000
1	2	3	1	0.0000	1.3164	0.2737	-2.5016	0.0000	0.0000	0.0000
6	! Nr of hydrogen bonds;at1;at2;at3;Rhb;Dehb;vhb1									
3	2	3		2.1653	-3.6983	1.7831	17.0964			
3	2	4		1.9036	-6.6449	1.4500	19.5000			
4	2	3		2.4625	-5.5217	1.4500	19.5000			
4	2	4		2.4924	-2.6497	1.4500	19.5000			
1	2	3		2.1397	-4.7669	1.4500	19.5000			
1	2	4		2.2924	-2.5509	1.4500	19.5000			

Chapter 4

Development of a Reactive Force Field for Photo-initiated Acrylate Polymerization in Volumetric Additive Manufacturing

My contribution to this work included all the computational work, analysis, and writing.

This chapter is currently in preparation for publication:

Huang, Y., Faller, R., Shusteff, M., 2022

Abstract

Light-driven and photo-curable polymer based additive manufacturing (AM) has a huge potential due to its highest resolution and precision. Unlike point- and layer-wise photopolymer AM technologies, volumetric AM (VAM) forms the entire part all at once, which produces parts much more rapidly and overcomes stair step surface artifacts and inhomogeneous mechanical properties. Acrylated radical chain-growth polymerized resins are widely used in photopolymer AM due to their fast kinetics, and they can also serve as the standard for developing other resin materials for volumetric AM technology. For successful control of the photopolymer resins, the molecular basis of the acrylate free-radical polymerization must be understood in detail. We present an optimized reactive force field (ReaxFF) molecular dynamics (MD) simulation force field for acrylate polymer resins to account for the radical polymerization thermodynamics and kinetics. The force field is trained against our extensive training set including density functional theory (DFT) calculations of reaction pathways along the radical polymerization from methyl

acrylate to methyl butyrate, bond dissociation energies, and structures and partial charges of all molecules and radicals. We also trained the force field against the reaction that should not happen during acrylate photopolymerization. The parametrization process utilizes a parallelized search algorithm, and the resulting model can describe polymer resin formation, crosslinking density, conversion rate, and residual monomers of the more complex acrylate mixtures.

4.1 Introduction

4.1.1 Acrylates as Photopolymer Resin

Photopolymerization-based techniques for additive manufacturing has attracted many interests and attentions in the field of biomedical devices, dentistry, tissue engineering, and drug delivery ^[1-4], due to their high precision ^[5], mature and commercialized photo-chemical innovations, and environmental and economic benefits ^[3,4]. The resins for photopolymerization-based additive manufacturing (AM) consist of photoinitiators and monomers/oligomers in liquid state. During the AM process, initiating species are generated by light irradiating the photoinitiators, and they react with monomers/oligomers and then drive the chain growth via radical or cationic reactions ^[3]. Because the resins can only be cured under light irradiation, the solid part can be easily separated from the liquid resin after the printing process. Acrylate monomers/oligomers are widely used as photocurable resin materials for commercial products due to their fast kinetics and compatibility with different form of commercial 3D printers, and they proceed via free-radical polymerization in the AM process (Figure 4.1). Some of the frequently used acrylates for AM includes poly(ethylene glycol) diacrylate (PEGDA), pentaerythritol

triacrylate (PETA), bisphenol A ethoxylate diacrylate (Bis-EDA), bisphenol A-glycidyl methacrylate (Bis-GMA), and urethane dimethacrylate (UDMA).

Although acrylates demonstrate effectiveness in AM photopolymerization, they still have limitations and more efforts are needed to improve their performances. Acrylate resins tend to shrink during polymerization, and the fast kinetics could result in less uniform networks and hinder the dissipation of shrinkage stress ^[3-5], causing brittleness and cracks. However, both kinetics and shrinkage can be tuned by molecular structure and weight ^[3]. Therefore, studies of the acrylate resin materials design and how it could affect the polymerization step and network structure could help improve the mechanical properties of the printed products and create more versatile application scenarios.

Another limitation of acrylates as photopolymerization AM resin is the oxygen inhibition, when the resin is in contact with air, and oxygen inhibits the free-radical polymerization, causing slow polymerization rate, low conversion, long induction period, and uncontrollable surface properties^[6]. Using Thiol-ene-based resins, which are relatively newly emerging, can alleviate the oxygen inhibition by their radical step-growth polymerization. Moreover, products printed from thiol-ene systems show good toughness, and thus making thiol-ene resins attractive AM candidates. Evaluating acrylate resins' polymerization kinetics and polymer network structure can also set standard for novel thiol-ene resins development. This work focuses on developing a computational model for acrylates free-radical polymerization that can be used for kinetics and network structure analysis for various acrylate resins.

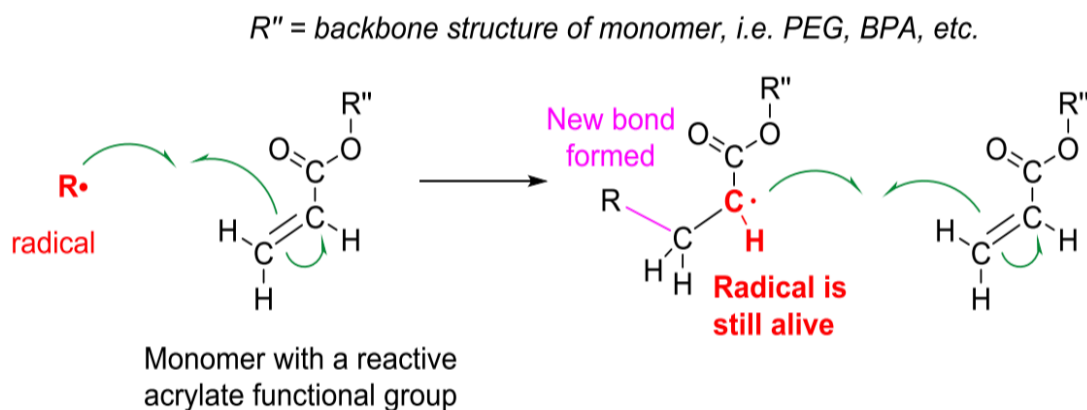


Figure 4.1: Generalized acrylate free-radical reaction.

4.1.2 ReaxFF For Acrylate Free-radical polymerization

Quantum chemistry based methods like density functional theory (DFT) calculation are powerful tools to describe chemical reactions on the atomistic scale. However, this type of calculations is very computationally expensive and thus limits the time scale and length scales. On the other hand, empirical force field methods like classical molecular dynamics (MD) simulation can study the system's dynamic evolution in nanoseconds with thousands of atoms, but losses the reaction information since the atoms connectivity are predefined. Therefore, in order to describe the radical chain-growth polymerization kinetics as well as the structural and rheological properties of the acrylate resin, we use ReaxFF-based reactive MD simulation that fills in the gap between quantum chemistry methods and classical empirical force field methods. ReaxFF is a reactive force field that allows bond formation and dissociation. The total system energy is comprised of bond order (BO) dependent energies include bond energies (E_{bond}), over-coordination penalty term (E_{over}), under-coordination stabilization term (E_{under}), lone-pair energies (E_{lp}), valence angle energies (E_{val}), and torsion energies (E_{torsion}), and non-bonded energies include Coulomb energy (E_{coulomb}) and van der Waals energy (E_{vdWaals})^[7]. Bond order is a smooth function of atom

distance, and the BO dependent energies are functions of BO. Coulomb energy is determined using a geometry dependent charge distribution scheme – electronegativity equalization method (EEM), and van der Waals energy accounts for short-range Pauli repulsion and long-range dispersion^[8,9].

The first ReaxFF force field was developed by van Duin et al. in 2001^[10] for a hydrocarbon system, and then improved for hydrocarbon combustion by Chenoweth, van Duin, and Goddard in 2008^[7,9]. The 2008-C/H/O force field contains H, O, and C that are the key elements in acrylates. Another ReaxFF force field Ti/N/F/C/O/H^[8] was developed from the 2008-C/H/O force field and contains more elements so it is compatible for more complex acrylates systems. However, running MD simulations with either the 2008-C/H/O force field or the Ti/N/F/C/O/H force field resulted in the reaction between the radical and the ketone group or the ether group instead of the reaction between the radical and the vinyl group in the liquid monomer (Figure 4.2). Therefore, we took the Ti/N/F/C/O/H force field as the starting parameter set and re-optimized the C/O/H parameters to better describe the acrylate free-radical polymerization. Parameters of other elements like Ti, N, and F were kept untouched because they did not participate in the reactions.

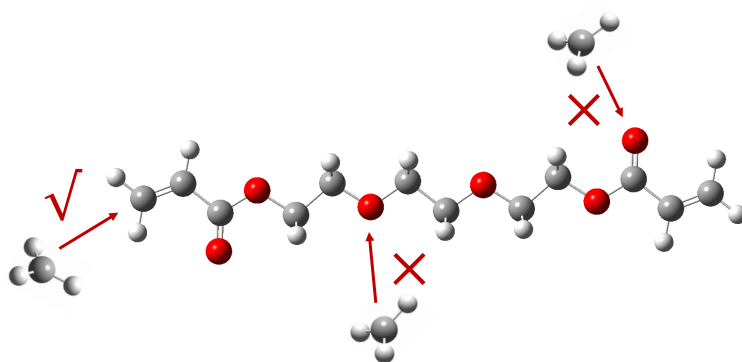


Figure 4.2: Correct and wrong reaction sites as shown in poly (ethylene glycol) diacrylate ($M_n=250$) (PEGDA).

4.2 Computational Methods

4.2.1 Training Set Generation

The optimization process has of two parts: training set generation and parametrization using parallel search algorithm ^[11] that determines the force field parameters by minimizing the penalty function (total weighted error):

$$P = \sum_i^{train\ set} \left(\frac{V_{i,train} - V_{i,react}}{\omega_i V_{i,train}} \right)^2 \quad (1)$$

where $V_{i,train}$ is the property value (eg. energy) from the training set, $V_{i,react}$ is the property value computed from ReaxFF parameters, and ω_i is the inverse weight that attributes the level of importance of each value to the whole parameterization.

The property values in the training set were all calculated from DFT using the computational chemistry software package Gaussian 16 ^[12]. In order to reduce computational expense and increase the transferability among different acrylates, we used the simplest radical reaction where CH_3 radical reacts with the vinyl group in methyl acrylate to produce methyl butyrate radical as the training reaction (Figure 4.3a). Energies of configurations along the reaction pathway were calculated and the energy differences between the transition state and other configurations, and between the product and reactant were stored in the training set as reaction energies. The energies along the reaction pathway of the unphysical reaction where the radical reacts with the ketone group in methyl acrylate (Figure 4.3b) was also included because we also trained the force field against the wrong reaction to prevent it from competing with the correct reaction. The radical reacting with the ether group was due to unphysically weak C-O bond and it could not happen if the C-O bond orders and bond strengths are correct, so this reaction was not included in the training set. C-C, C=C bonds at the active sites for both methyl acrylate and methyl

butyrate were scanned from very short distances to equilibrium distances and then to very large distances without relaxing the whole structure, and the potential energies along the scans were obtained. Same calculations were done for C-O and C=O bonds in methyl acrylate and methyl butyrate as well, since those bonds behave poorly with the non-optimized force field. C-H bond was also included to ensure that improving the C-C/C=C/C-O/C=O bonds didn't worsen the already good C-H bond. Because the spin multiplicity could change when the bonds were pulled to dissociation and ReaxFF does not treat spin multiplicity explicitly, both singlet state scan and triplet state scan were performed for bonds in methyl acrylate, and both doublet state scan and quartet state scan were performed for bonds in methyl butyrate in DFT calculations, and only the lower energy points of the two states for each scan were included in the training set as bond dissociation energies to train the ReaxFF parameters. Bond lengths, angles, and atomic partial charges of CH₃, methyl acrylate, methyl butyrate, and the transition state were also included in the training set. Other types of information, for example liquid state density, can also be included in the training set. But in this work, we decided to exclude these types of information from the training set and use them as validations after the parametrization process is done, to avoid overfitting. The hybrid Becke, 3-parameter, Lee–Yang–Parr functional B3LYP^[13,14] and 6311++g(d,p) basis set were used for all geometry optimizations and energy calculation. For consistency to how charge equilibrium is development for ReaxFF^[10], the DFT partial charges were calculated using Mulliken charge calculation population analysis^[15], using B3LYP/631g(d,p).

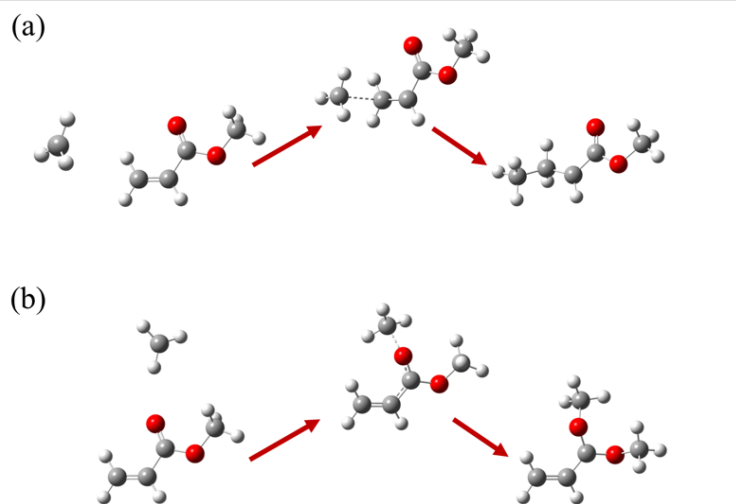


Figure 4.3: (a) Methyl radical reacts with the vinyl group in methyl acrylate to create the methyl butyrate radical; (b) Methyl radical reacts with the ketone group in methyl acrylate (wrong reaction).

4.2.2 Parallel Parametrization Scheme

All ReaxFF parametrization algorithms explore the parameter space and determine the optimum parameters that minimize the penalty function in a similar form as equation 1. Initially, ReaxFF parametrizations were done by the single-parameter parabolic-search algorithm^[16] that optimize the force field parameters one at a time by a parabolic extrapolation procedure. This is inefficient and requires lot of experience and intuition to determine which set of parameters to optimize first in order to speed up the process. Recent parametrization algorithms include Monte Carlo Simulated Annealing scheme^[17] and Genetic Algorithms^[18]. These algorithms enhance the exploration of the parameter space and allows global optimization, but the number of parameters should be limited as small as possible to make the procedure feasible^[11, 19]. A typical ReaxFF force field includes hundreds of parameters, and most of them are coupled, so choosing a smallest possible parameter set and determine which to optimize first are difficult. Therefore, we use the

parallelized search algorithm ^[11] developed by Deetz et al, to optimize our force field. This algorithm utilizes parallel computing. Each processor is assigned with a small list of parameters, each parameter is independently evaluated, and all parameters are updated simultaneously after each iteration. Therefore, faster parametrization can be achieved ^[11], and the likelihood of being entrapped in local minima is reduced. Not all parameters in the ReaxFF force field need to be optimized. Some of them are not directly related to the reactions of interest, and some are insensitive to the parametrization. Table 4.1 shows which parameters are chosen and why they are chosen. The optimization process is similar to the one in our previous work ^[20], and is described in the supporting information.

Parameters chosen		Reasons for chosen
Electrostatic	Electronegativity equalization method (EEM) parameters for C,H,O	Bonding prediction
Valence bond	All bond radii/order/dissociation energy, and under/over coordination energy parameters for C, H, O atoms and C-C, C-O, C-H pairs	Bonding prediction Reaction pathway Molecular structure
Valence angle	All angle parameters for C, H, O related angles in the molecules	Geometry prediction
Van der Waals	Vdw radii/dissociation energy/shielding for C, H, O	Bonding prediction

Table 4.1: Choice of force field parameters.

4.2.3 MD Simulations

All MD simulations were performed in LAMMPS ^[21,22], and all the initial configurations with different monomers were independently generated using Packmol ^[23] in which the molecules are randomly distributed, and atom overlaps were minimized. The initial box sizes were determined assuming the densities of the liquid monomers were the same as the documented experimental values. Then all the initial configurations were subjected to equilibration. During equilibration, a series of short trajectories with short time steps were performed and temperature

was carefully ramped up from 100 K to 300 K in NVT (canonical) ensemble followed by 500 ps run in NPT (isothermal-isobaric) ensemble at 300 K and 1 atm to eliminate any high-energy configurations and relax the systems across the periodic boundary. The liquid monomer properties were then taken from the equilibrated systems at ambient conditions. For radical polymerization reactions, CH₄ was first introduced to the monomers and once the systems were equilibrated, one of the hydrogen on CH₄ was removed to generate the CH₃· radical and initiate the reaction. Due to computer time considerations, reactions were run in NPT ensemble for 20 ns. For all the MD simulations, temperature was maintained using a Nosé-Hoover thermostat^[24, 25] with a damping parameter of 25 fs, and pressure was maintained using a Nosé-Hoover barostat with a damping parameter of 250 fs. Except during the temperature ramping process, time steps of 0.25 fs were used for all the NVT and NPT runs.

4.3 Results and Discussions

4.3.1 Force Field Development

The developed force field ff.C/H/O/N generally match the DFT partial charges, geometries (bonds and angles), bond dissociation energies, and reaction energies well. The error between DFT calculated and ReaxFF fitted averaged bond lengths and angles were smaller than 6%, and 3% respectively, as shown in Figure 4.5, proving that the developed ReaxFF ff.C/H/O/N reproduces charge and geometry well for the CH₃·, methyl acrylate, methyl butyrate, and transition state system.

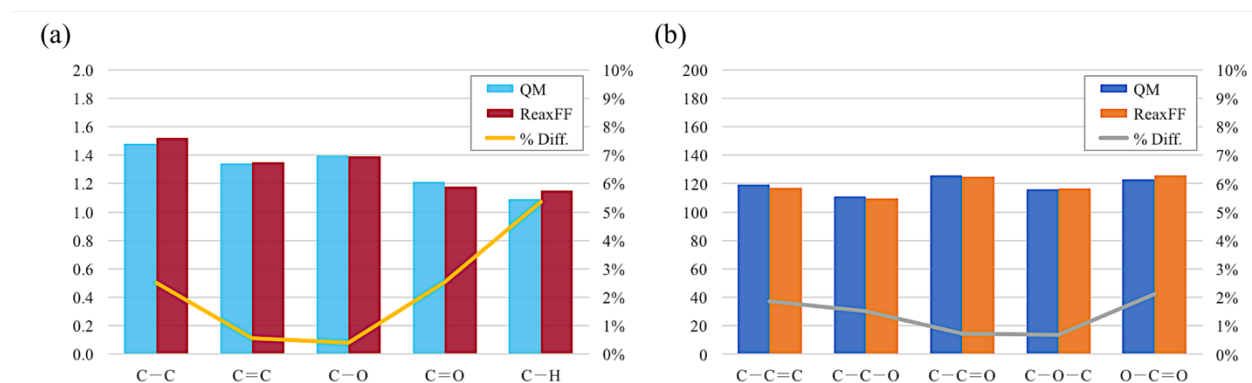


Figure 4.5: (a) Average bond length differences; (b) Average angle differences.

Bond dissociation energy (BDE) scans in figure 4.6 and energies along reaction pathways in figure 4.7 also show good agreement between DFT and ReaxFF. Including the energetics of the wrong reaction into the training process worsen the BDE of the C-O and C=O bonds, and vice versa. So, we compromised the accuracy such that the wrong reaction has a positive reaction energy and a similar barrier as the correct reaction, while still having reasonable BDE of the C-O and C=O bonds. The energy barriers calculated from DFT and ReaxFF of the correct reaction were 3.93 kcal/mol and 5.22 kcal/mol, respectively; those of the wrong reaction were 19.38 kcal/mol and 20.74 kcal/mol, respectively. The reaction energies calculated from DFT and ReaxFF of the correct reaction were -29.80 kcal/mol and -41.27 kcal/mol, respectively; those of the wrong reaction were 2.94 kcal/mol and 0.63 kcal/mol, respectively. The energetics results from the 2008-C/H/O force field and the force field before parametrization were also calculated for comparison. It was obvious that the developed ReaxFF force field outperformed both. All these comparisons show that the optimized ReaxFF force field can be used for acrylate radical polymerization reactions for our resins.

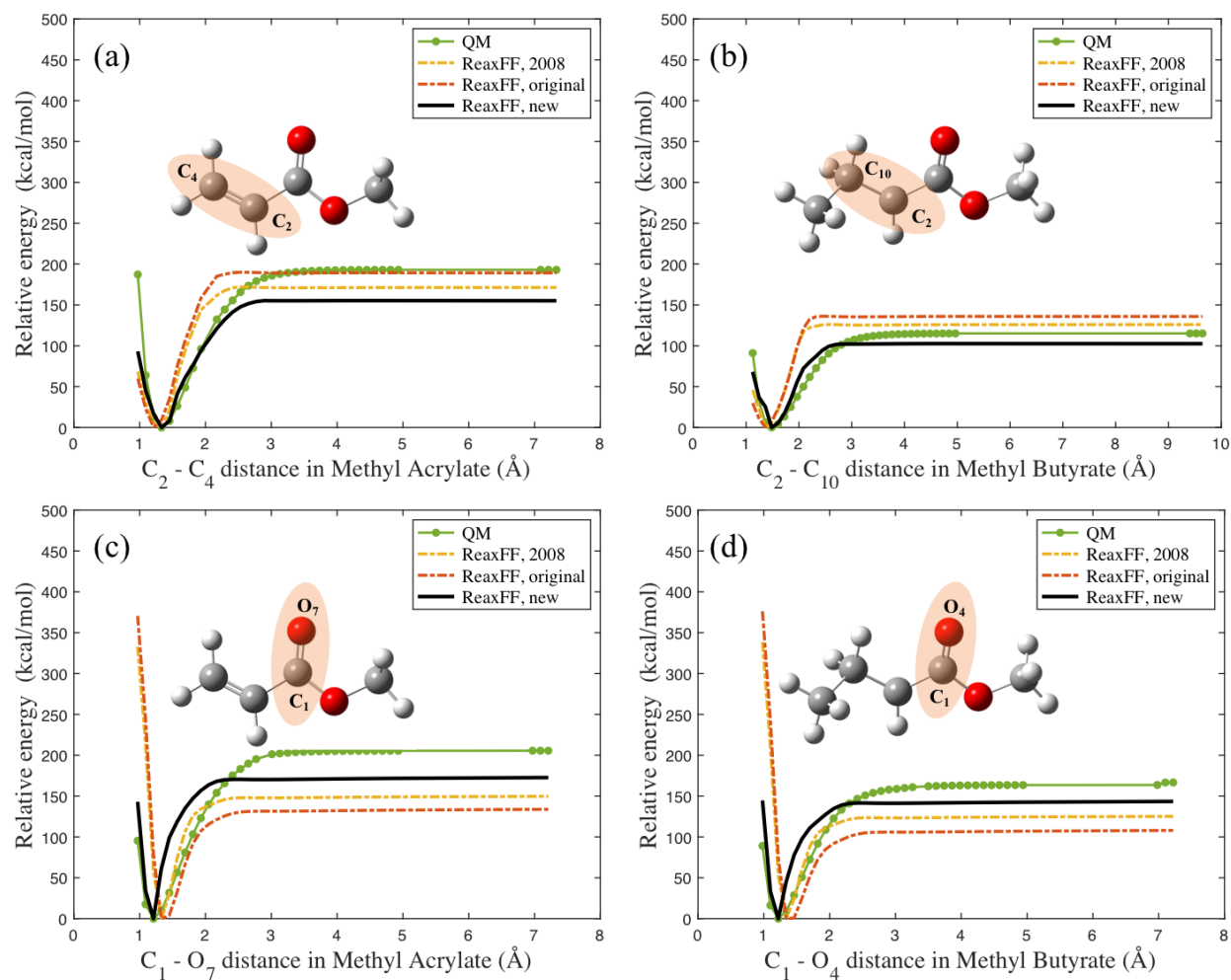


Figure 4.6: (a-d) Bond dissociation energy scans comparison between DFT and ReaxFF, “ReaxFF, new” is the developed force field, “ReaxFF, 2008” is the 2008-C/H/O force field, and “ReaxFF, original” is the force field before parametrization.

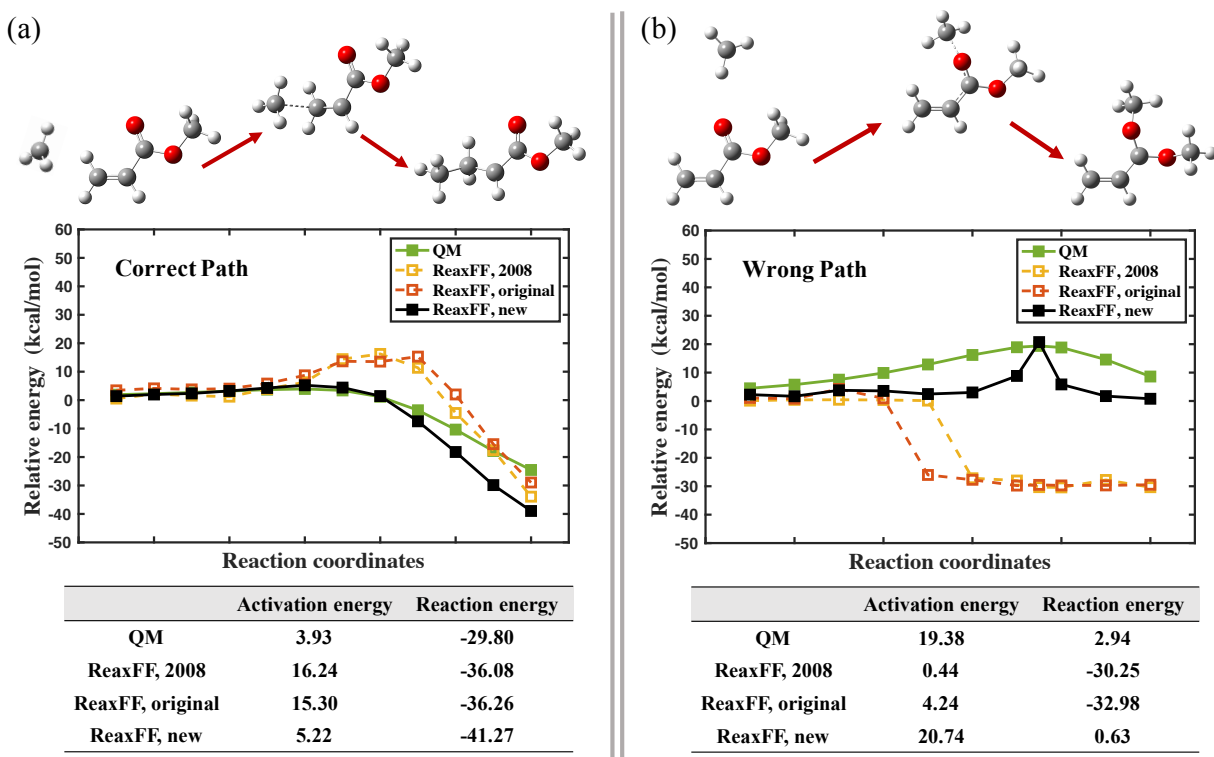


Figure 4.7: Intrinsic reaction coordinate from DFT vs ReaxFF with barrier and reaction energies. (a) Energies along the correct reaction path; (b) Energies along the wrong reaction path.

4.3.2 Liquid Monomers

Three different liquid acrylate monomers with different shapes, sizes, and number of functional groups were studied using the developed ff.C/H/O/N. Methyl acrylate (Figure 4.8a) is a short acrylate monomer with only one vinyl group, poly(ethylene glycol) diacrylate ($M_n=250$) (PEGDA) is a linear acrylate monomer with two vinyl group at each end (Figure 4.8b), and tris[2-(acryloyloxy)ethyl] isocyanurate (ICN triacrylate) is a branched acrylate monomer with three vinyl groups at the end of each branch (Figure 4.8c). Densities were calculated by running MD simulations of the three neat monomer systems with the developed ff.C/H/O/N at 300K and 1 atm. Three randomly generated initial configurations of each monomer were generated for the average

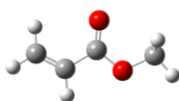
densities. Table 4.2 shows the ReaxFF calculated densities and the experimentally measured densities, and the errors were all smaller than 6%. The diffusion coefficients were calculated from the mean-square-displacement (MSD) of the monomers:

$$6Dt = \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle = MSD(t) \quad (2)$$

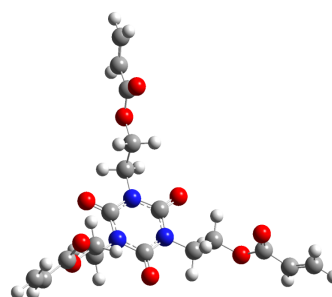
$$D = \frac{1}{6} \frac{d}{dt} MSD(t) \approx \frac{1}{6} [\text{slope of } MSD(t)] \quad (3)$$

and were also tabulated in table 4.2, which are close to the diffusion coefficients of short linear diacrylate HDDA, long linear diacrylate PEGDA, and branched triacrylate PETA calculated by an all-atom force field and a cutoff-based scheme by Karnes et al ^[26]. Methyl acrylate is the smallest species among the three and thus it diffuses the fastest. ICN-triacrylate is the largest one and it has three branches with three functional groups, and so it diffuses much slower than the two linear monomers.

(a) Methyl acrylate



(c) Tris[2-(acryloyloxy)ethyl] isocyanurate



(b) Poly(ethylene glycol) diacrylate (Mn=250)

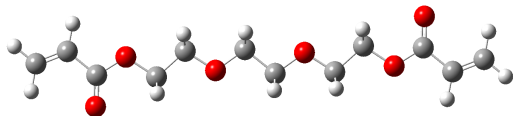


Figure 4.8: Chemical formulas and representations of (a) methyl acrylate, (b) PEGDA, and (c) ICN triacrylate.

	Density _{ReaxFF} (g/cc)	Density _{Expt.} (g/cc)	Error of Density	D _{ReaxFF} (cm ² /s)
Methyl acrylate	0.99±0.02	0.96	3.13%	3.7×10 ⁻⁶
PEGDA	1.11±0.01	1.12	0.89%	3.3×10 ⁻⁷
ICN-triacrylate	1.23±0.02	1.30	5.33%	4.7×10 ⁻⁸

Table 4.2: Densities and Diffusion coefficients of methyl acrylate, PEGD, and ICN-triacrylate.

Radial distribution functions $g(r)$ of the vinyl group in the three acrylates and the coordination $coord(r)$ were also calculated from LAMMPS for the three acrylates by equation 4 to investigate the local orderings of the three monomers and was plotted in figure 4.9:

$$g_{C_{v,A}-C_{v,B}} = \frac{1}{\eta_c} \left\langle \sum_{i=1}^N \delta(r - r_i) \right\rangle \quad (4)$$

where C_v is the end carbon in the vinyl group on the reference molecule A or the neighboring molecule B, η_c is a normalization constant to ensure that $g_{1-2}(r \rightarrow \infty) = 1$. The coordination $coord(r)$ were also calculated by multiplying the integral of $g(r)$ by the volume density of the end carbon in the vinyl group. Both $g(r)$ and $coord(r)$ were plotted in figure 8. The first maxima of $g(r)$ of all three monomers are at around 3.75 Å, and the second maxima of $g(r)$ of the three monomers are at around 7.25 Å, indicating two solvation shells. The peak heights of the first solvation shells are in the range from 1.3 – 1.5, smaller than highly polar molecular liquids like water, which is about 2.8^[27], meaning the local intramolecular order of the acrylates with different shapes and sizes is less than in highly polar liquids. The respective coordination number of vinyl group in the methyl acrylate, PEGDA, and ICN-triacrylate in the first solvation shell was 4.1, 3.4, and 2.1. The diffusive transport and steric access to the reactive site of methyl acrylate are greater

than the other two types of acrylate, and so faster free-radical polymerization was expected for methyl acrylate resin.

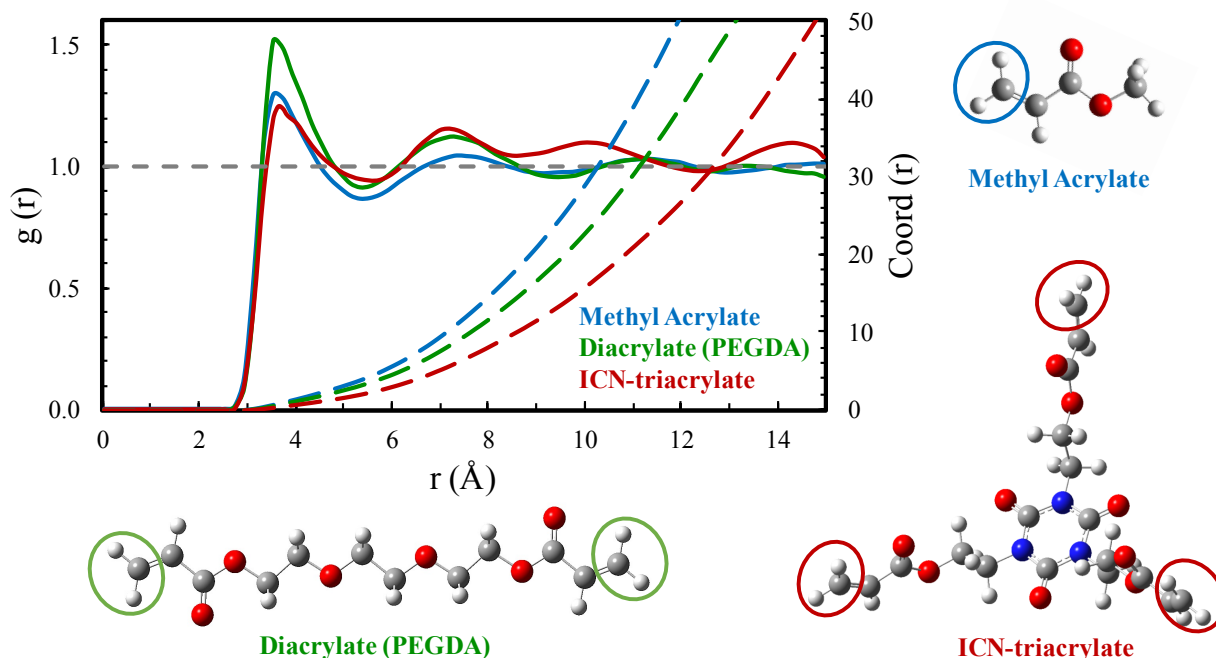


Figure 4.9: RDF of the vinyl group in acrylates. Solid lines are $g(r)$ of the three monomers, and dashed lines are $Coord(r)$ of the three monomers. The end carbons in the vinyl groups are circled for each monomer.

4.3.3 Free-radical Reactions

CH_3 radical was introduced to the monomer resin to start the polymerization process so that for every 40 monomers there was 1 CH_3 radical. 3 CH_4 was randomly distributed within a box of 120 methyl acrylates and the system was equilibrated at 300K and 1atm in NPT ensemble, and then temperature was ramped up to 700K and equilibrated again in NVT ensemble followed by ramping up pressure from 1 atm to 3000 atm in NPT ensemble. Higher temperature was used to speed up the reaction in limited simulation time, and the high pressure was chosen to prevent the system from evaporation and keep the density at the ambient value. The proton in each of the CH_4 was

then deleted to create 3 $\text{CH}_3\cdot$ radicals and run the reaction at 700K and 3000 atm in NPT ensemble for 20ns. All $\text{CH}_3\cdot$ radicals reacted with acrylate monomers and initiated the propagations with the first few hundreds of picoseconds, and two of the new radicals reacted with other radicals and terminated the free-radical propagation soon after. One of the radical from the first propagation kept reacting with its neighboring monomers until the radical from the third propagation formed, as seen in figure 4.10.

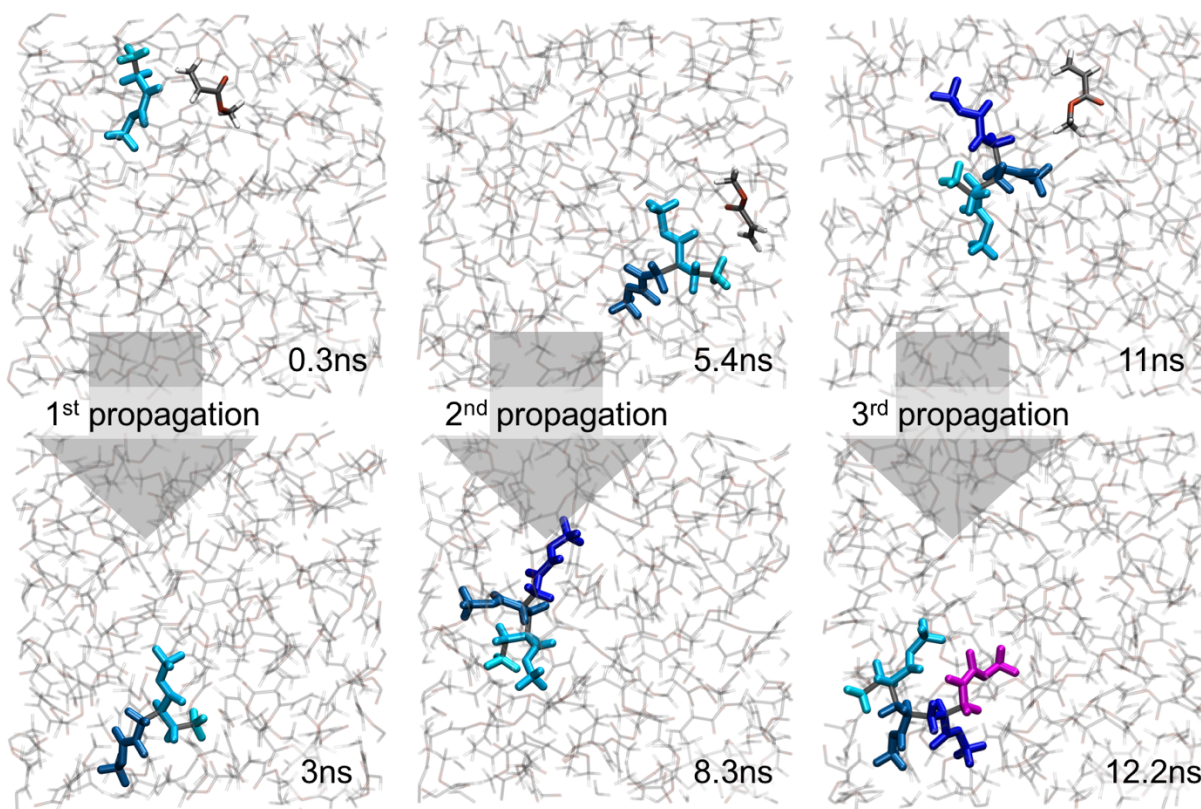


Figure 4.10: Snapshots of the free-radical polymerization started with methyl acrylate reacting with $\text{CH}_3\cdot$ radical. The reacting monomers and $\text{CH}_3\cdot$ radical that propagated until the third propagation were shown in colors.

Same MD simulations were also done for PEGDA and ICN-triacrylate, where 1 $\text{CH}_3\cdot$ radical and 40 acrylates were randomly distributed in the simulation box for each system, and ran at 700K and 3000atm in NPT ensemble. This kept the same $\text{CH}_3\cdot$ radical concentration with the methyl acrylate system while reducing the computational cost. Due to the differences in shape, size, diffusive transport and steric access to the reactive site, the PEGDA and ICN-triacrylate systems were expected to propagate slower than the methyl acrylate system. In 20ns, the radical initiated the reaction for both PEGDA and ICN-triacrylate systems, and the first propagation for both systems were seen as pictured in figure 4.11. Further propagations require longer simulation time, but since the local chemistry is the same as the first propagation, the model developed is capable of describing full polymerization given stronger computer power and longer simulation time.

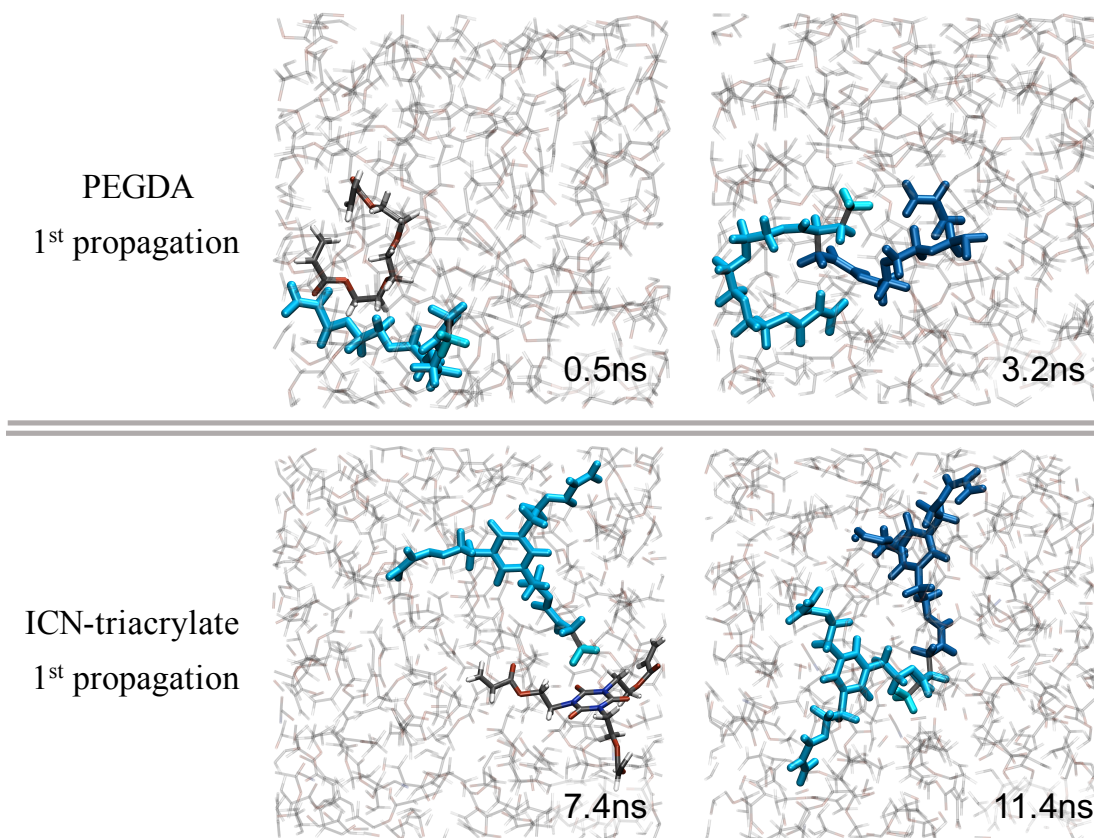


Figure 4.11: Snapshots of the free-radical polymerization started with $\text{CH}_3\cdot$ radical reacting with PEGDA (top) and ICN-triacrylate (bottom). The reacting monomers and $\text{CH}_3\cdot$ radical were shown in colors.

4.4 Conclusions

The developed ReaxFF force field ff.C/H/O/N reproduces atomic partial charges, geometries, and bond energies of the target acrylates very well. More importantly the reaction energy and barrier calculated from ReaxFF shows that the optimized force field can accurately describe the thermodynamics and kinetics of the acrylate free radical polymerization reaction, and it can be extended to other acrylates like PEGDA and ICN-triacrylate, with correct ReaxFF predicted densities and diffusion coefficients, and was able to analyze the diffusive transport and steric

access to the functional group and reactive site of different acrylates. Particularly, the thermodynamics and kinetics of the reactions that were not expected to happen were also included in the training process, and the parameters were trained against these reactions to prevent them from competing with the correct reactions, which was not reported in previous ReaxFF development works. This developed force field correctly described the initiation and the following propagations of the radical reacting with liquid acrylate monomers that have different size, shape, and number of functional groups. This enables the structure, mechanics, and reactivity analysis of the resin network after curing, and could help with the materials selection and resin design of the photo-initiated additive manufacturing including VAM, if having more computational resources. Future work will be done by running the simulations for longer time and compare the network structure and mechanical properties after curing among various acrylates.

Acknowledgements

This work was partially supported by Lawrence Livermore National Laboratory's Laboratory-Directed Research and Development (LDRD) funding, project 19-ERD-012 and partially performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

1. Wang, Martha O., et al. "Evaluating 3D-Printed biomaterials as scaffolds for vascularized bone tissue engineering." *Advanced Materials* 27.1 (2015): 138-144.
2. Rusling, James F. "Developing Microfluidic Sensing Devices Using 3D Printing." *ACS sensors* 3.3 (2018): 522-526.
3. Bagheri, Ali, and Jianyong Jin. "Photopolymerization in 3D printing." *ACS Applied Polymer Materials* 1.4 (2019): 593-611.
4. Zhang, Jing, and Pu Xiao. "3D printing of photopolymers." *Polymer Chemistry* 9.13 (2018): 1530-1540.
5. Quan, Haoyuan, et al. "Photo-curing 3D printing technique and its challenges." *Bioactive Materials* 5.1 (2020): 110-115.
6. Lee, T. Y., et al. "The effect of monomer structure on oxygen inhibition of (meth) acrylates photopolymerization." *Polymer* 45.18 (2004): 6155-6162.
7. Chenoweth, Kimberly, Adri CT Van Duin, and William A. Goddard. "ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation." *The Journal of Physical Chemistry A* 112.5 (2008): 1040-1053.
8. Kim, Sung-Yup, and Adri CT van Duin. "Simulation of titanium metal/titanium dioxide etching with chlorine and hydrogen chloride gases using the ReaxFF reactive force field." *The Journal of Physical Chemistry A* 117.27 (2013): 5655-5663.
9. Senftle, Thomas P., et al. "The ReaxFF reactive force-field: development, applications and future directions." *npj Computational Materials* 2.1 (2016): 1-14.
10. Van Duin, Adri CT, et al. "ReaxFF: a reactive force field for hydrocarbons." *The Journal of Physical Chemistry A* 105.41 (2001): 9396-9409.
11. Deetz, Joshua D., and Roland Faller. "Parallel optimization of a reactive force field for polycondensation of alkoxy silanes." *The Journal of Physical Chemistry B* 118.37 (2014): 10966-10978.
12. Frisch, M. J., et al. "Gaussian 16." (2016): 2016.
13. Becke, Axel D. "Becke's three parameter hybrid method using the LYP correlation functional." *J. Chem. Phys* 98 (1993): 5648-5652.
14. Lee, Ch, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti correlation energy formula into a functional of the electron density." *Phys. Rev. B* 37 (1988): 785-789.
15. Mulliken, Robert S. "Electronic population analysis on LCAO-MO molecular wave functions. I." *The Journal of Chemical Physics* 23.10 (1955): 1833-1840.
16. van Duin, Adri CT, Jan MA Baas, and Bastiaan Van De Graaf. "Delft molecular mechanics: a new approach to hydrocarbon force fields. Inclusion of a geometry-dependent charge calculation." *Journal of the Chemical Society, Faraday Transactions* 90.19 (1994): 2881-2895.
17. Iype, Eldhose, et al. "Parameterization of a reactive force field using a Monte Carlo algorithm." *Journal of computational chemistry* 34.13 (2013): 1143-1154.
18. Larsson, Henrik R., Adri CT van Duin, and Bernd Hartke. "Global optimization of parameters in the reactive force field ReaxFF for SiOH." *Journal of computational chemistry* 34.25 (2013): 2178-2189.

19. Hubin, Pierre O., et al. "Parameterization of the ReaxFF reactive force field for a proline-catalyzed aldol reaction." *Journal of computational chemistry* 37.29 (2016): 2564-2572.
20. Huang Y, Wexler AS, Bein KJ, Faller R. Development of a ReaxFF Force Field for Aqueous Phosphoenolpyruvate as a Novel Biomimetic Carbon Capture Absorbent. *The Journal of Physical Chemistry C*. 2022 May 26.
21. Plimpton, Steve, and Bruce Hendrickson. "Parallel molecular dynamics algorithms for simulation of molecular systems." 1995. 114-132.
22. Aktulga, Hasan Metin, et al. "Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques." *Parallel Computing* 38.4-5 (2012): 245-259.
23. Martínez, Leandro, et al. "PACKMOL: a package for building initial configurations for molecular dynamics simulations." *Journal of computational chemistry* 30.13 (2009): 2157-2164.
24. Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics*. 1984 Jul 1;81(1):511-9.
25. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*. 1985 Mar 1;31(3):1695.
26. Karnes JJ, Weisgraber TH, Oakdale JS, Mettry M, Shusteff M, Biener J. On the network topology of cross-linked acrylate photopolymers: a molecular dynamics case study. *The Journal of Physical Chemistry B*. 2020 Sep 22;124(41):9204-15.
27. Laury ML, Wang LP, Pande VS, Head-Gordon T, Ponder JW. Revised parameters for the AMOEBA polarizable atomic multipole water model. *The Journal of Physical Chemistry B*. 2015 Jul 23;119(29):9423-37.

4.5 Supporting Information 1: Training Set and Additional Training Results

Figure S1 shows the workflow of the optimization process. After the training set is generated, the atom types and coordinates were fed to LAMMPS to perform the MD simulations and the penalty function was calculated. For a parameter A_i with value a_i , the penalty function is evaluated for a_i , $a_i(1-\beta)$, and $a_i(1+\beta)$, where β is the scaling factor, the one that gives minimum penalty function was accepted. All parameters are independently evaluated and after each iteration, they are updated simultaneously. The scaling factor was first set to be 5%, and then gradually decreased for fine-tuning. The weights were assigned so that each section weighed similarly to the total error, but after the penalty function converged, the weights were tuned based on the importance level of each item in the training set, and continue the optimization procedure.

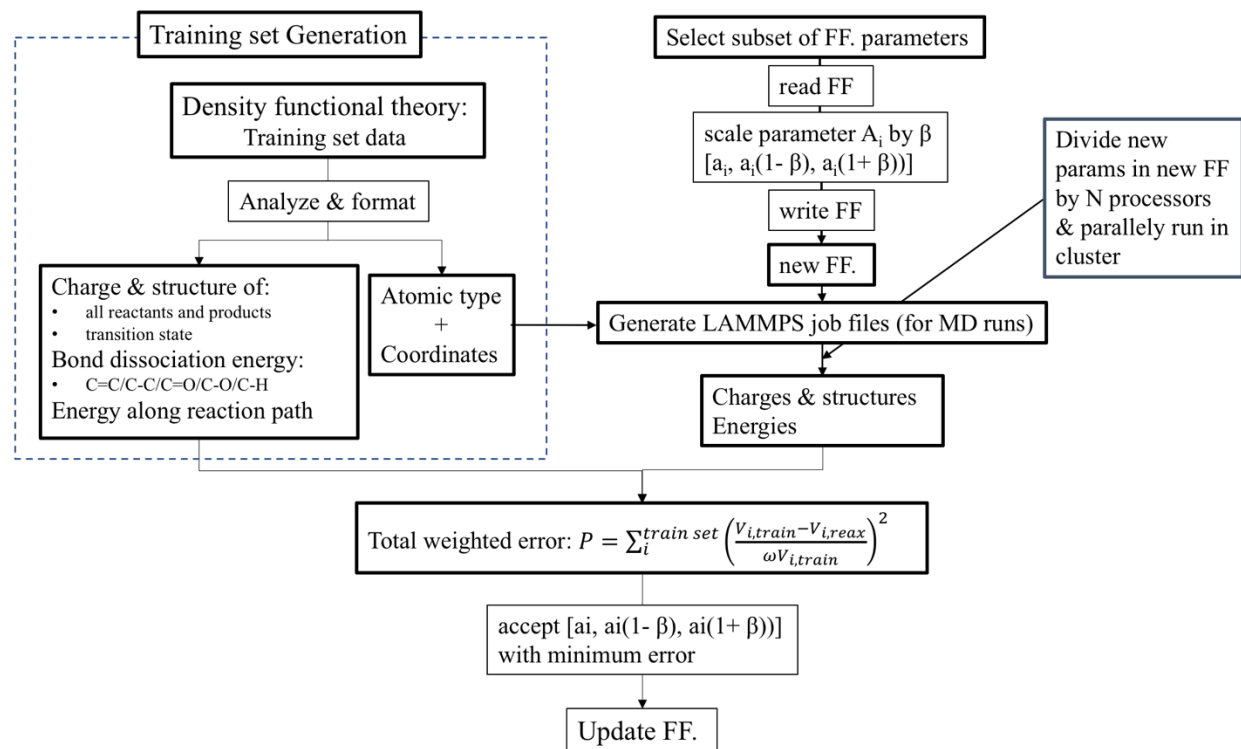


Figure S1: Workflow of the whole optimization process.

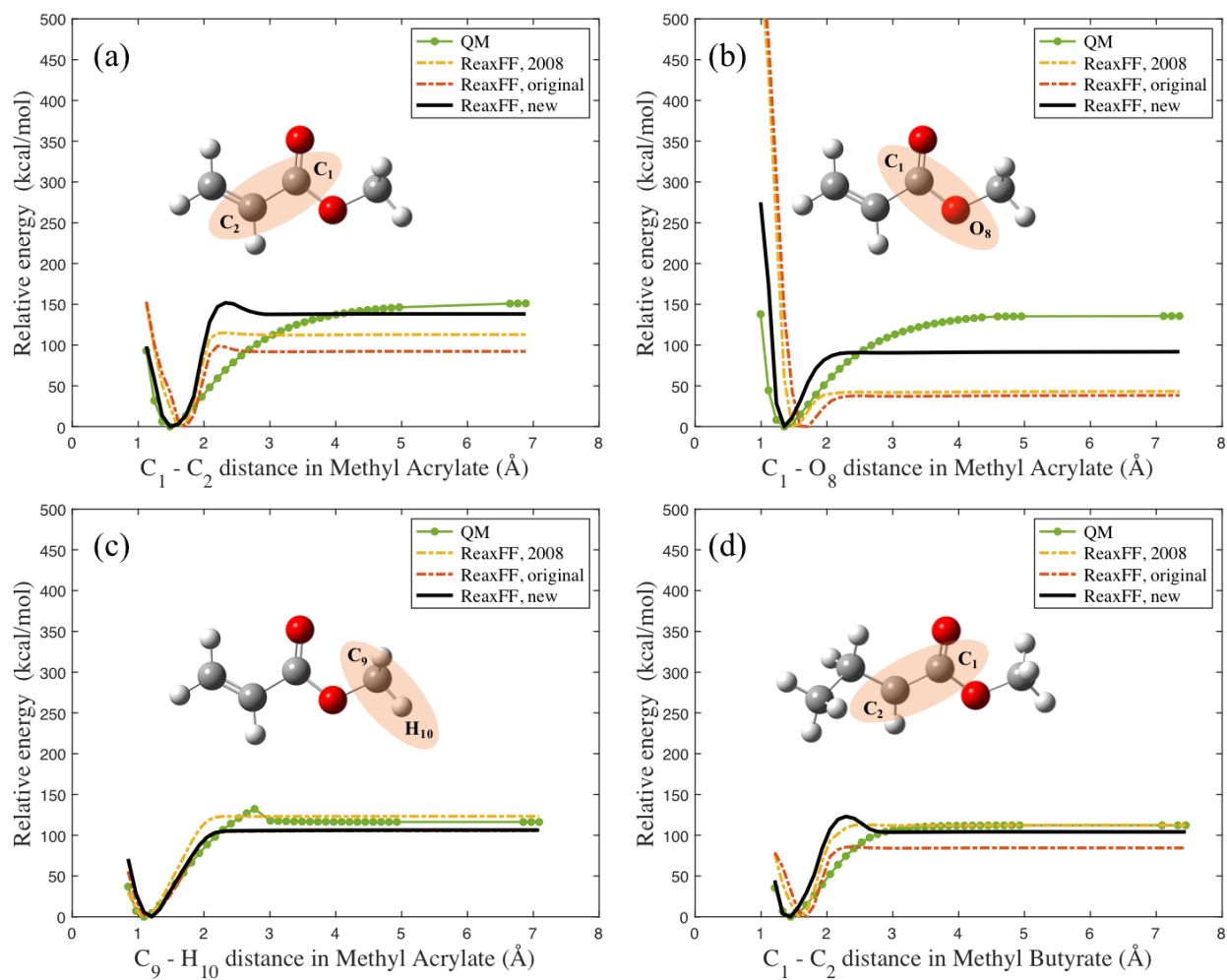


Figure S2: Bond dissociation energies of additional bonds.

4.6 Supporting Information 2: Final Force Field

Reactive MD-force field: C/H/O/N for acrylates photopolymerization

39 ! Number of general parameters
50.0000 !Overcoordination parameter
9.5469 !Overcoordination parameter
1.6725 !Valency angle conjugation parameter
1.7224 !Triple bond stabilisation parameter
6.8702 !Triple bond stabilisation parameter
60.4850 !C2-correction
1.0588 !Undercoordination parameter
4.6000 !Triple bond stabilisation parameter
12.1176 !Undercoordination parameter
13.3056 !Undercoordination parameter
-70.5044 !Triple bond stabilization energy
0.0000 !Lower Taper-radius
10.0000 !Upper Taper-radius
2.8793 !Not used
33.8667 !Valency undercoordination
6.0891 !Valency angle/lone pair parameter
1.0563 !Valency angle
2.0384 !Valency angle parameter
6.1431 !Not used
6.9290 !Double bond/angle parameter
0.3989 !Double bond/angle parameter: overcoord
3.9954 !Double bond/angle parameter: overcoord
-2.4837 !Not used
5.7796 !Torsion/BO parameter
10.0000 !Torsion overcoordination
1.9487 !Torsion overcoordination
-1.2327 !Conjugation 0 (not used)
2.1645 !Conjugation
1.5591 !vdWaals shielding
0.1000 !Cutoff for bond order (*100)
1.7602 !Valency angle conjugation parameter
0.6991 !Overcoordination parameter
50.0000 !Overcoordination parameter
1.8512 !Valency/lone pair parameter
0.5000 !Not used
20.0000 !Not used
5.0000 !Molecular energy (not used)
0.0000 !Molecular energy (not used)
0.7903 !Valency angle conjugation parameter
12 ! Nr of atoms; cov.r; valency;a.m;Rvdw;Evdw;gammaEEM;cov.r2;#
alfa;gammavdW;valency;Eunder;Eover;chiEEM;etaEEM;n.u.

cov r3;Elp;Heat inc.;n.u.;n.u.;n.u.;n.u.
 ov/un;vall;n.u.;val3,vval4

C	1.3955	4.0000	12.0000	1.7969	0.1837	0.9000	1.1172	4.0000
	10.2498	2.0812	4.0000	34.9131	79.5548	5.9666	7.0000	0.0000
	1.2235	0.0000	202.2908	9.4071	35.8021	12.8682	0.8563	0.0000
	-2.7534	2.3766	1.0564	4.0000	2.8904	0.0000	0.0000	0.0000
H	0.8975	1.0000	1.0080	1.3211	0.0977	0.8203	-0.1000	1.0000
	8.4286	31.6458	1.0000	0.0000	121.1250	3.7248	9.6093	1.0000
	-0.1000	0.0000	55.1878	3.1168	2.4182	0.0003	1.0698	0.0000
	-19.9435	4.1665	1.0338	1.0000	2.7371	0.0000	0.0000	0.0000
O	1.2264	2.0000	15.9990	2.3264	0.0999	1.0898	1.0548	6.0000
	9.4808	13.4988	4.0000	39.3984	116.0768	8.5000	8.3122	2.0000
	0.9049	0.4053	68.0152	3.3297	0.7821	0.0020	0.9745	0.0000
	-3.5477	2.7568	1.0493	4.0000	2.9956	0.0000	0.0000	0.0000
N	1.2333	3.0000	14.0000	1.9324	0.1376	0.7921	1.1748	5.0000
	10.0667	7.8431	4.0000	32.2482	100.0000	7.5795	6.3952	2.0000
	1.0433	27.4290	119.9837	1.9457	4.2874	3.4869	0.9745	0.0000
	-4.3875	2.6192	1.0183	4.0000	2.8793	0.0000	0.0000	0.0000
S	1.9405	2.0000	32.0600	2.0677	0.2099	1.0336	1.5479	6.0000
	9.9575	4.9055	4.0000	52.9998	112.1416	5.7824	8.2545	2.0000
	1.4601	9.7177	71.1843	5.7487	23.2859	12.7147	0.9745	0.0000
	-11.0000	2.7466	1.0338	6.2998	2.8793	0.0000	0.0000	0.0000
Mg	1.8315	2.0000	24.3050	2.2464	0.1806	0.5020	1.0000	2.0000
	10.9186	27.1205	3.0000	38.0000	0.0000	0.9499	5.6130	0.0000
	-1.3000	0.0000	127.9160	49.9248	0.3370	0.0000	0.0000	0.0000
	-1.0823	2.3663	1.0564	6.0000	2.9663	0.0000	0.0000	0.0000
P	1.5994	3.0000	30.9738	1.7000	0.1743	1.0000	1.3000	5.0000
	9.1909	14.9482	5.0000	0.0000	0.0000	1.6676	7.0946	0.0000
	-1.0000	25.0000	125.6300	0.2187	21.4305	15.1425	0.0000	0.0000
	-3.9294	3.4831	1.0338	5.0000	2.8793	0.0000	0.0000	0.0000
Na	1.8000	1.0000	22.9898	2.8270	0.1872	0.4000	-1.0000	1.0000
	10.0000	2.5000	1.0000	0.0000	0.0000	-0.9871	6.7728	0.0000
	-1.0000	0.0000	23.0445	100.0000	1.0000	0.0000	0.8563	0.0000
	-2.5000	3.9900	1.0338	8.0000	2.5791	0.0000	0.0000	0.0000
Ti	2.0254	4.0000	47.8800	2.2105	0.1574	0.6311	0.1000	4.0000
	12.7041	16.6482	4.0000	0.1000	0.0000	-1.3647	6.8406	0.0000
	-1.0000	0.0000	143.1770	27.6505	-0.0753	0.0064	0.8563	0.0000
	-15.0000	3.8359	1.0338	12.0000	2.2632	0.0000	0.0000	0.0000
Cl	1.7140	1.0000	35.4500	1.9139	0.2000	0.3500	-1.0000	7.0000
	11.5345	10.1330	1.0000	0.0000	0.0000	9.9704	6.1703	0.0000
	-1.0000	1.2769	143.1770	6.2293	5.2294	0.1542	0.8563	0.0000
	-10.2080	2.9867	1.0338	6.2998	2.5791	0.0000	0.0000	0.0000
F	1.2100	1.0000	18.9984	1.8601	0.1200	0.3000	-0.1000	7.0000
	11.5000	7.5000	4.0000	9.2533	0.2000	9.0000	15.0000	0.0000
	-1.0000	35.0000	1.5000	6.9821	4.1799	1.0561	0.0000	0.0000

-7.3000 2.6656 1.0493 4.0000 2.9225 0.0000 0.0000 0.0000
 X -0.1000 2.0000 1.0080 2.0000 0.0000 1.0000 -0.1000 6.0000
 10.0000 2.5000 4.0000 0.0000 0.0000 8.5000 1.5000 0.0000
 -0.1000 0.0000 -2.3700 8.7410 13.3640 0.6690 0.9745 0.0000
 -11.0000 2.7466 1.0338 4.0000 2.8793 0.0000 0.0000 0.0000
 58 ! Nr of bonds; Edis1;LPpen;n.u.;pbe1;pbo5;13corr;pbo6
 pbe2;pbo3;pbo4;Etrip;pbo1;pbo2;ovcorr
 1 1 149.9147 89.5187 77.8050 -0.7332 -0.4539 1.0000 35.7311 0.4147
 0.5061 -0.1000 9.5969 1.0000 -0.0701 6.3905 1.0000 0.0000
 1 2 169.0523 0.0000 0.0000 -0.6706 0.0000 1.0000 6.0000 0.7652
 5.4905 1.0000 0.0000 1.0000 -0.0474 6.5679 0.0000 0.0000
 2 2 153.3934 0.0000 0.0000 -0.4600 0.0000 1.0000 6.0000 0.7300
 6.2500 1.0000 0.0000 1.0000 -0.0790 6.0552 0.0000 0.0000
 1 3 140.9785 90.9238 67.0405 -0.3916 -0.2592 1.0000 11.5802 1.0000
 0.4934 -0.2648 7.3763 1.0000 -0.1428 5.3730 0.0000 0.0000
 3 3 142.2858 145.0000 50.8293 0.2506 -0.1000 1.0000 29.7503 0.6051
 0.3451 -0.1055 9.0000 1.0000 -0.1225 5.5000 1.0000 0.0000
 1 4 134.1215 140.2179 79.9745 0.0163 -0.1428 1.0000 27.0617 0.2000
 0.1387 -0.3681 7.1611 1.0000 -0.1000 5.0825 1.0000 0.0000
 3 4 130.8596 169.4551 40.0000 0.3837 -0.1639 1.0000 35.0000 0.2000
 1.0000 -0.3579 7.0004 1.0000 -0.1193 6.8773 1.0000 0.0000
 4 4 157.9384 82.5526 152.5336 0.4010 -0.1034 1.0000 12.4261 0.5828
 0.1578 -0.1509 11.9186 1.0000 -0.0861 5.4271 1.0000 0.0000
 2 3 160.0000 0.0000 0.0000 -0.5725 0.0000 1.0000 6.0000 0.5626
 1.1150 1.0000 0.0000 0.0000 -0.0920 4.2790 0.0000 0.0000
 2 4 185.3171 0.0000 0.0000 -0.3689 0.0000 1.0000 6.0000 0.2854
 7.6517 1.0000 0.0000 1.0000 -0.0408 6.0255 0.0000 0.0000
 1 5 128.7959 56.4134 39.0716 0.0688 -0.4463 1.0000 31.1766 0.4530
 0.1955 -0.3587 6.2148 1.0000 -0.0770 6.6386 1.0000 0.0000
 2 5 136.1049 0.0000 0.0000 -0.4669 0.0000 1.0000 6.0000 0.3803
 10.5730 1.0000 0.0000 1.0000 -0.1000 7.0000 1.0000 0.0000
 3 5 135.6998 220.0000 40.0000 0.5848 -0.2406 1.0000 22.1005 0.2335
 0.7069 -0.2681 8.3465 1.0000 -0.0922 5.4651 1.0000 0.0000
 4 5 0.0000 0.0000 0.0000 0.5000 -0.2000 1.0000 40.0000 0.3000
 0.4000 -0.2500 9.0000 1.0000 -0.1000 6.0000 1.0000 0.0000
 5 5 0.0000 0.0000 0.0000 0.2500 -0.5000 1.0000 20.0000 1.0000
 0.2500 -0.2500 10.0000 1.0000 -0.1000 8.0000 1.0000 0.0000
 2 6 58.6896 0.0000 0.0000 -0.0203 -0.1418 1.0000 13.1260 0.0230
 8.2136 -0.1310 0.0000 1.0000 -0.2692 6.4254 0.0000 24.4461
 3 6 87.0227 0.0000 43.3991 0.0030 -0.3000 1.0000 36.0000 0.0250
 0.0087 -0.2500 12.0000 1.0000 -0.0439 6.6073 1.0000 24.4461
 6 6 32.3808 0.0000 0.0000 -0.0076 -0.2000 0.0000 16.0000 0.2641
 4.8726 -0.2000 10.0000 1.0000 -0.0729 4.6319 0.0000 0.0000
 1 7 110.0000 92.0000 0.0000 0.2171 -0.1418 1.0000 13.1260 0.6000
 0.3601 -0.1310 10.7257 1.0000 -0.0869 5.3302 1.0000 0.0000

2	7	0.1466	0.0000	0.0000	0.2250	-0.1418	1.0000	13.1260	0.6000
		0.3912	-0.1310	0.0000	1.0000	-0.1029	9.3302	0.0000	0.0000
3	7	201.0058	194.1410	0.0000	1.0000	-0.5000	1.0000	25.0000	0.4873
		0.4358	-0.1571	15.8745	1.0000	-0.2431	6.3823	1.0000	0.0000
4	7	130.0000	0.0000	0.0000	0.2171	-0.1418	1.0000	13.1260	0.6000
		0.3601	-0.1310	10.7257	1.0000	-0.0869	5.3302	1.0000	0.0000
6	7	0.1000	0.0000	0.0000	0.2500	-0.5000	1.0000	35.0000	0.6000
		0.5000	-0.5000	20.0000	1.0000	-0.2000	10.0000	1.0000	0.0000
7	7	0.0000	0.0000	0.0000	0.2171	-0.5000	1.0000	35.0000	0.6000
		0.5000	-0.5000	20.0000	1.0000	-0.2000	10.0000	1.0000	0.0000
1	8	0.0000	0.0000	0.0000	-1.0000	-0.3000	1.0000	36.0000	0.7000
		10.1151	-0.3500	25.0000	1.0000	-0.1053	8.2003	1.0000	0.0000
2	8	0.0000	0.0000	0.0000	-1.0000	-0.3000	1.0000	36.0000	0.7000
		10.1151	-0.3500	25.0000	1.0000	-0.1053	8.2003	1.0000	0.0000
3	8	45.8933	0.0000	0.0000	-0.1511	-0.3000	1.0000	36.0000	0.3105
		5.8448	-0.3500	25.0000	1.0000	-0.0659	7.9140	1.0000	0.0000
4	8	0.0000	0.0000	0.0000	-1.0000	-0.3000	1.0000	36.0000	0.7000
		10.1151	-0.3500	25.0000	1.0000	-0.1053	8.2003	1.0000	0.0000
5	8	0.0000	0.0000	0.0000	-1.0000	-0.3000	1.0000	36.0000	0.7000
		10.1151	-0.3500	25.0000	1.0000	-0.1053	8.2003	1.0000	0.0000
6	8	0.0000	0.0000	0.0000	0.2500	-0.5000	1.0000	35.0000	0.6000
		0.5000	-0.5000	20.0000	1.0000	-0.2000	10.0000	1.0000	0.0000
7	8	0.0000	0.0000	0.0000	0.2500	-0.5000	1.0000	35.0000	0.6000
		0.5000	-0.5000	20.0000	1.0000	-0.2000	10.0000	1.0000	0.0000
8	8	64.4508	0.0000	0.0000	-0.3738	0.3000	0.0000	25.0000	0.2158
		0.9915	-0.4000	12.0000	1.0000	-0.0515	5.0000	0.0000	0.0000
4	6	50.0000	10.0901	0.0000	-1.0000	-0.3000	1.0000	36.0000	0.7058
		0.8567	-0.3487	17.4990	1.0000	-0.0794	8.2232	1.0000	0.0000
1	9	0.0000	0.0000	0.0000	-0.2872	-0.3000	1.0000	36.0000	0.0082
		1.7973	-0.2500	20.0000	1.0000	-0.2578	6.5219	1.0000	0.0000
2	9	0.0000	0.0000	0.0000	-0.2872	-0.3000	1.0000	36.0000	0.0082
		1.7973	-0.2500	20.0000	1.0000	-0.2578	6.5219	1.0000	0.0000
3	9	130.5629	37.6984	0.0000	0.9228	-0.3000	0.0000	36.0000	0.0850
		0.1150	-0.2818	16.1571	1.0000	-0.1343	6.8264	0.0000	0.0000
4	9	0.0000	0.0000	0.0000	-0.2872	-0.3000	1.0000	36.0000	0.0082
		1.7973	-0.2500	20.0000	1.0000	-0.2578	6.5219	1.0000	0.0000
5	9	0.0000	0.0000	0.0000	-0.2872	-0.3000	1.0000	36.0000	0.0082
		1.7973	-0.2500	20.0000	1.0000	-0.2578	6.5219	1.0000	0.0000
6	9	0.0000	0.0000	0.0000	-0.2872	-0.3000	1.0000	36.0000	0.0082
		1.7973	-0.2500	20.0000	1.0000	-0.2578	6.5219	1.0000	0.0000
7	9	0.0000	0.0000	0.0000	-0.2872	-0.3000	1.0000	36.0000	0.0082
		1.7973	-0.2500	20.0000	1.0000	-0.2578	6.5219	1.0000	0.0000
8	9	0.1000	0.0000	0.0000	0.2500	-0.5000	1.0000	35.0000	0.6000
		0.5000	-0.5000	20.0000	1.0000	-0.2000	10.0000	1.0000	0.0000
9	9	80.1930	0.0000	0.0000	-0.8469	-0.2000	0.0000	16.0000	0.2022

		0.7528	-0.1924	14.9725	1.0000	-0.0885	5.0000	0.0000	0.0000
1	10	0.0000	0.0000	0.0000	0.5000	-0.2000	0.0000	16.0000	0.5000
		1.0001	-0.2000	15.0000	1.0000	-0.1000	10.0000	0.0000	0.0000
2	10	98.9788	0.0000	0.0000	-0.0572	-0.2000	0.0000	16.0000	1.1523
		2.2822	-0.2000	15.0000	1.0000	-0.1093	5.1686	0.0000	0.0000
3	10	0.0000	0.0000	0.0000	0.5000	-0.2000	0.0000	16.0000	0.5000
		1.0001	-0.2000	15.0000	1.0000	-0.1000	10.0000	0.0000	0.0000
4	10	0.0000	0.0000	0.0000	0.5000	-0.2000	0.0000	16.0000	0.5000
		1.0001	-0.2000	15.0000	1.0000	-0.1000	10.0000	0.0000	0.0000
5	10	0.0000	0.0000	0.0000	0.5000	-0.2000	0.0000	16.0000	0.5000
		1.0001	-0.2000	15.0000	1.0000	-0.1000	10.0000	0.0000	0.0000
6	10	0.0000	0.0000	0.0000	0.5000	-0.2000	0.0000	16.0000	0.5000
		1.0001	-0.2000	15.0000	1.0000	-0.1000	10.0000	0.0000	0.0000
7	10	0.0000	0.0000	0.0000	0.5000	-0.2000	0.0000	16.0000	0.5000
		1.0001	-0.2000	15.0000	1.0000	-0.1000	10.0000	0.0000	0.0000
8	10	0.0000	0.0000	0.0000	0.5000	-0.2000	0.0000	16.0000	0.5000
		1.0001	-0.2000	15.0000	1.0000	-0.1000	10.0000	0.0000	0.0000
9	10	166.7325	0.0000	0.0000	-0.6307	-0.2000	0.0000	16.0000	0.5805
		0.6328	-0.2000	15.0000	1.0000	-0.0608	8.5378	0.0000	0.0000
10	10	0.2500	0.0000	0.0000	0.1803	-0.2000	0.0000	16.0000	0.3356
		0.9228	-0.2000	15.0000	1.0000	-0.1178	5.6715	0.0000	0.0000
1	11	237.8781	0.0000	0.0000	-0.7438	-0.5000	1.0000	35.0000	1.0460
		3.6661	-0.2500	15.0000	1.0000	-0.0800	5.4719	1.0000	0.0000
2	11	0.0000	0.0000	0.0000	-0.4643	0.0000	1.0000	6.0000	0.6151
		12.3710	1.0000	0.0000	1.0000	-0.1008	8.5980	0.0000	0.0000
3	11	0.0000	0.0000	0.0000	-0.4643	0.0000	1.0000	6.0000	0.6151
		12.3710	1.0000	0.0000	1.0000	-0.1008	8.5980	0.0000	0.0000
4	11	0.0000	0.0000	0.0000	-0.4643	0.0000	1.0000	6.0000	0.6151
		12.3710	1.0000	0.0000	1.0000	-0.1008	8.5980	0.0000	0.0000
5	11	0.0000	0.0000	0.0000	-0.4643	0.0000	1.0000	6.0000	0.6151
		12.3710	1.0000	0.0000	1.0000	-0.1008	8.5980	0.0000	0.0000
11	11	250.0765	0.0000	0.0000	0.2298	-0.3500	1.0000	25.0000	0.8427
		0.1167	-0.2500	15.0000	1.0000	-0.1506	7.3516	1.0000	0.0000
23	! Nr of off-diagonal terms; Ediss;Ro;gamma;rsigma;rpi;rpi2								
1	2	0.1236	1.3271	10.3132	1.1181	-1.0000	-1.0000		
2	3	0.0256	1.2241	10.8918	1.0668	-1.0000	-1.0000		
2	4	0.0687	1.5130	10.0094	0.9412	-1.0000	-1.0000		
1	3	0.1153	1.7413	8.8197	1.2162	0.9925	1.0637		
1	4	0.1447	1.8766	9.7990	1.3436	1.1885	1.1363		
3	4	0.1048	2.0003	10.1220	1.3173	1.1096	1.0206		
1	5	0.1408	1.8161	9.9393	1.7986	1.3021	1.4031		
2	5	0.0895	1.6239	10.0104	1.4640	-1.0000	-1.0000		
3	5	0.1962	1.7872	10.2319	1.4622	1.4025	-1.0000		
4	5	0.1505	1.9000	10.5104	1.8000	1.4000	-1.0000		
2	6	0.0100	1.6000	13.2979	1.8670	-1.0000	-1.0000		

3	6	0.0809	1.7000	11.4606	1.5177	-1.0000	-1.0000		
3	7	0.0534	1.7520	10.4281	1.8000	1.4498	-1.0000		
6	7	0.1801	1.8566	9.8498	0.1000	-1.0000	-1.0000		
3	8	0.0825	1.5904	11.3396	1.5905	-1.0000	-1.0000		
2	9	0.1750	1.7939	13.5000	0.0100	-1.0000	-1.0000		
3	9	0.1200	1.8000	10.5000	1.6526	1.4718	-1.0000		
9	10	0.1819	2.2000	12.2652	1.8478	1.4718	-1.0000		
1	9	0.2035	1.7007	11.4530	0.0100	-1.0000	-1.0000		
2	10	0.0376	1.6671	9.6285	1.2123	-1.0000	-1.0000		
3	10	0.1945	2.2766	11.2353	-1.0000	-1.0000	-1.0000		
1	11	0.1071	1.6243	11.0402	1.3176	-1.0000	-1.0000		
2	11	0.0431	1.7204	10.3632	0.5386	-1.0000	-1.0000		
85		! Nr of angles;at1;at2;at3;Thetao,o;ka;kb;pv1;pv2							
1	1	1	59.0573	30.7029	0.7606	0.0000	0.7180	6.2933	1.1244
1	1	2	65.7758	14.5234	6.2481	0.0000	0.5665	0.0000	1.6255
2	1	2	70.2607	25.2202	3.7312	0.0000	0.0050	0.0000	2.7500
1	2	2	0.0000	0.0000	6.0000	0.0000	0.0000	0.0000	1.0400
1	2	1	0.0000	3.4110	7.7350	0.0000	0.0000	0.0000	1.0400
2	2	2	0.0000	27.9213	5.8635	0.0000	0.0000	0.0000	1.0400
1	1	3	53.9517	7.8968	2.6122	0.0000	3.0000	58.6562	1.0338
3	1	3	76.9627	44.2852	2.4177	-25.3063	1.6334	-50.0000	2.7392
1	1	4	78.5538	21.4381	7.4715	0.0000	1.1046	50.0000	1.5275
3	1	4	73.9544	12.4661	7.0000	0.0000	1.1046	0.0000	1.1880
4	1	4	89.3168	20.2660	7.5000	0.0000	1.1046	0.0000	1.5403
2	1	3	65.0000	16.3141	5.2730	0.0000	0.4448	0.0000	1.4077
2	1	4	74.2929	31.0883	2.6184	0.0000	0.1000	0.0000	1.0500
1	2	4	0.0000	0.0019	6.3000	0.0000	0.0000	0.0000	1.0400
1	3	1	72.6199	42.5510	0.7205	0.0000	2.9294	0.0000	1.3096
1	3	3	81.9029	32.2258	1.7397	0.0000	0.9888	68.1072	1.7777
1	3	4	82.4890	31.4554	0.9953	0.0000	3.0000	0.0000	1.0783
3	3	3	80.7324	30.4554	0.9953	0.0000	3.0000	50.0000	1.0783
3	3	4	84.3637	31.4554	0.9953	0.0000	3.0000	0.0000	1.0783
4	3	4	89.7071	31.4554	0.9953	0.0000	3.0000	0.0000	1.1519
1	3	2	70.1101	13.1217	4.4734	0.0000	0.8433	0.0000	3.0000
2	3	3	75.6935	50.0000	2.0000	0.0000	1.0000	0.0000	1.1680
2	3	4	75.6201	18.7919	0.9833	0.0000	0.1000	0.0000	1.0500
2	3	2	85.8000	9.8453	2.2720	0.0000	2.8635	0.0000	1.5800
1	4	1	81.4699	7.2318	1.2608	0.0000	3.0000	0.0000	1.2127
1	4	3	103.3204	33.0381	0.5787	0.0000	3.0000	0.0000	1.2127
1	4	4	50.0000	25.0250	4.7651	0.0000	3.0000	0.0000	1.2028
3	4	3	74.1978	42.1786	1.7845	-18.0069	3.0000	0.0000	1.2127
3	4	4	74.8600	43.7354	1.1572	-0.9193	3.0000	0.0000	1.2127
4	4	4	75.0538	14.8267	5.2794	0.0000	3.0000	0.0000	1.2127
1	4	2	68.2294	29.6576	1.0533	0.0000	0.3481	0.0000	1.5443
2	4	3	81.3686	40.0712	2.2396	0.0000	0.3481	0.0000	1.5443

2	4	4	83.0104	43.4766	1.5328	0.0000	0.3481	0.0000	1.5443
2	4	2	79.6336	17.7917	3.7832	0.0000	0.0222	0.0000	2.0238
1	2	3	0.0000	25.0000	3.0000	0.0000	1.0000	0.0000	1.0400
1	2	4	0.0000	0.0019	6.0000	0.0000	0.0000	0.0000	1.0400
1	2	5	0.0000	0.0019	6.0000	0.0000	0.0000	0.0000	1.0400
3	2	3	0.0000	15.0000	2.8900	0.0000	0.0000	0.0000	2.8774
3	2	4	0.0000	0.0019	6.0000	0.0000	0.0000	0.0000	1.0400
4	2	4	0.0000	0.0019	6.0000	0.0000	0.0000	0.0000	1.0400
2	2	3	0.0000	8.5744	3.0000	0.0000	0.0000	0.0000	1.0421
2	2	4	0.0000	0.0019	6.0000	0.0000	0.0000	0.0000	1.0400
1	1	5	74.4180	33.4273	1.7018	0.1463	0.5000	0.0000	1.6178
1	5	1	79.7037	28.2036	1.7073	0.1463	0.5000	0.0000	1.6453
2	1	5	63.3289	29.4225	2.1326	0.0000	0.5000	0.0000	3.0000
1	5	2	85.9449	38.3109	1.2492	0.0000	0.5000	0.0000	1.1000
1	5	5	85.6645	40.0000	2.9274	0.1463	0.5000	0.0000	1.3830
2	5	2	83.8555	5.1317	0.4377	0.0000	0.5000	0.0000	3.0000
2	5	5	97.0064	32.1121	2.0242	0.0000	0.5000	0.0000	2.8568
3	5	3	81.0926	30.2268	6.4132	-5.4471	2.5968	0.0000	3.0000
1	5	3	70.0000	35.0000	3.4223	0.0000	1.3550	0.0000	1.2002
1	3	5	57.3353	41.0012	1.0609	0.0000	1.3000	0.0000	3.0000
3	3	5	83.9753	31.0715	3.5590	0.0000	0.8161	0.0000	1.1776
2	3	5	89.8843	17.5000	3.3660	0.0000	2.0000	0.0000	2.0734
2	6	2	0.0000	49.8261	0.2093	0.0000	2.0870	0.0000	2.2895
2	2	6	0.0000	40.0366	3.1505	0.0000	1.1296	0.0000	1.1110
6	2	6	0.0000	0.5047	0.8000	0.0000	0.8933	0.0000	4.6650
2	6	6	0.0000	8.7037	0.0827	0.0000	3.5597	0.0000	1.1198
3	6	3	0.0000	9.2317	0.1000	0.0000	1.0000	0.0000	1.0920
6	3	6	0.0008	25.0000	8.0000	0.0000	1.0000	0.0000	3.0000
2	3	6	66.0423	5.0000	1.0000	0.0000	1.0000	0.0000	1.2500
2	6	3	0.0000	0.5000	0.1000	0.0000	1.0000	0.0000	3.0000
3	3	6	70.0000	20.0000	1.0000	0.0000	1.0000	0.0000	1.2500
3	7	3	90.0000	18.4167	0.6799	-8.0000	0.1310	0.0000	2.2321
2	3	7	72.6004	9.6150	0.8905	0.0000	3.5473	0.0000	1.0400
3	3	7	60.0000	40.0000	4.0000	0.0000	1.0000	0.0000	1.0400
3	2	7	0.0000	10.0000	1.0000	0.0000	1.0000	0.0000	1.0400
6	3	7	41.0995	3.2207	7.3523	0.0000	0.1101	0.0000	1.0947
7	3	7	62.1312	7.5931	0.1000	0.0000	0.5154	0.0000	2.1744
1	3	7	74.1394	8.5687	1.7132	0.0000	-0.6553	0.0000	2.2323
2	7	3	75.0000	25.0000	2.0000	0.0000	1.0000	0.0000	1.2500
3	7	7	70.0000	25.0000	2.0000	0.0000	1.0000	0.0000	1.2500
3	9	3	90.0000	30.4624	2.1468	0.0000	0.0500	0.0000	1.9485
9	3	9	90.0000	5.7486	5.0000	0.0000	2.0000	0.0000	1.1000
10	9	10	70.9022	29.8228	3.9292	0.0000	0.1917	0.0000	2.0357
9	10	9	48.6200	40.0000	5.0000	0.0000	0.0500	0.0000	1.1000
3	3	9	62.9344	15.0215	4.3743	0.0000	0.6168	0.0000	1.1673

3	9	9	33.7127	8.0623	3.4580	0.0000	0.0500	0.0000	2.6065	
2	3	9	90.0000	9.7766	8.0000	0.0000	0.0505	0.0000	1.7257	
1	3	9	90.0000	7.6430	0.2869	0.0000	1.8647	0.0000	4.0000	
3	2	10	0.0000	0.0100	0.0100	0.0000	0.0000	0.0000	1.1456	
11	1	11	77.8443	49.0744	5.9913	0.0000	0.7835	0.0000	2.3020	
1	11	1	0.0000	19.9962	3.2299	0.0000	2.1012	0.0000	1.1537	
1	11	11	0.0000	25.0000	1.0000	0.0000	1.0000	0.0000	1.0400	
11	1	2	69.6421	10.0000	2.0000	0.0000	1.0000	0.0000	1.0400	
57	! Nr of torsions;at1;at2;at3;at4;;V1;V2;V3;V2(BO);vconj;n.u;n									
1	1	1	1	-0.2500	34.7453	0.0288	-6.3507	-1.6000	0.0000	0.0000
1	1	1	2	-0.2500	29.2131	0.2945	-4.9581	-2.1802	0.0000	0.0000
2	1	1	2	-0.2500	31.2081	0.4539	-4.8923	-2.2677	0.0000	0.0000
1	1	1	3	1.2799	20.7787	-0.5249	-2.5000	-1.0000	0.0000	0.0000
2	1	1	3	1.9159	19.8113	0.7914	-4.6995	-1.0000	0.0000	0.0000
3	1	1	3	-1.4477	16.6853	0.6461	-4.9622	-1.0000	0.0000	0.0000
1	1	3	1	0.4816	19.6316	-0.0057	-2.5000	-1.0000	0.0000	0.0000
1	1	3	2	1.2044	80.0000	-0.3139	-6.1481	-1.0000	0.0000	0.0000
2	1	3	1	-2.5000	31.0191	0.6165	-2.7733	-2.9807	0.0000	0.0000
2	1	3	2	-2.4875	70.8145	0.7582	-4.2274	-3.0000	0.0000	0.0000
1	1	3	3	-0.3566	10.0000	0.0816	-2.6110	-1.9631	0.0000	0.0000
2	1	3	3	-1.4383	80.0000	1.0000	-3.6877	-2.8000	0.0000	0.0000
3	1	3	1	-1.1390	78.0747	-0.0964	-4.5172	-3.0000	0.0000	0.0000
3	1	3	2	-2.5000	70.3345	-1.0000	-5.5315	-3.0000	0.0000	0.0000
3	1	3	3	-2.0234	80.0000	0.1684	-3.1568	-2.6174	0.0000	0.0000
1	3	3	1	1.1637	-17.3637	0.5459	-3.6005	-2.6938	0.0000	0.0000
1	3	3	2	-2.1289	12.8382	1.0000	-5.6657	-2.9759	0.0000	0.0000
2	3	3	2	2.5000	-22.9397	0.6991	-3.3961	-1.0000	0.0000	0.0000
1	3	3	3	2.5000	-25.0000	1.0000	-2.5000	-1.0000	0.0000	0.0000
2	3	3	3	-2.5000	-2.5103	-1.0000	-2.5000	-1.0000	0.0000	0.0000
3	3	3	3	-2.5000	-25.0000	1.0000	-2.5000	-1.0000	0.0000	0.0000
0	1	2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	2	2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	2	3	0	0.0000	0.1000	0.0200	-2.5415	0.0000	0.0000	0.0000
0	1	1	0	0.0000	50.0000	0.3000	-4.0000	-2.0000	0.0000	0.0000
0	1	4	0	1.7932	141.5515	0.9686	-4.2368	-1.9727	0.0000	0.0000
0	2	4	0	-1.5000	0.1032	0.0100	-5.0965	0.0000	0.0000	0.0000
0	3	4	0	1.1397	61.3225	0.5139	-3.8507	-3.0000	0.0000	0.0000
0	4	4	0	0.7265	44.3155	1.0000	-4.4046	-2.0000	0.0000	0.0000
4	1	4	4	-0.0949	8.7582	0.3310	-7.9430	-2.0000	0.0000	0.0000
0	1	5	0	3.3423	30.3435	0.0365	-2.7171	0.0000	0.0000	0.0000
0	5	5	0	-0.0555	-5.0000	0.1515	-2.2056	0.0000	0.0000	0.0000
0	2	5	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	3	5	3	2.5000	2.5000	0.2237	-10.0000	-1.0000	0.0000	0.0000
0	3	5	0	-2.5000	50.0000	-0.5000	-10.0000	-1.0000	0.0000	0.0000
0	6	6	0	0.0000	0.0000	0.1200	-2.4426	0.0000	0.0000	0.0000

0	2	6	0	0.0000	0.0000	0.1200	-2.4847	0.0000	0.0000	0.0000
0	3	6	0	0.0000	0.0000	0.1200	-2.4703	0.0000	0.0000	0.0000
1	1	3	3	-0.0002	20.1851	0.1601	-9.0000	-2.0000	0.0000	0.0000
1	3	3	1	0.0002	80.0000	-1.5000	-4.4848	-2.0000	0.0000	0.0000
3	1	3	3	-0.1583	20.0000	1.5000	-9.0000	-2.0000	0.0000	0.0000
1	1	1	7	-0.3232	14.3871	0.1823	-9.8682	-1.7255	0.0000	0.0000
7	1	1	7	-0.1452	50.0000	-0.1915	-8.0773	-1.7255	0.0000	0.0000
0	1	7	0	4.0000	45.8264	0.9000	-4.0000	0.0000	0.0000	0.0000
0	7	7	0	4.0000	45.8264	0.9000	-4.0000	0.0000	0.0000	0.0000
2	1	3	7	-1.5000	18.9285	0.3649	-6.1208	0.0000	0.0000	0.0000
2	3	7	3	1.5000	-1.0000	0.2575	-6.2100	0.0000	0.0000	0.0000
1	3	7	3	-1.4375	-0.8700	0.9861	-2.5424	0.0000	0.0000	0.0000
7	3	7	3	-1.5000	21.5086	-1.0000	-4.8869	0.0000	0.0000	0.0000
2	1	3	9	0.0000	84.3556	0.1000	-3.1953	0.0000	0.0000	0.0000
1	1	3	9	0.0000	51.0461	0.1059	-7.2043	0.0000	0.0000	0.0000
2	3	9	3	-0.2500	0.0100	-0.5000	-4.6984	0.0000	0.0000	0.0000
1	1	1	11	0.5000	0.1000	0.4683	-11.5274	-1.7255	0.0000	0.0000
2	1	1	11	0.0000	49.3871	0.2000	-10.5765	-1.7255	0.0000	0.0000
11	1	1	11	-0.5000	95.4727	-0.2080	-4.8579	-1.7255	0.0000	0.0000
0	1	11	0	4.0000	45.8264	0.9000	-4.0000	0.0000	0.0000	0.0000
0	11	11	0	4.0000	45.8264	0.8897	-4.0000	0.0000	0.0000	0.0000
9	! Nr of hydrogen bonds;at1;at2;at3;Rhb;Dehb;vhb1									
3	2	3		2.1200	-3.5800	1.4500	19.5000			
3	2	4		1.6787	-3.9601	1.4500	19.5000			
4	2	3		1.5585	-3.9305	1.4500	19.5000			
4	2	4		1.9336	-5.8831	1.4500	19.5000			
3	2	5		1.5000	-2.0000	1.4500	19.5000			
4	2	5		1.5000	-2.0000	1.4500	19.5000			
5	2	3		1.5000	-2.0000	1.4500	19.5000			
5	2	4		1.5000	-2.0000	1.4500	19.5000			
5	2	5		1.5000	-2.0000	1.4500	19.5000			

Chapter 5

Development and simulation of fully glycosylated molecular models of ACE2-Fc fusion proteins and their interaction with the SARS-CoV-2 spike protein binding domain

My contribution to this work was running MD simulations and performing stability analyses for the AF^{GG}/SpFr system.

This chapter originally appeared in the literature as:

Bernardi A, Huang Y, Harris B, Xiong Y, Nandi S, McDonald KA, Faller R. Development and simulation of fully glycosylated molecular models of ACE2-Fc fusion proteins and their interaction with the SARS-CoV-2 spike protein binding domain. PLoS One. 2020 Aug 5;15(8):e0237295.

Abstract

We develop fully glycosylated computational models of ACE2-Fc fusion proteins which are promising targets for a COVID-19 therapeutic. These models are tested in their interaction with a fragment of the receptor-binding domain (RBD) of the Spike Protein S of the SARS-CoV-2 virus, via atomistic molecular dynamics simulations. We see that some ACE2 glycans interact with the S fragments, and glycans are influencing the conformation of the ACE2 receptor. Additionally, we optimize algorithms for protein glycosylation modelling in order to expedite future model development. All models and algorithms are openly available.

5.1 Introduction

As of June 29, 2020 more than 10 Million people have been confirmed to be infected with coronavirus disease 2019 (COVID-19) which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This zoonotic pandemic has disrupted society worldwide on a peacetime-unprecedented scale. It also spurred a wide range of scientific endeavors to attack the various aspects of this disease. As the disease spreads there is a critical need for tools that enable the strategic design of biopharmaceutical countermeasures. We are here addressing computationally a molecular approach to aid in the design of a specific class of potential COVID-19 countermeasures.

The genomic sequence of the virus responsible for COVID-19, SARS-CoV-2, was made available in January 2020 ¹, providing critical information on the primary amino acid sequences of potential targets. A particularly important target is the SARS-CoV-2 spike (S) protein that is responsible for the first step in the viral infection process, binding to human cells via the angiotensin converting enzyme 2 (hACE2) receptor. The conserved expression and interaction of ACE2 indicates a wide range of hosts (human and non-human) for SARS-CoV-2 ². The S protein contains two domains S1 and S2 on each monomer. It is a homotrimer with each monomer comprised of 1281 amino acids. The monomers are expected to be highly glycosylated with 22 N-linked glycosylation sequons and 4 O-linked predicted glycosylation sites ³, although only 16 N-linked glycosylation sites were observed in a cryo-EM map of S produced in HEK293F cells ⁴. Very recently, Watanabe et al. performed site-specific glycoform analysis of full-length trimeric S protein made recombinantly in transfected HEK293F cells ⁵. Their analysis showed high occupancy at all 22 sites, with about 14 sites classified as complex, 2 sites as oligomannose,

and the remaining sites containing mixtures of oligomannose, hybrid and complex glycan structures. Seven of the sites with complex glycoforms, including the 2 sites on the RBD, also had a high degree (>95%) of core fucosylation. Viral coat proteins are often glycosylated which helps pathogens evade the host immune system, modulate access of host proteases, and can enhance cellular attachment through modification of protein structure and/or direct participation at the viral coat protein/cell receptor interface. These glycans are, however, only partially resolved in the experimental structure such that a computational approach is helpful to predict their behavior.

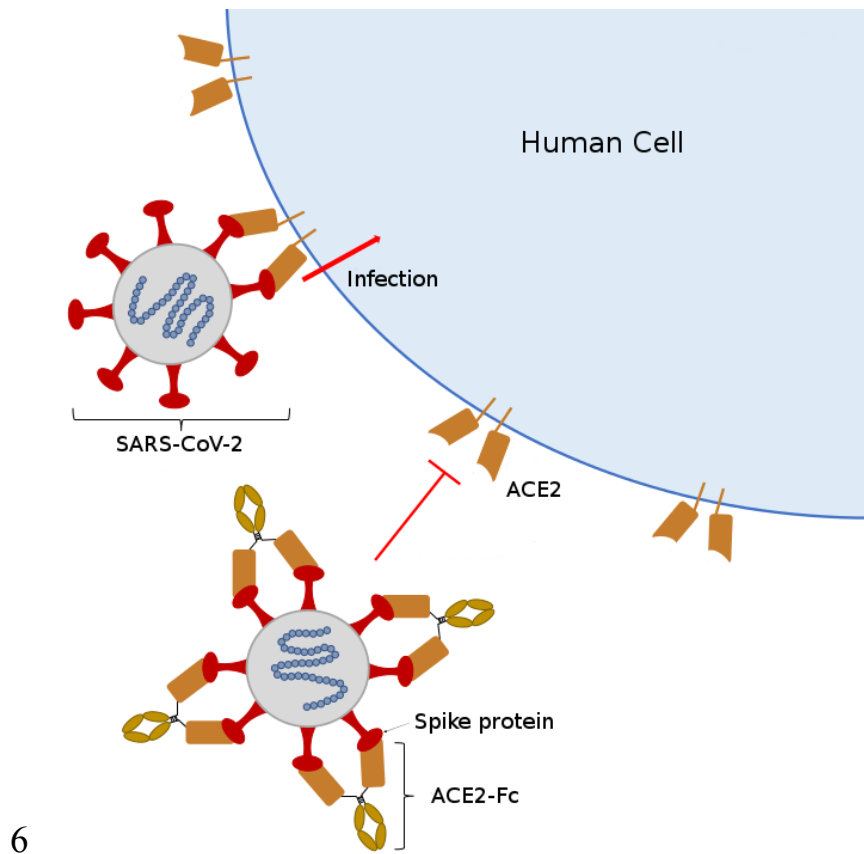


Figure 5.1: Proposed strategy for SARS-CoV-2 neutralization by ACE2-Fc immunoadhesin. ACE2-Fc binds to the spike (S) protein on the virus and blocks binding to the human cellular receptor ACE2, preventing cellular entrance of SARS-CoV-2.

The human ACE2 protein is a 788 amino acid integral membrane protein with seven N-linked glycosylation sites in the extracellular domain. The binding kinetics between the SARS-CoV-2 spike protein and the hACE2 receptor will depend on the 3D structures of both molecules and their molecular interactions which may be impacted by glycosylation⁶⁻⁸, as has been observed for other glycosylated viral spike proteins and their human receptors. Knowledge of the spike protein and ACE2 amino acid sequences have led to the commercial availability of the spike protein, ACE2, and various fragments of these, with and without purification/fusion tags, produced recombinantly in various expression hosts including Human embryonic kidney cells (HEK293), insect cells, Chinese Hamster Ovary cells (CHO), and *E. coli*. While the availability of recombinant sources for S and ACE2 glycoproteins have greatly contributed to our understanding of the structure and interactions between these proteins, it is important to recognize that glycosylation of recombinant S and ACE proteins will depend on the host cell⁹, the recombinant protein, as well as production^{10,11} and purification methods¹². As molecular models and molecular dynamics simulations can describe the interactions of proteins with glycans and the modulation of protein structure by glycans^{13,14} they are powerful tools to assess the significance of glycosylation on 3D structure and binding site interactions between the SARS-CoV-2 spike protein and the human ACE2 receptor, and to design novel biotherapeutics including optimizing glycosylation.

A promising strategy for the design of COVID-19 therapeutic proteins is a fusion of the extracellular domain of ACE2, the human receptor for SARS-CoV-2, with the Fc region of human immunoglobulin, IgG1, by a linker separating the two domains¹⁵. The neutralization strategy behind ACE2-Fc is shown in Figure 5.1¹⁵. This therapeutic design is often called an immunoadhesin, a chimeric protein combining the ligand-binding region of the cell receptor with

the constant domain of an immunoglobulin ¹⁶ . These chimeric molecules form dimers through disulfide bonds between Fc domains; this bivalency increases the affinity for the ligand. The human ACE2 receptor has been shown to be the primary receptor that SARS-CoV-2 uses for entry into and infection of human cells ^{17,18} , although the binding site is distinct from the catalytic domain of ACE2. With an ACE2-Fc immunoadhesin the ACE2 portion can act as a circulating “bait or decoy” to bind the SARS-CoV-2 spike protein preventing it from entering human cells while the Fc region confers longer circulatory half-life, provides effector functions of the immune system to clear the virus, and allows simple well-established purification using Protein A affinity chromatography. Immunoadhesins are a distinct class of antivirals that can be used prophylactically as well as post-infection and differ from both vaccines and antibodies. Unlike vaccines, they are not intended to elicit an immune response to the viral infection, and unlike antibody therapies, their design is greatly simplified since once the cellular receptor for viral entry is identified the immunoadhesin can be quickly designed and produced.

This strategy also precludes the coronavirus mutating to escape binding with the ACE2-Fc protein, as it would also lose binding affinity to the native ACE2 human cell receptor resulting in a less pathogenic virus. The re-emergent SARS-CoV-1 virus in 2003-2004 had a lower affinity for human ACE2 resulting in less severe infection and no secondary transmissions ¹⁹ . In this strategy the exogenous ACE2 would compensate for decreased ACE2 levels in the lungs during infection, contributing to the treatment of acute respiratory distress, and potentially reduce inflammation and reactive oxygen species in the lung ²⁰ . Most importantly, recombinant ACE2-Fc fusion proteins, with native ACE2 catalytic activity as well as a mutant version with lower ACE2 catalytic activity, produced using transfection of HEK293, have shown high affinity binding

to the SARS-CoV-2 spike protein and to potently neutralize SARS-CoV-2 *in vitro*²¹. Simulations are an ideal tool to optimize such a construct and guide the experimental production of ACE2-Fc.

Glycans are branched, flexible chains of carbohydrates that explore a much wider range of conformations at equilibrium conditions than the protein chain itself as the latter is typically not dynamically changing strongly from its folded form as that would affect its functionality. The faster dynamics of glycans complicates the structural and conformational characterization of glycans in laboratory experiments²². In atomistic molecular dynamics (MD) simulations, glycan conformations can be straight-forwardly analyzed to obtain structural information, as glycan dynamics are much closer to the computational timescale than the protein dynamics. However, neighboring glycans can interact with each other and essentially lock each other in which can lead to very slow equilibration into the correct conformation¹³. Therefore, algorithms are needed to generate realistic glycan configurations as glycans are regularly not fully resolved experimentally. Consequently, only a few simulations of related fully glycosylated proteins available²³⁻²⁶ among them recently a proposed glycosylated model of the Spike protein²⁷. Very recently a short simulation of the Spike protein with glycosylation has been published which is enabling longer studies²⁸. Our group has made significant progress in the field of glycan modeling in recent years^{13,14,29}.

N-glycan structure is highly heterogeneous, and the relative abundance of glycans depends on the expression system for glycoprotein production. Plant-based transient expression systems are well-suited to produce recombinant ACE2-Fc under the current COVID-19 pandemic given high production speeds. Two glycovariants of ACE2-Fc are simulated in this work: one is targeted

for ER retention with high mannose glycoforms, and the second is targeted for secretion with plant complex glycoforms. These glycovariants are currently being expressed and purified at UC Davis.

In order to properly understand the interaction between the spike protein and the variant ACE2 receptors bound to its fusion partner the glycosylation of both entities needs to be taken into account. The few existing computational studies of ACE2 interaction with the spike protein we are aware of are using aglycosylated proteins³⁰⁻³². Also, molecular docking studies have been performed with the older related SARS-CoV-1 virus protein implicated in the SARS epidemic in the early 2000s³³. We develop *in silico* models to predict the 3D structure of two glycosylated ACE2-Fc variants. Additionally, we evaluate the interactions between these two ACE2-Fc variants and a glycosylated spike protein fragment (SpFr) which contains the receptor binding domain of the SARS-CoV-2 spike protein.

5.2 Materials and Methods

5.2.1 Sequences and Initial Structure

ACE2-Fc is a homodimer of ACE2 bound to Fc via a synthetic linker. Two sequence variants are used in this work to model ACE2-Fc. The ACE2 and Fc domains N- and C-terminal residues for both variants, respectively, are as follows: ACE2, 18Q-740S (NCBI ID: NP_001358344.1); Fc, 109C-330K (UniProt ID: P01857). Variant 1 (Sequence Seq1 in Supporting Information; 960 amino acids) contains a C-terminal SEKDEL tag, which is used to express predominantly ER-retained proteins with high-mannose glycoforms in plant-based expression systems. Variant 2 (Sequence Seq2 in Supporting Information; 954 amino acids) does not use a C-terminal SEKDEL

tag, and will express standard plant glycoforms in plant-based expression systems. Variant 2 has two mutations: H357N and H361N. These mutations are used to deactivate the standard function of ACE2, by preventing the coordination of Zn^{2+} in the active site ²¹. The ACE2-Fc variants contain 18 disulfide bonds, with 4 of them being interchain. Table ST1 (in Supporting Information) describes the disulfide linkages. The variants also contain 8 N-glycosylation sites per monomer. Each peptidase domain of the ACE2-Fc variants is capable of binding one SARS-CoV-2 SpFr (Sequence Seq3 in Supporting Information; 183 amino acids), which contains one glycosylation site. The ACE2-Fc/SpFr structure is depicted in Figure 5.2. Zoomed views of the ACE2/SpFr interface are shown in Figure SF4. The coordinated Zn^{2+} site is shown in Figure SF1. All 3D structures are rendered with VMD ³⁴.

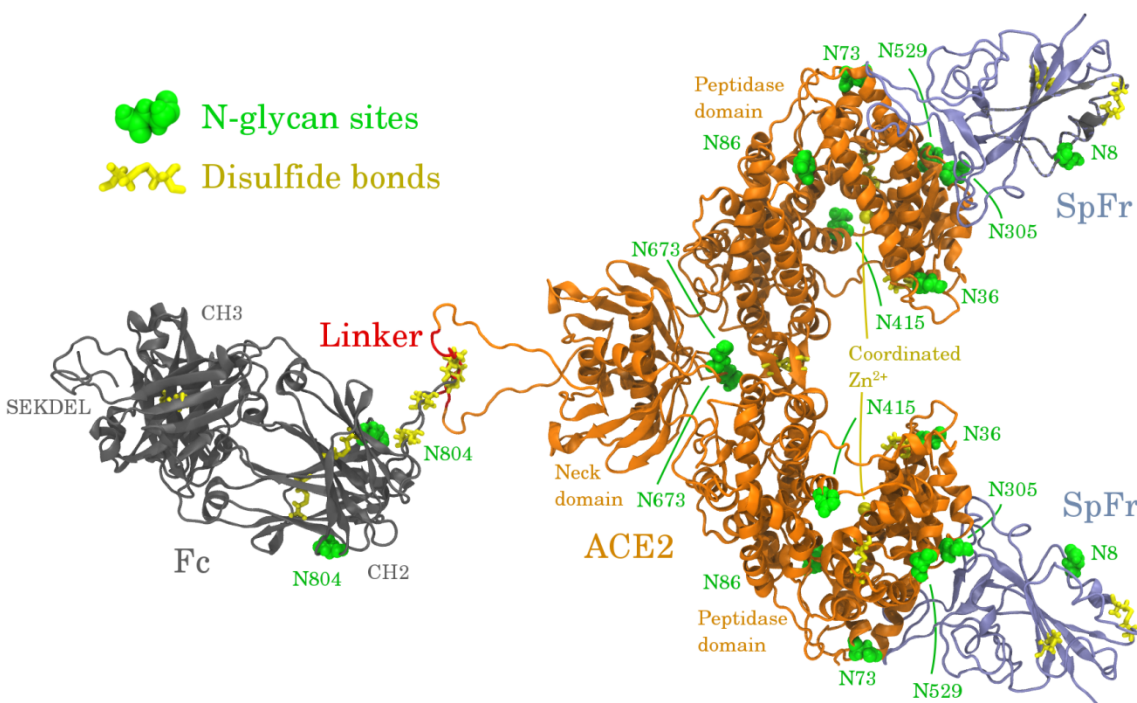


Figure 5.2: Infographic of the ACE2-Fc variant 1 homodimer bound to two SpFr.

5.2.2 Simulated Systems

ACE2-Fc variant 1 will express high-mannose type glycans when synthesized in plants, while variant 2 will express standard plant glycans. Additionally, SARS-CoV-2 SpFr will exhibit its own glycosylation depending on the host cell; here we assume common mammalian-like glycosylation. Our simulations employed uniform glycosylation profiles to approximate these glycosylation profiles. ACE2-Fc variant 1 is fully glycosylated with Man8 glycans, variant 2 is fully glycosylated with GnGnXF3 glycans the latter is consistent with a recent experimental study³⁵, and the SpFr is glycosylated with ANaF^{6,36}. Figure 5.3 shows these glycans using the Consortium of Functional Glycomics nomenclature.

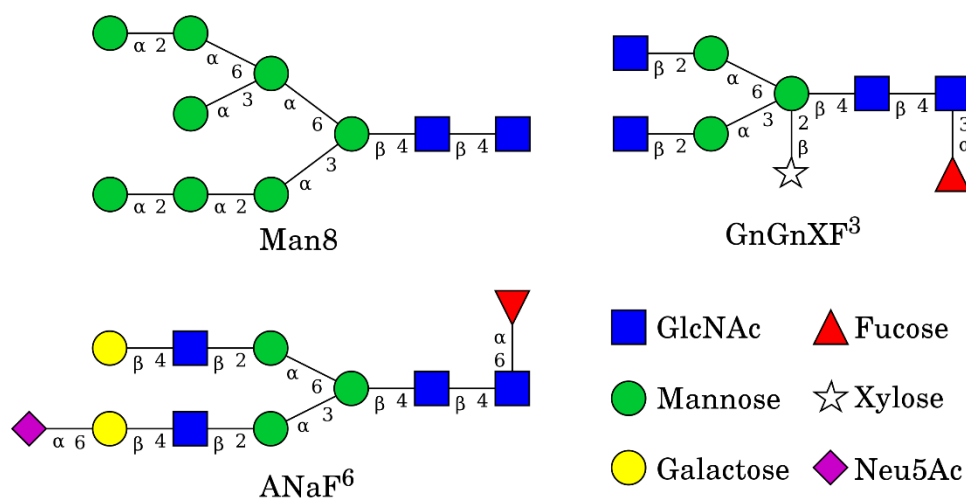


Figure 5.3: Glycans used in the simulated systems. All structures were built using GlycanBuilder³⁷.

Four systems containing ACE2-Fc variants were simulated in this work. The first system contains ACE2-Fc variant 1 with Man8 glycans. The second system contains ACE2-Fc variant 2 with GnGnXF³ glycans. The third and fourth systems are the immunoadhesins of the first and

second systems with the SARS-CoV-2 SpFr bound, respectively. The SpFr is always glycosylated with ANaF⁶. Table 5.1 summarizes the simulated systems.

Table 5.1: Description of simulated systems.

System ID	ACE2-Fc Sequence	ACE2-Fc Glycosylation	SpFr bound?	ACE2/SpFr ref. PDB
AF ^{M8}	Variant 1	Man8	no	6M17
AF ^{GG}	Variant 2	GnGnXF ³	no	6M17
AF ^{M8} /SpFr	Variant 1	Man8	yes	6M18
AF ^{GG} /SpFr	Variant 2	GnGnXF ³	yes	6M18

This work is largely made possible due to the recent cryogenic electron microscopy work that resolves the ACE2-B⁰AT1 and ACE2-B⁰AT1/SpFr structures, corresponding to PDB codes 6M18 and 6M17, respectively³⁸. The ACE2 and ACE2/SpFr domains were taken from these structures and fused to the Fc domain (PDB 3SGJ)³⁹. The Zn²⁺ and coordinating residues in 6M17 and 6M18 are poorly coordinated in these structures. The conformation of these residues along with a coordinating water were instead taken from PDB 1R42⁴⁰. Histidine protonation states for each system were determined using Reduce⁴¹, and are summarized in Table ST2.

5.2.3 Simulation Procedure

The simulation procedure includes the following steps:

1. Fuse ACE2 with Fc to ACE2-Fc using Modeller⁴²
2. Model Zn²⁺ and coordinating residues with MCPB.py⁴³
3. Attach glycans using glycam.org⁴⁴
4. Merge structures from 2. and 3. using github.com/austenb28/MCPB_Glycam_merge⁴⁵
5. Generate topology files using AmberTools⁴³
6. Convert topology files to Gromacs format using Acypype^{29,46}

7. Perform rigid energy minimization (EM) of glycans using github.com/austenb28/GlyRot

47

8. Perform EM (emtol = 1000 kJ/mol/nm)

9. Solvate and add ions

10. Perform 10 ps constant volume (NVT) (dt = 0.2 fs, T = 310 K)

11. Perform EM (emtol = 1000 kJ/mol/nm)

12. Perform 100 ps NVT (dt = 2 fs, T = 310 K)

13. Perform 100 ps constant pressure (NPT) (dt = 2 fs, T = 310 K, P = 1 atm)

14. Perform 75 ns production NPT (dt = 2 fs, T = 310 K, P = 1 atm)

Steps 2 and 4 are only required for AF^{M8} and AF^{M8}/SpFr, since they contain the coordinated Zn²⁺ sites. Steps 4 and 7 exhibit new, publicly available software under GNU General Public Licenses. GlyRot has previously been used to model glycosylated butyrylcholinesterase and CMG2-Fc^{13,14}. Forcefield topologies were generated using the AmberFF14SB⁴⁸ forcefield for protein atoms, the Glycam06-j⁴⁹ forcefield for glycan atoms, and the SPC/E water model⁵⁰. Steps 8 through 14 are performed using the Gromacs suite⁵¹⁻⁵³. Systems were solvated in rectangular boxes such the minimum distance between the solute and periodic boundary is 1.2 nm. A rectangular box (for size see Table ST3) was found to be sufficient for 75 ns; longer simulations may require a larger cubic box if the solute rotates significantly. A reduced timestep NVT in step 10 is required to relax solute-solvent contacts. Steps 10-13 used position restraints on the protein atoms. All simulations were performed at 310K and 1 atm with the Velocity Rescale thermostat⁵⁴ and Parinello-Rahman barostat⁵⁵ using time constants of 0.1 ps and 2 ps, respectively. All water bonds are constrained with SETTLE⁵⁶; all other bonds are constrained with LINCS⁵⁷. A 1 nm

cutoff was used for short-range nonbonded interactions. Particle Mesh Ewald was used to model long-range electrostatics⁵⁸. Table ST3 contains additional information on system sizes and solvation. Each system was simulated using one compute node with 16 cores. Simulations averaged 2.9 ns/day for systems without the SpFr, and 2.0 ns/day for systems with the SpFr.

5.3 Results and Discussion

Figure 5.3 shows the starting configurations generated as described above (left) and the configurations after MD for 75 ns (right) of all simulated systems. Systems of this size will not fully equilibrate in 75 ns, but evidence of structural stability and concerted motion can still be observed. This is in agreement with a recent equilibration study of a fully glycosylated Spike protein²⁸. All systems exhibit varying length of the flexible linker domain between ACE2 and Fc during simulation. The domain separation can be quantified by analyzing the center of mass distance between the ACE2 and Fc ordered domains, shown in Figure SF2. AF^{M8} and AF^{M8}/SpFr exhibit clearly more shortening of the linker domain than AF^{GG} and AF^{GG}/SpFr, possibly due to the difference in glycosylation in the Fc domain, which is closest to the flexible linker region. AF^{M8}/SpFr has the shortest distance between the ACE2 and Fc domains after 75 ns, which is consistent with its final configuration shown in Figure 5.3. In the AF^{M8}/SpFr and AF^{GG}/SpFr systems, ACE2 glycans near the ACE2/SpFr interface form contacts across the binding interface, indicating that glycosylation may significantly affect binding kinetics. Additionally, glycans on SpFr that were initially oriented away from the protein are reoriented towards ACE2 after 75 ns. The structure of the ordered domains of ACE2-Fc and SpFr appear to retain structural

stability. As expected, the glycans, on the other hand, show significant reorientation, as the configurational dynamics of glycans is faster than proteins¹³.

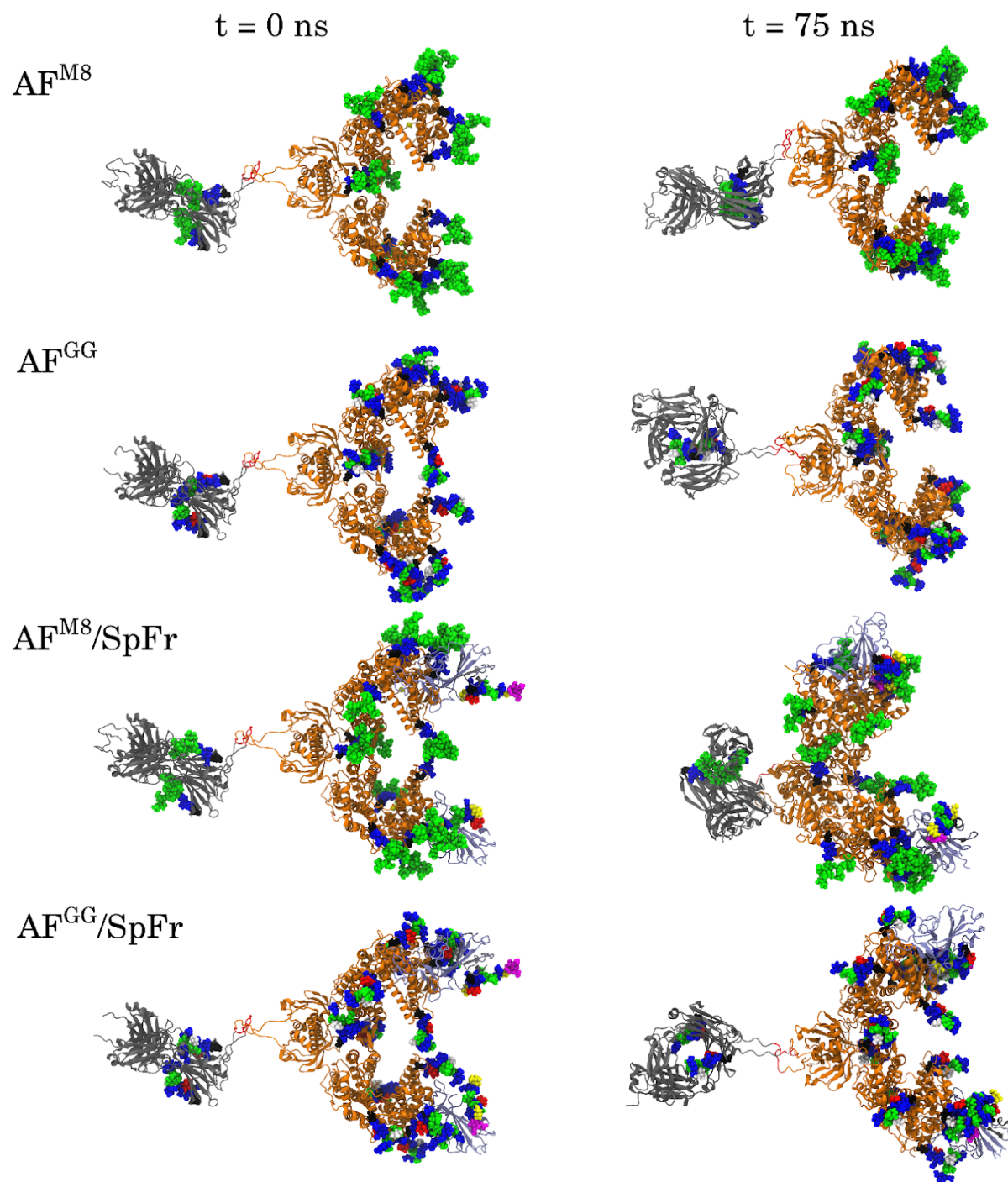


Figure 5.3: Initial (left) and 75 ns simulated (right) configurations of all systems.

To quantitatively assess structural stability, the root mean square deviation (RMSD) of the ordered domains of ACE2-Fc are shown in Figure 5.4. All profiles exhibit dynamics near or below 2.5 Å, indicating no major unfolding events have occurred. Conformational trending occurs when the RMSD increases from the initial and decreases towards the final. Conformational trending is evident in the ACE2 domain of all systems. Conformational trending is less evident for the Fc domains, except for the AF^{GG}/SpFr system, which exhibits significant conformational trending during the first 20 ns. This difference could indicate that GnGnXF³ glycosylation in the Fc domain of the AF^{GG}/SpFr promoted a conformational change in the Fc domain. Backbone RMSD profiles for the SpFr are provided in Figure SF3. The SpFr domains show RMSD profiles with significant conformational trending, potentially due to contacts with nearby glycans.

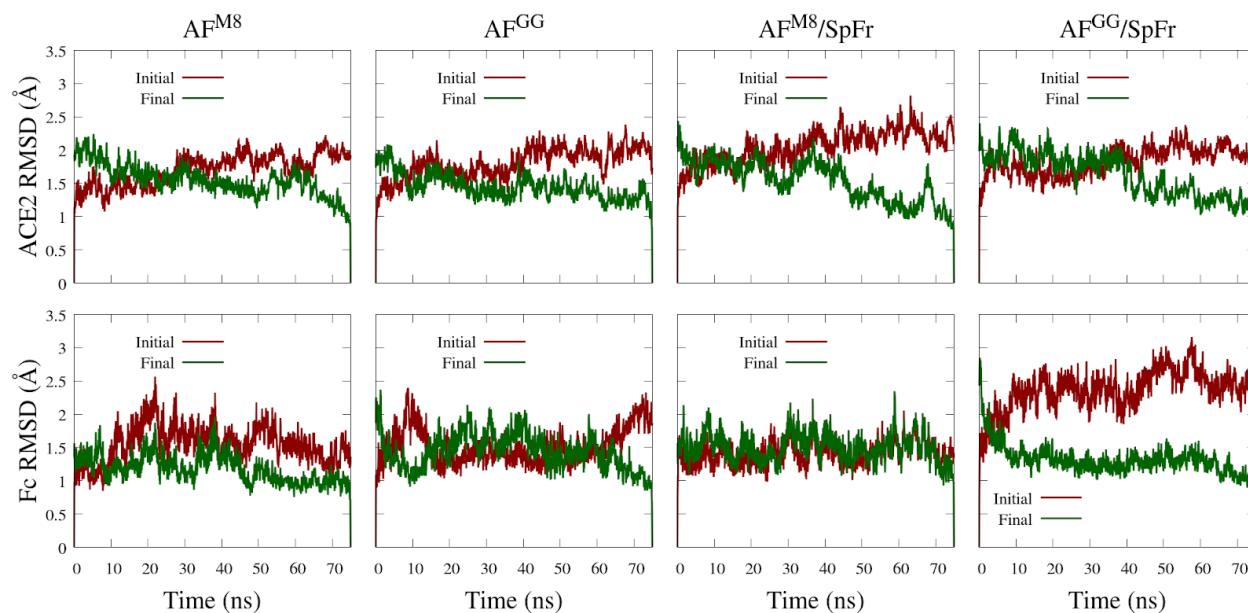


Figure 5.4: Backbone RMSD profiles of ACE2 (top) and Fc (bottom) ordered domains referenced to initial and final simulation configurations. ACE2: residues 4-707. Fc: residues 745-950. (see SI for sequences)

5.4 Conclusions

We have developed fully glycosylated models of ACE2-Fc immunoadhesins with and without interactions to glycosylated SARS-CoV-2 spike protein fragments. Atomic resolution models can be used to help guide the development of ACE2 and/or ACE2-Fc therapeutics for COVID-19 and potentially other coronavirus borne diseases.

We found that glycosylations affects protein structure, and potentially ACE2/SpFr binding. It is not yet clear how important these differences are, but they must be treated carefully when designing ACE2-Fc variants. The work exhibited here provides a direct avenue for collaborations between experimental and computational researchers.

All models developed here are freely available for researchers and future COVID-19 related simulations. Simulations with a wider variety of glycosylations as well as for longer times are in progress and will be reported in the future. The open-source workflows and tools that have been generated for glycoprotein simulations will be useful for general simulations of glycosylated systems. We hope that glycosylation becomes a standard variable in protein molecular simulations in the near future.

Acknowledgements

We thank Priya Shah for stimulating discussions. Simulations were performed on the HPC1 computing cluster at UC Davis.

References

1. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-9.
2. Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. *Biochemical and Biophysical Research Communications*. 2020;525(1):135-40.
3. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nature Medicine*. 2020;26(4):450-2.
4. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Velesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020;181(2):281-92.e6.
5. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*. 2020:eabb9983.
6. Ohuchi M, Ohuchi R, Feldmann A, Klenk HD. Regulation of receptor binding affinity of influenza virus hemagglutinin by its carbohydrate moiety. *Journal of Virology*. 1997;71(11):8377-84.
7. Fenouillet E, Gluckman JC, Bahraoui E. Role of N-linked glycans of envelope glycoproteins in infectivity of human immunodeficiency virus type 1. *Journal of Virology*. 1990;64(6):2841-8.
8. Goffard A, Callens N, Bartosch B, Wychowski C, Cosset F-L, Montpellier C, et al. Role of N-Linked Glycans in the Functions of Hepatitis C Virus Envelope Glycoproteins. *Journal of Virology*. 2005;79(13):8400-9.
9. Goh JB, Ng SK. Impact of host cell line choice on glycan profile. *Critical Reviews in Biotechnology*. 2018;38(6):851-67.
10. Ehret J, Zimmermann M, Eichhorn T, Zimmer A. Impact of cell culture media additives on IgG glycosylation produced in Chinese hamster ovary cells. *Biotechnol Bioeng*. 2019;116(4):816-30.
11. Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology*. 2009;19(9):936-49.
12. Zhang P, Woen S, Wang T, Liao B, Zhao S, Chen C, et al. Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discovery Today*. 2016;21(5):740-65.
13. Bernardi A, Kirschner KN, Faller R. Structural analysis of human glycoprotein butyrylcholinesterase using atomistic molecular dynamics: The importance of glycosylation site ASN241. *PLOS ONE*. 2017;12(11):e0187994.
14. Xiong Y, Karuppanan K, Bernardi A, Li Q, Kommineni V, Dandekar AM, et al. Effects of N-Glycosylation on the Structure, Function, and Stability of a Plant-Made Fc-Fusion Anthrax Decoy Protein. *Frontiers in Plant Science*. 2019;10(768).
15. Kruse R. Therapeutic strategies in an outbreak scenario to treat the novel coronavirus originating in Wuhan, China [version 2; peer review: 2 approved]. 2020;9(72).
16. Wycoff K, Maclean J, Belle A, Yu L, Tran Y, Roy C, et al. Anti-infective immunoadhesins from plants. *Plant biotechnology journal*. 2015;13(8):1078-93.
17. 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. 2020:eabb2762.

18. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. 2020:eabb2507.
19. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 2005;24(8):1634-43.
20. Verdecchia P, Cavallini C, Spanevello A, Angeli F. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. *Eur J Intern Med.* 2020:S0953-6205(20)30151-5.
21. Lei C, Fu W, Qian K, Li T, Zhang S, Ding M, et al. Potent neutralization of 2019 novel coronavirus by recombinant ACE2-Ig. *bioRxiv.* 2020:2020.02.01.929976.
22. Petrescu AJ, Petrescu SM, Dwek RA, Wormald MR. A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology.* 1999;9(4):343-52.
23. Lu D, Yang C, Liu Z. How Hydrophobicity and the Glycosylation Site of Glycans Affect Protein Folding and Stability: A Molecular Dynamics Simulation. *The Journal of Physical Chemistry B.* 2012;116(1):390-400.
24. Jo S, Qi Y, Im W. Preferred conformations of N-glycan core pentasaccharide in solution and in glycoproteins. *Glycobiology.* 2015;26(1):19-29.
25. Harbison A, Fadda E. An atomistic perspective on antibody-dependent cellular cytotoxicity quenching by core-fucosylation of IgG1 Fc N-glycans from enhanced sampling molecular dynamics. *Glycobiology.* 2019.
26. Yanaka S, Yogo R, Inoue R, Sugiyama M, Itoh SG, Okumura H, et al. Dynamic Views of the Fc Region of Immunoglobulin G Provided by Experimental and Computational Observations. *Antibodies.* 2019;8(3):39.
27. Grant OC, Montgomery D, Ito K, Woods RJ. 3D Models of glycosylated SARS-CoV-2 spike protein suggest challenges and opportunities for vaccine development. *bioRxiv.* 2020:2020.04.07.030445.
28. Woo H, Park S-J, Choi YK, Park T, Tanveer M, Cao Y, et al. Developing a Fully-glycosylated Full-length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *The Journal of Physical Chemistry B.* 2020.
29. Bernardi A, Faller R, Reith D, Kirschner KN. ACPYPE update for nonuniform 1–4 scale factors: Conversion of the GLYCAM06 force field from AMBER to GROMACS. *SoftwareX.* 2019;10:100241.
30. Cheng P, Zhengdan Z, Yulong S, Xiaoyu W, Kaijie M, Yanqing Y, et al. Exploring the Binding Mechanism and Accessible Angle of SARS-CoV-2 Spike and ACE2 by Molecular Dynamics Simulation and Free Energy Calculation. *chemrxiv.* 2020:11877492.v1.
31. Brielle ES, Schneidman-Duhovny D, Linial M. The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. *bioRxiv.* 2020:2020.03.10.986398.
32. Nami B, Ghanaeian A, Ghanaeian K, Nami N. The Effect of ACE2 Inhibitor MLN-4760 on the Interaction of SARS-CoV-2 Spike Protein with Human ACE2: A Molecular Dynamics Study. *ChemRxiv.* 2020; Preprint:<https://doi.org/10.26434/chemrxiv.12159945>.v1.
33. Zhang Y, Zheng N, Hao P, Cao Y, Zhong Y. A molecular docking model of SARS-CoV S1 protein in complex with its receptor, human ACE2. *Computational Biology and Chemistry.* 2005;29(3):254-7.
34. Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. *J Molec Graphics.* 1996;14(1):33-8.

35. Shajahan A, Archer-hartmann SA, Supekar NT, Gleinich AS, Heiss C, Azadi P. Comprehensive characterization of N- and O- glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2. *bioRxiv*. 2020:2020.05.01.071688.
36. Fisher P, Spencer H, Thomas-Oates J, Wood AJ, Ungar D. Modeling Glycan Processing Reveals Golgi-Enzyme Homeostasis upon Trafficking Defects and Cellular Differentiation. *Cell Reports*. 2019;27(4):1231-43.e6.
37. Ceroni A, Dell A, Haslam SM. The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code for Biology and Medicine*. 2007;2(1):3.
38. 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(6485):1444-8.
39. Ferrara C, Grau S, Jäger C, Sondermann P, Brünker P, Waldhauer I, et al. Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. *Proceedings of the National Academy of Sciences*. 2011;108(31):12669-74.
40. Towler P, Staker B, Prasad SG, Menon S, Tang J, Parsons T, et al. ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis. *J Biol Chem*. 2004;279(17):17996-8007.
41. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation¹Edited by J. Thornton. *Journal of Molecular Biology*. 1999;285(4):1735-47.
42. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*. 2016;54(1):5.6.1-5.6.37.
43. Case DA, Betz RM, Cerutti DS, T.E. Cheatham I, Darden TA, Duke RE, et al. AMBER 2016. University of California, San Francisco; 2016.
44. Woods Group. GLYCAM Web Athens, GA: Complex Carbohydrate Research Center, University of Georgia; 2005-2020 [Available from: <http://glycam.org>].
45. Bernardi A. MCPB_Glycam_merge: GitHub; 2020 [Available from: https://github.com/austenb28/MCPB_Glycam_merge].
46. Sousa da Silva AW, Vranken WF. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*. 2012;5(1):367.
47. Bernardi A. GlyRot: GitHub; 2020 [Available from: <https://github.com/austenb28/GlyRot>].
48. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*. 2015;11(8):3696-713.
49. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. Journal of Computational Chemistry*. 2008;29(4):622-55.
50. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective Pair Potentials. *J Phys Chem*. 1987;91(24):6269-71.
51. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013;29(7):845-54.
52. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *J Comp Chem*. 2005;26(16):1701-18.

53. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19-25.
54. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*. 2007;126(1):014101.
55. Parrinello M, Rahman A. Strain fluctuations and elastic constants. *J Chem Phys*. 1982;76(5):2662-6.
56. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*. 1992;13(8):952-62.
57. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem*. 1997;18(12):1463-72.
58. Essman U, Perela L, Berkowitz ML, Darden HLT, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103(19):8577 - 92.

5.5 Supporting Information

Variant 1: ACE2WildType(18-740)-SSERKCCVE-IgG1Fc(109-330)- SEKDEL

NCBI Reference Sequence ID: NP_001358344.1 (ACE2); UniProtKB Sequence ID: P01857

(IgG1Fc_human corresponds to amino acid residues 109 - 330)

PDB codes: 6M17 or 6M18 (ACE2), 1R42 (Zn + coordinating residues + coordinating water),

3SGJ (Fc)

```
1 - QSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNAGD - 50
51 - KWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSSEDKSKRLN - 100
101 - TILNTMSTIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNERLWAWES - 150
151 - WRSEVGKQLRPLYEEYVVLKNEMARANHYEDYGDYWRGDYEVNGVDGYDY - 200
201 - SRGQLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYISPIGCLPAHLL - 250
251 - GDMWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFVS - 300
301 - VGLPMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKGFRIILMCTKVITMD - 350
351 - DFLTAHHEMGGHIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPKH - 400
401 - LKSIIGLLSPDFQEDNTEINFLKQALTIVGTLPTFTYMLEKWRWVFKGE - 450
451 - IPKDQWMKKWEMKREIVGVVEPVPHDETYCDPASLFHVSNDYSFIRYYT - 500
501 - RTLYQFQFQEALCQAAKHEGPHKCDISNTEAGQKLFNMLRLGKSEPW - 550
551 - LALENVVGAKNMNRPLLNYFEPLFTWLKDNKNSFVGWSTDWSPYADQS - 600
601 - IKVRISLKSALGDKAYEWNENMYLFRSSVAYAMRQYFLKVNQMILFGE - 650
651 - EDVRVANLKPRI SFNFVVTAPKNSVSDIIPRTEVEKAIRMSRSRINDAFRL - 700
701 - NDNSLEFLGIQPTLGPPNQPPVSSSERKCCVECPPCPAPELLGGPSVFLF - 750
751 - PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRE - 800
801 - EQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQP - 850
851 - REPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKT - 900
901 - TPPVLDSDGSFFLYSKLTVDKSRWQQGNVVFSCSVMHEALHNHYTQKLSL - 950
951 - SPGKSEKDEL
```

Sequence Seq1. ACE2-Fc variant 1 sequence. ACE2 in orange, linker in red, Fc in grey.

Glycosylation sites highlighted in green. Coordinating Zn²⁺ residues highlighted in pink.

Variant 2: ACE2Mutant(18-740,H374N,H378N)-SSERKCCVE-IgG1Fc(109-330)

NCBI Reference Sequence ID: NP_001358344.1 (ACE2), UniProtKB Sequence ID: P01857

(IgG1Fc_human corresponds to amino acid residues 109 – 330)

PDB codes: 6M17 or 6M18 (ACE2), 3SGJ (Fc)

1 - QSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTITEENVQNMNAGD - 50
 51 - KWSAFLKEQSTLAQMYPLQEIQLTVKLQLQALQNGSSVLSSEDKSKRLN - 100
 101 - TILNTMSTIYSTGKVCNPDNPQECLELLEPGLNEIMANSLDYNERLWAWES - 150
 151 - WRSEVGKQLRPLYEYVVLKNEMARANHYEDYGDYWRGDYEVNGVDGYDY - 200
 201 - SRGQLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYISPIGCLPAHLL - 250
 251 - GDMWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFVS - 300
 301 - VGLPMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKGFRLMCTKVMTD - 350
 351 - DFLTAHNEMGNIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPKH - 400
 401 - LKSIIGLLSPDFQEDETEINFLKQALTIVGTLPTFTYMLEKWRWMVFKGE - 450
 451 - IPKDQWMKKWEMKREIVGVPEVPHDETYCDPASLFHVSNDYSFIRYYT - 500
 501 - RTLYQFQFQEALCQAAKHEGPLHKCDISSTEAGQKLFNMLRLGKSEPWT - 550
 551 - LALENVVGAKNMNVRPLLNYFEPLFTWLKQNKNSFVGWSTDWSPYADQS - 600
 601 - IKVRISLKSALGDKAYEWNDEMFLFRSSVAYAMRQYFLKVKNQMIILFGE - 650
 651 - EDVRVANLKPRI SFNFVVTAPKVSVDIIPRTEVEKAIRMSRSRINDAFRL - 700
 701 - NDNSLEFLGIQPTLGPPNQPPVSSSERKCCVECPPCPAPELLGGPSVFLF - 750
 751 - PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRE - 800
 801 - EQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQP - 850
 851 - REPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKT - 900
 901 - TPPVLDSDGSFFLYSKLTVDKSRWQQGNV FSCSVMHEALHNHYTQKSLSL - 950
 951 - SPGK

Sequence Seq2. ACE2-Fc variant 2 sequence. ACE2 in orange, linker in red, Fc in grey.

Glycosylation sites highlighted in green. Mutated residues highlighted in pink.

SpFr (crystallized residues only):

NCBI Reference Sequence ID: YP_009724390.1(SpFr corresponds to amino acid residues 336 – 518)

PDB codes: 6M17

1 - CPFGEVFAATRFASVYAWNKRKISNCVADYSVLVNSASFSTFKCYGVSPT - 50
 51 - KLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIA - 100
 101 - WNSNNLDSKVGGNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEG - 150
 151 - FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELL

Sequence Seq3. Spike fragment (SpFr) sequence. Glycosylation site highlighted in green.

Table ST1. Table of disulfide bonds. Interchain disulfide bonds are specified using A and B.

ACE2-Fc				SF	
CYS 1	CYS 2	CYS 1	CYS 2	CYS 1	CYS 2
116	124	733A	733B	1	26
327	344	736A	736B	44	97
513	525	768	828		

729A	729B	874	932		
730A	730B				

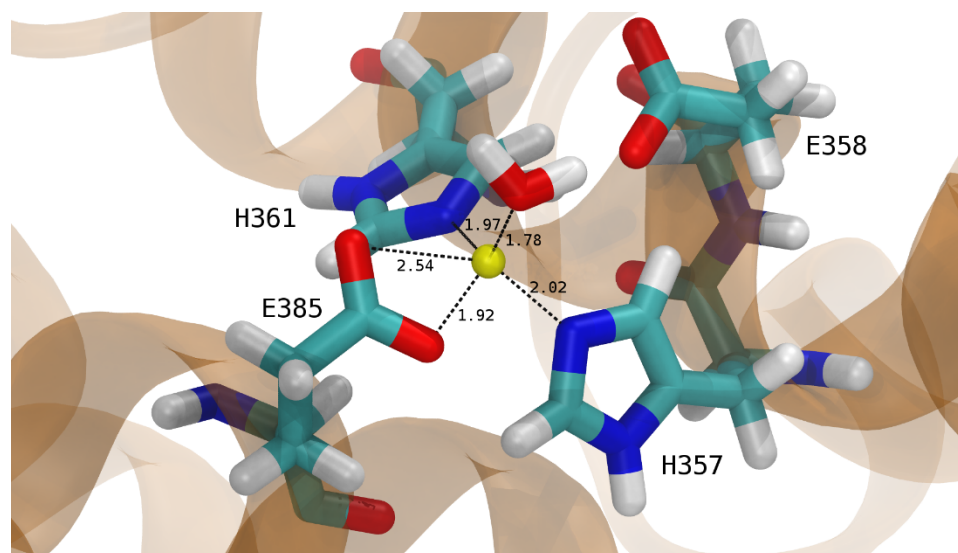


Figure SF1. Coordinating residues of the Zn^{2+} (yellow) active site of variant 1. Coordinating bonds with their distances in angstroms are labeled.

Table ST2. List of delta-protonated histidines for simulated systems. All other histidines are epsilon-protonated. Indices are for ACE2-Fc; SpFr has no histidines.

AF ^{M8}	AF ^{GG}	AF ^{M8} /SpFr	AF ^{GG} /SpFr
H17	H17	H356	H356
H357	H942	H357	H942
H361		H361	
H942		H942	

Table ST3. Additional simulation details.

System ID	Initial Box dimensions (nm x nm x nm)	# waters	# Na ⁺	# Cl ⁻
AF ^{M8}	18.0 x 15.5 x 22.0	188975	619	573
AF ^{GG}	17.6 x 16.3 x 22.7	197092	644	598
AF ^{M8} /SF	18.0 x 16.7 x 24.2	222602	721	677

AF ^{GG} /SF	17.9 x 16.7 x 24.2	221872	718	674
----------------------	-----------------------	--------	-----	-----

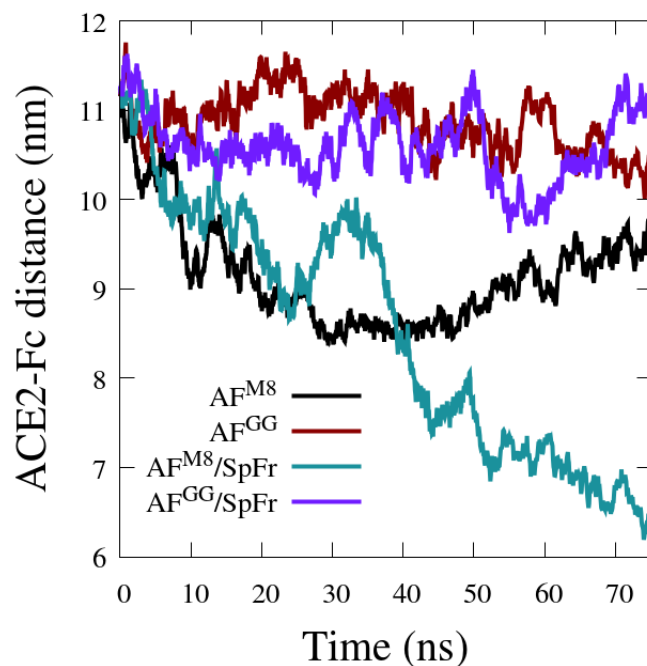


Figure SF2. Center of mass distance between the ordered domains of ACE2 and Fc. ACE2: residues 4-707. Fc: residues 745-950.

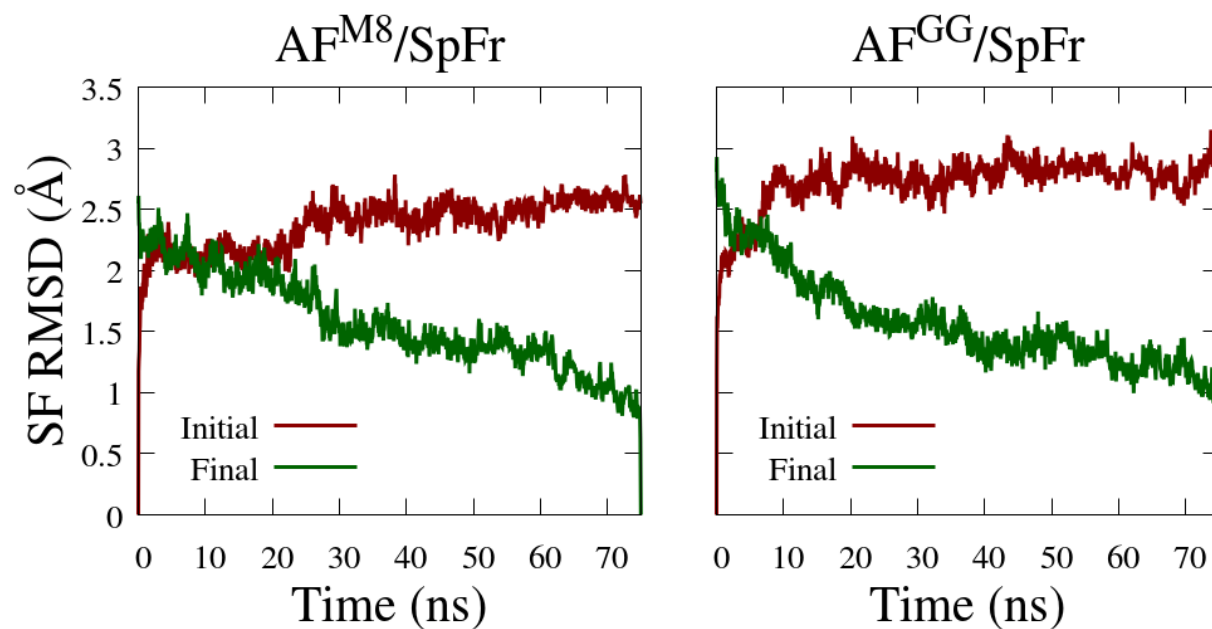


Figure SF3. Backbone RMSD profiles of the SpFr referenced from initial and final simulation configurations.

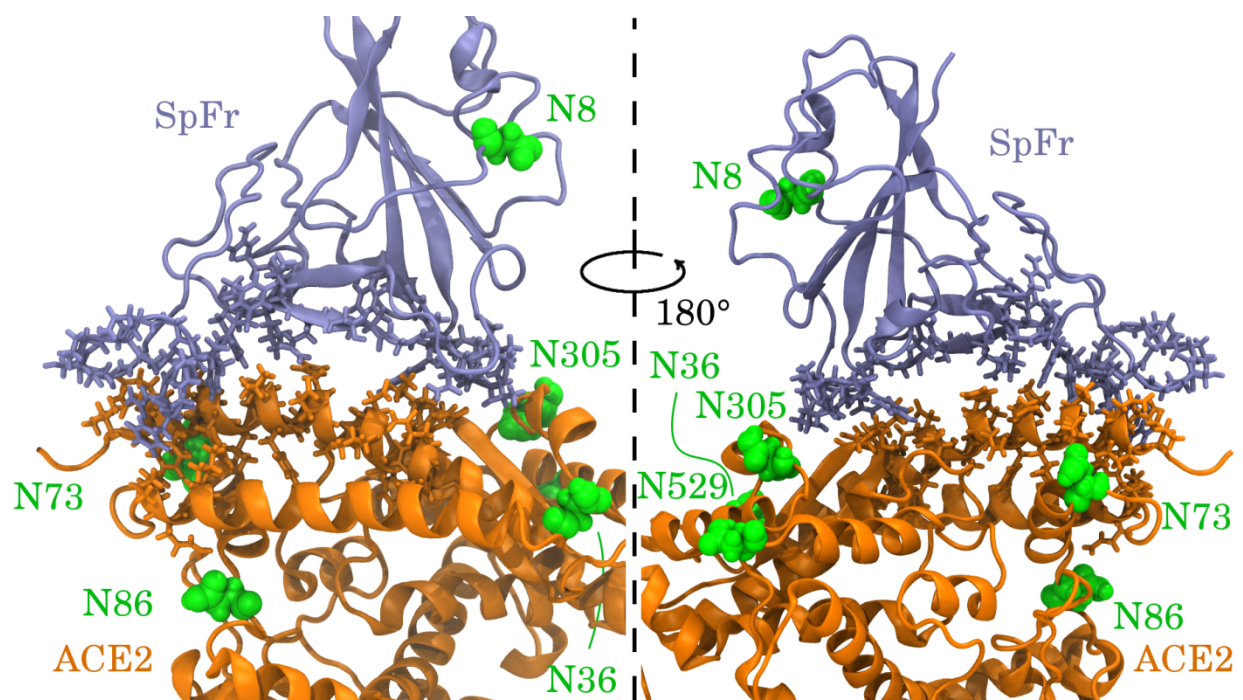


Figure SF4. Front (left) and back (right) zoom views of the interface between the ACE2 and SpFr domains. Interfacial residues are shown in licorice.

Chapter 6

SARS-Cov-2 Spike binding to ACE2 is stronger and longer ranged due to glycan interaction

My contribution to this work was setting up the MD simulations procedure, running simulations and performing analyses for the A2 fragments.

This chapter originally appeared in the literature as:

Huang Y, Harris BS, Minami SA, Jung S, Shah PS, Nandi S, McDonald KA, Faller R. SARS-CoV-2 spike binding to ACE2 is stronger and longer ranged due to glycan interaction. Biophysical journal. 2022 Jan 4;121(1):79-90.

Abstract

Highly detailed steered Molecular Dynamics simulations are performed on differently glycosylated receptor binding domains of the SARS-CoV-2 spike protein. The binding strength and the binding range increases with glycosylation. The interaction energy rises very quickly with pulling the proteins apart and only slowly drops at larger distances. We see a catch slip type behavior where interactions during pulling break and are taken over by new interactions forming. The dominant interaction mode are hydrogen bonds but Lennard-Jones and electrostatic interactions are relevant as well.

Statement of Significance

Glycosylation of the receptor binding domain of the Spike protein of SARS-CoV-2 as well as the ACE2 receptor leads to stronger and longer ranged binding interactions between the proteins. Particularly, at shorter distances the interactions are between residues of the proteins themselves whereas at larger distances these interactions are mediated by the glycans.

6.1 Introduction

As of July 2021, more than 182 million people globally have been confirmed to be infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus infectious disease 2019 (COVID-19). This zoonotic pandemic has disrupted society and spurred a wide range of scientific endeavors to improve our knowledge of coronaviruses and address the crisis. As the disease spreads and in order to prepare for potential future events there is a critical need for understanding the interaction of the virus with proteins involved in infection and immune clearance, or with proteins used as potential countermeasures or for the purpose of improved tests. Here, we study the interactions between SARS-CoV-2 and the human receptor responsible for binding using a molecular dynamics approach and validate it experimentally.

The SARS-CoV-2 spike (S) protein is a major structural protein and is therefore involved in many interactions. Through the receptor binding domain (RBD), S binds to the human angiotensin converting enzyme 2 (hACE2 or ACE2) receptor on the cell surface and initiates infection. There has been significant effort directed at understanding this interaction both

experimentally and computationally¹⁻⁷. Such studies are critical for the development of more efficient tests and therapeutics including vaccines.

Viral structural proteins like S are often glycosylated to help pathogens evade the host immune system, modulate access to proteases, and enhance the cellular attachment through modification of protein structure and/or direct participation at the virus-host interface⁸⁻¹⁴. Furthermore, many mammalian viruses use glycans on cell-surface glycoproteins or glycolipids as receptors¹⁵. Despite the important role of glycans in virus-host interactions, the glycans themselves are often only partially resolved in experimental structures generated from experimental techniques such as CryoEM¹⁶. Computational modeling of these glycans is therefore helpful in predicting their behavior and structural contributions.

S is a trimer where each monomer is expected to be highly glycosylated with 22 N-linked glycosylation sequons and 4 O-linked predicted glycosylation sites¹⁷. Only 16 N-linked glycosylation sites were observed in a cryo-EM map of S produced in HEK293F cells¹⁸. A study by Watanabe *et al.* (2020) determined site-specific glycoform analysis of full-length trimeric S protein made in HEK293F cells¹⁶. In another study of S glycosylation patterns including O glycosylation were determined¹⁹. In a similar vein, it has recently been argued that glycosylation can have influences post-vaccination and for vaccine resistance²⁰. Yet, the influence of glycosylation on the S-ACE2 interaction has been studied to a lesser extent^{21,22}. We address this gap in knowledge in the current study to reveal how glycans modulate the interaction of S with ACE2.

We expect that, as both S and ACE2 are glycosylated, the interaction is possibly modulated by the glycans. Few computational studies explicitly take the glycosylation of the receptor and/or the virus into account²³⁻²⁶. This is true in general as glycosylation has only very recently become

a stronger focus in simulations ²⁷⁻³¹ . One previous study has addressed the free energy of binding between the RBD and ACE2, including the impact of protein glycosylation ³² . However, previous studies were limited to a single simple glycan model, and did not study interactions of glycans or the influence of different complex glycan distributions beyond pulling force and protein contacts. Additional studies have shown experimentally and computationally that the RBD and ACE2 have different binding strength and dissociation rates when they are glycosylated vs non-glycosylated ^{33,34} . However, previous computational efforts often used simpler models for the glycans. We earlier developed a fully glycosylated model for the SARS-CoV-2 RBD and ACE2 proteins with different glycosylation patterns ² . We extend this model here to explore how a combination of complex glycans impact the energy and duration of binding. This is particularly important to improve rapid tests where viral antigens may be made in a variety of hosts with different glycan distributions.

In our previous study, we modeled ACE2 combined with the Fc domain as a therapeutic decoy. The extracellular domain of ACE2 was fused with the Fc region of human immunoglobulin, IgG1 ⁷ . The fusion ACE2 to the Fc domain of IgG1 has several advantages as a therapeutic decoy since it increases circulatory half-life and facilitates purification through the use of the common Protein A affinity chromatography platform. This served to neutralize the S protein on the virus and block the S protein's binding to cellular ACE2 for virus entry. ACE2-Fc was also modeled with plant glycosylation patterns. Due to the anticipated demand for high-speed production of the recombinant ACE2-Fc, plant-based transient expression systems are well-suited for rapid production. Plant cells can readily produce glycoproteins with either native, plant glycosylation ³⁵ or with modified human-like glycoforms through genetic manipulation ³⁶ . We simulated two plant glycovariants of ACE2-Fc in our previous work: Variant 1 was targeted for ER retention with high

mannose glycoforms, and Variant 2 was targeted for secretion with plant complex glycoforms. Since heterologous glycoproteins can be retained in the ER by adding a C-terminal H/KDEL-tag and the formation of Man8GlcNAc2 (Man8) N-glycans is typical for H/KDEL-tagging³⁷, Variant 1 was fully glycosylated with MAN8 glycans. Variant 2 was fully glycosylated with GlcNAc2XylFucMan3GlcNAc2 (GnGnXF³) that is a standard plant glycoform, and the S protein fragment was glycosylated with ANaF⁶. Figure 6.1 shows the glycans used in our systems. In our previous study we simulated the influence of the two glycoforms on the interaction of S protein and the specific recombinant ACE2-Fc fusion protein. We expect that the glycosylation influence is not restricted to the fusion proteins. In this study we focus on the contribution of these different glycosylation patterns on the protein-protein interactions via hydrogen bonding, interaction energies, and determine the corresponding free energies.

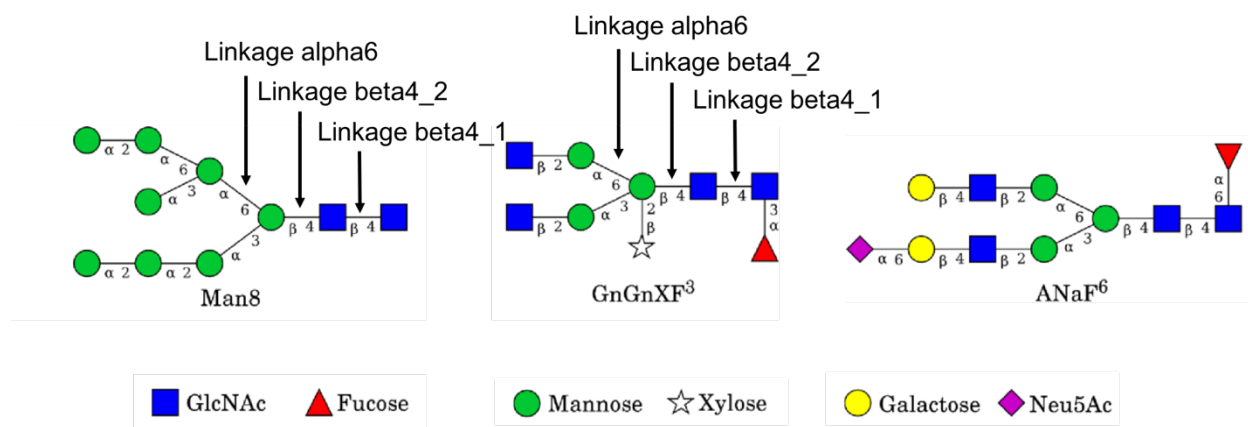


Figure 6.1: Glycans used in the simulations, adapted from previous work², with linkages of interest in MAN8 and GnGnXF³ glycans for dynamic analysis.

6.2 Materials and Methods

6.2.1 Simulation

Binding between the receptor binding domain of spike (RBD) and ACE2 receptor was determined using steered molecular dynamics, also known as the pulling of proteins³⁸. The starting atomic coordinates for all pulling systems were taken from the final 75 ns configurations of our previous paper². In that paper two sequence variants of ACE2-Fc were used to model the interaction between ACE2-F_C and SARS-CoV-2 RBD. Variant 1 (AF^{M8}/SpFr) contained a C-terminal SEKDEL tag which is used for ER retained proteins to express high mannose glycoforms and Variant 2 (AF^{GG}/SpFr) which does not contain the SEKDEL tag and expresses standard plant glycoforms. ACE2-B0AT1 and ACE2-B0AT1/SpFr structures were obtained from the protein data base. These structures had been determined using cryo-electron microscopy (PDB codes 6M18 and 6M17³⁹). These structures were fused to the Fc domain (PDB 3SGJ⁴⁰). The Zn²⁺ coordinating residues and water were taken from structure PDB 1R42⁴¹ in the case of Variant 1 ACE2. Variant 2 has 2 mutations that prevent Zn²⁺ coordination. The presence of zinc in protein structures is still actively being studied to determine its role in adjusting binding specificity^{42,43}. It has been demonstrated that Zn²⁺ plays a role in stabilizing some protein structures and can aid in the formation of biological oligomers^{42,43}. The final frame of the 75 ns trajectories for both ACE2-Fc/SpFr Variants was selected, and proteins were trimmed at residue 780 ALA (Figure 6.2) to make the pulling simulations a manageable 851 residues with glycans and 780 residues without glycans for AF^{M8}/SpFr, and 845 residues with glycans and 780 residues without glycans for AF^{GG}/SpFr. Because the system changed, the force field files had to be regenerated using AmberTools⁴⁴ as described previously². Briefly, the molecules were trimmed and glycans were

removed, then Man8 glycans were reattached to the truncated Variant 1 of ACE2, GnGnXF³ to the truncated Variant 2 of ACE2 and ANaF⁶ to the SpFr in both variants using Glycam.org⁴⁵. The coordinating Zn²⁺ was reattached to truncated and glycosylated Variant 1 using MCPB.py⁴⁶. Special care was taken to align the shortened original coordinates and the newly generated force field. Truncations from Variant 1 and Variant 2 that remained aglycosylated for both ACE2 and RBD were also studied to compare the influence of glycosylation on binding. The truncated systems were named A1Fr^{M8}/SpFr, A1Fr/SpFr, A2Fr^{GG}/SpFr, and A2Fr/SpFr, respectively. All amino acid sequences are available in supporting information and all S-S bridges are retained in our simulations.

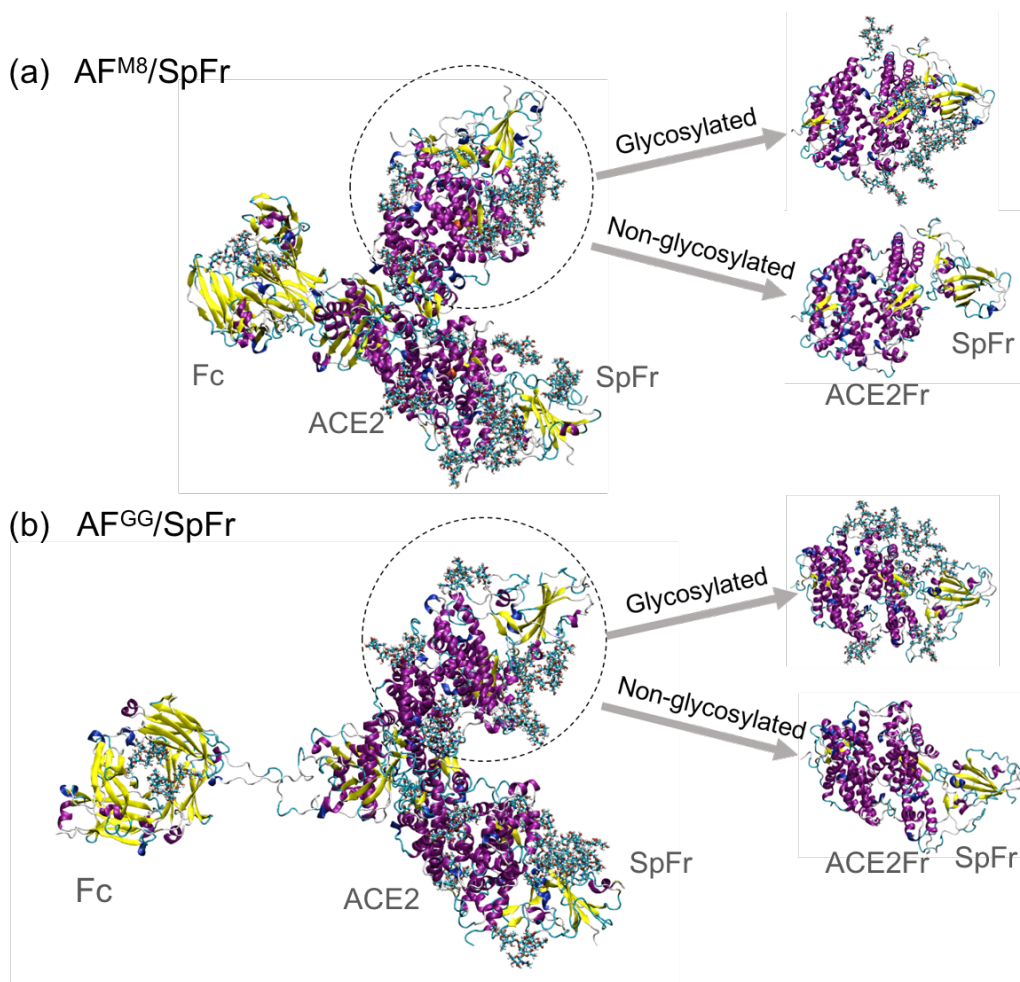


Figure 6.2: Schematics of generating the different initial structures. The full recombinantly made Ace2-FC systems on the left are used in the BLI experiments for determining binding affinities to the RBD, whereas the four truncated systems on the right containing only a fragment of ACE2 are modeled in simulations. From top to bottom, the truncated systems correspond to A1Fr^{M8}/SpFr, A1Fr/SpFr, A2Fr^{GG}/SpFr, and A2Fr/SpFr, respectively.

After the initial structures and corresponding force fields were generated, the proteins were rotated so that the pulling direction was along one of the principal axes, and the simulation boxes were expanded to 10 x 10 x 26 nm for A1Fr^{M8}/SpFr and A1Fr/SpFr, and 10 x 10 x 30 nm for A2Fr^{GG}/SpFr and A2Fr/SpFr so that the spike RBD fragments did not experience interactions with the ACE2 fragments across periodic boundaries during pulling. Then the new box was solvated with 80,271 water molecules and 24 Na⁺ as counter ions for A1Fr^{M8}/SpFr, 80,764 waters and 23 Na⁺ cations for A1Fr/SpFr, 93,541 waters and 26 Na⁺ cations for A2Fr^{GG}/SpFr, 93,989 waters and 25 Na⁺ cations for A2Fr/SpFr. Energy minimizations were performed until the convergence criteria were met (emtol = 1,000 kJ/mol/nm), followed by a 100 ps constant volume (NVT) (dt = 2 fs, T = 310 K) and a 100 ps constant pressure (NPT) (dt = 2 fs, T = 310 K, P = 1 atm), to equilibrate the systems. All simulations for equilibration were performed at 310 K and 1 atm with the Velocity Rescale thermostat⁴⁷ and Parrinello-Rahman barostat⁴⁸. All water bonds were constrained with SETTLE⁴⁹, and all other bonds were constrained with LINCS⁵⁰. Box expansion, solvation, and equilibration were performed using the Gromacs suite version 2019.1⁵¹.

Pulling simulations were then performed to study the free energy of binding as well as the structural arrangement of the separating proteins during interaction. For both variants, the ACE2 fragment was set to be immobile but deformable, whereas the spike RBD fragment (also flexible)

was pulled away from the ACE2 fragment. Pull simulations were performed under NpT conditions using a 2 fs timestep, a pull coordination spring constant of 1000 kJ/mol/nm², a Nose-Hoover thermostat⁵² at 310 K, and a Parrinello-Rahman barostat at 1 atm.

A total of 36 pulling simulations were performed at three different pulling rates (1 nm/ns, 5 nm/ns and 10 nm/ns) on the four truncated structures using Gromacs 2019.1⁵¹. Each structure was pulled at each rate 3 times for sampling purposes. The starting configuration was the same for each independent run, but the random seed for the velocities in each run was randomly assigned, resulting in independent behaviors. This approach clearly generated independent runs as seen in Figure 6.3. Systems were pulled over a distance of 8 nm until full separation (no interaction) was achieved (see Figure 6.3).

Hydrogen bonds were analyzed using the built-in Gromacs bond command⁵¹ with a default cutoff distance of 3.5 Angstroms. This command was used to generate the hydrogen bonds and Lennard Jones contacts as a function of time as well as a hydrogen bond interaction bitmap and corresponding index file of the different interactions. The hydrogen bonding interaction bitmap was recreated in python using matplotlib⁵³ in order to add labels for donor acceptor pairs and calculate the percent occupancy of hydrogen bonds across the simulation (script information available in SI). Short range Lennard Jones and Coulombic interaction energies were calculated from the Gromacs .edr file by specifying energy groups on the ACE2 and RBD using the gmx energy command⁵⁴.

6.2.2 Experiment

6.2.2.1 Protein Deglycosylation

ACE2-Fc (Acro Biosystems, Newark, DE, AC2-H5257) and RBD (Sino Biological, Chesterbrook, PA, 40592-V08B) deglycosylation was performed using Remove-iT PNGase F (Bio-Rad, Hercules, CA). Samples with PNGase F were incubated at 310 K for 5 hours. PNGase F was then removed by incubating the samples in chitin magnetic beads according to manufacturer instructions (New England Biolabs, Ipswich, MA). Deglycosylation of proteins was confirmed via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). 8 μ L of Laemmli sample buffer (Bio-Rad) and 2 μ L β -mercaptoethanol (Bio-Rad) were added to 30 μ L of sample. Samples were heated at 368 K for 5 minutes, then run on Mini-PROTEAN TGX Stain-Free Precast Gels (Bio-Rad) at 200V for 36 minutes. Gels were imaged using a ChemiDoc Imaging System (Bio-Rad).

6.2.2.2 Biolayer Interferometry

Anti-hIgG-Fc (AHC) biosensors (FortéBio, Fremont, CA) were used to immobilize ACE2-Fc by immersing the biosensors in solution containing 100 nM ACE2-Fc for 10 minutes. The Octet RED384 was used to obtain response measurements for protein association and dissociation. Two-fold serial dilutions of RBD were tested, from 250 nM to 7.81 nM. Data were collected for 60 seconds for the baseline, 400 seconds for association, and 800 seconds for dissociation. The experiment was performed at 299 K.

FortéBio Data Analysis Software version 8.1.0.53 was used for data processing and analysis. From the raw data, reference well values were subtracted, the y-axes were aligned to baseline, inter-step correction was applied for alignment to dissociation, and Savitzky-Golay

Filtering⁵⁵ was used for smoothing. Using a 1:1 binding model, steady-state analysis was performed on the response average from 390-395 seconds. From the binding affinities of glycosylated and deglycosylated ACE2-Fc and RBD, the change in binding energy following deglycosylation of ACE2-Fc and RBD was calculated as:

$$\Delta G_{\text{non-glycosylated}} - \Delta G_{\text{glycosylated}} = RT \ln \left(\frac{K_{D,\text{non-glycosylated}}}{K_{D,\text{glycosylated}}} \right)$$

6.3 Results

Figure 6.3 presents the pull force as a function of the pull distance between the ACE2 fragments and RBD for different glycosylation states at 1 nm/ns pulling rate. The pull distances are calculated based on the centers of mass for the ACE2 fragments and RBD but normalized to start from 0 nm to highlight differences between configurations. Pull force vs pull distance plots for higher pulling rates can be seen in supplemental information (Figure S1). Fundamentally, we see that for all conditions under study there is an immediate sharp increase in force when pulling the two proteins away from each other indicating strong local binding between the ACE2 binding domain and RBD. After going through a peak in force, the force drops off at increasing distance but with a clearly smaller slope than the initial increase. As expected, the pull force increases with pulling rate (blue, orange, green lines in Figure S1) such that the lowest force is most relevant for comparison to experiments. Importantly, for the same fragment the peak force is clearly higher by ~250 kJ/mol/nm at 1nm/ns, with glycosylation than without. This indicates an overall stronger binding of the glycoproteins than their aglycosylated counterparts for both types of glycosylation simulated. Additionally, the force curves are much broader for the glycosylated structures as compared to the aglycosylated ones indicating the presence of glycans extends the range for

binding in addition to strengthening it. Also, the force is longer ranged (only at larger distances does it reach zero) which indicates that the glycans which extend away from the proteins contribute to the binding at longer distances. As shown in Figures 6.3a and 6.3b the aglycosylated structures return to baseline at roughly 2.5 nm of pulling distance. Importantly the glycosylated structures in Figure 6.3a and 6.3b have an extended window of pulling force of 2-3 nm for A1Fr^{M8}/SpFr, and a smaller difference of roughly 1 nm for A1Fr^{GG}/SpFr when compared to their aglycosylated counterparts. This indicates both Man8 and GnGnXF glycans increase binding strength, and binding range, but the type of glycan affects both the strength and interaction distance of the specific binding.

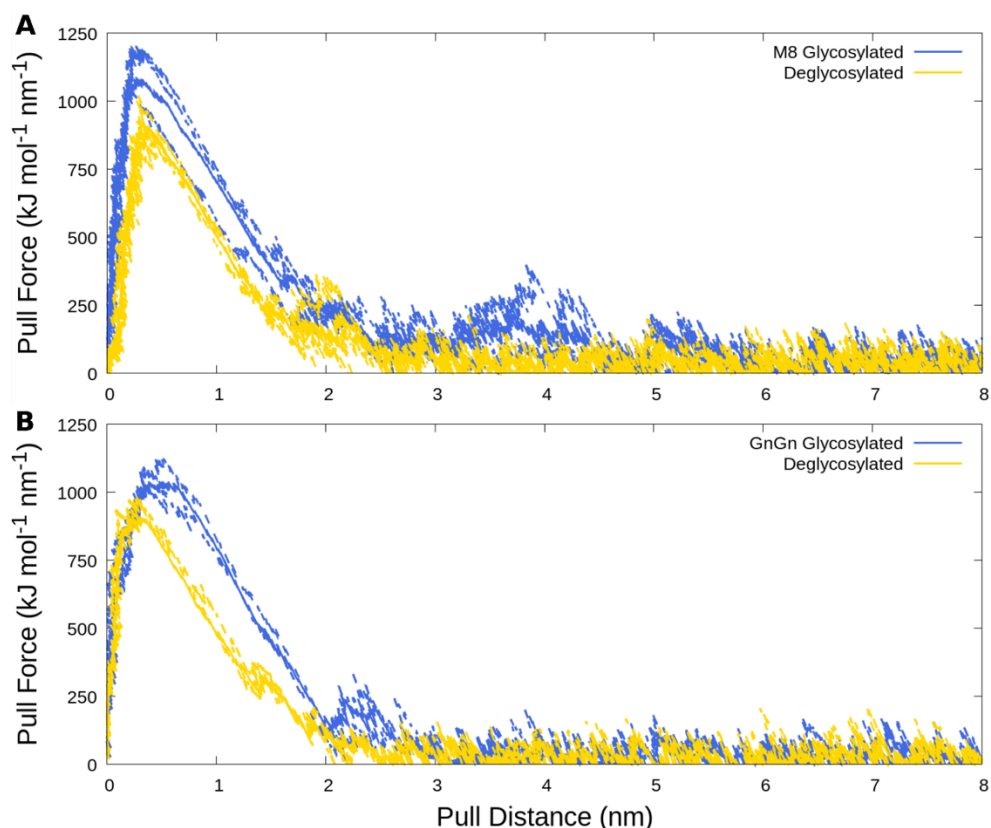


Figure 6.3: Traces of pull force versus pull distance. A) Man8 glycosylated A1Fr^{M8}/SpFr and aglycosylated structure A1Fr/SpFr B) GnGnXF glycosylated A2Fr^{GG}/SpFr. and aglycosylated Structure A2Fr/SpFr. Blue

lines correspond to glycosylated structures, gold to deglycosylated. Dashed lines are individual replicas, solid lines are averages.

To further characterize the extension of binding interactions, Figure 6.4 shows hydrogen bonding interaction maps between the ACE2 and RBD proteins. Figure 6.4a and 6.4c are for A1Fr^{M8}/SpFr and A2Fr^{GG}/SpFr respectively while 6.4b and 6.4d are the corresponding aglycosylated versions. (Full scale images with donor:acceptor pairs labeled are available in Figures S2-S5) The y-axis contains information about the donor and acceptor pair for the hydrogen bond and the x-axis corresponds to simulation time. Interaction types are colored and sorted according to the interaction type: protein-protein interactions are colored as white, protein-glycan as yellow, and glycan-glycan as magenta. Hydrogen bonding is clearly a major interaction mode between proteins. It is interesting that in A1Fr^{M8}/SpFr (Figure 6.4a) the predominant interactions involve glycans directly while for A2Fr^{GG}/SpFr (Figure 6.4c) the predominant interactions are protein-protein interactions which are indirectly strengthened by glycosylation. This indirect protein-protein strengthening is most clearly seen when comparing occupancy calculated from these heatmaps as shown in the tables in Figure 6.5 and Figures S6-9. There are multiple binding regimes as a function of time for the two glycosylated structures; this is more pronounced in the A1Fr^{M8}/SpFr case. This behavior manifests itself due to the original active hydrogen bonds in the complex releasing, but other hydrogen bonds catch and eventually release at larger distances before complete unbinding is seen. This catch-slip behavior is particularly attributable to the glycans, as the H-bonds present at longer distance are particularly ones involving glycans, either protein-glycan or direct glycan-glycan bonding. Both non-glycosylated structures shown in Figure

6.4b and 6.4d express maps of similar protein-protein interactions, though the A2Fr/SpFr shown in Figure 6.4d contains many more interactions as indicated by the increased number of rows.

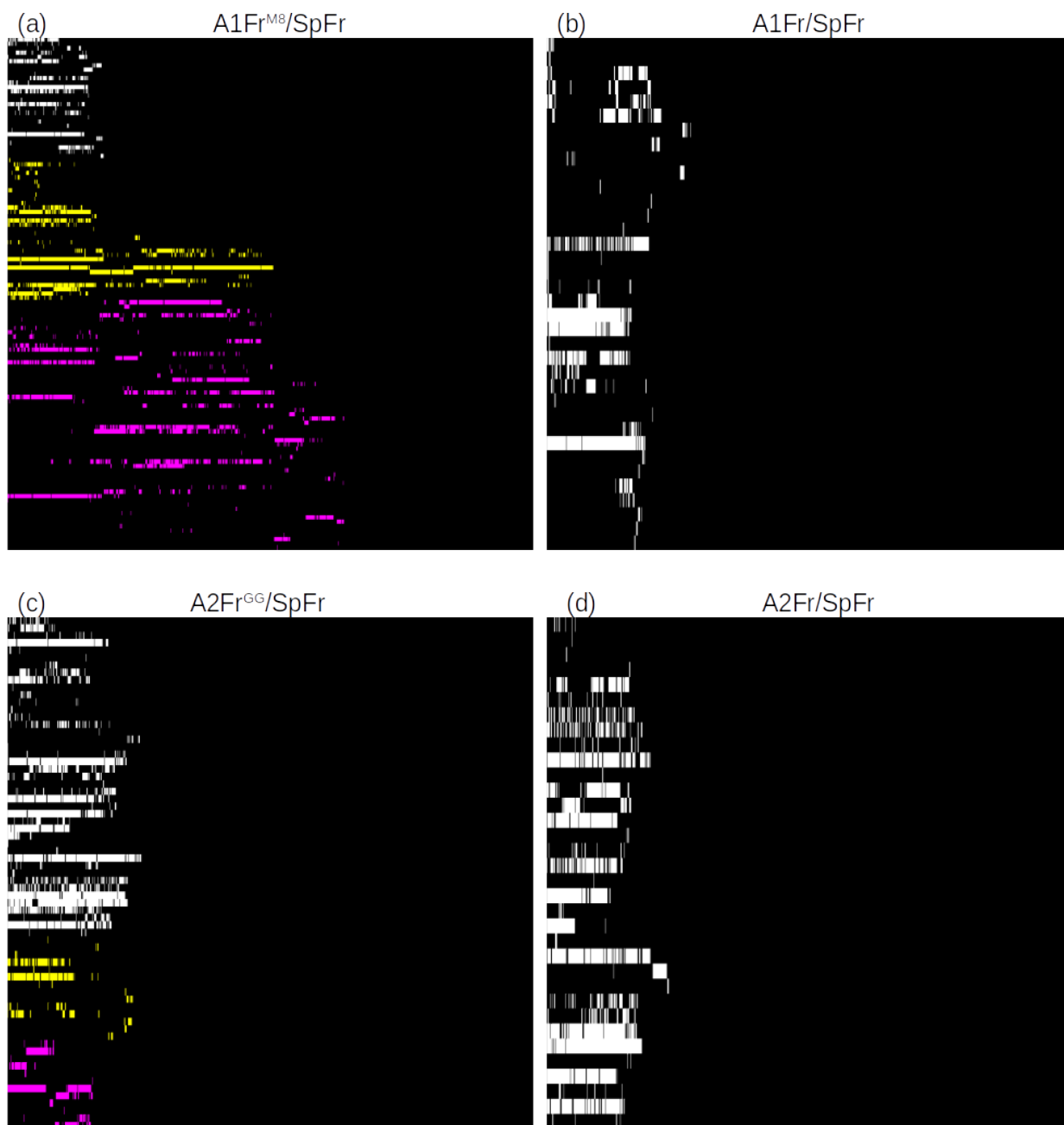


Figure 6.4: Hydrogen bond interactions vs simulation time. A) Man8 Glycosylated A1Fr^{M8}/SpFr. B) Aglycosylated Structure A1Fr/SpFr. C) GnGnXF³ Glycosylated A2Fr^{GG}/SpFr. D) Aglycosylated Structure

A2Fr/SpFr. Colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.

Figure 6.5 shows the configurations where RBD with and without ANaF⁶ started to be pulled away from the ACE2 fragment for the 4 different systems. The top 5 hydrogen bonds by occupancy, i.e. the fraction of time a given hydrogen bond is active, and their corresponding donor:acceptor pairs are highlighted. (Top 25 hydrogen bonds by occupancy for the 4 different configurations are available in Figures S6-S9) A1Fr^{M8}/SpFr clearly shows the predominant interactions are between the RBD glycan and ACE2 glycan and between the RBD glycan and the ACE2 protein, while for A2Fr^{GG}/SpFr the predominant interactions are between the protein backbones. It is also interesting to note that the predominant interactions in A2Fr^{GG}/SpFr are the protein-protein interactions. The strongest glycan interaction for A2Fr^{GG}/SpFr are not found until hydrogen bond #9 ranked by occupancy (Figure S8) while the top 3 hydrogen bonds ranked by occupancy involve glycans for A1Fr^{M8}/SpFr. A1Fr^{M8}/SpFr also clearly shows a different starting orientation than A2Fr^{GG}/SpFr, with minor changes in ACE2 structure and obvious rotation in the RBD with direct glycan-glycan interaction. These minor structural and orientational differences are also seen in the aglycosylated structures. Interacting groups for the hydrogen bonding shown follow AMBER nomenclature⁵⁶. The first letter corresponds to element with subsequent letters and numbers being linkage bookkeeping. For example, N, NZ, and NE2 all refer to nitrogen with different linkages, while O and its variants refer to Oxygen.

Figure 6.6 shows how the different structures of MAN8 and GnGnXF³ affect the hydrogen bonding regime. Although MAN8 and GnGnXF³ have similar size (223 atoms vs 222 atoms), their shapes are very different. MAN8 is relatively flatter comparing to GnGnXF³, making it bend less

flexibly. Therefore, when MAN8 is close to ANaF⁶, they interact in a side-by-side fashion, whereas when GnGnXF³ is close to ANaF⁶, they interact in a head-to-head fashion, forming less hydrogen bonds than the MAN8/GnGnXF³ pair.

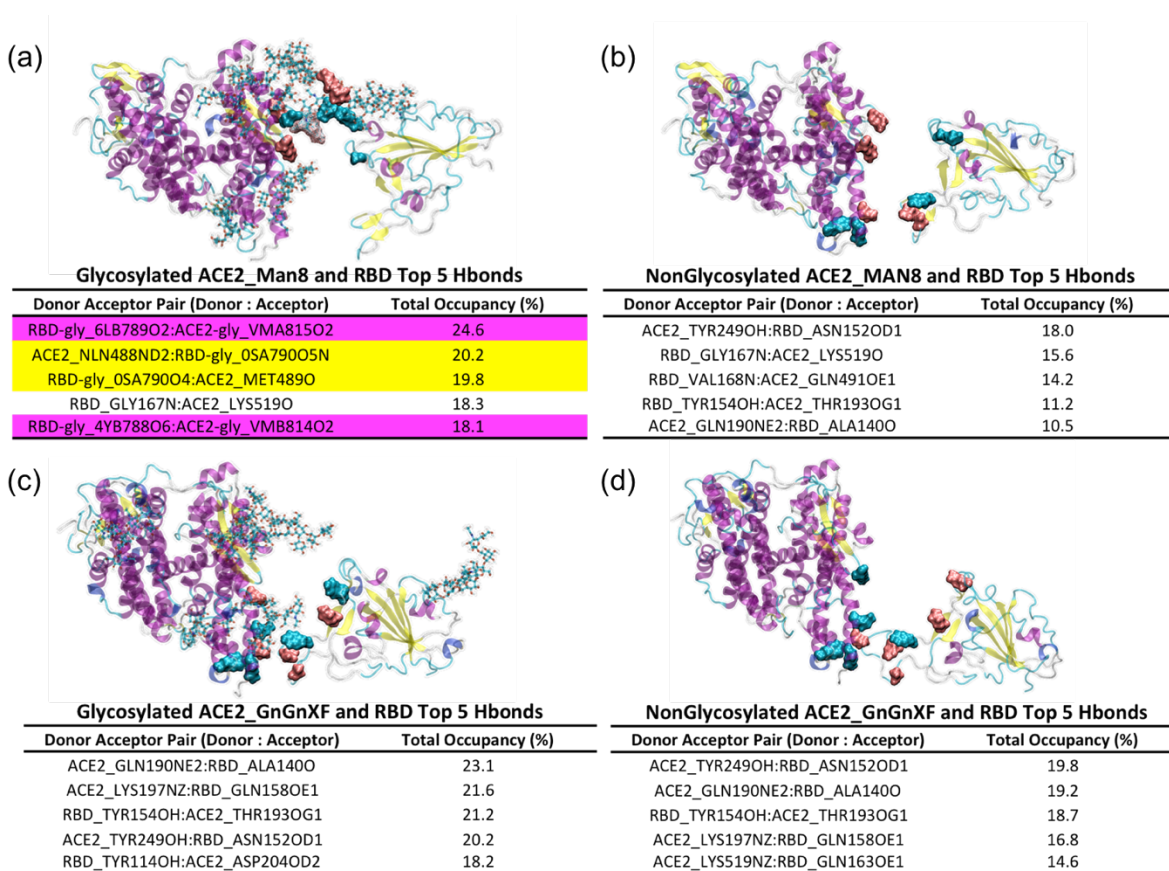


Figure 6.5: Top 5 hydrogen bond donor:acceptor pairs and occupancy. A) Man8 Glycosylated A1Fr^{M8}/SpFr. B) Aglycosylated Structure A1Fr/SpFr. C) GnGnXF³ Glycosylated A2Fr^{GG}/SpFr. D) Aglycosylated Structure A2Fr/SpFr. Table colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan. On the 4 configurations, residues highlighted with blue indicate donors, and pink indicate acceptors.

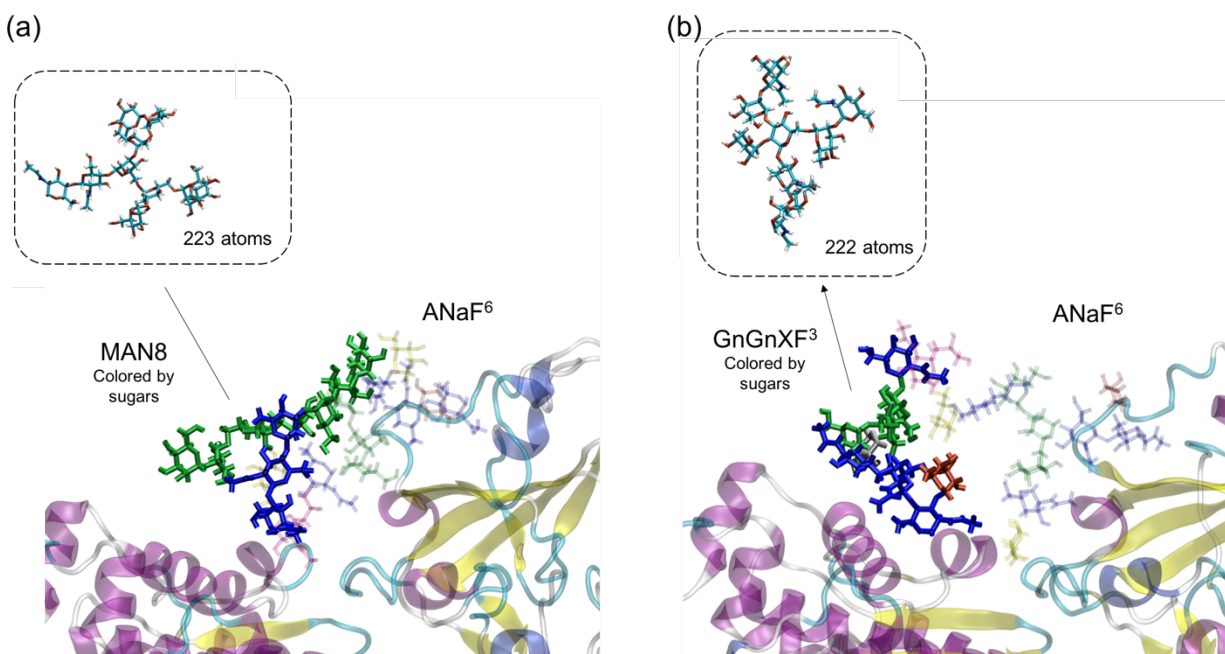


Figure 6.6: Different structures and hydrogen bonding regimes of MAN8 and GnGnXF³ when interacting with ANaF⁶ on RBD. A) MAN8 that interacts with ANaF⁶. B) GnGnXF³ that interacts with ANaF⁶. Inserts: shape and size of the MAN8 and GnGnXF³ without bending towards ANaF⁶. The glycans attached to proteins were colored by different sugars: Blue: GlcNAc; Green: Mannose; Yellow: Galactose; Red: Fucose; Silver: Xylose; Purple: Neu5Ac.

An autocorrelation function (ACF) analysis was performed for the angles and dihedrals of interest in both glycosylations, MAN8 and GnGnXF³, to further study the flexibility of the different glycans. These different flexibilities might be able to explain some of the emerging hydrogen bonding patterns. The angles and dihedrals chosen for the analysis are the ones between sugars, i.e., at the linkages. Figure 6.1 shows the linkages of interest; the angles and dihedrals at linkage beta4_1, beta4_2, and alpha6 of the glycans at the 6 glycosylation sites on the ACE2 fragment in A1Fr^{M8}/SpFr and A2Fr^{GG}/SpFr at positions N219, N256, N269, N488, N598, N712, were studied. We specifically focused on glycans at N488 for both systems as it interacts with

ANaF⁶ on RBD. To improve statistics, trajectories from the previous 75 ns runs² were used for the ACF analysis. Figure 6.7 shows the angle and dihedral motions for both MAN8 and GnGnXF³ at glycosylation sites N219, N269, and N488. ACF results for glycans at all 6 sites are available in Figures S10, S11. Glycans on sites N219 and N269 show typical ACF behaviors of all glycans that do not directly interact with ANaF⁶ on RBD. Comparing the angle motion with dihedral motion for both glycans, ACF_{Angle} decreases significantly whereas ACF_{Dihedral} decrease slowly, indicating that angle motions are more favored for glycans and dihedral motions are constrained (alpha6 at N269 in MAN8 is the only exception where two motions are similarly favored). Comparing ACF of the different linkages, ACF of linkage alpha6 decreases much faster than the 2 beta4 linkages, indicating that linkage alpha6, which is the linkage to the branches, is the most flexible linkage. Comparing ACF of MAN8 and GnGnXF³, ACF_{Angle} and ACF_{Dihedral} of MAN8 decrease either at similar rate or slower than those of GnGnXF³ with very few exceptions (angle: N219_beta4-2, N598_beta4-2 (Figure S10); dihedral: N219_alpha6, N256_beta4-1 (Figure S11), indicating that MAN8 is generally less flexible than GnGnXF³ for the angle and dihedral motions at linkage beta4_1, beta4_2, and alpha6. The glycans at N488 are the ones interacting with ANaF⁶ on RBD. All angle motions and dihedral motions of MAN8 at N488 are less flexible than for GnGnXF³, which further proves that side-by-side hydrogen bonding fashion with ANaF⁶ is favored by MAN8 resulting in more hydrogen bonds between glycans before pulling, whereas a head-to-head arrangement is favored by GnGnXF³ resulting in less hydrogen bonds between glycans before pulling. In addition, the angle motions of glycans at N488 are more constrained than those of glycans at N219, and the dihedral motion of glycans at N488 are more constrained than those of glycans at N269, indicating that glycans at N488 are generally constrained because they are connected to the protein on one end, and interacting with ANaF⁶ on the other end.

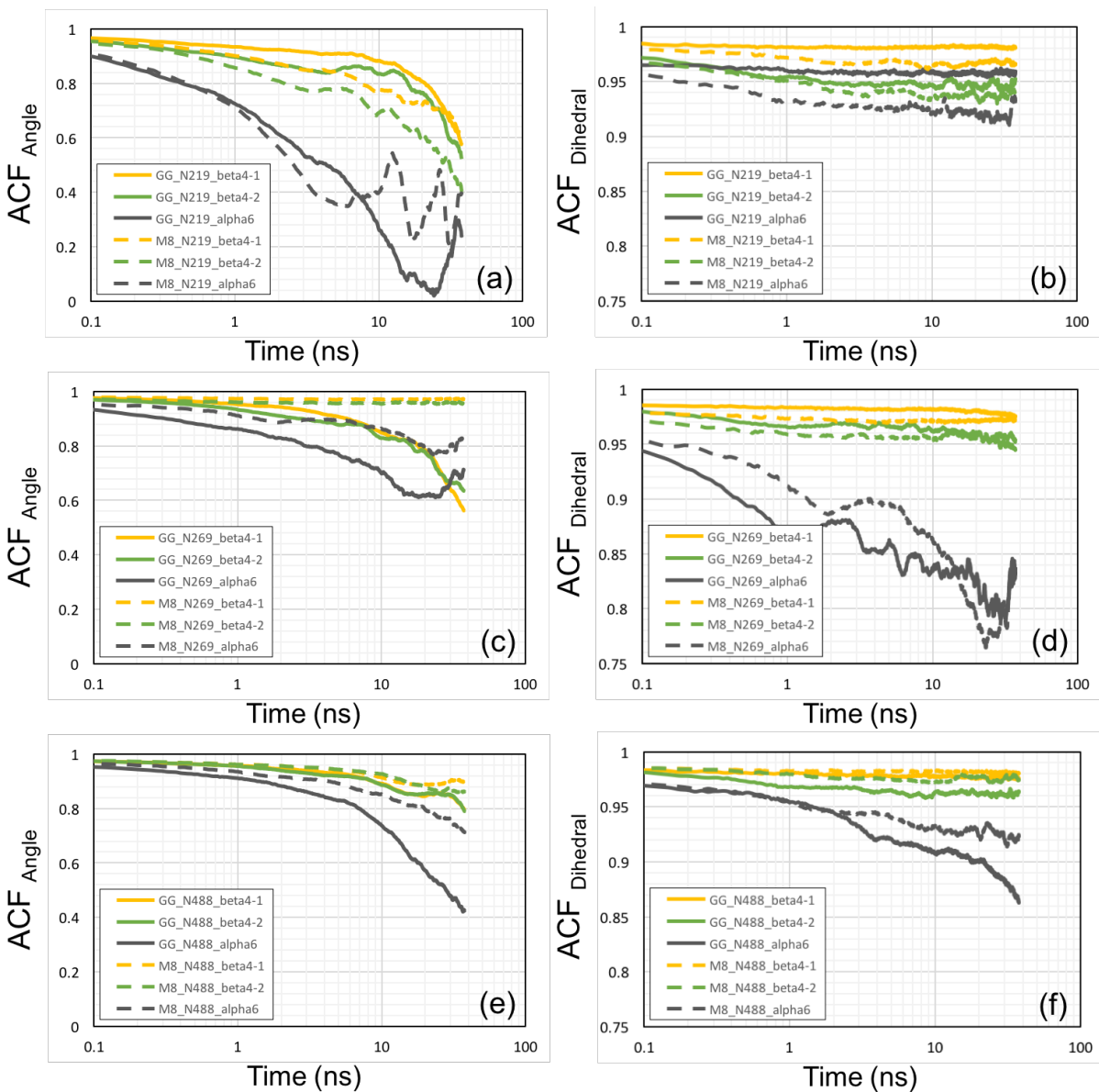


Figure 6.7: Autocorrelation function analysis of angles and dihedrals at linkage beta4₁, beta4₂, and alpha6 for MAN8 and GnGnXF³ at ACE2 fragment glycosylation sites in semi-log lots. Glycans at N219 (a-b) and at N269 (c-d) shows typical behaviors, and glycan at N488 (e-f) are the ones directly interacting with ANaF⁶ on RBD. Dashed lines are the dynamic motions of MAN8, and solid lines are the dynamic motions of GnGnXF³.

In addition to hydrogen bonding, we find that electrostatic and Lennard Jones interactions contribute to the binding between ACE2 and RBD. These interactions are plotted in Figure 6.8 with subplots 8a-d corresponding to the same variants as before. The y-axis corresponds to the interaction energy between the ACE2 and RBD groups with the yellow line corresponding to Coulombic interactions and blue being short range Lennard Jones energies. Interestingly, it appears that at very short distances the electrostatic interaction is more important (more negative interaction potential) than the Lennard Jones interaction; this reverses at intermediate distances (1-2 nm from close contact) where the two lines cross for most of the systems. In some cases, there is a recrossing before the lines essentially merge and the interaction dies out. The glycosylated systems show a similar extension in interaction energies as in the hydrogen bonds, roughly 2-3 nm for the A1Fr^{M8}/SpFr and 1 nm for A2Fr^{GG}/SpFr. A1 variants demonstrate a differently shaped interaction curve than A2 variants for both glycosylated and aglycosylated systems, this can also be attributable to difference in starting orientation and zinc coordination.

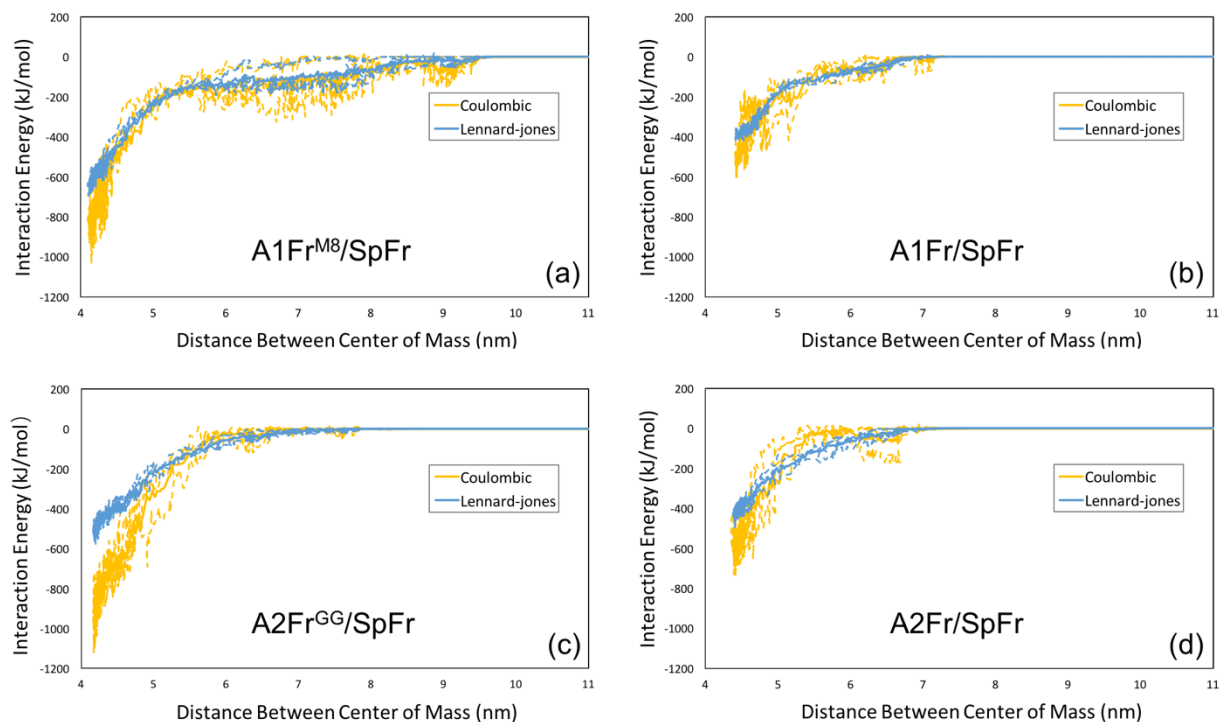


Figure 6.8: Lennard Jones and electrostatic energies. a) Man8 Glycosylated A1Fr^{M8}/SpFr. b) Aglycosylated Structure A1Fr/SpFr. c) GnGnXF Glycosylated A2Fr^{GG}/SpFr. d) Aglycosylated Structure A2Fr/SpFr.

Visual inspection of the starting configurations of the two systems shows a difference in RBD alignment in the binding pocket. To evaluate if this difference was due to a rocking motion of the RBD or was caused by differences in the glycans a principal component analysis (PCA) was performed on the trajectories from our previous publication² to determine the dominant motions of the RBD. Results of the PCA are presented in Figure 6.9 and S12-S15. Figures 6.9a and 6.9b show still structures with arrows indicating direction of projected motion from the dominant principal component. Corresponding video files are available in SI along with time dependence and pair-wise plots of principal components (Figures S12-S15). Figure 9a shows the motion of the

spike fragment from A1Fr^{M8}/SpFr is a scissoring between helices and oscillation of the turn at the top of the structure. Figure 6.9b shows a similar motion, but the oscillation of the turn is missing due to the formation of a helix at that site. This structural change comes from the stable structure after 75 ns simulation due to differences between the glycans and ACE2 interaction. Figures 6.9c and d show cumulative variance vs number of principal components for A1Fr^{M8}/SpFr and A2Fr^{GG}/SpFr respectively. This clearly shows that most of the variance is explained by the first principal component (~90% and ~96% for A1Fr^{M8}/SpFr and A2Fr^{GG}/SpFr, respectively).

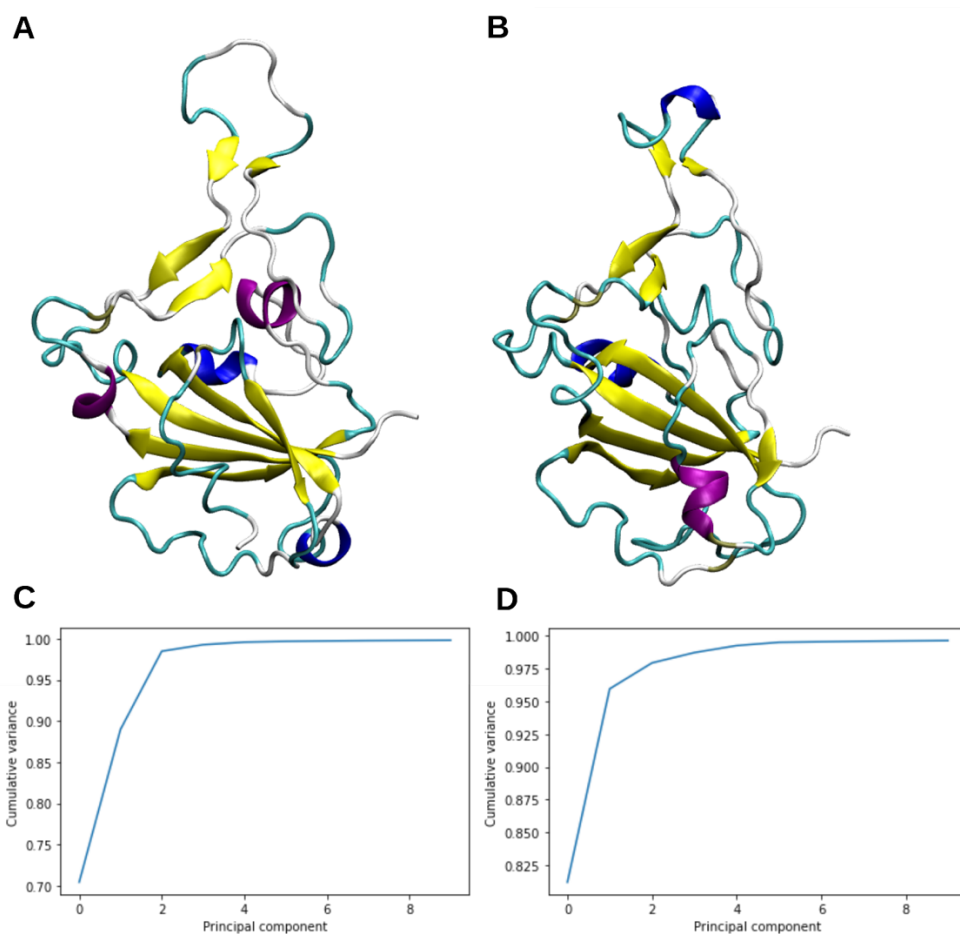


Figure 6.9: First Principal Component (PC1) projected motion and cumulative variance. A) PC1 projected motion for A1Fr^{M8}/SpFr. B) PC1 projected motion for A2Fr^{GG}/SpFr C) Principal component cumulative

variance A1Fr^{M8}/SpFr. D) Principal component cumulative variance A2Fr^{GG}/SpFr. Arrows indicate contraction.

To determine whether changes in binding affinity due to deglycosylation can be observed experimentally, we performed biolayer interferometry using ACE2-Fc and RBD with and without removal of N-glycans. Biolayer interferometry is an optical technique that measures biomolecular interactions by detecting changes in the interference pattern of reflected light from a surface before and after binding⁵⁷. The response is measured as a shift in wavelength in units of nm. Figure 6.10a shows that deglycosylation of proteins via PNGase F treatment results in slightly lower bands on an SDS-PAGE gel, as expected from the smaller protein sizes following glycan removal. We then performed biolayer interferometry on ACE2-Fc and RBD that are either both deglycosylated or glycosylated (Figure 6.10b-d). To do this, ACE2-Fc was immobilized onto a biosensor using the Fc tag and placed in a solution containing the RBD analyte. Steady state analysis was performed on the response using a 1:1 Langmuir binding model, where the response indicates the shift in interference patterns caused by analyte binding (Figure 6.10d). Glycosylated ACE2-Fc and RBD have a binding affinity, K_D , of 30 nM, which is similar to values reported by other groups (34,58). Deglycosylation of ACE2-Fc and RBD results in a 2- to 3-fold increase in binding affinity to 77 nM. From the increase in binding affinity, the magnitude of the binding energy decreases by 2.3 kJ/mol following removal of N-linked glycans. This is consistent with our simulation results that predicts that less pulling force is required to break the protein interactions after deglycosylation.

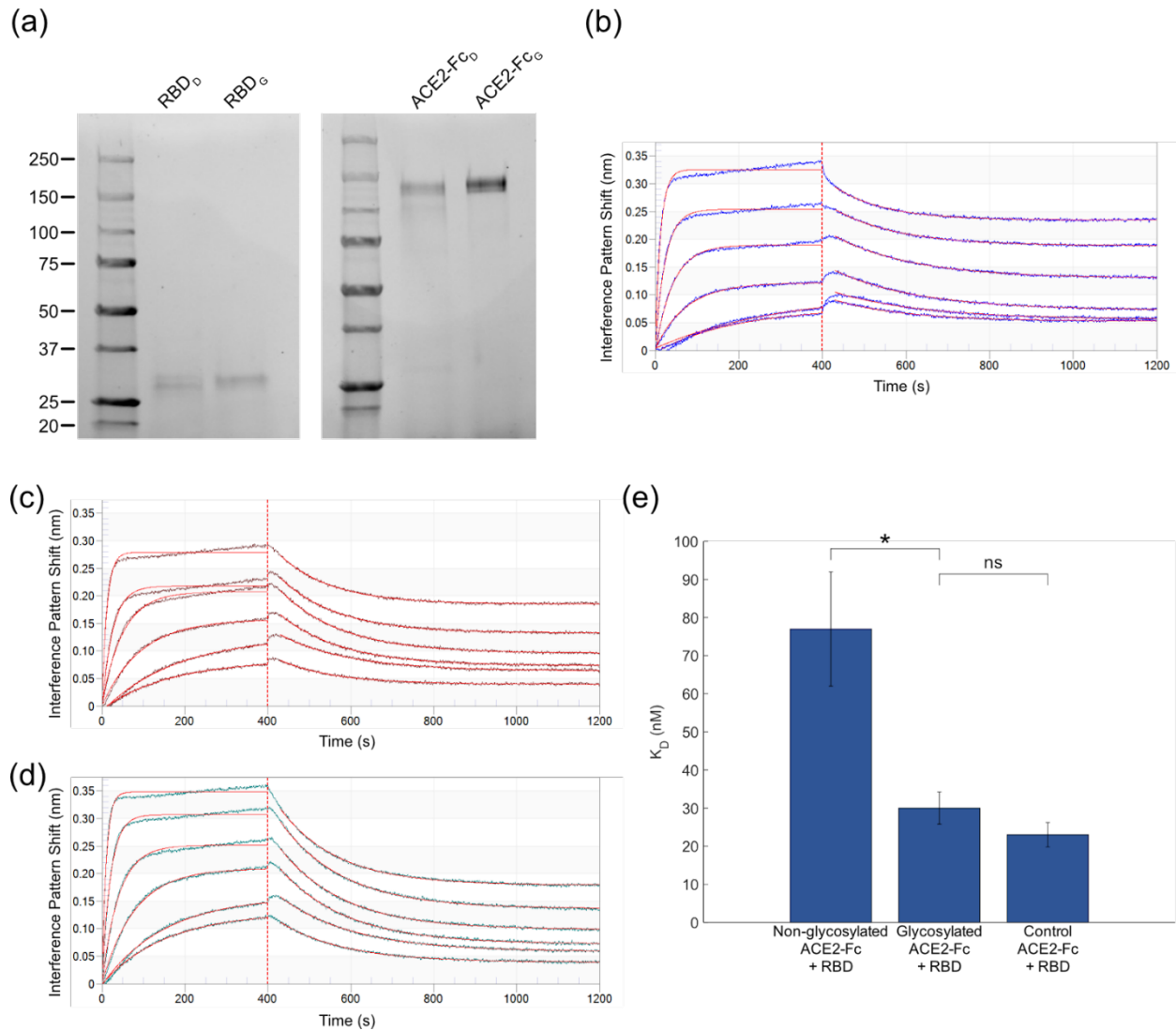


Figure 6.10: Biolayer Interferometry on glycosylated and deglycosylated ACE2-Fc and RBD. (a) SDS-PAGE on ACE2-Fc and RBD with and without PNGase F treatment. A total of 1 μ g of protein is loaded onto each lane. Subscript D indicates deglycosylated proteins, and subscript G indicates glycosylated proteins. (b, c, d) Biolayer interferometry response for (b) deglycosylated ACE2-Fc and RBD, (c) glycosylated ACE2-Fc and RBD, and (d) glycosylated ACE2-Fc and RBD without glyco buffer 2 and incubation at 37°C. Red lines are the fits to the raw data shown in blue, brown, and green, respectively. Error bars represent standard error. * indicates $p < 0.05$. “ns” indicates not significant ($p > 0.05$). Probability values were calculated using one-way ANOVA followed by Tukey’s test.

6.4 Discussions

Detailed mechanistic studies of binding interaction events can improve our understanding of how specific changes to proteins affect binding strength. Differences in binding dissociation rate could have implications in infectivity⁵⁹⁻⁶¹. Viral protein and host receptor interactions are complex due to the interplay between interaction types, different degrees of motion during a binding event, as well as the role of glycans in shielding or strengthening receptor binding. SARS-CoV-2 spike protein and ACE2 interactions are no different. Understanding the implications of different glycans on the binding behavior of spike could prove useful as more variants emerge with potentially different glycosylation patterns. Recent studies have shown experimentally and computationally that the ACE2 and RBD of coronavirus spike fragments have different binding strengths and dissociation rates when they are glycosylated vs non-glycosylated.^{33,34}

Previous computational efforts focused on the binding difference between SARS-CoV-1 and SARS-CoV-2 with glycan interactions modeled by a generic pentasaccharide³². Their analysis focused on the difference in binding strengths and protein contacts between RBD^{CoV1} and RBD^{CoV2}. Our results are in alignment with this trend of stronger interactions caused by the glycans but go further in the analysis of the mechanisms behind this stronger interaction and evaluate more realistic glycan models.

First, our results clearly show that the glycans result in stronger and longer ranged interactions that get extended by a catch-slip mechanism between the glycans, i.e., a hydrogen bond breaks and another one at larger distance takes its place. This catch-slip behavior is clearly seen in the hydrogen bonding maps shown in Figure 6.4. The catch-slip behavior is a result of the original hydrogen bond interactions that are present relaxing and then reforming later. Analysis of A1Fr^{M8}/SpFr in Figure 6.4a clearly shows the relaxation and reformation of glycan contributed

hydrogen bonds. This behavior can be attributed to the increased flexibility of the glycans which increases the ability for these late-stage hydrogen bonds to form due to both increased contacts and increased ability to extend through solution. The different structures of MAN8 and GnGnXF³ also contribute to the different hydrogen bond interactions between an ACE2 glycan and RBD glycan. The flatter MAN8 allows more hydrogen bonds between MAN8 and ANaF⁶, therefore causing more glycan-glycan and glycan-protein interactions during pulling for A1Fr^{M8}/SpFr than for A2Fr^{GG}/SpFr. Angle and dihedrals motions are less flexible for MAN8 than for GnGnXF³, especially for the MAN8 and GnGnXF³ glycans that directly interact with ANaF⁶, proving that MAN8 is more constrained by the hydrogen bonds between MAN8 and ANaF⁶. The hydrogen bond map of A2Fr^{GG}/SpFr in Figure 6.4c shows that there is a present, but less pronounced, hydrogen bond formation between the glycans. The distance extension is seen clearly in the pull force vs center of mass distances (Figure 6.3) as well as the interaction energies vs center of mass distances (Figure 6.8), where the glycosylated structures have their interaction distance extended by as much as 2 nm. This extension can be clearly attributed to the glycans when compared against the hydrogen bond map in Figure 6.4.

Second, an analysis of hydrogen bond occupancy elucidates that the glycans not only result in secondary binding motifs, but also strengthen and extend the existing protein-protein interactions. This is most clearly seen in the % occupancy numbers for the A2Fr^{GG}/SpFr structure, with an increase of several percent in most of the top hydrogen bonds. This trend is also present in A1Fr^{M8}/SpFr when looking at the top protein-protein interactions such as RBD-GLY167:ACE2-LYS519 showing an increase of over 3%. This strengthening of the protein-protein hydrogen bonds may be a result of the extra stabilization in the RBD structure provided by the glycan. That the glycans strengthen the interactions is consistent with our biolayer interferometry results. A

frequent interaction point of interest is the N-glycosylation site ASN90 on ACE2 and GLN409 and THR415 of the spike RBD. Our results suggest a strong interaction in a nearby site ACE2-TYR249 (equivalent to TYR83 in standard numbering) and RBD-ASN152 (equivalent to ASN 487) for all variants studied. This interaction agrees with previous results suggesting a long interaction at this site due to the flexibility of the RBD loop³². It is interesting to note that this interaction is seemingly not affected by the glycan as it pertains to % occupancy.

It is necessary to comment on the difference in starting orientation of the RBD and the ACE2 between the two different starting truncations. By taking the final structure of the simulations from our previous study, it was possible that this resulted in a lower probability starting orientation. A principal component analysis was performed (Fig 10) to verify that the starting structures were truly the dominant orientation from our previous paper and not just an unlucky snapshot of a less favorable state. These results show that the dominant motion from the highest principal component is scissoring of helices and oscillation of a turn and not the rocking of the spike fragment. This suggests that the structure was stable in the ACE2 binding pocket and that the difference in starting structure is due to the differences between glycosylation and the effect of Zn²⁺ on the stability of ACE2. Figures 6.10 a, b clearly show the structural changes resulting from these interactions. These structural changes result in differences in the interaction behavior as seen by a slight 1nm extension of interaction energies as shown in Figure 6.10 b,d.

6.5 Conclusions

We have expanded on our previously developed model of fully glycosylated ACE2-Fc and SARS-CoV-2 spike protein fragments through the investigation of the binding strength and role of glycosylation on binding between these groups. This investigation provides further evidence

that the binding between SARS-CoV-2 spike and ACE2 receptor are aided by the glycosylation on each protein. We found that for multiple complex glycan types the interactions between RBD and ACE2 were strengthened and longer ranged. Protein-protein interactions were extended due to the increased stability provided by the glycans and binding strength is affected by a catch-slip behavior between the glycans. These computational results were corroborated by experimental evidence that the magnitude of the binding energy is decreased for deglycosylated proteins. Further work in analyzing the larger fragments of spike will be necessary for a more realistic model of RBD stability in order to address effects of mutations.

References

1. Liu, Z., X. Xiao, X. Wei, J. Li, J. Yang, H. Tan, J. Zhu, Q. Zhang, J. Wu, and L. Liu. 2020. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *Journal of Medical Virology*. 92(6):595-601.
2. Bernardi, A., Y. Huang, B. Harris, Y. Xiong, S. Nandi, K. A. McDonald, and R. Faller. 2020. Development and simulation of fully glycosylated molecular models of ACE2-Fc fusion proteins and their interaction with the SARS-CoV-2 spike protein binding domain. *PLOS ONE*. 15(8):e0237295.
3. Chan, K. K., D. Dorosky, P. Sharma, S. A. Abbasi, J. M. Dye, D. M. Kranz, A. S. Herbert, and E. Procko. 2020. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science*. 369(6508):1261-1265.
4. Nelson, G., O. Buzko, P. Spilman, K. Niazi, S. Rabizadeh, and P. Soon-Shiong. 2021. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. *bioRxiv*.2021.2001.2013.426558.
5. Mehdipour, A. R., and G. Hummer. 2021. Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. *Proceedings of the National Academy of Sciences*. 118(19):e2100425118.
6. Prajapat, M., N. Shekhar, P. Sarma, P. Avti, S. Singh, H. Kaur, A. Bhattacharyya, S. Kumar, S. Sharma, A. Prakash, and B. Medhi. 2020. Virtual screening and molecular dynamics study of approved drugs as inhibitors of spike protein S1 domain and ACE2 interaction in SARS-CoV-2. *Journal of Molecular Graphics and Modelling*. 101:107716.
7. Davidson, A. M., J. Wysocki, and D. Battle. 2020. Interaction of SARS-CoV-2 and Other Coronavirus With ACE (Angiotensin-Converting Enzyme)-2 as Their Main Receptor. *Hypertension*. 76(5):1339-1349.
8. Vigerust, D. J., and V. L. Shepherd. 2007. Virus glycosylation: role in virulence and immune interactions. *Trends in Microbiology*. 15(5):211-218.
9. Ohtsubo, K., and J. D. Marth. 2006. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*. 126(5):855-867.
10. Phillips, M., E. Nudelman, F. Gaeta, M. Perez, A. Singhal, S. Hakomori, and J. Paulson. 1990. ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand, sialyl-Lex. *Science*. 250(4984):1130-1132.
11. Sperandio, M., C. A. Gleissner, and K. Ley. 2009. Glycosylation in immune cell trafficking. *Immunological Reviews*. 230(1):97-113.
12. Watanabe, Y., T. A. Bowden, I. A. Wilson, and M. Crispin. 2019. Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 1863(10):1480-1497.
13. Banerjee, N., and S. Mukhopadhyay. 2016. Viral glycoproteins: biological role and application in diagnosis. *VirusDisease*. 27(1):1-11.
14. Bagdonaite, I., and H. H. Wandall. 2018. Global aspects of viral glycosylation. *Glycobiology*. 28(7):443-467.
15. Thompson, A. J., R. P. de Vries, and J. C. Paulson. 2019. Virus recognition of glycan receptors. *Current Opinion in Virology*. 34:117-129.

16. Watanabe, Y., J. D. Allen, D. Wrapp, J. S. McLellan, and M. Crispin. 2020. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*. eabb9983.
17. Andersen, K. G., A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry. 2020. The proximal origin of SARS-CoV-2. *Nature Medicine*. 26(4):450-452.
18. Walls, A. C., Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veelsler. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 181(2):281-292.e286.
19. Sanda, M., L. Morrison, and R. Goldman. 2021. N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Analytical Chemistry*. 93(4):2003-2009.
20. Fernández, A. 2021. Glycosylation of SARS-CoV-2 Steers Evolutionary Outcomes in the Postvaccination Phase. *ACS Pharmacology & Translational Science*. 4(1):410-412.
21. Allen, J. D., Y. Watanabe, H. Chawla, M. L. Newby, and M. Crispin. 2021. Subtle Influence of ACE2 Glycan Processing on SARS-CoV-2 Recognition. *Journal of Molecular Biology*. 433(4):166762.
22. Nguyen, K., S. Chakraborty, R. A. Mansbach, B. Korber, and S. Gnanakaran. 2021. Exploring the Role of Glycans in the Interaction of SARS-CoV-2 RBD and Human Receptor ACE2. *Viruses*. 13(5):927.
23. Sztain, T., S.-H. Ahn, A. T. Bogetti, L. Casalino, J. A. Goldsmith, E. Seitz, R. S. McCool, F. L. Kearns, F. Acosta-Reyes, S. Maji, G. Mashayekhi, J. A. McCammon, A. Ourmazd, J. Frank, J. S. McLellan, L. T. Chong, and R. E. Amaro. 2021. A glycan gate controls opening of the SARS-CoV-2 spike protein. *bioRxiv*.2021.2002.2015.431212.
24. Woo, H., S.-J. Park, Y. K. Choi, T. Park, M. Tanveer, Y. Cao, N. R. Kern, J. Lee, M. S. Yeom, T. I. Croll, C. Seok, and W. Im. 2020. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *The Journal of Physical Chemistry B*. 124(33):7128-7137.
25. Choi, Y. K., Y. Cao, M. Frank, H. Woo, S.-J. Park, M. S. Yeom, T. I. Croll, C. Seok, and W. Im. 2021. Structure, Dynamics, Receptor Binding, and Antibody Binding of the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane. *Journal of Chemical Theory and Computation*. 17(4):2479-2487.
26. Casalino, L., Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, and R. E. Amaro. 2020. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science*. 6(10):1722-1734.
27. Vuorio, J., I. Vattulainen, and H. Martinez-Seara. 2017. Atomistic fingerprint of hyaluronan–CD44 binding. *PLOS Computational Biology*. 13(7):e1005663.
28. Vuorio, J., J. Škerlová, M. Fábry, V. Veverka, I. Vattulainen, P. Řezáčová, and H. Martinez-Seara. 2021. N-Glycosylation can selectively block or foster different receptor–ligand binding modes. *Scientific Reports*. 11(1):5239.
29. Bernardi, A., K. N. Kirschner, and R. Faller. 2017. Structural analysis of human glycoprotein butyrylcholinesterase using atomistic molecular dynamics: The importance of glycosylation site ASN241. *PLOS ONE*. 12(11):e0187994.
30. Xiong, Y., K. Karuppanan, A. Bernardi, Q. Li, V. Kommineni, A. M. Dandekar, C. B. Lebrilla, R. Faller, K. A. McDonald, and S. Nandi. 2019. Effects of N-Glycosylation on the Structure, Function, and Stability of a Plant-Made Fc-Fusion Anthrax Decoy Protein. *Frontiers in Plant Science*. 10(768).

31. Lee, H. S., Y. Qi, and W. Im. 2015. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Scientific Reports*. 5(1):8926.
32. Cao, W., C. Dong, S. Kim, D. Hou, W. Tai, L. Du, W. Im, and X. F. Zhang. 2021. Biomechanical characterization of SARS-CoV-2 spike RBD and human ACE2 protein-protein interaction. *Biophys J*. 120(6):1011-1019.
33. Shang, J., G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, and F. Li. 2020. Structural basis of receptor recognition by SARS-CoV-2. *Nature*. 581(7807):221-224.
34. Yang, J., S. J. L. Petitjean, M. Koehler, Q. Zhang, A. C. Dumitru, W. Chen, S. Derclaye, S. P. Vincent, P. Soumillion, and D. Alsteens. 2020. Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor. *Nature Communications*. 11(1):4541.
35. Gomord, V., P. Chamberlain, R. Jefferis, and L. Faye. 2005. Biopharmaceutical production in plants: problems, solutions and opportunities. *Trends in Biotechnology*. 23(11):559-565.
36. Castilho, A., L. Neumann, P. Gattinger, R. Strasser, K. Vorauer-Uhl, T. Sterovsky, F. Altmann, and H. Steinkellner. 2013. Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants. *PLOS ONE*. 8(1):e54836.
37. De Meyer, T., and A. Depicker. 2014. Trafficking of endoplasmic reticulum-retained recombinant proteins is unpredictable in *Arabidopsis thaliana*. *Frontiers in Plant Science*. 5(473).
38. Izrailev, S., S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten. 1998. Steered Molecular Dynamics. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*. P. Deuffhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel, editors. Springer, Berlin.
39. Yan, R., Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. 367(6485):1444-1448.
40. Ferrara, C., S. Grau, C. Jäger, P. Sondermann, P. Brünker, I. Waldhauer, M. Hennig, A. Ruf, A. C. Rufer, M. Stihle, P. Umaña, and J. Benz. 2011. Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. *Proceedings of the National Academy of Sciences*. 108(31):12669-12674.
41. Towler, P., B. Staker, S. G. Prasad, S. Menon, J. Tang, T. Parsons, D. Ryan, M. Fisher, D. Williams, N. A. Dales, M. A. Patane, and M. W. Pantoliano. 2004. ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis *. *J Biol Chem*. 279(17):17996-18007.
42. Padjasek, M., A. Kocyła, K. Kluska, O. Kerber, J. B. Tran, and A. Krężel. 2020. Structural zinc binding sites shaped for greater works: Structure-function relations in classical zinc finger, hook and clasp domains. *Journal of Inorganic Biochemistry*. 204:110955.
43. Gamsjaeger, R., C. K. Liew, F. E. Loughlin, M. Crossley, and J. P. Mackay. 2007. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends in Biochemical Sciences*. 32(2):63-70.
44. Case, D. A., R. M. Betz, D. S. Cerutti, I. T.E. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A.

- Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, and P. A. Kollman (2016). AMBER 2016. University of California, San Francisco.
45. Woods Group. 2005-2020. 2020.
 46. Bernardi, A. 2020.
 47. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*. 126(1):014101.
 48. Parinello, M., and A. Rahman. 1982. Strain fluctuations and elastic constants. *J. Chem. Phys.* 76(5):2662-2666.
 49. Miyamoto, S., and P. A. Kollman. 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*. 13(8):952-962.
 50. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem*. 18(12):1463-1472.
 51. Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 1-2:19-25.
 52. Nosé, S. 1984. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*. 81(1):511-519.
 53. Hunter, J. D. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. 9(3):90-95, doi: 10.1109/MCSE.2007.55.
 54. Reinhardt, M., and H. Grubmüller. 2021. GROMACS implementation of free energy calculations with non-pairwise Variationally derived Intermediates. *Computer Physics Communications*. 264:107931.
 55. Savitzky, A., and M. J. E. Golay. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*. 36(8):1627-1639.
 56. Ponder, J. W., and D. A. Case. 2003. Force Fields for Protein Simulations. In *Advances in Protein Chemistry*. Academic Press, pp. 27-85.
 57. Abdiche, Y., D. Malashock, A. Pinkerton, and J. Pons. 2008. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Analytical Biochemistry*. 377(2):209-217.
 58. Bayarri-Olmos, R., A. Rosbjerg, L. B. Johnsen, C. Helgstrand, T. Bak-Thomsen, P. Garred, and M.-O. Skjoedt. 2021. The SARS-CoV-2 Y453F mink variant displays a pronounced increase in ACE-2 affinity but does not challenge antibody neutralization. *J Biol Chem*. 296, doi: 10.1016/j.jbc.2021.100536.
 59. Ozono, S., Y. Zhang, H. Ode, K. Sano, T. S. Tan, K. Imai, K. Miyoshi, S. Kishigami, T. Ueno, Y. Iwatani, T. Suzuki, and K. Tokunaga. 2021. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nature Communications*. 12(1):848.
 60. Liu, H., Q. Zhang, P. Wei, Z. Chen, K. Aviszus, J. Yang, W. Downing, C. Jiang, B. Liang, L. Reynoso, G. P. Downey, S. K. Frankel, J. Kappler, P. Marrack, and G. Zhang. 2021. The basis of a more contagious 501Y.V1 variant of SARS-CoV-2. *Cell Research*. 31(6):720-722.
 61. Steckbeck, J. D., I. Orlov, A. Chow, H. Grieser, K. Miller, J. Bruno, J. E. Robinson, R. C. Montelaro, and K. S. Cole. 2005. Kinetic Rates of Antibody Binding Correlate with

Neutralization Sensitivity of Variant Simian Immunodeficiency Virus Strains. *Journal of Virology*. 79(19):12311-12320.

6.6 Supporting Information

6.6.1 Pull Force vs Pull distance

As described in the manuscript, we analyzed pulling speeds of 10 nm/ns, 5nm/ns, and 1 nm/ns. Here we present the pull force vs pull distance plots for these three different pull rates. As expected pull force increases with pulling rate (blue, orange, green), and the peak force is clearly higher in the glycosylated states by roughly 250, 500, and 600 kJ /mol /nm.

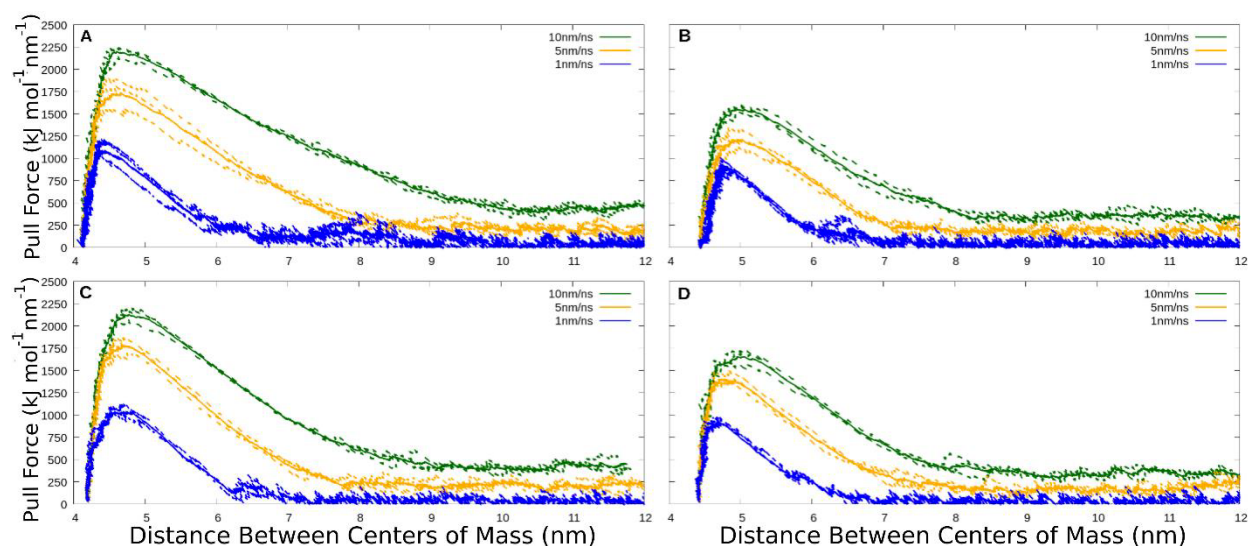


Figure S1: Traces of pull force versus distance. A) Man8 Glycosylated A1Fr^{M8}/SpFr. B) Non-glycosylated Structure A1Fr/SpFr. C) GnGnXF Glycosylated A2Fr^{GG}/SpFr. D) Non-glycosylated Structure A2Fr/SpFr. Different colors indicate different pull rate. Dashed lines are individual replicas, solid lines are averages.

6.6.2 Hydrogen Bonding Script

Hydrogen bonding maps are generated from a python script using external packages, numpy, pandas, matplotlib, gromacs and seaborn. The gromacs python package is used to load the .xpm bitmap generated from the gromacs hbond command and save it as a python array. The log and index files from the gromacs hbond command are then sorted and used to generate labels

for the previously generated array. For plotting simplicity, the corresponding array and labels are converted to a pandas data frame and plotted using the seaborn heatmap. The transformations made before plotting include a % occupancy calculation attained by calculating the number of 1s in the array divided by the number of columns in the row and multiplied by 100, and some conditional data frame rearrangement based on interaction type. This rearrangement was used to generate different colors for each interaction type by either multiplying values by -1 or 2 depending on the interaction involved.

6.6.3 Hydrogen Bonding Maps

As described in the manuscript we analyzed the hydrogen bonding interactions between ACE2 and Spike RBD proteins. Interactions were calculated and mapped as described above. Manuscript figure 4 was shown without labels due to lack of available resolution. Full scale images with labels are shown below in Figure S1-S4. Data is from 1 nm/ns pull rate for each system. Colors indicate interaction type. White: protein-protein, yellow: protein-glycan, and magenta: glycan-glycan.

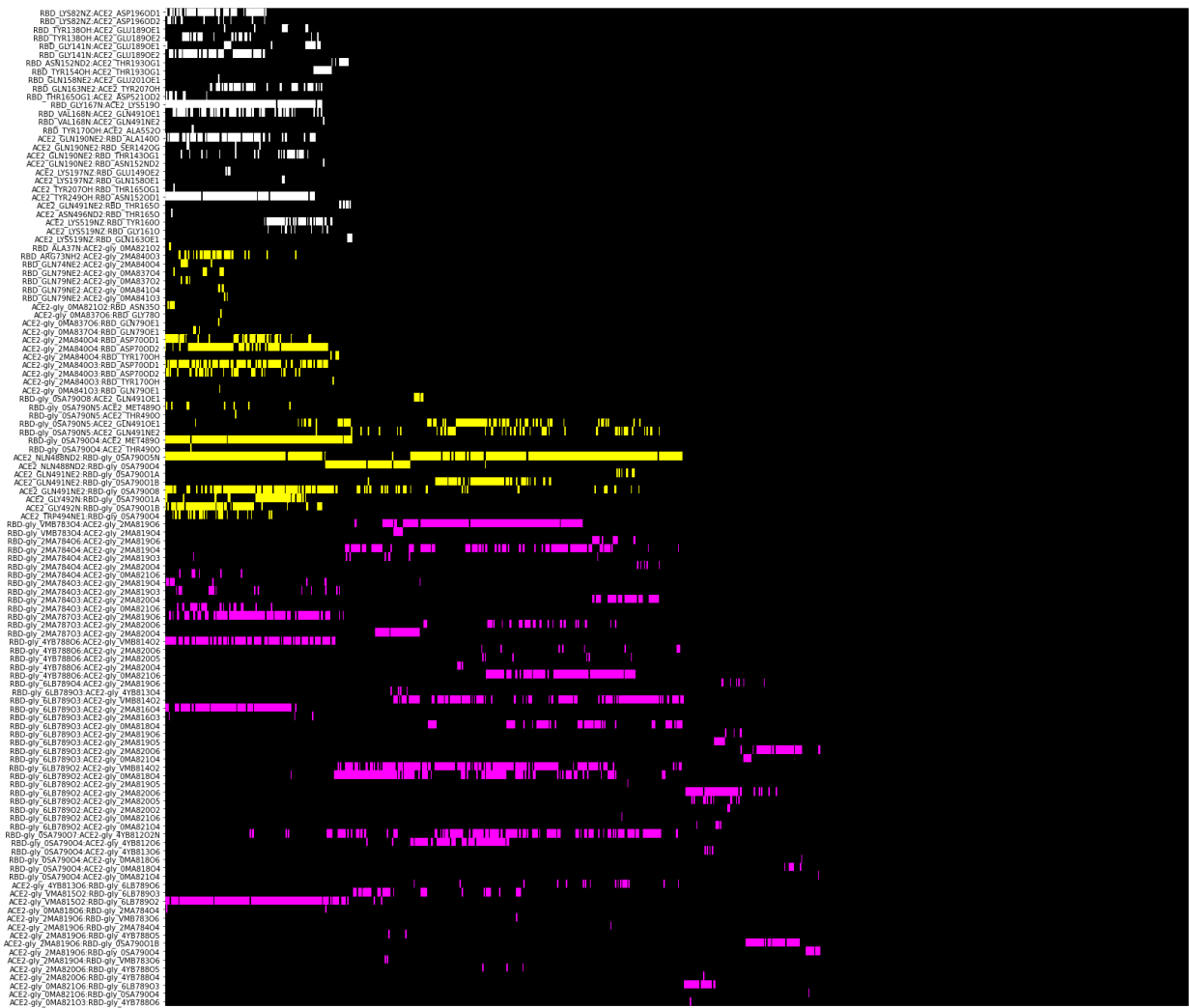


Figure S2. Hydrogen bonding donor:acceptor pairs vs simulated time for A1Fr^{M8}/SpFr. 1 ns / nm pulling speed. Colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.

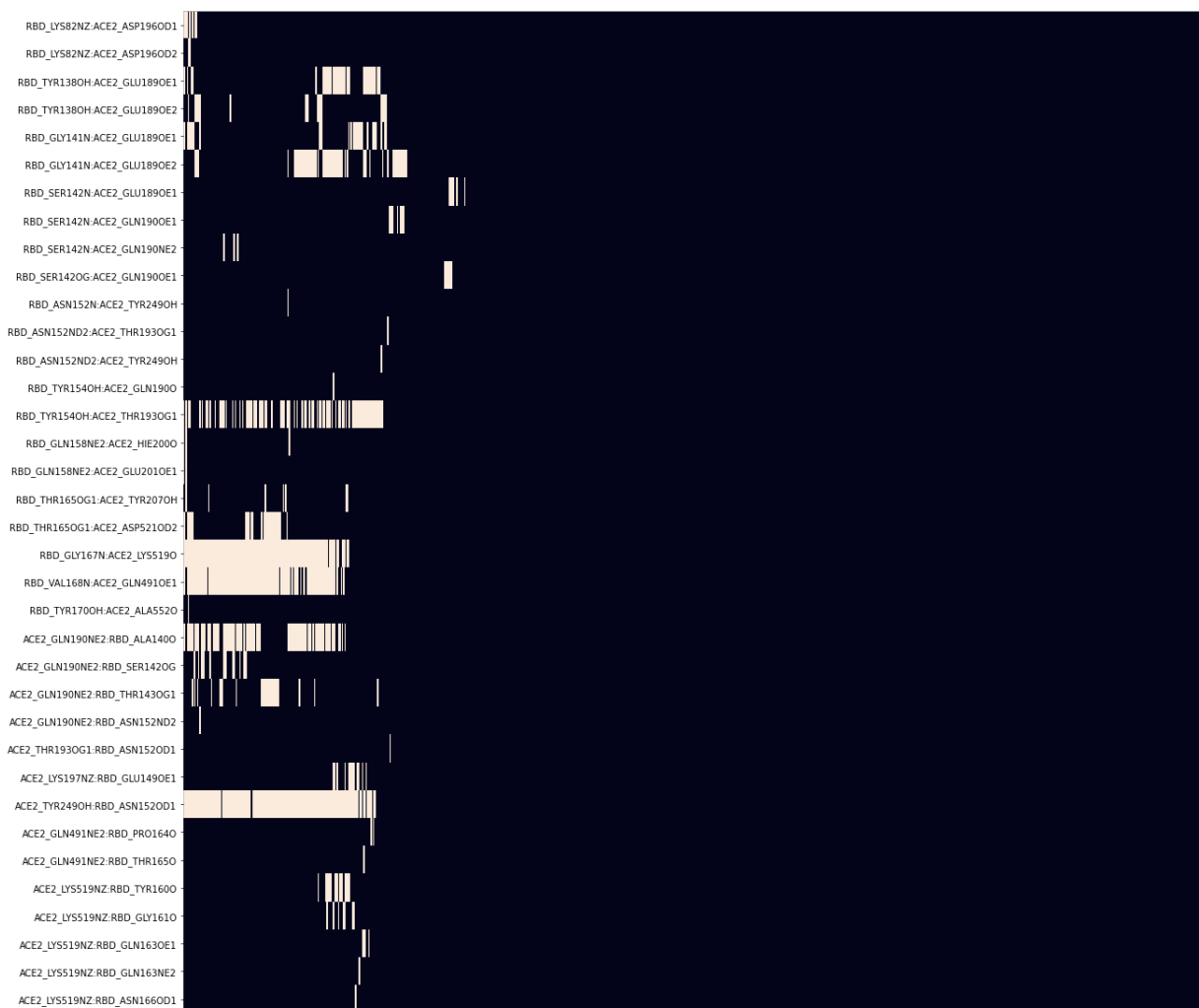


Figure S3. Hydrogen bonding donor:acceptor pairs vs simulated time for A1Fr/SpFr. 1 ns / nm pulling speed. Colors indicate interaction type: White: protein-protein.

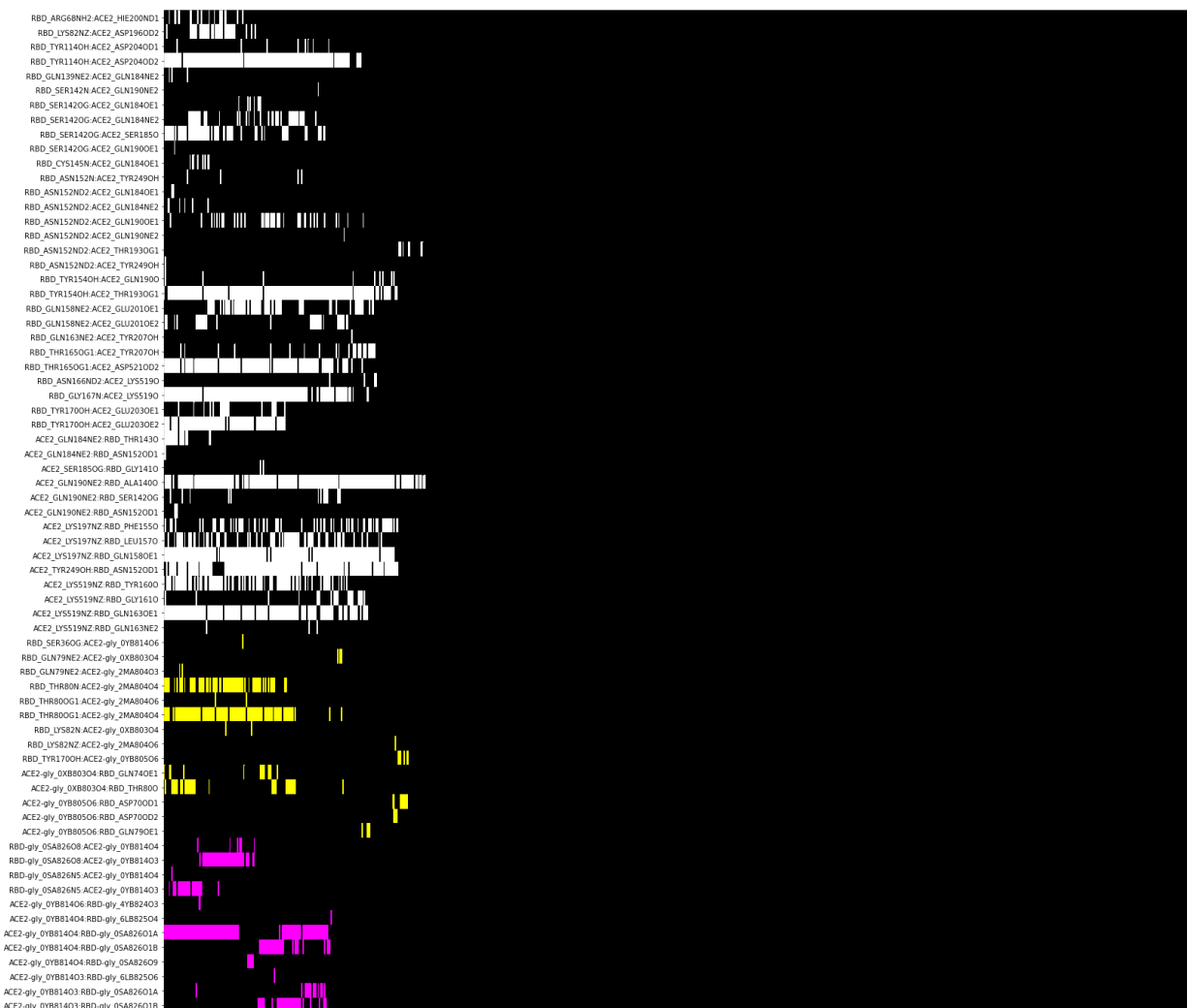


Figure S4. Hydrogen bonding donor:acceptor pairs vs simulated time for A2Fr^{GG}/SpFr. 1 ns / nm pulling speed. Colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.

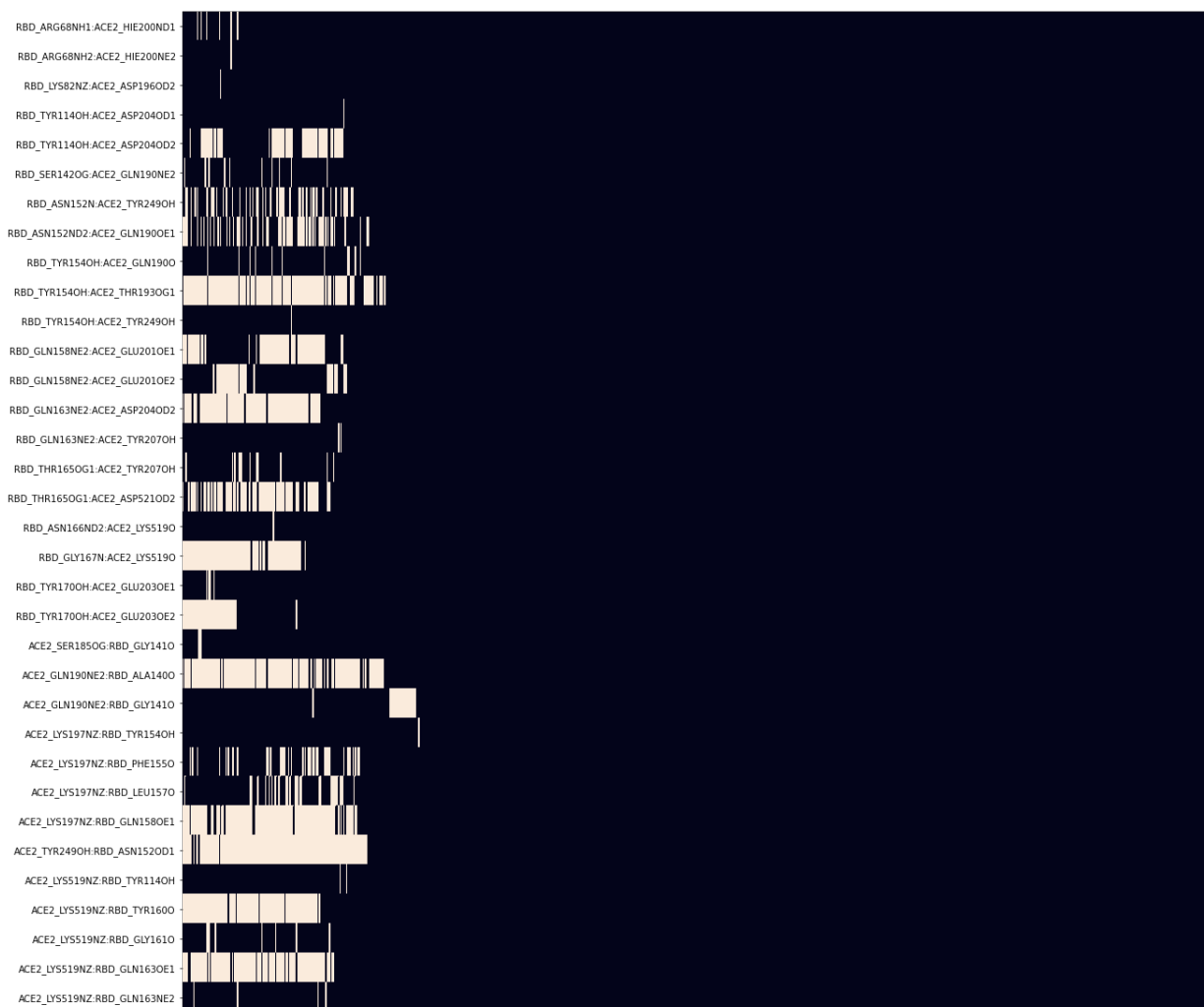


Figure S5. Hydrogen bonding donor:acceptor pairs vs simulated time for A2Fr/SpFr. 1 ns / nm pulling speed. Colors indicate interaction type: White: protein-protein.

6.6.4 Hydrogen Bond Occupancy

As described in the manuscript, the top 25 hydrogen bonds by occupancy were listed for all four system, and the top 5 donor (blue):acceptor (pink) pairs by occupancy were highlighted in the four configurations where RBD with and without AnaF6 started to be pulled away from the

ACE2 fragment. Figure S5-S8 correspond to A1Fr^{M8}/SpFr, A1Fr/SpFr, A2Fr^{GG}/SpFr, A2Fr/SpFr respectively.

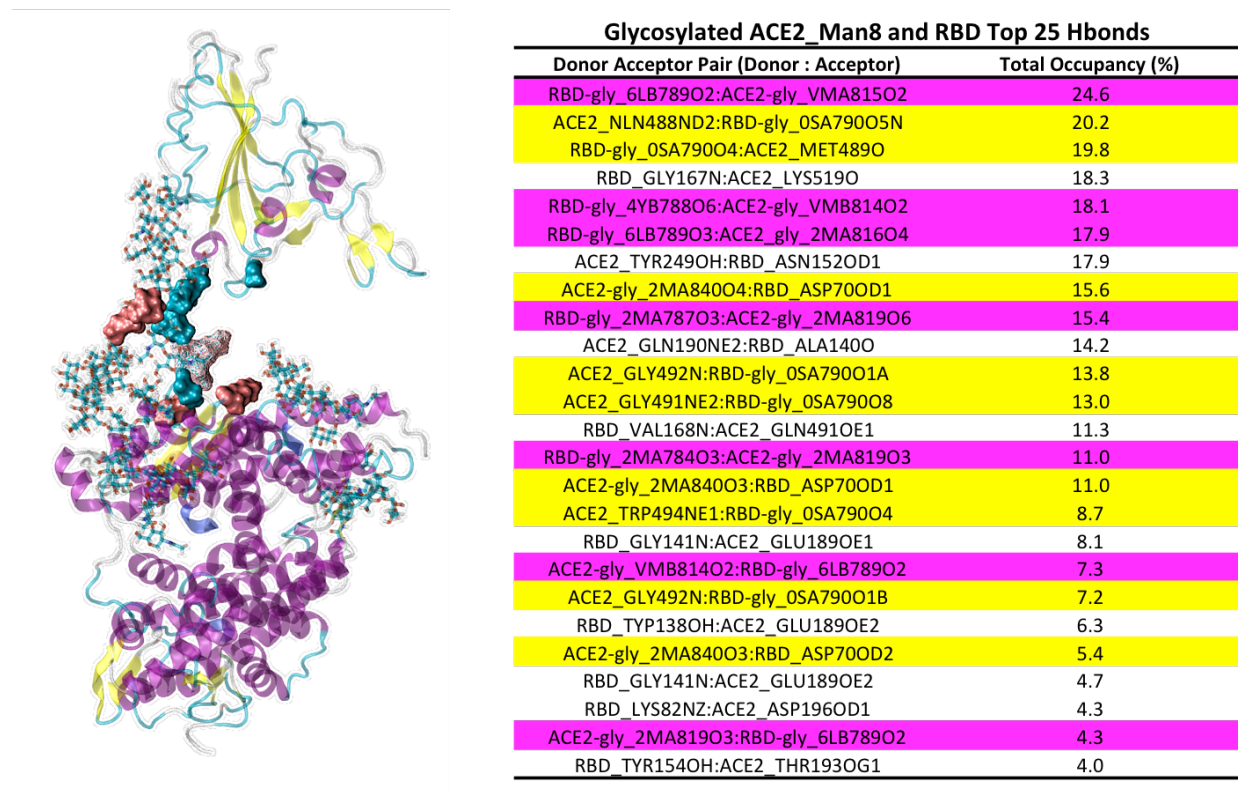
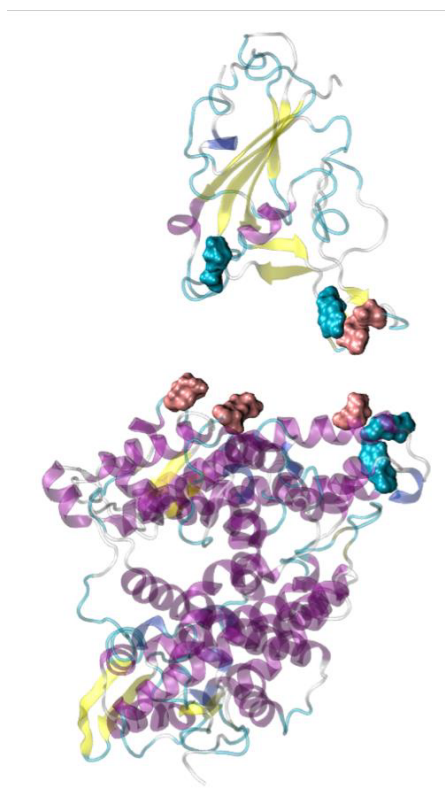


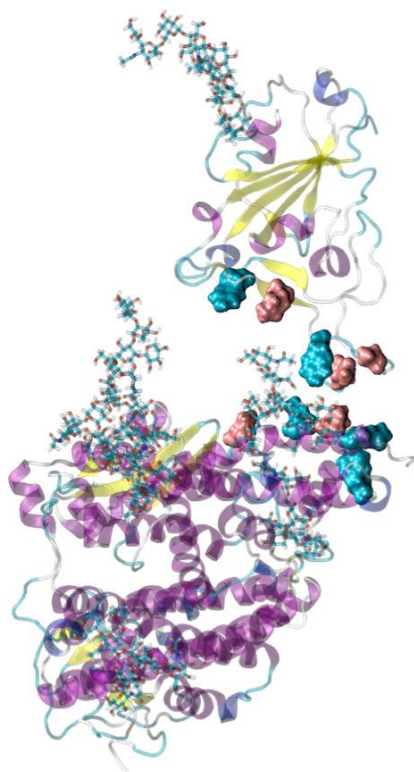
Figure S6. Hydrogen bond donor:acceptor pairs and occupancy for Man8 glycosylated A1Fr^{M8}/SpFr. Table colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.



NonGlycosylated ACE2_MAN8 and RBD Top 25 Hbonds

Donor Acceptor Pair (Donor : Acceptor)	Total Occupancy (%)
ACE2_TYR249OH:RBD_ASN152OD1	18.0
RBD_GLY167N:ACE2_LYS519O	15.6
RBD_VAL168N:ACE2_GLN491OE1	14.2
RBD_TYR154OH:ACE2_THR193OG1	11.2
ACE2_GLN190NE2:RBD_ALA140O	10.5
RBD_GLY141N:ACE2_GLU189OE2	7.4
RBD_TYR138OH:ACE2_GLU189OE1	4.8
RBD_GLY141N:ACE2_GLU189OE1	3.7
RBD_THR165OG1:ACE2_ASP521OD2	3.6
ACE2_GLN190NE2:RBD_THR143OG1	3.1
RBD_TYR138OH:ACE2_GLU189OE2	2.5
ACE2_LYS519NZ:RBD_TYR160O	2.1
ACE2_GLN190NE2:RBD_SER142OG	2.0
ACE2_LYS197NZ:RBD_GLU149OE1	1.6
RBD_SER142N:ACE2_GLN190OE1	1.1
RBD_LYS82NZ:ACE2_ASP196OD1	1.0
RBD_SER142OG:ACE2_GLN190OE1	0.9
RBD_THR165OG1:ACE2_TYR207OH	0.9
ACE2_LYS519NZ:RBD_GLY161O	0.9
RBD_SER142N:ACE2_GLU189OE1	0.7
ACE2_LYS519NZ:RBD_GLN163OE1	0.5
RBD_SER142N:ACE2_GLN190NE2	0.4
RBD_GLN158NE2:ACE2_HIE200O	0.4
RBD_LYS82NZ:ACE2_ASP196OD2	0.2
RBD_GLN158NE2:ACE2_GLU201OE1	0.2

Figure S7. Hydrogen bond donor:acceptor pairs and occupancy for non-glycosylated structure A1Fr/SpFr. Table colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.



Glycosylated ACE2_GnGnXF and RBD Top 25 Hbonds	
Donor Acceptor Pair (Donor : Acceptor)	Total Occupancy (%)
ACE2_GLN190NE2:RBD_ALA140O	23.1
ACE2_LYS197NZ:RBD_GLN158OE1	21.6
RBD_TYR154OH:ACE2_THR193OG1	21.2
ACE2_TYR249OH:RBD_ASN152OD1	20.2
RBD_TYR114OH:ACE2_ASP204OD2	18.2
RBD_GLY167N:ACE2_LYS519O	17.4
ACE2_LYS519NZ:RBD_GLN163OE1	17.2
RBD_THR165OG1:ACE2_ASP521OD2	16.2
ACE2-gly_0YB814O4:RBD-gly_0SA826O1A	11.9
RBD_THR80OG1:ACE2-gly_2MA804O4	11.7
RBD_TYR170OH:ACE2_GLU203OE2	11.0
ACE2_LYS519NZ:RBD_TYR160O	11.0
ACE2_LYS197NZ:RBD_LEU157O	8.6
RBD_THR80N:ACE2-gly_2MA804O4	8.0
ACE2_LYS197NZ:RBD_PHE155O	7.9
RBD_GLN158NE2:ACE2_GLU201OE1	7.7
RBD_SER142OG:ACE2_SER185O	7.6
RBD_SER142OG:ACE2_GLN184NE2	5.0
RBD-gly_0SA826O8:ACE2-gly_0YB814O3	4.9
RBD_ASN152ND2:ACE2_GLN190OE1	4.7
RBD_LYS82NZ:ACE2_ASP196OD2	4.4
RBD_GLN158NE2:ACE2_GLU201OE2	4.1
ACE2-gly_0YB814O3:RBD-gly_0SA826O1B	4.0
ACE2-gly_0XB803O4:RBD_THR80O	3.9
ACE2-gly_0YB814O4:RBD-gly_0SA826O1B	3.6

Figure S8. Hydrogen bond donor:acceptor pairs and occupancy for GnGnXF³ glycosylated A2Fr^{GG}/SpFr. Table colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.

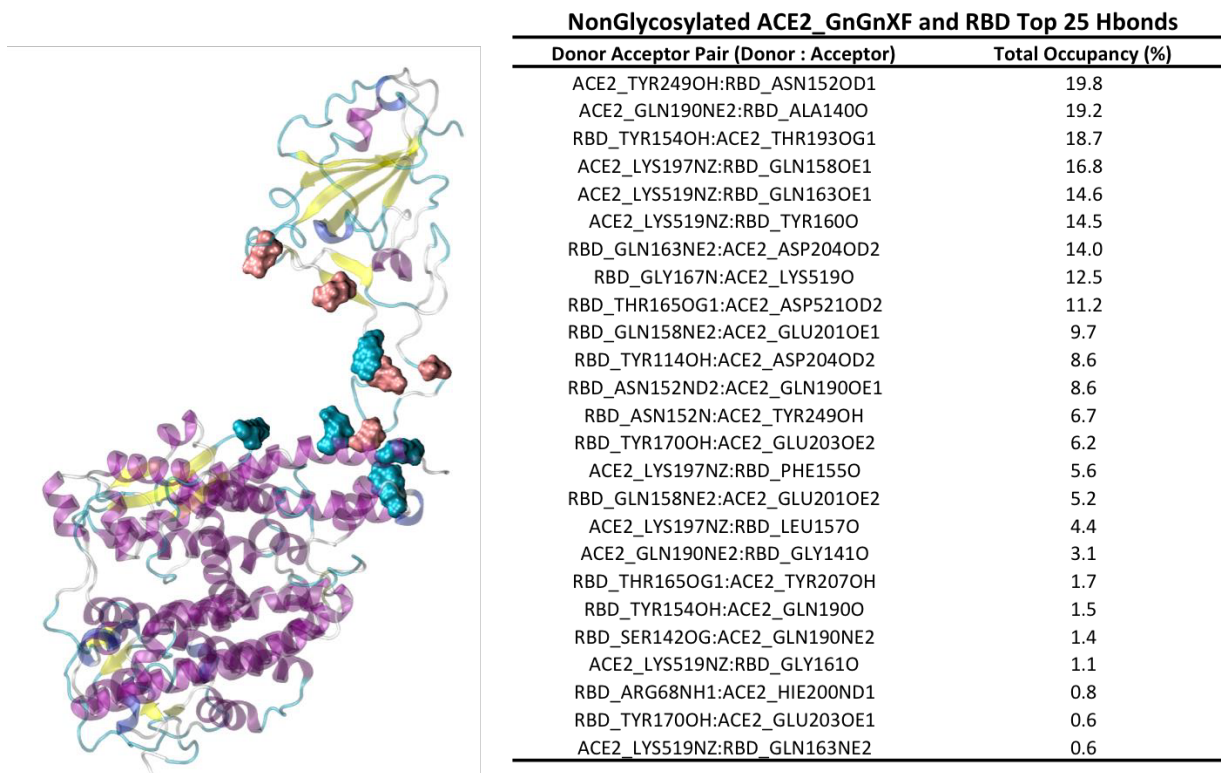


Figure S9. Hydrogen bond donor:acceptor pairs and occupancy for non-glycosylated structure A2Fr/SpFr. Table colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.

6.6.5 Angle Correlation Functions and Dihedral Correlation Functions

As discussed in the manuscript, both glycosylated systems A1Fr^{M8}/SpFr and A2Fr^{GG}/SpFr have 6 glycosylation sites on the ACE2 fragment: N219, N256, N269, N488, N598, N712. Angle autocorrelation functions (ACF) and dihedral autocorrelation functions were calculated at glycan linkages beta4_1, beta4_2, and alpha6 at all 6 glycosylation sites for both systems. Figure S9 shows all the angle ACF semi-log plots, and Figure S10 shows all the dihedral ACF semi-log plots.

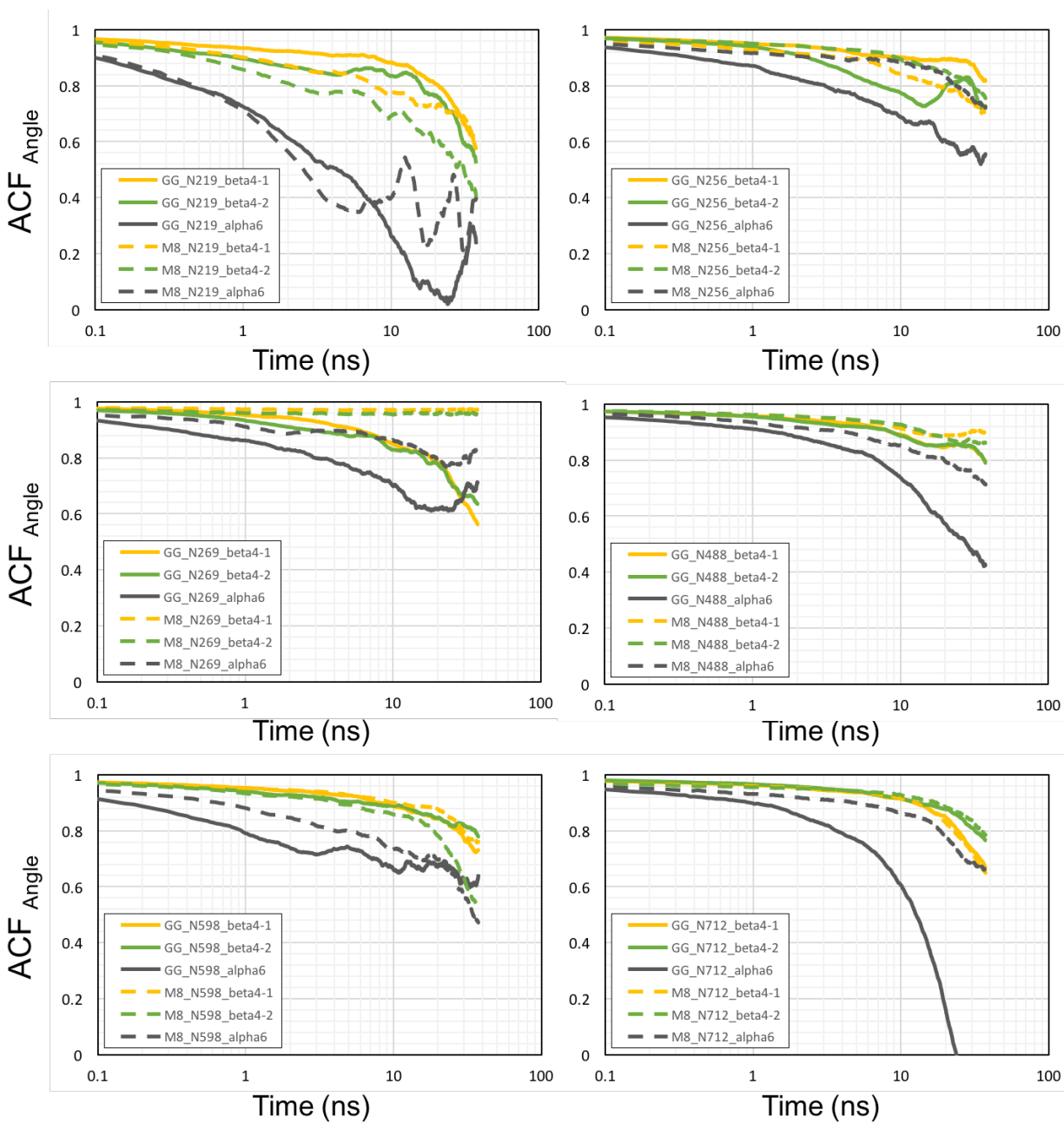


Figure S10. Autocorrelation function analysis of angles at linkage beta4_1, beta4_2, and alpha6 of MAN8 and GnGnXF³ at ACE2 fragment glycosylation sites in semi-log plots. Dashed lines are the dynamic motions of MAN8, and solid lines are the dynamic motions of GnGnXF³.

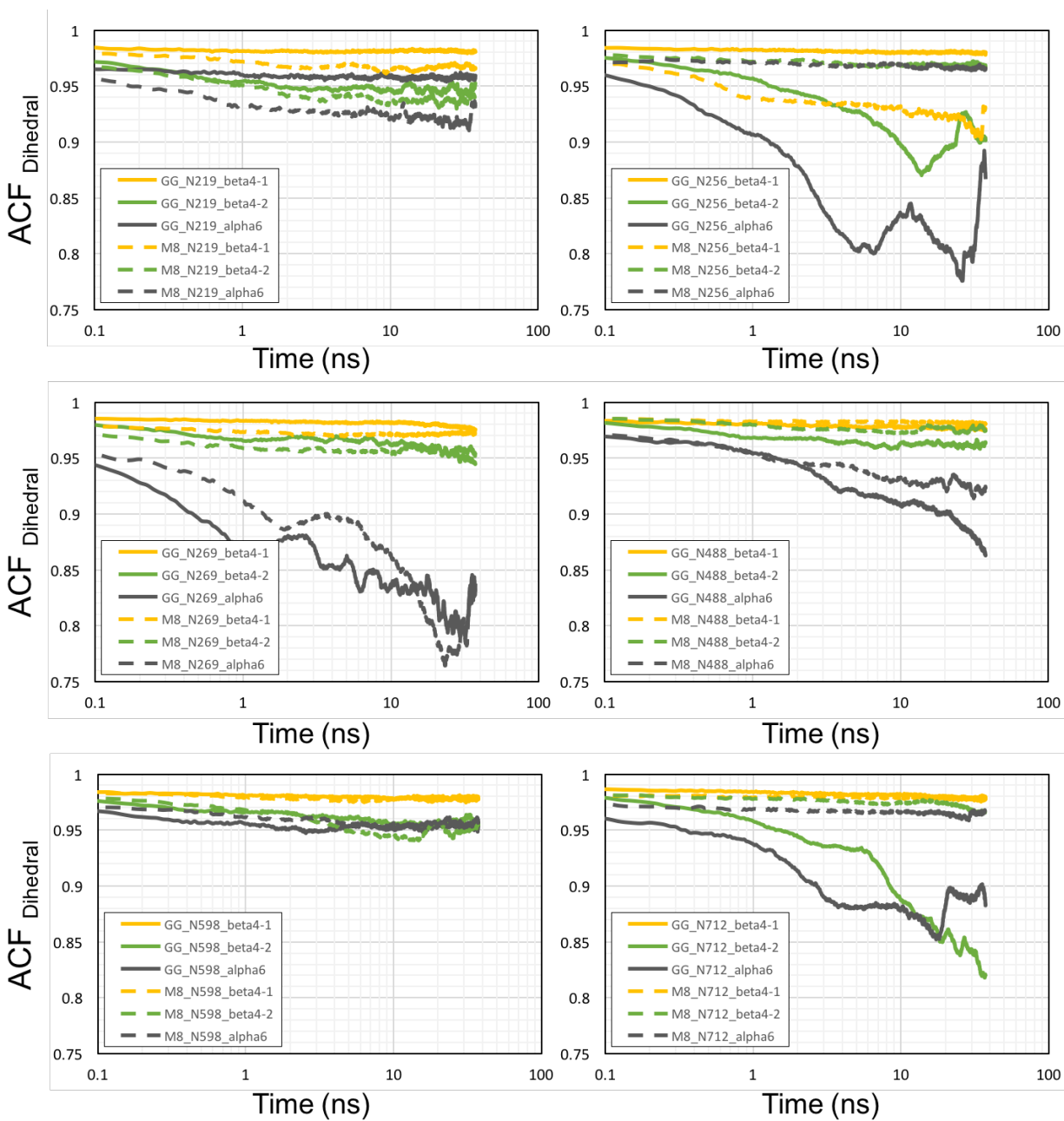


Figure S11. Autocorrelation function analysis of dihedrals at linkage beta4_1, beta4_2, and alpha6 of MAN8 and GnGnXF³ at ACE2 fragment glycosylation sites in semi-log plots. Dashed lines are the dynamic motions of MAN8, and solid lines are the dynamic motions of GnGnXF³.

6.6.6 Principal Components

As described in the manuscript, PCA was performed on the trajectories from our previous publication to determine the dominant motion of the RBD. The results in the manuscript show that roughly 90% and 95% of the variance in the motion was explained by the first component for the A1 and A2 variants respectively. Here we present the first 5 components, responsible for over 99% of the variance in both systems as a pair-wise interaction plot in Figure S11 and S12 for A1 and A2 respectively, and as PC vs time in figure S13 and S14. Simulation videos are included as supplementary files.

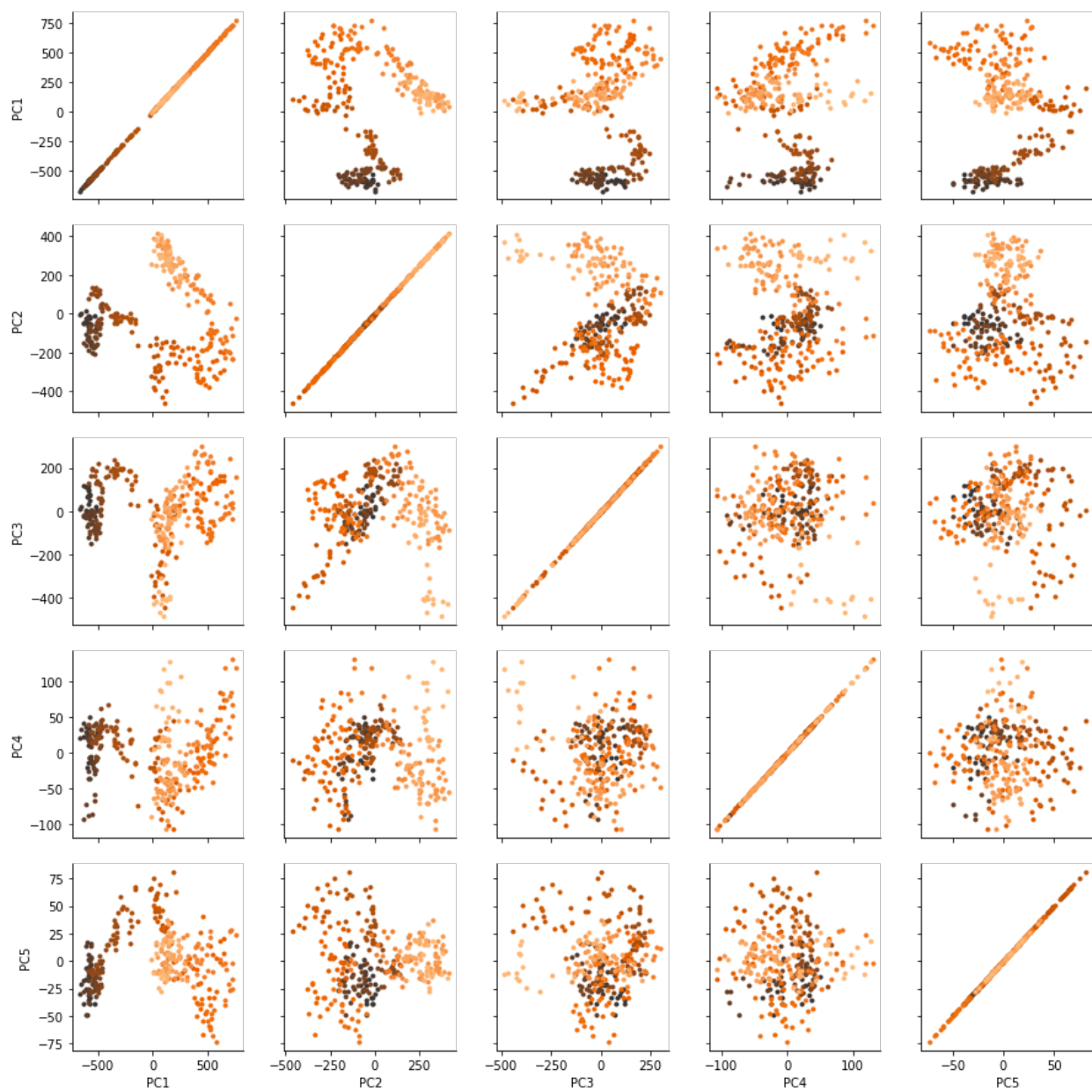


Figure S12. Principal component pair-wise interaction map for A1 variant system. First 5 principal components are shown. Color corresponds to time and goes black to dark orange.

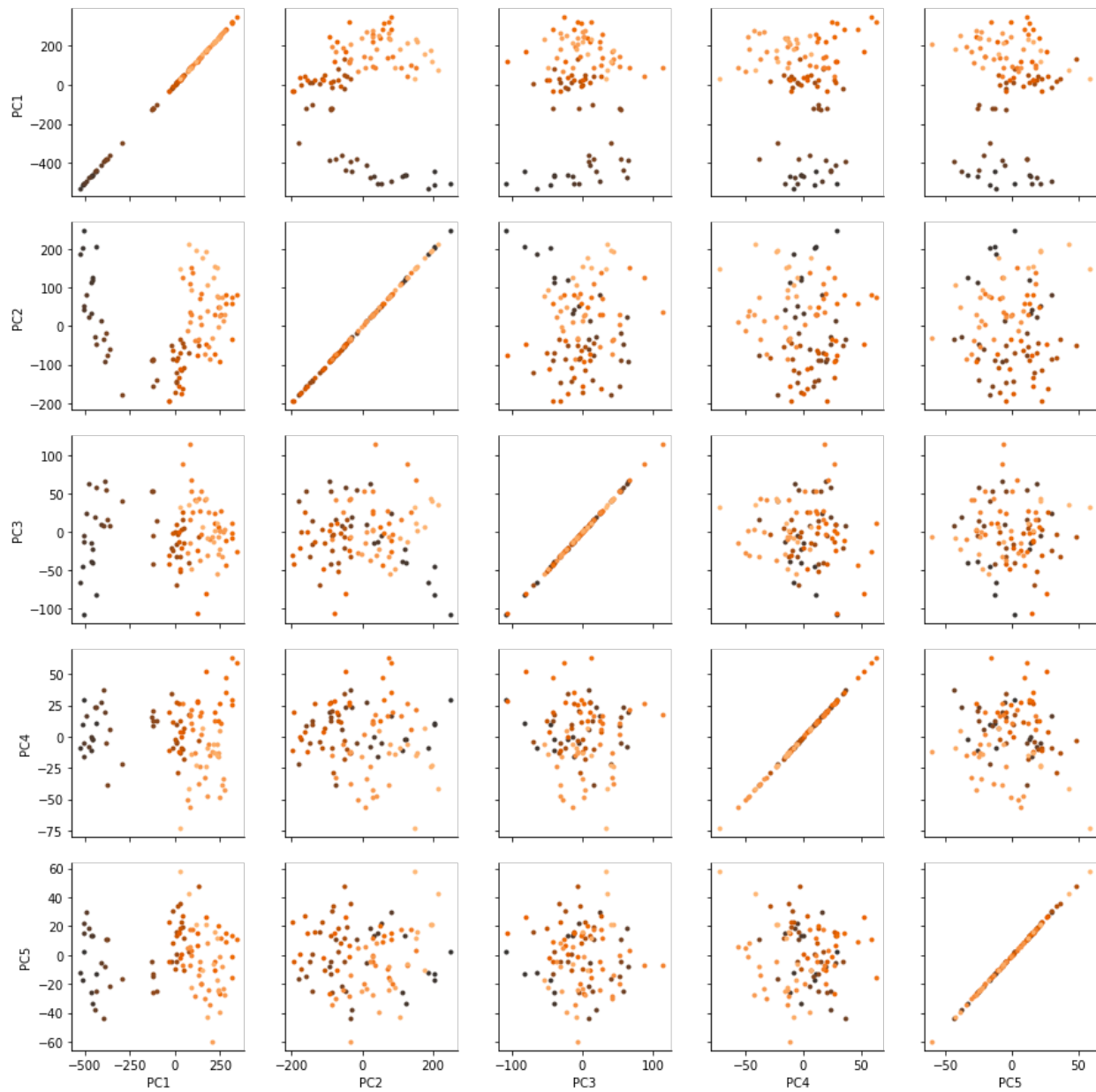


Figure S13. Principal component pair-wise interaction map for A2 variant system. First 5 principal components are shown. Color corresponds to time and goes black to dark orange.

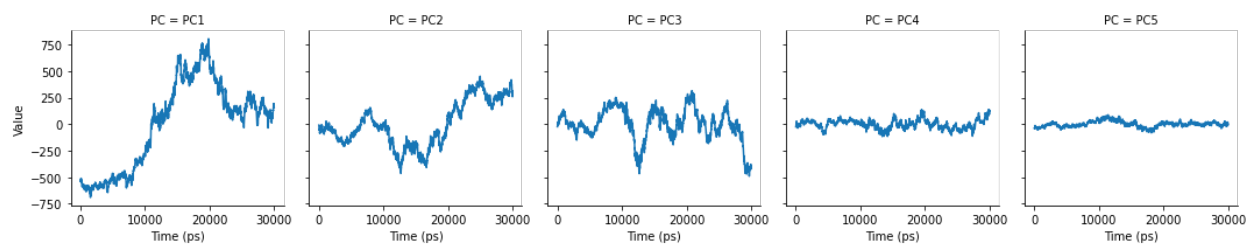


Figure S14. Principal component vs time for A1 variant system.

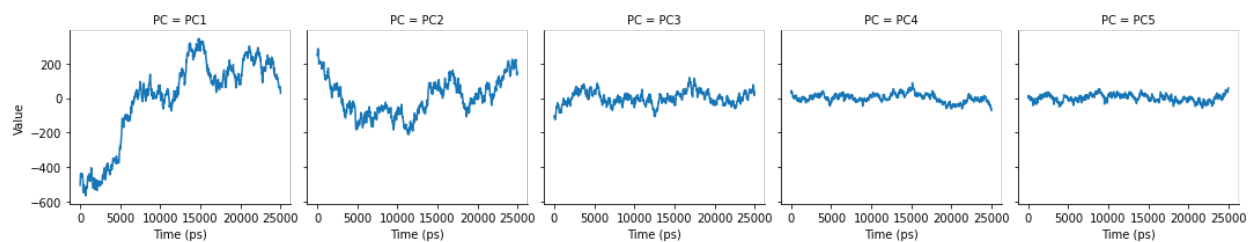


Figure S15. Principal component vs time for A2 variant system.

Chapter 7

Production of novel SARS-CoV-2 Spike truncations in Chinese hamster ovary cells leads to high expression and binding to antibodies

My contribution to this work was the simulations and analyses of T4 truncation, RBD and RBD without HIS tag.

This chapter originally appeared in the literature as:

Minami SA, Jung S, Huang Y, Harris BS, Kenaston MW, Faller R, Nandi S, McDonald KA, Shah PS. Production of novel SARS-CoV-2 Spike truncations in Chinese hamster ovary cells leads to high expression and binding to antibodies. *Biotechnology Journal*. 2022 Sep;17(9):2100678.

Abstract

SARS-CoV-2 Spike is a key protein that mediates viral entry into cells and elicits antibody responses. Its importance in infection, diagnostics, and vaccinations has created a large demand for purified Spike for clinical and research applications. Spike is difficult to express, prompting modifications to the protein and expression platforms to improve yields. Alternatively, Spike receptor binding domain (RBD) is commonly expressed with higher titers, though it has lower sensitivity in serological assays. Here, we improve transient Spike expression in Chinese hamster ovary (CHO) cells. We demonstrate that Spike titers increase significantly over the expression period, maximizing at 14 mg/L at day 7. In comparison, RBD titers peak at 54 mg/L at day 3.

Next, we develop 8 Spike truncations (T1-T8) in pursuit of a truncation with high expression and antibody binding. The truncations T1 and T4 express at 130 mg/L and 73 mg/L, respectively, which are higher than our RBD titers. Purified proteins were evaluated for binding to antibodies raised against full-length Spike. T1 has similar sensitivity as Spike against a monoclonal antibody and even outperforms Spike for a polyclonal antibody. These results suggest T1 is a promising Spike alternative for use in various applications.

7.1 Introduction

The emergence of coronavirus infectious disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has resulted in over 250 million infections and 5 million deaths globally since November 2019. Major aspects of containing this global pandemic are surveillance (large-scale and rapid asymptomatic testing) and herd immunity (immunity achieved in a large portion of the population with protective antibodies resulting from vaccination or natural infection). Many of these containment efforts require generating large amounts of viral glycoproteins. Consequently, the COVID-19 pandemic has highlighted the critical need for rapid, scalable, and cost-effective production of recombinant glycoproteins for use as antigens in diagnostic kits, research reagents, and even the active pharmaceutical ingredient in protein-based vaccines.

For SARS-CoV-2, diagnosis and vaccination strategies involve scalable production of the Spike glycoprotein. Spike is the structural protein responsible for protecting the viral genome and for entry into cells. Spike contains the S1 and S2 domains, which mediate host receptor binding and membrane fusion, respectively¹. The receptor binding domain (RBD) of Spike lies within the

S1 domain (Figure 7.1A). Spike is a major antigen and the primary target for antibody binding. Consequently, immunoassays to assess immunity of individuals or a community require a SARS-CoV-2 antigen, most commonly the Spike protein. Protein-based SARS-CoV-2 vaccines also rely on delivering Spike protein with adjuvant for immunization ².

A major limitation to scaling these approaches is generating Spike protein at high titers in a cost-effective manner. Several forms of full-length Spike have been produced in mammalian cell lines, including modifications to increase stability and expression, but titers remain at a low range of approximately 5-30 mg/L, with one report of 150 mg/L ³⁻⁵. A possible alternative is to express the Spike RBD, which can have expression levels of an order of magnitude higher than those of Spike, but is less sensitive than Spike in serological assays ³. This suggests that RBD may not have the same activity as Spike for such applications. Mutational scanning has been performed on RBD, which resulted in higher expression and stability ^{6,7}. Rational structure-based approaches have also been used to improve stability of full-length Spike ⁴. Nonetheless, identifying sequence-independent methods to increase expression is essential for comparisons with existing variants.

In this work, we transiently express Spike and RBD in Chinese hamster ovary (CHO) cells. To find high expressing and antibody binding forms of Spike, we also design and express 8 truncations of Spike, which include the RBD and additional residues. Two of these truncations express at high levels. Using simulation and experimentation, we find that one of the high-producing truncations also has more structural similarity to full-length Spike than the other and has higher binding to anti-Spike antibodies. Taken together, these truncations may provide an additional avenue for lower cost production of COVID-19 biologics with improved expression and antibody binding.

7.2 Methods

7.2.1 Plasmids

pCAGGS-Spike and pCAGGS-RBD were gifted from Florian Krammer and coworkers³. Spike and RBD both contain an N-terminal signal peptide for secretion and a hexahistidine (6x His) tag for purification. Spike-1 and RBD-1 contain the signal sequence MFVFLVLLPLVSSQ. Spike-2 and RBD-2 contain the signal sequence MEFGLSWLFLVAILKGVQC.⁸ Spike has two stabilizing mutations (K983P and V984P), and its polybasic furin site has been removed (RRAR to R). Truncations T1-T8 were synthesized (GenScript, Piscataway, NJ) with overhangs for insertion into pCAGGS vectors (Table S1). Truncations were inserted into pCAGGS vectors via Gibson Assembly of pCAGGS-RBD digested with XbaI and XhoI. Spike truncations were designed by adding increments of approximately 50 amino acids to the N- and/or C-termini of RBD. Each truncation includes an N-terminal signal peptide and a C-terminal 6x His tag. Possible structural and binding motifs for the truncations were determined with PredictProtein⁹.

7.2.2 Cell Culture and Transfection

ExpiCHO-S cells (Thermo Fisher Scientific, Waltham, MA) were maintained in a 125 mL vented shake flask with 30 mL of culture in ExpiCHO Expression Medium (Thermo Fisher Scientific). Cells were cultured in a humidified incubator at 37°C and 8% CO₂, on a 19 mm shaking diameter orbital shaker at 120 rpm (Ohaus, Parsippany, NJ).

For transfection in 125 mL shake flasks, cultures were transfected using the Expifectamine CHO Transfection Kit (Thermo Fisher Scientific), following manufacturer instructions for the Standard Protocol. For time course experiments, 0.5 mL of culture was harvested each day. Viable cell densities were measured using trypan blue and a TC20 automated cell counter (Bio-Rad,

Hercules, California). Samples were harvested by centrifuging at 300 rcf for 5 min and collecting the supernatant. For samples to undergo purification, entire cultures were centrifuged at 4,000 rcf for 20 minutes at 4°C and filtered through 0.22 µm filters. For transfection in 2.0 mL 96-well deep well blocks (Genesee Scientific, El Cajon, CA), 0.8 mL of cells at 6×10^6 cells/mL were plated on the day of transfection. Cells were cultured on a 3 mm shaking diameter orbital shaker at 900 rpm (Benchmark Scientific, Sayreville, NJ) and transfected according to manufacturer instructions. Samples were harvested 5 days post-transfection by centrifuging the cultures at 300 rcf for 5 minutes and collecting the supernatant.

7.2.3 Protein Purification and Concentration

Filtered samples were column purified using an AKTA Pure fast protein liquid chromatography (FPLC) system with a 5 mL prepacked Ni Sepharose HP column (Cytiva, Marlborough, MA), using imidazole to elute the proteins¹⁰. Samples were loaded onto the column at a flow rate of 5 mL/min, the resin was washed for 10 column volumes (CV), and proteins were eluted using imidazole. Detailed procedures are available in Supporting Information. Purified proteins were dialyzed against phosphate-buffered saline (PBS) using dialysis cassettes at 4°C (Thermo Fisher Scientific). Spike was dialyzed with a 20 kDa molecular weight cutoff (MWCO) membrane. RBD, T1, and T4 were dialyzed with 10 kDa MWCO membranes. Dialyzed samples were concentrated using centrifugal filter units (Millipore Sigma, Burlington, MA) at 4,000 rcf for 20 minutes at 4°C. Spike was concentrated using centrifugal filter units with a MWCO of 30 kDa. RBD, T1, and T4 were concentrated with 10 kDa MWCO centrifugal filter units.

7.2.4 SDS-PAGE and Western Blot

Samples from time course experiments and the truncation screening were prepared for sodium dodecyl-sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) by adding 12 μL of NuPAGE LDS Sample Buffer (Thermo Fisher Scientific) and 3 μL of tris(2-carboxyethyl)phosphine (Thermo Fisher Scientific) to 30 μL of sample. Mixtures were heated at 95°C for 10 minutes and 10 μL of samples were loaded into gels cast in-house, with a 12% acrylamide resolving layer and 4% acrylamide stacking layer. Samples were run through the gel for 15 minutes at 115 V, then 50 minutes at 150 V. Proteins were transferred onto polyvinylidene difluoride membranes in a wet sandwich and membranes were blocked using 5% non-fat milk. Membranes were stained overnight at 4°C with a 1:1000 diluted mouse anti-his primary antibody (MCA1396, RRID:AB_322084, Bio-Rad) and then for 1 hour at room temperature with a 1:4000 diluted rabbit anti-mouse HRP secondary antibody (SouthernBiotech Cat# 6170-05, RRID:AB_2796243, Birmingham, AL). Membranes were developed using Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) and imaged using an Amersham Imager 600 (Cytiva).

Purified samples were analyzed by SDS-PAGE with a method previously described¹¹. Images of the gels were taken using a ChemiDoc Imaging system (Bio-Rad), and proteins were transferred onto nitrocellulose membranes using Trans-Blot Turbo Packs (Bio-Rad) and Trans-Blot Turbo System (Bio-Rad). Membranes were blocked overnight in 1% casein, stained with 1:1000 diluted mouse anti-his primary antibody and stained with 1:4000 diluted rabbit anti-mouse secondary antibody. The chemiluminescent reactions were performed using Clarity ECL Substrate (Bio-Rad).

Concentrations for purified proteins were estimated using a combination of ELISA, Bradford Assay, and scanning densitometry on SDS-PAGE gels. Spike and RBD concentrations

were first calculated using sandwich ELISA, and purified T1 and T4 concentrations were determined using Bradford Assay. Next, 1 µg of proteins, as determined by the two methods, were loaded into each lane of a 4%-20% gradient stain-free gel (Bio-Rad). Dilutions of RBD standard from 1.5 µg to 0.5 µg were also loaded into the gel. Samples were run at 200 V for 36 minutes and imaged using a ChemiDoc imaging system (Bio-Rad). A standard curve was generated via densitometry through ImageJ, and primary band intensities for the samples were interpolated to quantify concentrations.

7.2.5 Enzyme-linked Immunosorbent Assay (ELISA)

Sandwich ELISAs were performed to quantify purified Spike and RBD and crude supernatants. 1:1000 mouse anti-his capture antibody in PBS was coated onto Immulon 2 HB 96-well plates (Thermo Fisher Scientific) at 4°C overnight. Plates were blocked with 200 µL/well PBS with 3% BSA for 30 minutes. Plates were loaded with serial dilutions of purified protein samples or crude supernatants. Plates were incubated with 1:1000 rabbit anti-RBD primary antibody (Sino Biological Cat# 40592-R001, RRID:AB_2857936, Wayne, PA), then 1:6000 or 1:4000 goat anti-rabbit, HRP secondary antibody (SouthernBiotech Cat# 4030-05, RRID:AB_2687483) in PBS with 1% BSA for purified or crude proteins, respectively. Plates were developed with 1-step Turbo TMB-ELISA Substrate Solution (Thermo Fisher Scientific) and 2N HCl. Absorbance at 450 nm was measured using a Spectramax 250 spectrophotometer (Molecular Devices, San Jose, CA). Plates were washed 3 times with 200 µL/well PBS with 0.05% Tween20 (PBS-T) between each step and incubations were performed using volumes of 100 µL/well for 1 hour at room temperature unless specified otherwise. Standard curves for quantifying Spike and RBD were generated using serial dilutions of Sf9 insect Spike (NR-52308, BEI Resources, Manassas, VA) and HEK293F human RBD (NR-52366, BEI Resources), respectively.

Indirect ELISAs were performed to assess the sensitivities of CHO-expressed proteins to a human anti-Spike monoclonal antibody CR3022 (NR-52392, BEI Resources, RRID:AB_2848080) and a rabbit anti-Spike polyclonal antibody (PAb, eEnzyme, SCV2-S-100, RRID:AB_2893135, Gaithersburg, MD). For CR3022, antigens were first coated onto plates at 4°C overnight. After blocking, serial dilutions of CR3022 in PBS with 1% BSA were loaded from 100 ng/well. Plates were loaded with 100 µL/well goat anti-hIgG, HRP secondary antibody at 1:4000 in PBS containing 1% BSA. For the PAb, 3-fold serial dilutions starting at 400 ng/well of rabbit anti-Spike primary antibody were used (PAb, SCV2-S-100, eEnzyme), and a 1:4000 goat anti-rabbit IgG, HRP secondary antibody was used instead.

7.2.6 Bradford Assay

Bradford assays were performed to quantify the concentration of total soluble protein (TSP) by using a protein assay dye reagent (Bio-Rad). For each BSA standard, sample, and diluted sample, 10 µL/well of sample and 190 µL/well of Bradford dye were loaded into 96-well plates. After incubating for 10 minutes at room temperature, the absorbances of standards and samples were measured at 450 nm and 590 nm¹², using a Spectramax M4 spectrophotometer (Molecular Devices). Standard curves for quantifying samples were generated by using serial dilutions of BSA from 0-0.5 mg/mL with 0.05 mg/mL steps.

7.2.7 Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) and Sequence Alignment

Sequences of purified T1 and T4 were obtained via LC-MS/MS. 10 µg of T1 and 20 µg of T4 were run on a 4%-20% gradient SDS-PAGE gel. Bands were extracted and submitted to the Genome Center Proteomics Core at the University of California, Davis. Briefly, proteins were

digested with trypsin and analyzed on a Dionex UltiMate 3000 RSLC system (Thermo Fisher Scientific) using a PepSep (PepSep, Denmark) ReproSil 8 cm 150 μm I.D. C18 column with 1.5 μm particle size (120 \AA pores). Mass spectra analysis is described in Supporting Information. Searches were conducted against the known sequences of T1 and T4, and alignments were performed using Multiple Alignment using Fast Fourier Transform ¹³.

7.2.8 Circular Dichroism (CD)

Concentrated samples were prepared for CD analysis by diluting 150 μg of protein in 50% PBS and 50% CD buffer (25mM of phosphate and 40mM of NaF). Single spectrum data were obtained using a Jasco J-715 CD spectrometer (Jasco, Easton, MD). Data were analyzed using BeStSel ¹⁴. Spectra of buffer were subtracted before analysis. To obtain secondary structure data for the PDB Spike structure, the PDB file 6VXX was analyzed using the STRIDE server ¹⁵.

7.2.9 Simulations

Starting configurations for molecular dynamics (MD) simulations were obtained by trimming the full Spike protein structure obtained from the protein data bank (6VXX). Structures were reduced to a single monomer and cut at the amino acid sequences corresponding to RBD, T1, T3, and T4. 6x His tags were added using modeler ¹⁶, which modifies amino acid sequences of proteins. The new His-tagged structures were prepared and had glycans attached using Glycam ¹⁷. The N-glycosylation sites of RBD and the RBD portion of all truncations had the glycoform FA2 attached. T1 contained no additional N-glycosylation sites, T3 contained an additional FA3 glycoform, and T4 contained an additional M5 glycoform. Amber ff14SB and Glycam06 forcefields ^{18,19} were used and generated using acpype.py following the method shown previously ^{4, 20, 21}. Simulations were conducted using the Gromacs 2019.1 suite with similar energy

minimization procedure as in previous simulations ²²⁻²⁴ including ones involving glycosylated RBD. Simulation runs after equilibration were carried out for 100 ns.

7.3 Results

7.3.1 Expression and Purification of Spike and RBD

We first compared the expression of Spike and RBD in ExpiCHO-S cells transfected in 96-well format. Spike and RBD were expressed with N-terminal secretion signals and C-terminal 6x His tags for downstream purification (Figure 7.1A). We also replaced the previously tested secretion signal with an alternative secretion signal that had relatively consistent performance compared to other signal sequences when tested across a variety of secreted proteins ⁸ to determine whether it affects expression and secretion of Spike and RBD (Spike-1, Spike-2, RBD-1, and RBD-2, Figure 7.1B).

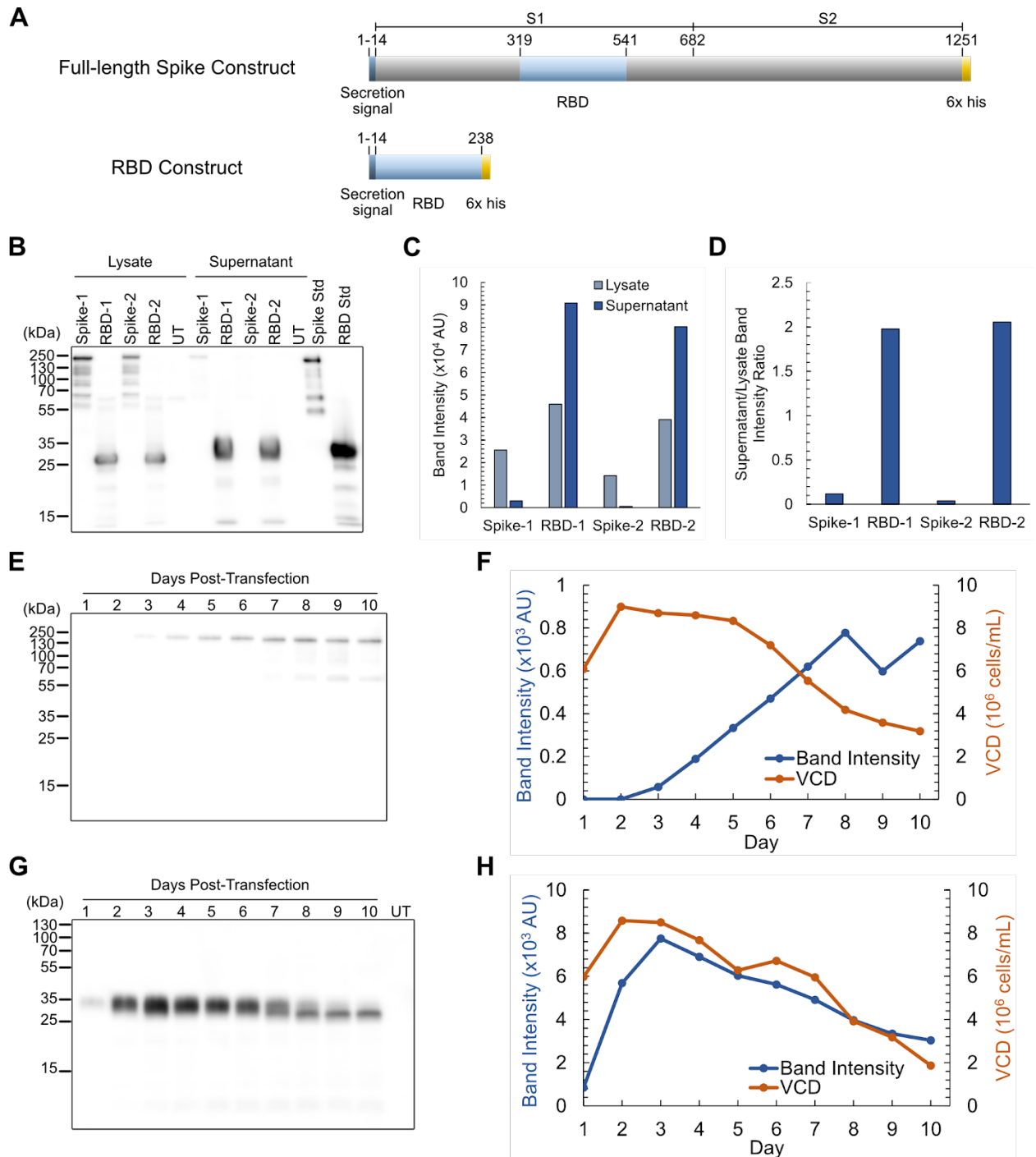


Figure 7.1: Transient Spike and RBD production in CHO cells. (A) Diagram of full-length Spike (1257 aa) and RBD (244 aa) constructs. Residues are labeled starting from the beginning of the secretion signal. (B) Western blot and (C) densitometry comparing two secretion signals for Spike and RBD. (D) Ratio of band intensities of supernatants and lysates. (E) Western blot and (F) densitometry on Spike expression time

course. (G) Western blot and (H) densitometry on RBD expression time course. Abbreviations: aa (amino acids); untransfected (UT); standard (std); viable cell density (VCD).

Relative amounts of protein in the supernatant and cell lysate were determined by western blot 5 days post-transfection (Figure 7.1B and 7.1C). Spike had significantly lower expression than RBD and was particularly less abundant in the supernatant. Comparison of ratios of supernatants over lysates also showed that Spike is significantly retained in the cells compared to RBD (Figure 7.1D). Both the expression and supernatant/lysate ratio remained the same for Spike and RBD with either signal sequence, indicating that low expression and high retention of Spike in the cells may be due to the protein sequence itself, and not a consequence of the tag used. All following experiments were performed with Spike-1 and RBD-1, hereafter referred to as Spike and RBD, respectively.

Next, cultures were scaled up to 25 mL and a time course experiment was performed to determine the optimal harvest time for maximum titers. Cells were transfected with plasmids encoding Spike and RBD, and a sample of the supernatant was collected every 24 hours over 10 days. Western blots were performed on the supernatants and band intensities were plotted over time (Figure 7.1E-H). Spike concentration in the supernatant increased steadily until 7 days post-transfection, after which time it remained stable (Figure 7.1E and 7.1F). In comparison, RBD concentration in the supernatant peaked at day 3, then decreased (Figure 7.1 G and 7.1H). From these results, Spike and RBD harvests were determined to be 7 and 3 days post-transfection, respectively.

To produce large quantities of Spike and RBD for purification and downstream analysis, 150 mL of supernatants were prepared from pooled 25 mL cultures. Crude titers of Spike and RBD

were measured using sandwich ELISAs on filtered crude, yielding 14 mg/L and 54 mg/L, respectively (Supporting Information Figure S1). The crude supernatants were purified through FPLC (Figure 7.2 and Supporting Information Figure S2). For purification of Spike, SDS-PAGE revealed bands in elution fractions E3, E4, and E5 (Figure 7.2A), which were confirmed by western blot to include immunoreactive bands consistent with full-length Spike (Figure 7.2B). For RBD, SDS-PAGE (Figure 7.2C) and western blot (Figure 7.2D) showed bands in elution fractions E3, E4, and E7. FPLC-purified samples were dialyzed using PBS and concentrated through centrifugal filter units. E3 fractions of both proteins were used for subsequent experiments.

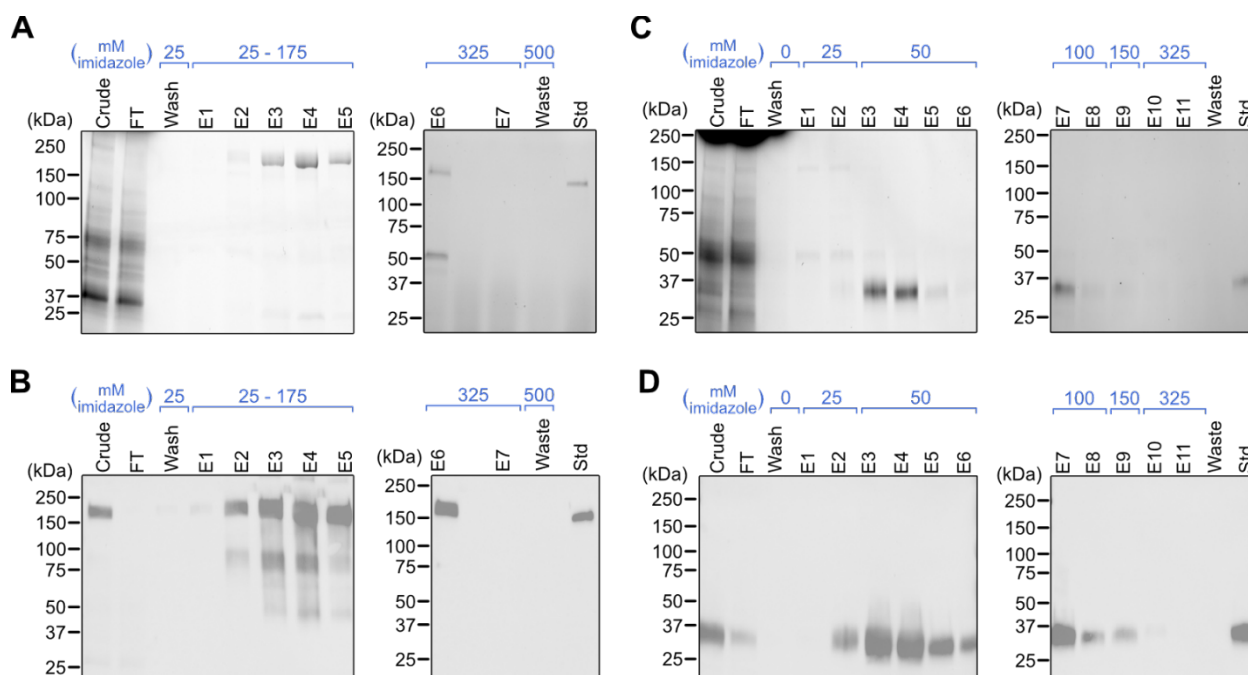


Figure 7.2: Purification of Spike and RBD. (A) SDS-PAGE and (B) western blot on Spike fractions. (C) SDS-PAGE and (D) western blot on RBD fractions. FT samples were collected during sample loading onto the column. FT and wash samples were pooled from multiple fractions at equal volumes. Abbreviations: flow-through (FT); elution (E); standard (std).

7.3.2 Novel Truncations to Improve Protein Titrers

Given the difficulty in expressing full-length Spike but its higher sensitivity in serological assays compared to RBD,³ we sought to determine whether a truncation falling between RBD and Spike could balance the higher expression of RBD and the higher antibody sensitivity of Spike. Since RBD is important for binding to the human receptor ACE2 and is a target of antibodies,^{26, 27} truncations were designed by starting with the RBD and adding amino acids on the N- and/or C-termini. We developed eight truncations, T1–T8, by adding increments of approximately 50 amino acids on the N- or C-terminal ends of RBD (Figure 7.3A). These increments were determined such that starting and ending residues do not interrupt major secondary structures present in Spike.²⁸⁻³⁰ For instance, T1 contains an additional predicted beta-sheet compared to RBD, and T4 contains an additional predicted beta-sheet compared to T1. We also expressed the full S1 subunit of Spike. Secretion signals and 6x His tags were added to N- and C-termini, respectively.

We first screened the truncations for expression levels. Cells were transfected in 96-well format, and lysates and supernatants were analyzed by western blot 5 days post-transfection (Figure 7.3B). T1 and T4 had the highest expression, as well as the highest supernatant/lysate signals (Figure 7.3C, D). In particular, T1 had even higher expression and relative secretion than RBD did. Given that S1 has previously been studied,³¹ T1 and T4 were selected for scale-up and purification. To determine optimal harvest dates for T1 and T4, an expression time course was performed as previously described for Spike and RBD. For T1, western blot and densitometry (Figure 7.3E, F) showed that expression peaked 3 days post-transfection with a single band. For T4, expression peaked after 4 days post-transfection, with degradation bands also increasing after this point (Figure 7.3G, H). Therefore, T1 and T4 harvest dates were determined to be 3 and 4 days post-transfection, respectively.

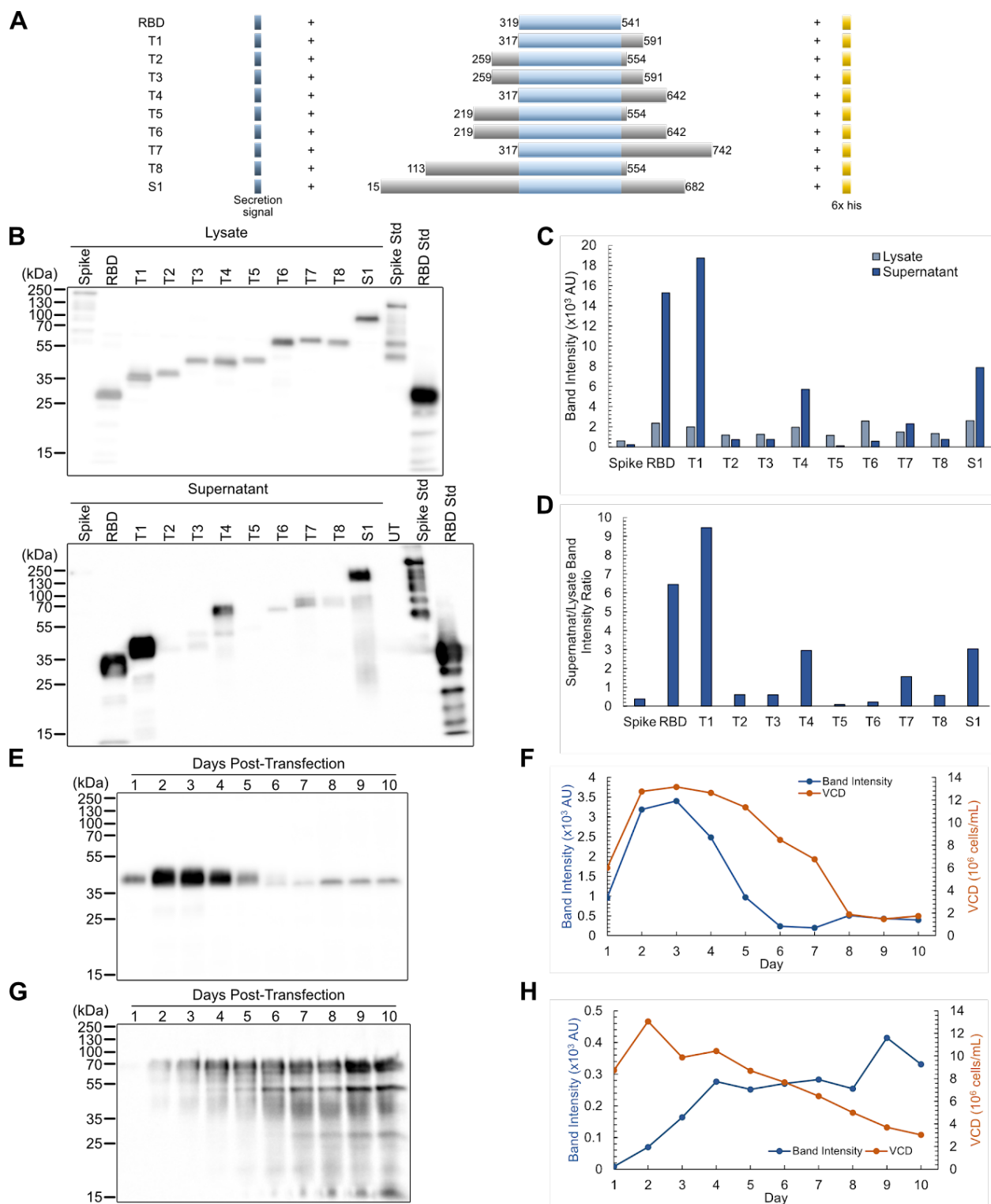


Figure 7.3: Expression of Spike truncations. (A) Construct diagram of Spike truncations. Residue numbers are relative to the position within full-length Spike, including the secretion tag. (B) Western blot and (C)

densitometry on truncations for lysates and supernatants. **(D)** Ratio of band intensities of supernatants over lysates. **(E)** Western blot and **(F)** densitometry of T1 expression over time. **(G)** Western blot and **(H)** densitometry of T4 expression over time. Abbreviations: untransfected (UT); standard (std); viable cell density (VCD).

For purification, transfections with T1 and T4 were performed in a total of 150 mL of culture each, and samples were purified through FPLC (Figure 7.4 and Supporting Information Figure S3). Two-step purifications were performed to obtain pure T1 and T4, which were prepared for further characterization through dialysis and spin-column concentration. The single band detected for T1 and two bands detected for T4 through western blot were cut out of an SDS-PAGE gel and submitted for proteomic analysis. Coverage for T1 and T4 was high (Supporting Information Figure S4), and the two bands of T4 on the SDS-PAGE gel suggest different post-translational modifications. For further characterization and binding assays, concentrations of purified samples were determined using SDS-PAGE densitometry (Supporting Information Figure S5). Quantifying crude titers via ELISA, we found that T1 and T4 expressed at 130 and 73 mg L⁻¹, respectively, which are higher than our RBD titers (Supporting Information Figure S6).

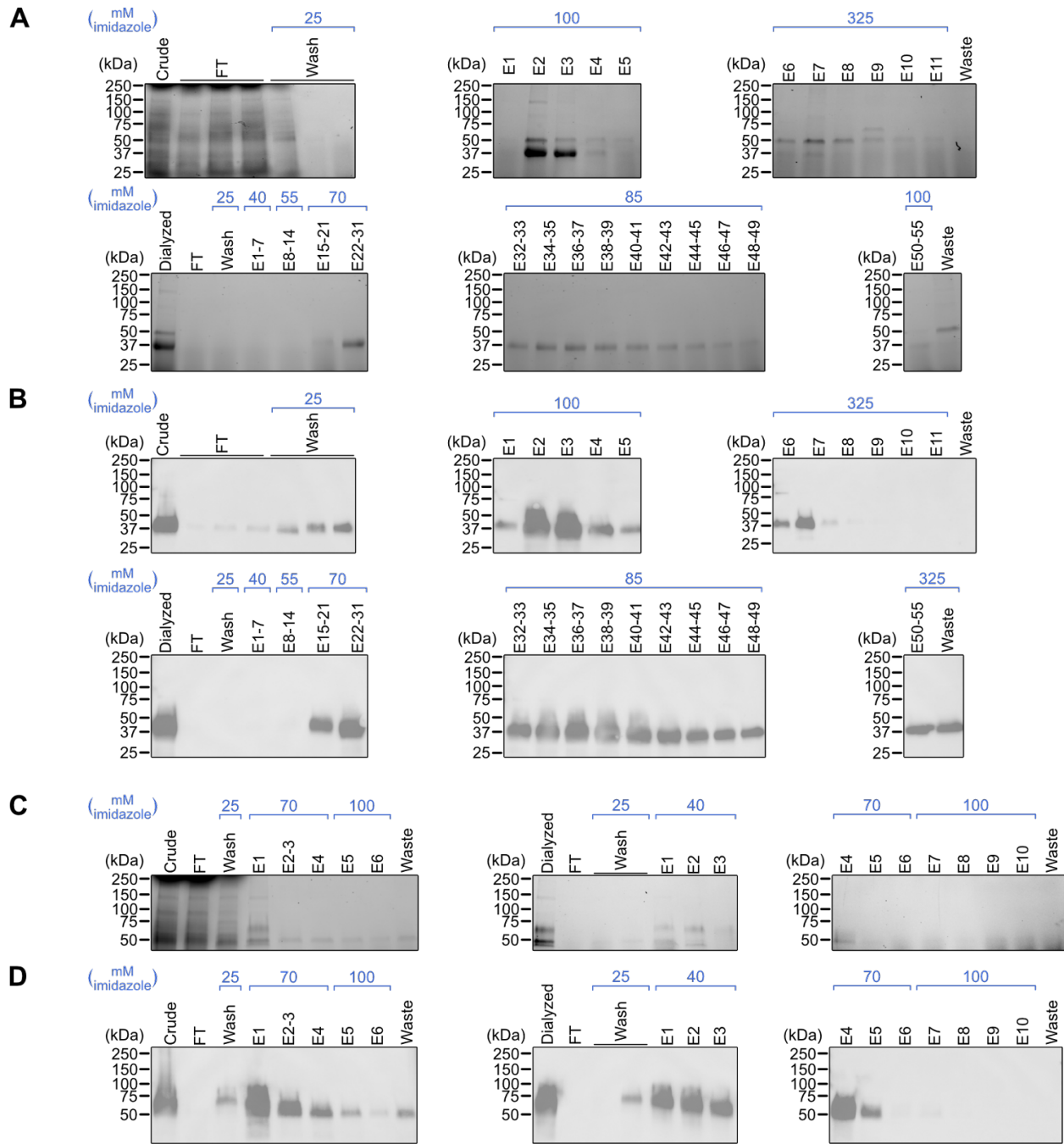


Figure 7.4: Purification of T1 and T4. (A) SDS-PAGE and (B) western blot on T1 fractions from crude purification and repurification. (C) SDS-PAGE and (D) western blot on T4 fractions from crude purification and repurification. FT and wash samples were pooled from multiple fractions at equal volumes. Abbreviations: flow-through (FT); elution (E).

7.3.3 Binding Sensitivities against Antibodies

The activities of the CHO-expressed proteins were evaluated via indirect ELISAs with antibodies against Spike. First, the monoclonal antibody CR3022 was tested, which binds to the receptor binding domain of Spike²⁷. Serial dilutions of CR3022 were bound to Spike, RBD, T1, and T4 (Figure 7.5A). Binding sensitivities were compared by taking the areas under the curves (Figure 7.5B). To compare with another recombinant source of Spike, Sf9 insect Spike was also used in the assays.

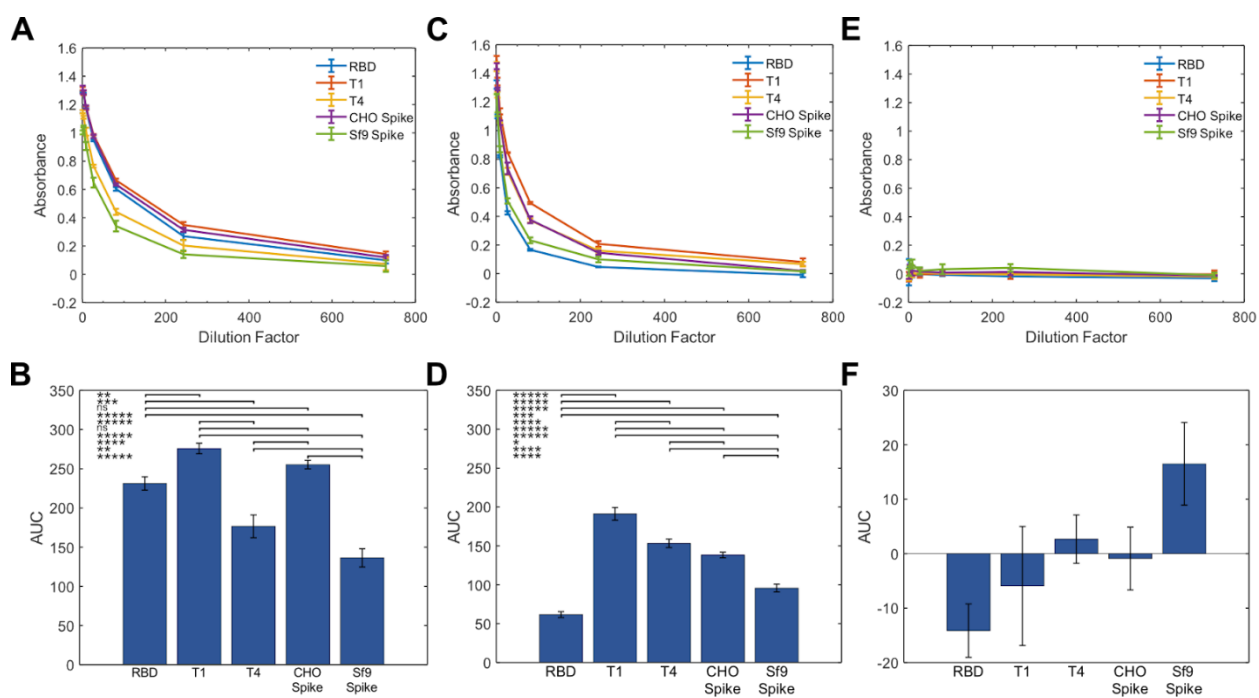


Figure 7.5: Binding assays of Spike truncations with anti-Spike antibodies. (A) Absorbance as a function of dilution factor of CR3022. (B) AUC calculated from (A). (C) Absorbance as a function of dilution factor of an anti-Spike PAb. (D) AUC calculated from (C). (E) Absorbance against dilution factor of a rabbit normal IgG antibody. (F) AUC calculated from (E). Error bars represent \pm SD of technical triplicates. p-values were calculated using a one-way ANOVA followed by Tukey's Test. * indicates $p < 5 \times 10^{-2}$, ** $p <$

5×10^{-3} , *** $p < 5 \times 10^{-4}$, **** $p < 5 \times 10^{-5}$, and ***** $p < 5 \times 10^{-6}$. Abbreviations: area under curve (AUC); ns (not significant).

CHO-expressed Spike had higher binding to CR3022 than Sf9-expressed Spike. This may be due to differences in folding or glycosylation between the CHO- and insect-expressed proteins. Among the CHO-expressed proteins, T1 had higher binding to CR3022 than RBD did and is comparable to the performance of Spike. T4 had lower signal but still outperformed Sf9-expressed Spike. Next, serial dilutions of a polyclonal antibody raised against full-length Spike were tested (Figure 7.5C and 7.5D). Given that PABs may recognize multiple binding epitopes in a protein, larger forms of Spike were expected to have higher performance. Strikingly, however, T1 and T4 had very high signal across dilutions of the antibody, and T1 outperformed full-length CHO-expressed Spike. The increased sensitivities were not due to non-specific binding of T1 and T4 to rabbit antibodies, since a control rabbit IgG did not produce significant signal (Figure 7.5E and 7.5F).

7.3.4 Structural Characterization of Truncations

To determine whether structural similarities are maintained between the truncations and the relevant regions of Spike, structures of T1 and T4 were predicted using MD. Snapshots of simulated structures of RBD, T1, and T4 at 0 ns and 100 ns simulation times are shown (Figure 7.6A-D). The RBD portion of all structures remained stable during this time. T1 and T4 showed similarly stable secondary structures in the additional residues at the bottom of the structure. The more flexible turn features curled in and stabilized over the course of the trajectory. To quantify this behavior, the root mean squared deviation (RMSD) of the whole structures and RBD subdomains for each truncation were evaluated (Figure 7.6E and 7.6F). These RMSD plots show

deviation relative to initial structures and provide further evidence that the RBD subdomains are stable or reach a stable structure early in the simulation trajectory. The T1 and T4 RMSD plots show more conformational change, likely due to the flexibility of the turn features observed in the snapshots. An interesting observation is that the RBD with a 6x His tag appeared to be more stable compared to the RBD without the tag. Based on this RMSD data, T1 appeared to stabilize the RBD in line with this 6x His tag, while the T4 structure aligned more closely to the RBD without a 6x His tag.

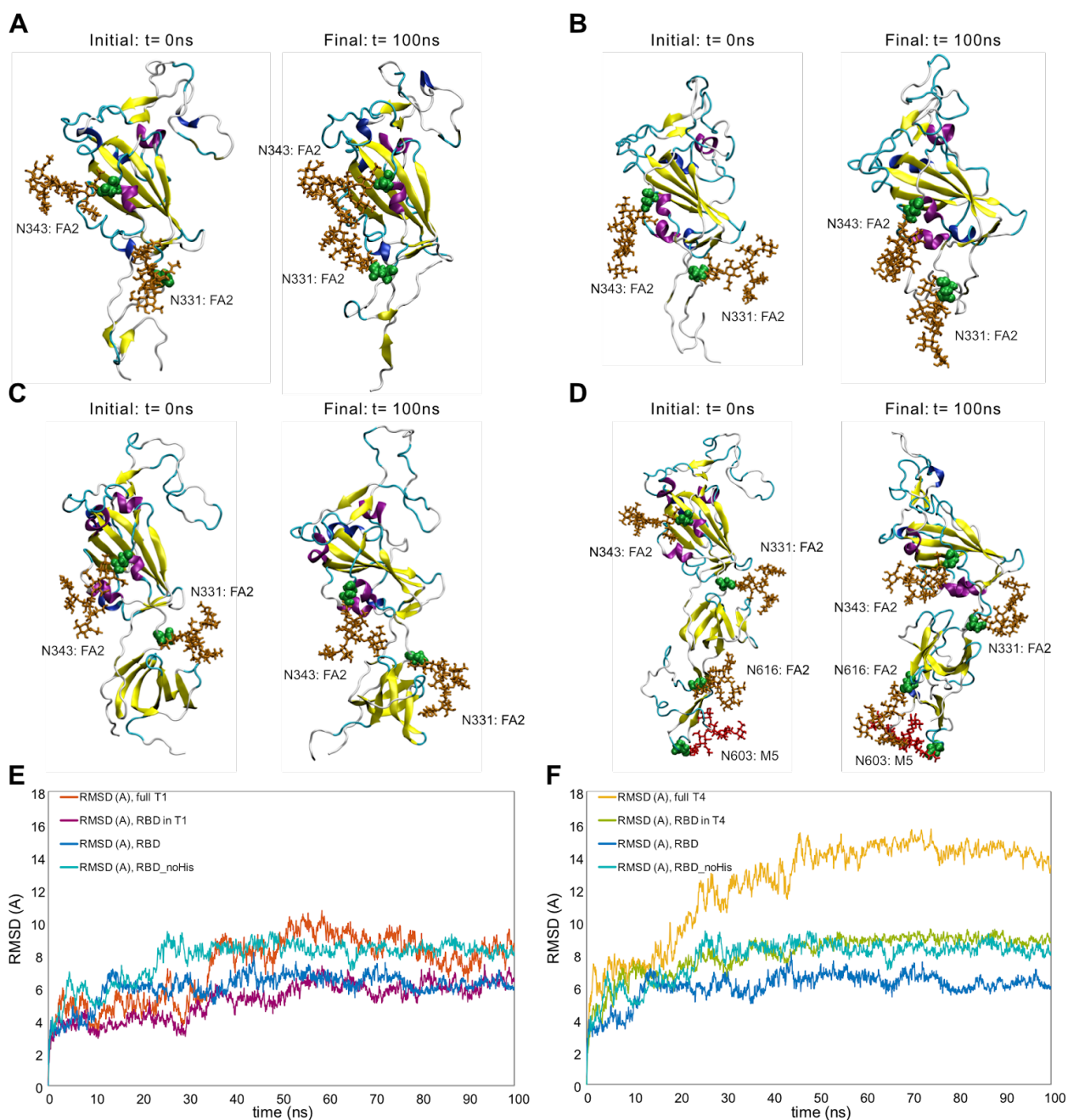


Figure 7.6: MD structural stability snapshots and analysis. MD snapshots are visualized for (A) RBD, (B) RBD without the 6x His tag (RBD_noHis), (C) T1, and (D) T4 at 0 ns and 100 ns. Backbone RMSD profiles of (E) full T1 and T1 RBD subdomain and (F) full T4 and T4 RBD subdomains are compared against RBD with and without His tag referenced to initial configurations. Green spheres represent the glycosylated asparagine residue, brown sticks represent FA2 glycans, and red sticks represent M5 glycans.

We hypothesize that some truncations did not express well because of structural differences. To explore this idea, we compared T1 and T3, which only differ by ~50 amino acids at the N-terminus but had vastly different expression (Figure 7.3B and Figure 7.3C). MD was used to determine whether structural differences may have caused the discrepancy in expression. RMSD analysis showed that T3 had much higher RMSD compared to T1 (Supporting Information Figure S7). Visualization of T3 revealed that the difference in RMSD was due to the additional FA3 glycan binding to its own RBD, which could contribute to low expression. Removal of the FA3 glycan from T3 resulted in a secondary structure that matched more closely to T1 and a more stable RBD within T3 (Supporting Information Figure S7). It is possible that other truncations also had incorrect folding.

Experimentally, secondary structure compositions of CHO-expressed Spike, RBD, T1, and T4 were obtained using circular dichroism (CD). $\Delta\epsilon$ values were obtained, which is a measure of the difference in absorbance of left- and right-circularly polarized light. Using the BeStSel server, $\Delta\epsilon$ as a function of wavelength was analyzed to predict the secondary structure compositions. The distributions of observed secondary structures were similar for most proteins (Figure 7.7A and Supporting Information Figure S8). CHO Spike and Sf9 Spike had very similar compositions, suggesting high structural similarity. RBD and T1 also had similar compositions. T4 was slightly dissimilar, with low beta sheet content compared to other proteins. CD-analyzed proteins were also compared to a structure of Spike determined through cryo-electron microscopy (PDB 6VXX).²⁶ 6VXX had similar alpha helix and beta sheet content as CHO and Sf9 Spike, but had much higher turn content and lower “other” content, which includes coils, bends, irregular loops, β -bridges, 3_{10} helices, and π -helices.

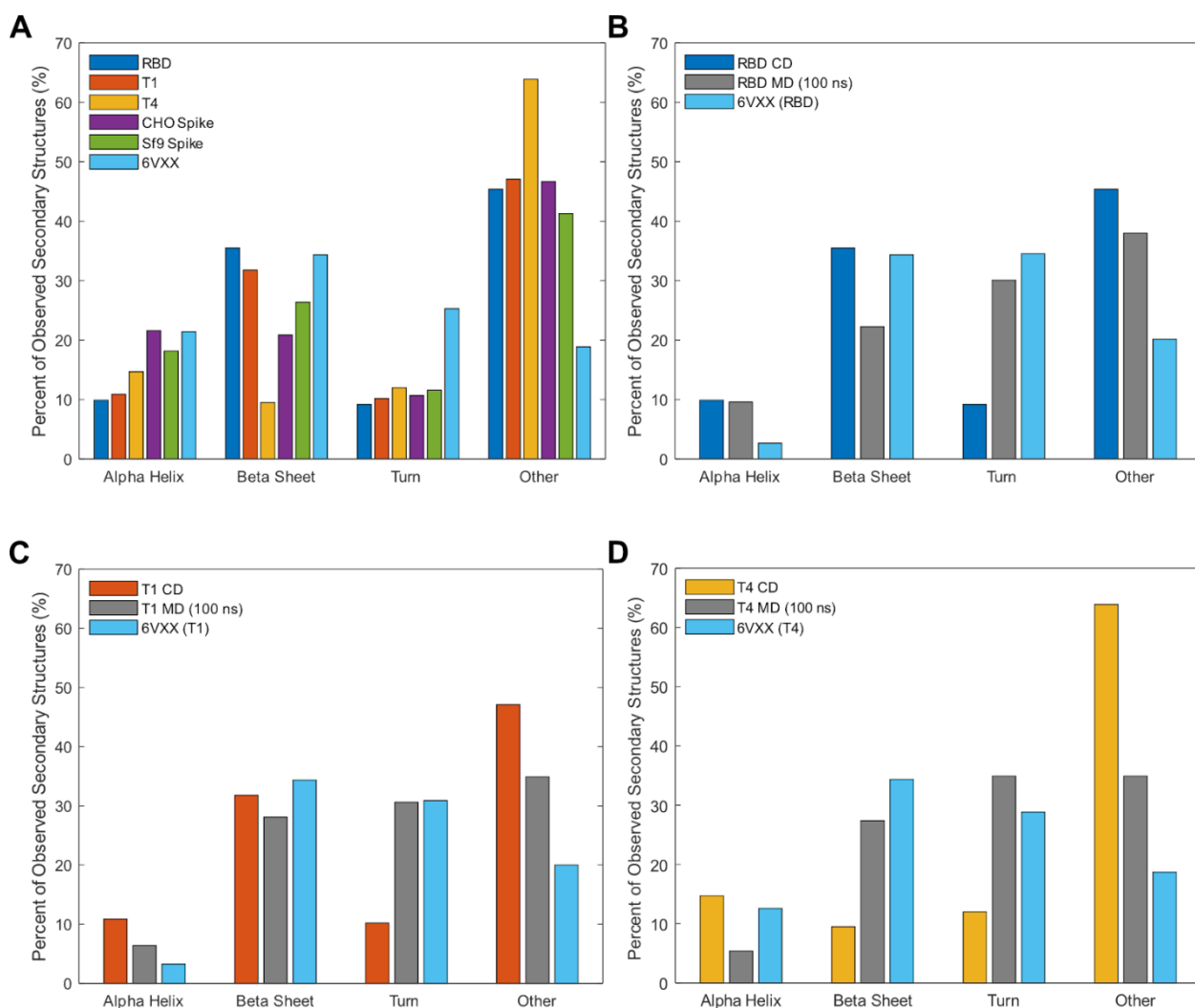


Figure 7.7: Structural composition of Spike and truncations produced in CHO cells. (A) Overall secondary structure compositions of CHO-expressed proteins and 6VXX. (B) Comparison of RBD MD to CD data and the RBD region of 6VXX. (C) Comparison of T1 MD to CD data and the T1 region of 6VXX. (D) Comparison of T4 MD to CD data and the T4 region of 6VXX. MD structural data represent proteins including the 6x His tags with the final structural compositions at 100 ns.

CD-derived structural information was also compared with MD secondary structures for RBD, T1, and T4 determined using DSSP³² (Figure 7.7B-D). Truncated 6VXX structures containing the relevant residues were also included, which represent structural composition had

truncating Spike not resulted in any structural changes. For all three proteins, 6VXX and MD structures had high similarity, with CD-derived structures having lower turn content. Overall, MD and CD results suggest that T1 and T4 retain accurate RBD structure, and consistent with their high sensitivities in the ELISAs against anti-Spike antibodies.

7.4 Discussion

Production of Spike fragments is important for its use in diagnostics, protein subunit vaccines, and research. In addition, high binding affinities of the Spike fragments are critical in these applications. Several approaches have been used to increase Spike yields, including stabilizing mutations⁴, comparison of different cell lines⁵, and optimization of production conditions, such as temperature shifts³³. Here, we expressed full-length Spike and RBD transiently in CHO cells to determine the intracellular and extracellular production kinetics. In addition, we developed 8 truncations in pursuit of a truncation which exhibits both high expression and binding to antibodies.

The regions of Spike that cause lower expression and higher sensitivity compared to RBD are not known, but the initial screen of the truncations showed that T1 is highly expressed and secreted compared to other truncations, with T4 following at much lower titers (Figure 7.3C, D). This suggests that residues downstream of T1 may be contributing to decreased titers. Comparing T1 to T2 and T3, residues upstream of T1 also appear to decrease titers. The additional residues in T2, T3, and T4 contain predicted glycosylation sites, which may introduce avenues for protein retention such as incomplete glycosylation. This is supported by the lysate proteins running at their expected molecular weights and the supernatant proteins much higher, though protein size did not appear to correlate with relative retention in the cell (Figure 7.3B). In contrast, T1 only contains

the same glycosylation sites as RBD and was found in the crude at much higher titers compared to other truncations. Interestingly, the MD simulations of T3 suggest that lower stability may result from unexpected intramolecular glycan–protein interactions for fully glycosylated truncations (Supporting Information Figure S7). Potential glycosylation sites of all truncations are labeled in Supporting Information Table S1 and depicted in structures in Figure 7.6 and Supporting Information Figure S7.

In the ELISA sensitivity assays for CR3022 and the PAb, CHO-expressed Spike has higher AUCs for both antibodies compared to Sf9 Spike (Figure 7.5). The discrepancy may be due to potentially different glycosylation profiles between both proteins, which would be consistent with the idea that CHO-expressed proteins tend to have more human-like glycosylation patterns³⁴. We also found that CHO Spike produces higher signal than RBD when probed with the PAb, consistent with results from serological assays³. This was also expected because polyclonal antibodies target multiple epitopes, and full-length Spike may contain more binding epitopes than RBD. Surprisingly, T1 and T4 have higher sensitivities to the PAb, outperforming full-length Spike. One possibility is that T1 and T4 contain an additional epitope, not present on RBD, that has high affinity but is sterically hindered when additional residues are present. This may also be the reason for the higher performance of T1 over T4. Another possibility is that the different Spike truncations may have different glycosylation profiles, which can influence structure and antibody binding sensitivities.³⁵ Ongoing investigations include glycoproteomics analysis of the Spike truncations to determine whether glycosylation profiles are different among the truncations and may influence binding. Additional avenues of investigating the reason for the high affinities of the truncations include visualization of binding through cryo-electron microscopy and analysis of binding kinetics and thermodynamics through methods such as biolayer interferometry and steered MD. Finally,

we note that the western blot of purified T4 revealed two major bands, where the nonhomogeneity of T4 may indicate different folding or glycosylation (Figure 7.4D). Considering the potential characterization studies discussed and product quality, we believe the heterogeneity of T4 further points to T1 as the ideal truncation.

7.5 Conclusions

We expressed SARS-CoV-2 Spike and RBD in CHO cells and optimized harvest dates. Additionally, we expressed 8 new truncations and found that T1 and T4 have high expression and secretion, where T1 has even higher expression than RBD. T1 and T4 also have higher binding sensitivity to a Spike polyclonal antibody compared to Spike. Overall, T1 had the highest performance in all expression and binding experiments conducted in this work. Its high expression and sensitivity suggest T1 may be a promising Spike alternative in research and clinical applications. Further work is needed to understand why T1 has higher affinity to antibodies and whether the higher affinity translates to assays with convalescent sera.

Acknowledgments

We would like to acknowledge the Protein Structure and Dynamics Core, Biochemistry and Molecular Medicine, University of California, Davis for obtaining CD spectral data of purified proteins. We would like to acknowledge the Proteomics Core Facility, University of California, Davis for proteomics analysis on T1 and T4. We thank Nitin Beesabathuni and Oanh Pham for scientific and editorial feedback on the manuscript. Computer simulations were performed on the hpc1/hpc2 servers at UC Davis. YH and RF were partially supported by the National Science

Foundation under grant no. CBET 1911267. BSH was partially supported by LLNL's LDRD program, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

1. Huang, Y., Yang, C., Xu, X., Xu, W., & Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: Potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica*, 41(9), 1141–1149.
2. Heath, P. T., Galiza, E. P., Baxter, D. N., Boffito, M., Browne, D., Burns, F., Chadwick, D. R., Clark, R., Cosgrove, C., Galloway, J., Goodman, A. L., Heer, A., Higham, A., Iyengar, S., Jamal, A., Jeanes, C., Kalra, P. A., Kyriakidou, C., McAuley, D. F., ... Toback, S. (2021). Safety and efficacy of NVX-CoV2373 Covid-19 vaccine. *New England Journal of Medicine*, 385, 1172–1183.
3. Amanat, F., Stadlbauer, D., Strohmaier, S., Nguyen, T. H. O., Chromikova, V., McMahon, M., Jiang, K., Arunkumar, G. A., Jurczyszak, D., Polanco, J., Bermudez-Gonzalez, M., Kleiner, G., Aydiillo, T., Miorin, L., Fierer, D. S., Lugo, L. A., Kojic, E. M., Stoeber, J., Liu, S. T. H., ... Krammer, F. (2020). A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nature Medicine*, 26(7), 1033–1036.
4. Hsieh, C.-L., Goldsmith, J. A., Schaub, J. M., DiVenere, A. M., Kuo, H.-C., Javanmardi, K., Le, K. C., Wrapp, D., Lee, A. G., Liu, Y., Chou, C.-W., Byrne, P. O., Hjorth, C. K., Johnson, N. V., Ludes-Meyers, J., Nguyen, A. W., Park, J., Wang, N., Amengor, D., ... McLellan, J. S. (2020). Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science (New York, NY)*, 369(6510), 1501–1505.
5. Stuiblé, M., Gervais, C., Lord-Dufour, S., Perret, S., L'Abbé, D., Schrag, J., St-Laurent, G., & Durocher, Y. (2021). Rapid, high-yield production of full-length SARS-CoV-2 spike ectodomain by transient gene expression in CHO cells. *Journal of Biotechnology*, 326, 21–27.
6. Smaoui, M. R., & Yahyaoui, H. (2021). Unraveling the stability landscape of mutations in the SARS-CoV-2 receptor-binding domain. *Science Reports*, 11(1), 9166.
7. Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veisler, D., & Bloom, J. D. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*, 182(5), 1295–1310.e20.
8. Haryadi, R., Ho, S., Kok, Y. J., Pu, H. X., Zheng, L., Pereira, N. A., Li, B., Bi, X., Goh, L.-T., Yang, Y., & Song, Z. (2015). Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. *PLoS ONE*, 10(2), e0116878.
9. Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N., & Rost, B. (2014). PredictProtein—An open resource for online prediction of protein structural and functional features. *Nucleic Acids Research*, 42(W1), W337–W343.
10. Esposito, D., Mehalko, J., Drew, M., Snead, K., Wall, V., Taylor, T., Frank, P., Denson, J.-P., Hong, M., Gulten, G., Sadtler, K., Messing, S., & Gillette, W. (2020). Optimizing high-yield production of SARS-CoV-2 soluble spike trimers for serology assays. *Protein Expression and Purification*, 174, 105686.

11. Xiong, Y., Karuppanan, K., Bernardi, A., Li, Q., Kommineni, V., Dandekar, A. M., Lebrilla, C. B., Faller, R., McDonald, K. A., & Nandi, S. (2019). Effects of N-glycosylation on the structure, function, and stability of a plant-made Fc-fusion anthrax decoy protein. *Frontiers in Plant Science*, 10, 768.
12. Ernst, O., & Zor, T. (2010). Linearization of the Bradford Protein Assay. *Journal of Visualized Experiments*, 38, 1918.
13. Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166.
14. Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., & Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences*, 112(24), E3095–E3103.
15. Frishman, D., & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23(4), 566–579.
16. Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, 54(1), 5.6.1–5.6.37.
17. Woods, R. (2005). GLYCAM. Complex Carbohydrate Research Center, University of Georgia.
18. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: A generalizable biomolecular force field. *Carbohydrates*.
19. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696–3713.
20. Bernardi, A., Kirschner, K. N., & Faller, R. (2017). Structural analysis of human glycoprotein butyrylcholinesterase using atomistic molecular dynamics: The importance of glycosylation site ASN241. *PLOS ONE*, 12(11), e0187994.
21. Bernardi, A., Faller, R., Reith, D., & Kirschner, K. N. (2019). ACPYPE update for nonuniform 1–4 scale factors: Conversion of the GLYCAM06 force field from AMBER to GROMACS. *SoftwareX*, 10, 100241.
22. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25.
23. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7), 845–854.
24. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16), 1701–1718.
25. Huang, Y., Harris, B. S., Minami, S. A., Jung, S., Shah, P., Nandi, S., McDonald, K., & Faller, R. (2021). SARS-Cov-2 Spike binding to ACE2 is stronger and longer ranged due to glycan interaction (p. 2021.07.15.452507).

26. Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 Spike glycoprotein. *Cell*, 181(2), 281–292.e6.
27. Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C. D., So, R. T. Y., Lv, H., Mok, C. K. P., & Wilson, I. A. (2020). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science*, 368(6491), 630–633.
28. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858.
29. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., & Ferrin, T. E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*, 7(1), 339.
30. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), 70–82.
31. Ren, W., Sun, H., Gao, G. F., Chen, J., Sun, S., Zhao, R., Gao, G., Hu, Y., Zhao, G., Chen, Y., Jin, X., Fang, F., Chen, J., Wang, Q., Gong, S., Gao, W., Sun, Y., Su, J., He, A., ... Sun, L. (2020). Recombinant SARS-CoV-2 spike S1-Fc fusion protein induced high levels of neutralizing responses in nonhuman primates. *Vaccine*, 38(35), 5653–5658.
32. Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12), 2577–2637.
33. Johari, Y. B., Jaffé, S. R. P., Scarrott, J. M., Johnson, A. O., Mozzanino, T., Pohle, T. H., Maisuria, S., Bhayat-Cammack, A., Lambiase, G., Brown, A. J., Tee, K. L., Jackson, P. J., Wong, T. S., Dickman, M. J., Sargur, R. B., & James, D. C. (2021). Production of trimeric SARS-CoV-2 spike protein by CHO cells for serological COVID-19 testing. *Biotechnology and Bioengineering*, 118(2), 1013–1021.
34. Esko, J. D., & Stanley, P. (2015). Glycosylation mutants of cultured mammalian cells. In A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, & P. H. Seeberger (Eds.), *Essentials of glycobiology* (3rd ed.). Cold Spring Harbor Laboratory Press.
35. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, 369(6501), 330–333.

7.6 Supporting Information

Table S1. Protein sequences of CHO-expressed Spike truncations. Signal sequences are highlighted in gray, and potential N-linked glycosylation sites are highlighted in green. Percentages indicate percent coverage of full-length Spike construct, including signal sequences and 6x His tags.

Truncation	Amino acid sequence, including secretion signal and 6x His tag
T1 (23.5%)	MFVFLVLLPLVSSQNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWN RKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLCFTNVYADSFVIRGDE VRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFR KSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY RVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL PFQQFGRDIADTTDAVRDPQTLEILDITPCSHHHHHH
T2 (25.1%)	MFVFLVLLPLVSSQTAGAAAYVGYLQPRTFLLKYNE ^N GTITDAVDCALDPL SETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLCFTNVYADSF VIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLY YRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGV GYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTE HHHHHH
T3 (28.1%)	MFVFLVLLPLVSSQTAGAAAYVGYLQPRTFLLKYNE ^N GTITDAVDCALDPL SETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLCFTNVYADSF VIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLY YRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGV GYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTE SNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSHHHHHH
T4 (27.5%)	MFVFLVLLPLVSSQNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWN RKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLCFTNVYADSFVIRGDE VRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFR KSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY RVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN ^N TSNQVAVLYQ DV ^N CTEVPVAIHADQLTPTWRVYSTGSNVHHHHHH
T5 (28.3%)	MFVFLVLLPLVSSQGFSALEPLVDLPIGNITRFQTLALHRSYLTPGDSSSG WTAGAAAYVGYLQPRTFLLKYNE ^N GTITDAVDCALDPLSETKCTLKSFTVE KGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNC VADYSVLYNSASFSTFKCYGVSPTKLNLCFTNVYADSFVIRGDEVQRQIAPG QTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKP FERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSF ELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTEHHHHHH
T6 (35.3%)	MFVFLVLLPLVSSQGFSALEPLVDLPIGNITRFQTLALHRSYLTPGDSSSG WTAGAAAYVGYLQPRTFLLKYNE ^N GTITDAVDCALDPLSETKCTLKSFTVE KGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNC

	VADYSVLYNSASFSTFKCYGVSPTKLNLCFTNVYADSFVIRGDEVQRQIAPG QTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKP ERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSF ELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFG RDIADTTDAVRDPQTLEILDITPCSFSGGVSVITPGTNTSNQVAVLYQDVNCTE VPVAIHADQLTPTWRVYSTGNSVHHHHHH
T7	MFVFLVLLPLVSSQNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWN RKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLCFTNVYADSFVIRGDE VRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFR KSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY RVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL PFQQFGRDIADTTDAVRDPQTLEILDITPCSFSGGVSVITPGTNTSNQVAVLYQ DVNCTEVPVAIHADQLTPTWRVYSTGNSVFNQTRAGCLIGAHEVNSYECDIP IGAGICASYQTQTNSPASVASQSIIAYTMSLGAENSVAYSNSNSIAIPTNFTISV TTEILPVSMTKTSVDCTMYICGDHHHHHH
T8 (35.5%)	MFVFLVLLPLVSSQKTQSLNIVNATNVVIKVFCEFCNDPFLGVYYHKNNKS WMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFNIDGYFK IYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTGDS SSGWTAGAAAYYVGYLQPRTFLLKYNEGTITDAVDCALDPLSETKCTLSFTV EKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRISN CVADYSVLYNSASFSTFKCYGVSPTKLNLCFTNVYADSFVIRGDEVQRQIAP GQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKP FERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLS FELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTEHHHHHH
S1 (54.7%)	MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQ DLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTKSNIRGWI FGTTLDLSDKTQSLNIVNATNVVIKVFCEFCNDPFLGVYYHKNNKSWMESEF RVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFNIDGYFKIYSKHTP INLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTGDS SSGWTAGAAAYYVGYLQPRTFLLKYNEGTITDAVDCALDPLSETKCTLSFTVEKGIYQ TSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRISNCVADYSV LYNSASFSTFKCYGVSPTKLNLCFTNVYADSFVIRGDEVQRQIAPGQTGKIA DYNLYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDISTE IQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPA TVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTT DAVRDPQTLEILDITPCSFSGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHA DQLTPTWRVYSTGNSVFNQTRAGCLIGAHEVNSYECDIPIGAGICASYQTQ TNSPAHHHHHH

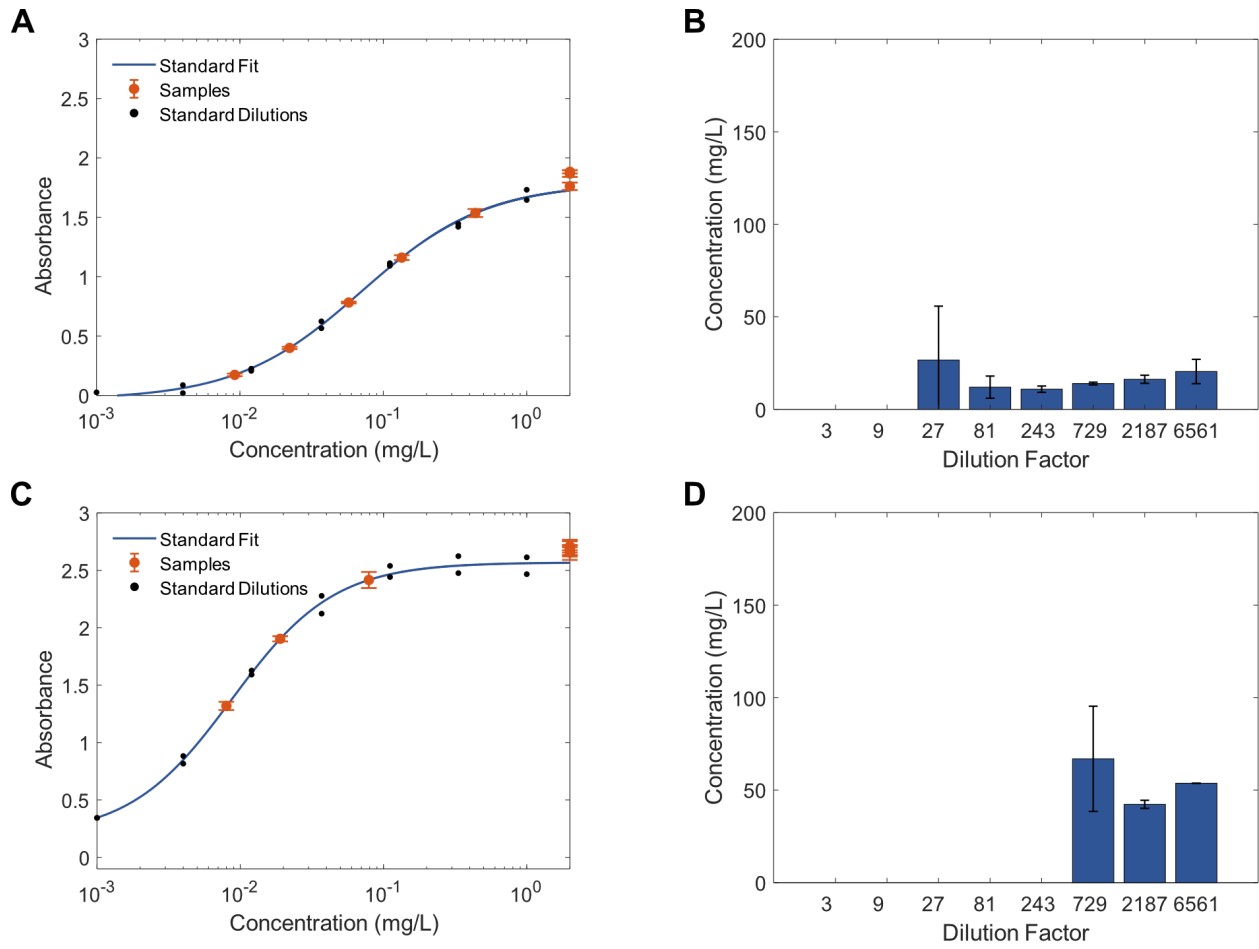


Figure S1. Measurement of crude titers for Spike and RBD via sandwich ELISA. (A) Dilutions of crude Spike plotted against the Spike standard curve. (B) Back-calculated concentrations of Spike crude titers. (C) Dilutions of crude RBD plotted against the RBD standard curve. (D) Back-calculated concentrations of RBD crude titers. Error bars represent \pm SD of technical triplicates.

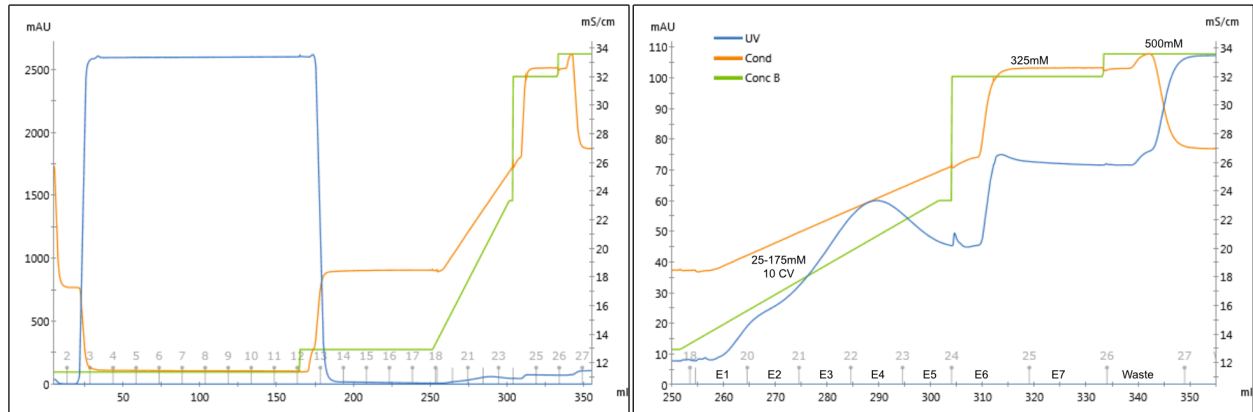
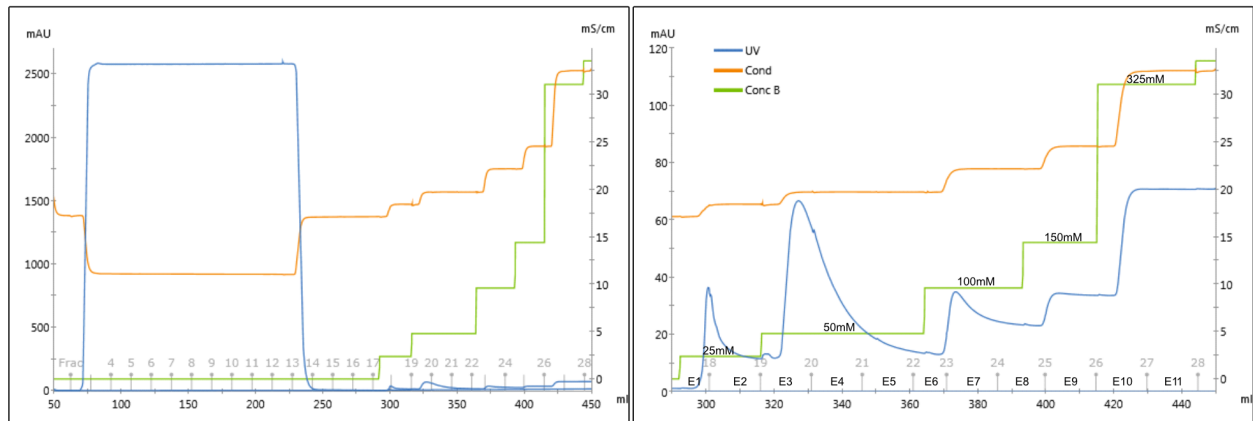
A**B**

Figure S2. Spike and RBD chromatograms plotted over volume flowed through the system. (A) Chromatogram of entire Spike purification (left) and of zoomed in elution fractions (right). 150 mL of sample was loaded and washed with PBS containing 25 mM imidazole. A continuous gradient was applied from 25 mM-175 mM imidazole over 10 CV to elute Spike. Fractions E3 and E4 were collected. (B) Chromatogram of entire RBD purification (left) and of zoomed in elution fractions (right). 150 mL of sample was loaded and washed for 10 CV with PBS. A step gradient was applied for elution with 6 CV steps at 25 mM, 50 mM, 100 mM, 150 mM, and 325 mM imidazole. Fractions E3 and E4 were collected. Y-axis on the left is UV absorbance and Y-axis on the right is conductivity.

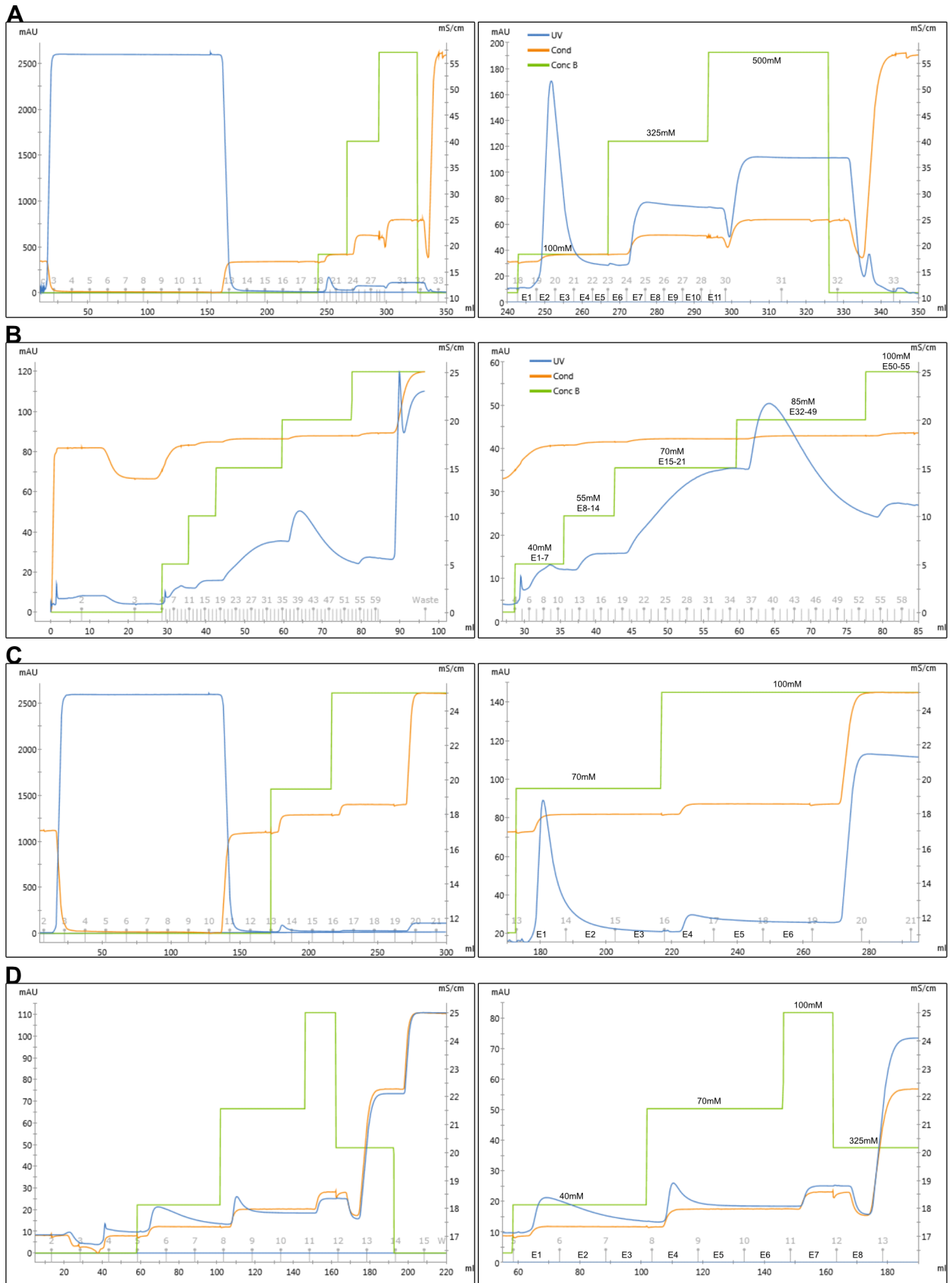


Figure S3. T1 and T4 purification chromatograms. (A) Chromatogram of entire T1 purification (left) and of zoomed in elution fractions (right). First, 150 mL of sample was loaded and washed with PBS containing 25 mM imidazole. Step gradients were applied at 100 mM and 325 mM imidazole to elute T1. (B) Fractions E2 and E3 from the first purification were combined, dialyzed and re-purified using a step gradient with steps at 40 mM, 55 mM, 70 mM, 85 mM, and 100 mM imidazole. Fractions eluting from 40-85 mM imidazole were combined. (C) Chromatogram of entire T4 purification (left) and of zoomed in elution fractions (right). First, 120 mL of sample was loaded, and the resin was washed with PBS containing 25 mM imidazole. A step gradient at 70 mM and 100 mM imidazole was used to elute T4. (D) T4 from the first purification was dialyzed and re-purified using step gradients at 40 mM, 70 mM, 100 mM, 325 mM, and 500 mM imidazole. 40 mM imidazole fractions and the first 70 mM imidazole fraction were combined and used for further characterization.

A

```
Ref_T1_Seq      NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISNCVADYSVLYNSASFST
T1_Coverage     -FRVQPTESIVR-----FASVYAWNKRKISNCVADYSVLYNSASFST
                *****
                *****

Ref_T1_Seq      FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAW
T1_Coverage     FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAW
                *****

Ref_T1_Seq      NSNNLDSKVGGNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYG
T1_Coverage     NSNNLDSKVGGNLYLR--KSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYG
                *****

Ref_T1_Seq      FQPTNGVGYPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNGLTGTGVLTESN
T1_Coverage     FQPTNGVGYPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNGLTGTGVLTESN
                *****

Ref_T1_Seq      KKFLPFQQFGRDIADTTDAVRDPQTEILDITPCSHHHHHH
T1_Coverage     KKFLPFQQFGRDIADTTDAVRDPQTEILDITPCSHHHHHH
                *****
```

B

```
Ref_T4_Seq      NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISNCVADYSVLYNSASFST
T4_Top_Coverage --RVQPTESIVR-----FASVYAWNKRKISNCVADYSVLYNSASFST
                *****
                *****

Ref_T4_Seq      FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAW
T4_Top_Coverage FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAW
                *****

Ref_T4_Seq      NSNNLDSKVGGNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYG
T4_Top_Coverage NSNNLDSKVGGNLYLR--KSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYG
                *****

Ref_T4_Seq      FQPTNGVGYPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNGLTGTGVLTESN
T4_Top_Coverage FQPTNGVGYPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNGLTGTGVLTESN
                *****

Ref_T4_Seq      KKFLPFQQFGRDIADTTDAVRDPQTEILDITPCFSGGVSITPGTNTSNQVAVLYQDVN
T4_Top_Coverage KKFLPFQQFGRDIADTTDAVRDPQTEILDITPCS-----
                *****

Ref_T4_Seq      CTEVPVAIHADQLTPTWRVYSTGSNVHHHHHH
T4_Top_Coverage -----HHHHHH
                *****
```

C

```
Ref_T4_Seq      NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISNCVADYSVLYNSASFST
T4_Bottom_Cover -FRVQPTESIVR-----FASVYAWNKRKISNCVADYSVLYNSASFST
                *****
                *****

Ref_T4_Seq      FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAW
T4_Bottom_Cover FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAW
                *****

Ref_T4_Seq      NSNNLDSKVGGNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYG
T4_Bottom_Cover NSNNLDSKVGGNLYLR--KSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYG
                *****

Ref_T4_Seq      FQPTNGVGYPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNGLTGTGVLTESN
T4_Bottom_Cover FQPTNGVGYPYRVVLSFELLHAPATVCGPKK-----CVNFNGLTGTGVLTESN
                *****

Ref_T4_Seq      KKFLPFQQFGRDIADTTDAVRDPQTEILDITPCFSGGVSITPGTNTSNQVAVLYQDVN
T4_Bottom_Cover KKFLPFQQFGRDIADTTDAVRDPQTEILDITPCS-----
                *****

Ref_T4_Seq      CTEVPVAIHADQLTPTWRVYSTGSNVHHHHHH
T4_Bottom_Cover -----HHHHHH
                *****
```

Figure S4. Shotgun proteomics on T1 and T4. Coverage of (A) T1, (B) T4 top band, and (C) T4 bottom band against full sequences. Tandem mass spectra were extracted by MS Convert (ProteoWizard). Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analyzed using X! Tandem (The GPM, thegpm.org; version X! Tandem Alanine (2017.2.1.4)). X! Tandem was set up to search the Uniprot human database and known T1 and T4 sequences assuming the digestion enzyme trypsin. X! Tandem was searched with a fragment ion mass tolerance of 20 PPM and a parent ion tolerance of 20 PPM. Carbamidomethyl of cysteine and selenocysteine was specified in X! Tandem as a fixed modification. Glu->pyro-Glu of the n-terminus, ammonia-loss of the n-terminus, gln->pyro-Glu of the n-terminus, deamidated of asparagine and glutamine, oxidation of methionine and tryptophan and dioxidation of methionine and tryptophan were specified in X! Tandem as variable modifications. Scaffold (version Scaffold_4.9.0, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 98.0% probability by the Scaffold Local FDR algorithm. Peptide identifications were also required to exceed specific database search engine thresholds. Protein identifications were accepted if they could be established at greater than 5.0% probability to achieve an FDR less than 5.0% and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003) Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

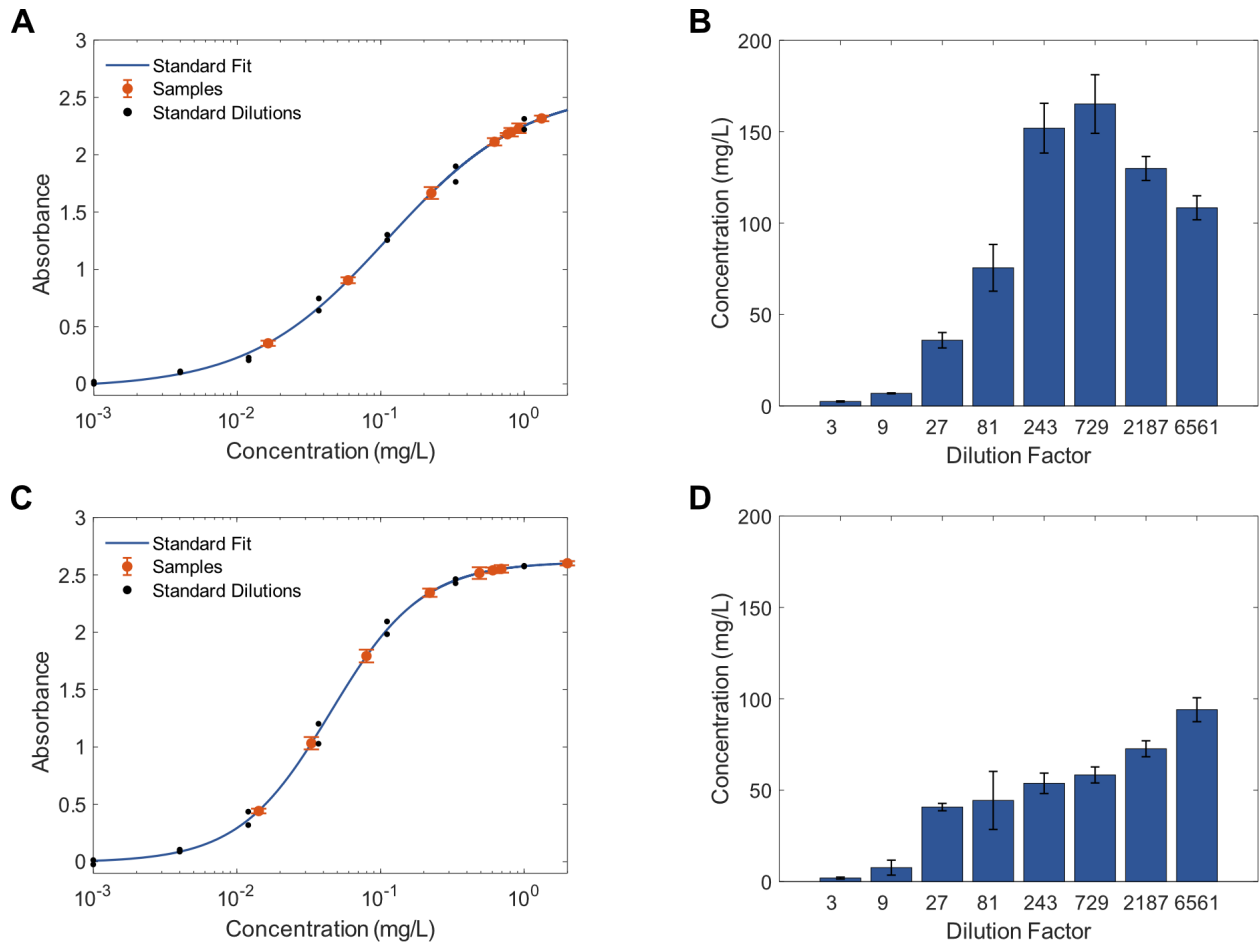


Figure S5. Measurement of crude titers for T1 and T4 via sandwich ELISA. (A) Dilutions of crude T1 plotted against the T1 standard curve. (B) Back-calculated concentrations of T1 crude titers. (C) Dilutions of crude T4 plotted against the T4 standard curve. (D) Back-calculated concentrations of T4 crude titers. Error bars represent \pm SD of technical triplicates.

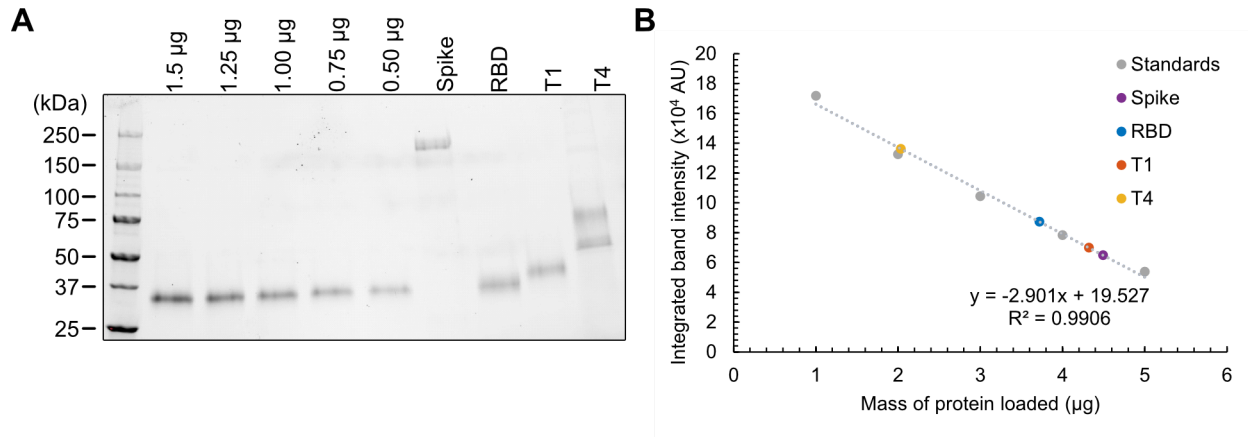


Figure S6. Quantification of purified proteins. (A) SDS-PAGE and (B) quantification for purified proteins and serial dilutions of RBD obtained from BEI Resources. A standard curve was prepared using a linear fit to serial dilutions of the standard protein.

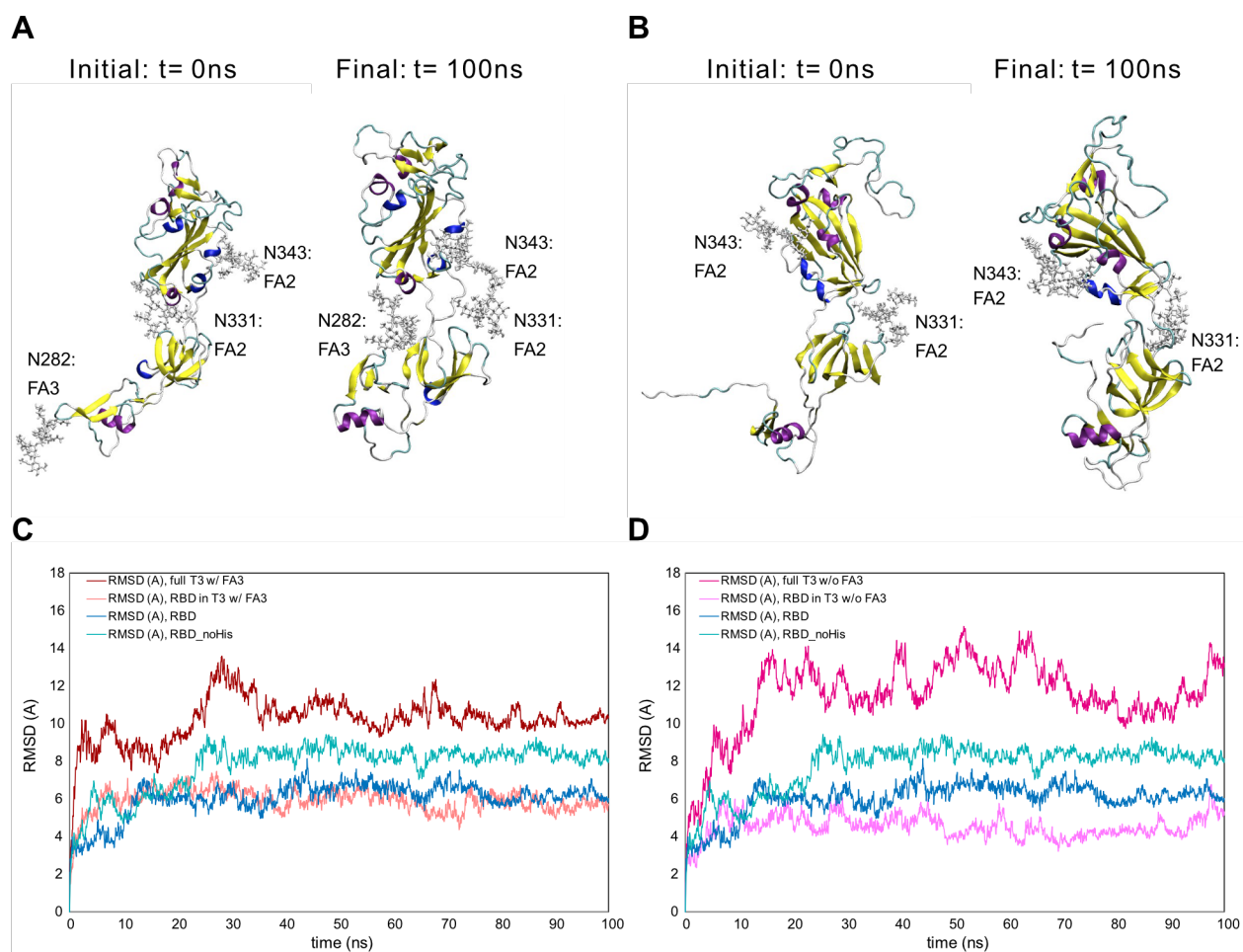


Figure S7. Molecular dynamics structural stability snapshots and analysis of T3. MD snapshots are visualized for **(A)** T3 with FA3 glycan and **(B)** T3 without FA3 glycan. **(C)** Backbone RMSD profiles of full T3 and T3 RBD subdomain with FA3 glycan. **(D)** Backbone RMSD profiles of full T3 and T3 RBD subdomain without FA3 glycan. Profiles of **(C)** and **(D)** include RBD with and without the 6x His tag referenced to initial configurations for comparison.

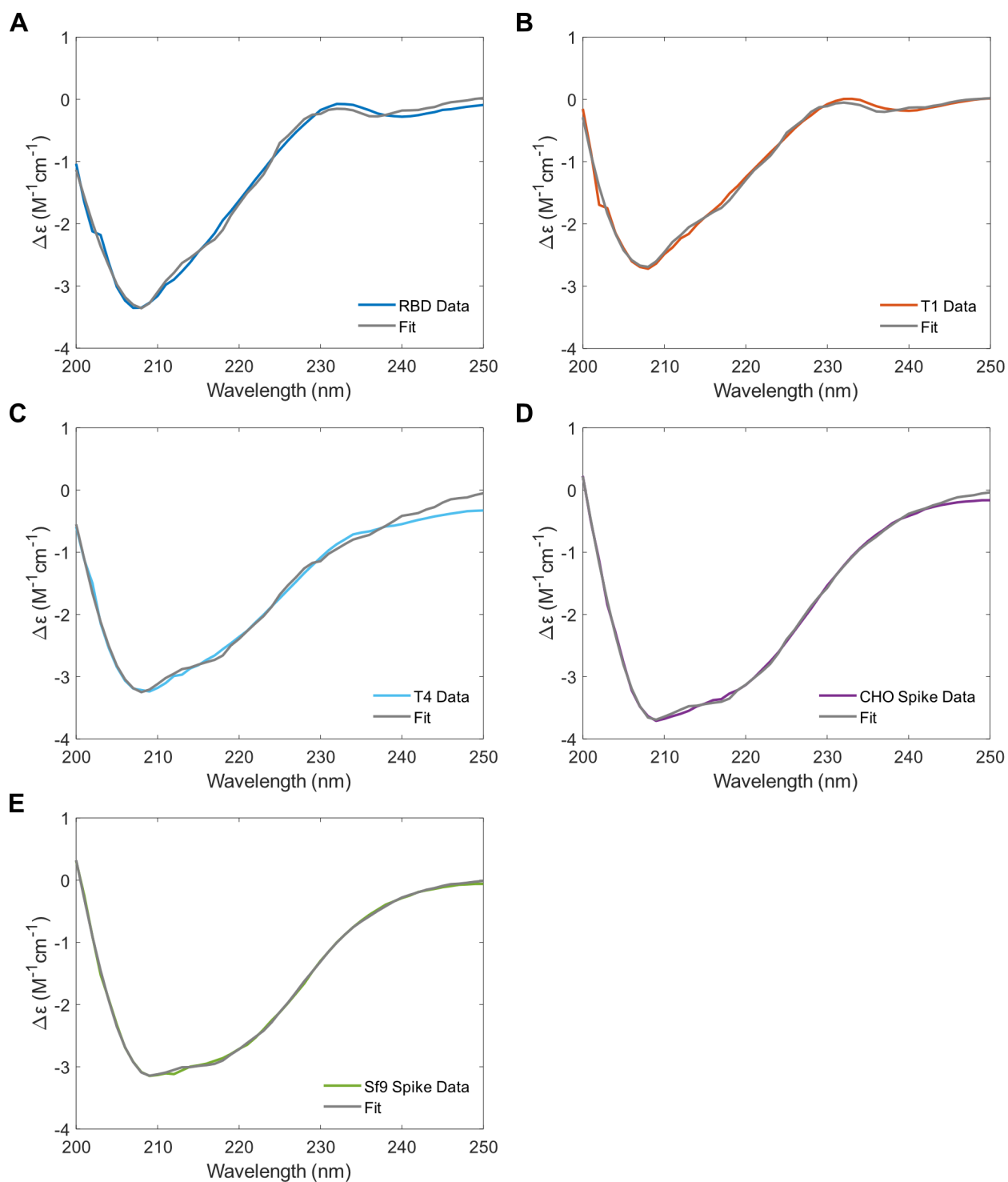


Figure S8. Raw spectral data on proteins analyzed via circular dichroism. $\Delta\epsilon$ is plotted against wavelength for (A) RBD, (B) T1, (C) T4, (D) CHO Spike, and (E) Sf9 Spike.

Chapter 8

Summary and Outlook

In this work we demonstrated two classes of molecular systems, molecules or materials under chemical reactions and glycosylated proteins, that can be studied through quantum mechanics calculations and molecular dynamics simulations. For the chemical reaction studies, we used two different reactive potentials, both are bond order dependent empirical force fields that require input training data from density functional theory calculations but with very different approaches to handle interatomic interactions. The reactive potentials used in this work fill in the gap between quantum mechanical methods and classical dynamical methods to enable describing charge transfer, bonding, and chemical reactivity of systems in evolution, while keeping the computational cost reasonable and manageable. One Tersoff potential and two ReaxFF force fields were developed, and they were applied to various applications including atomic layer etching, carbon capture, and photopolymer resin for volumetric additive manufacturing. The microscopic properties and reaction mechanisms evaluated and studied can provide insights to materials design and processing in the macroscopic world. The force field developed for atomic layer etching was able to describe the surface at nanoscale and predict qualitatively the correlations between the chlorination and/or Ar bombardment energies and the penetration depth, etching efficiency, etching species, etc., which provided theoretical guidance when tuning processing parameters. The ReaxFF force field developed for novel biomimetic carbon capture absorbent could successfully describe the reactions hypothesized and proved PEP can be a CO₂ absorbent to substitute MEA. Together with DFT calculations, we could then design better absorbents from PEP molecule. The

ReaxFF force field developed for acrylate free-radical polymerization provided a model to study such reaction, and since acrylates were often used as 3D printing resins, it could help with resin materials design, and set the bench mark for more novel photopolymer resins. Our model development and computational studies on glycoproteins offered a molecular model for glycosylated S protein and human ACE2, which has significant value since S protein are normally glycosylated to evade host immune system and bind human receptor more strongly, but such computational studies were rare. Using our computational model, we were able to study how glycans affected binding and stability of the proteins, and help evaluate and design spike truncations that could possibly advance to vaccine candidates.

As demonstrated, MD simulations is a powerful and versatile tool to study the atomistic scale properties and mechanisms of molecules and materials in various systems, and predict their macroscopic behaviors. However, due to the computational cost, its limitations are also obvious. In addition to acquiring better computational resources, there are also many ways to improve the modeling techniques and performances. For example, multiscale simulation techniques could be used. Coarse graining could be used together with atomistic simulation like reactive MD simulations to investigate different aspects of the system at different length scale. More advanced reactive force field, like accelerated ReaxFF, could also be used to speed up the reactions. Experiments should also be included for future work to compare with the computational results. In this work, experiments were only used to validate some of the thermodynamic and diffusion properties and reaction mechanisms (eg. FTIR). In the future, picosecond spectroscopy could be used to observe reactions at comparable time scale and verify the reactive simulation findings. For protein simulations, one of the limitations is the lack of availability of protein structures. Utilizing AI predicted structures, like the ones predicted by AlphaFold, and incorporating machine learning

techniques might empower the modeling and design of therapeutic proteins. To get better conformational sampling, enhanced simulation methods such as replica exchange molecular dynamics could be used in the future. In addition, the current fully glycosylated model could also be extended to more virus and human receptor systems to study the glycosylation effects for therapeutic and diagnostic designs.

Bibliography

1. Parr RG. Density functional theory of atoms and molecules. In *Horizons of quantum chemistry 1980* (pp. 5-15). Springer, Dordrecht.
2. Fiolhais C, Nogueira F, Marques MA, editors. *A primer in density functional theory*. Springer Science & Business Media; 2003 Jun 11.
3. Koch W, Holthausen MC. *A chemist's guide to density functional theory*. John Wiley & Sons; 2015 Nov 18.
4. Kohn W, Becke AD, Parr RG. Density functional theory of electronic structure. *The Journal of Physical Chemistry*. 1996 Aug 1;100(31):12974-80.
5. Car R, Parrinello M. Unified approach for molecular dynamics and density-functional theory. *Physical review letters*. 1985 Nov 25;55(22):2471.
6. Gao J, Thompson MA. *Combined quantum mechanical and molecular mechanical methods*. Washington, DC: American Chemical Society; 1998 Dec 28.
7. Comba P, Hambley TW, Martin B. *Molecular modeling of inorganic compounds*. John Wiley & Sons; 2009 Jul 10.
8. Allen MP, Tildesley DJ. *Computer simulation of liquids*. Oxford university press; 2017 Aug 15.
9. Schlick T. *Molecular modeling and simulation: an interdisciplinary guide*. New York: Springer; 2010 Aug 25.
10. Gilmer GH, Huang H, Roland C. Thin film deposition: fundamentals and modeling. *Computational Materials Science*. 1998 Nov 1;12(4):354-80.
11. Farrow CL, Billinge SJ. Relationship between the atomic pair distribution function and small-angle scattering: implications for modeling of nanoparticles. *Acta Crystallographica Section A: Foundations of Crystallography*. 2009 May 1;65(3):232-9.
12. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron*. 2018 Sep 19;99(6):1129-43.
13. Vriend G. WHAT IF: a molecular modeling and drug design program. *Journal of molecular graphics*. 1990 Mar 1;8(1):52-6.
14. Jones, R.O. Density functional theory: Its origins, rise to prominence, and future. *Reviews of modern physics*. 2015;87(3):897.
15. Geerlings P, De Proft F, Langenaeker W. Conceptual density functional theory. *Chemical reviews*. 2003 May 14;103(5):1793-874.
16. Burke K. Perspective on density functional theory. *The Journal of chemical physics*. 2012 Apr 21;136(15):150901.
17. Alder B.J., Wainwright T.E. Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics*. 1959 Aug;31(2):459-66.
18. Hollingsworth S.A., Dror R.O. Molecular dynamics simulation for all. *Neuron*. 2018 Sep 19;99(6):1129-43.
19. Du J, Rimsza J.M. Atomistic computer simulations of water interactions and dissolution of inorganic glasses. *npj Materials Degradation*. 2017 Nov 13;1(1):1-2.
20. Parr R.G. Density functional theory of atoms and molecules. In *Horizons of quantum chemistry 1980* (pp. 5-15). Springer, Dordrecht.

21. Koch W., Holthausen M.C. A chemist's guide to density functional theory. John Wiley & Sons; 2015 Nov 18.
22. Sholl D.S., Steckel J.A. Density functional theory: a practical introduction. John Wiley & Sons; 2011 Sep 20.
23. Combes J.M., Duclos P, Seiler R. The born-oppenheimer approximation. InRigorous atomic and molecular physics 1981 (pp. 185-213). Springer, Boston, MA.
24. Born M., Heisenberg W. Zur quantentheorie der molekeln. InOriginal Scientific Papers Wissenschaftliche Originalarbeiten 1985 (pp. 216-246). Springer, Berlin, Heidelberg.
25. Kohn W., Sham L.J. Self-consistent equations including exchange and correlation effects. Physical review. 1965 Nov 15;140(4A):A1133.
26. Vosko S.H, Wilk L., Nusair M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Canadian Journal of physics. 1980 Aug 1;58(8):1200-11.
27. Vosko S.H., Wilk L. Influence of an improved local-spin-density correlation-energy functional on the cohesive energy of alkali metals. Physical Review B. 1980 Oct 15;22(8):3812.
28. Becke A.D. Perspective: Fifty years of density-functional theory in chemical physics. The Journal of chemical physics. 2014 May 14;140(18):18A301.
29. Perdew J.P., Chevary J.A., Vosko S.H., Jackson K.A., Pederson M.R., Singh D.J., Fiolhais C. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. Physical review B. 1992 Sep 15;46(11):6671.
30. Perdew J.P., Burke K., Ernzerhof M. Generalized gradient approximation made simple. Physical review letters. 1996 Oct 28;77(18):3865.
31. Perdew J.P., Chevary J.A., Vosko S.H., Jackson K.A., Pederson M.R., Singh D.J., Fiolhais C. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. Physical review B. 1992 Sep 15;46(11):6671.
32. Becke A.D. Density-functional thermochemistry. III The role of exact exchange J Chem Phys 98: 5648–5652.
33. Lee C., Yang W., Parr R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical review B. 1988 Jan 15;37(2):785.
34. Gkeka P. Molecular dynamics studies of peptide-membrane interactions: Insights from coarse-grained models.
35. Verlet L. Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Physical review. 1967 Jul 5;159(1):98.
36. Swope W.C., Andersen H.C., Berens P.H., Wilson K.R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. The Journal of chemical physics. 1982 Jan 1;76(1):637-49.
37. Allen M.P., Tildesley D.J.. Computer simulation of liquids. Oxford university press; 2017 Aug 15.
38. Gibbs J.W. Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics. C. Scribner's sons; 1902.
39. Evans D.J., Holian B.L. The nose–hoover thermostat. The Journal of chemical physics. 1985 Oct 15;83(8):4069-74.

40. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *The Journal of chemical physics*. 2007 Jan 7;126(1):014101.
41. Parrinello M, Rahman A. Crystal structure and pair potentials: A molecular-dynamics study. *Physical review letters*. 1980 Oct 6;45(14):1196.
42. Jensen L.F., Grønbech-Jensen N. Accurate configurational and kinetic statistics in discrete-time Langevin systems. *Molecular Physics*. 2019 Sep 17;117(18):2511-26.
43. Ho C.H., Chen C.O., Chen C.L.. A Study on Sputtering of Copper Seed Layer for Interconnect Metallization via Molecular Dynamics Simulation. *Applied Sciences*. 2021 Oct 18;11(20):9702.
44. Frenkel D, Smit B. *Understanding molecular simulation: from algorithms to applications*. Elsevier; 2001 Oct 19.
45. Schall J.D., Mikulski P.T., Ryan KE, Keating P.L., Knippenberg M.T., Harrison J.A. Reactive empirical bond-order potentials. *Encyclopedia of Nanotechnology*. 2012:2210-21.
46. Tersoff J.J. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Physical review B*. 1989 Mar 15;39(8):5566.
47. Van Duin A.C., Dasgupta S, Lorant F, Goddard W.A. ReaxFF: a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*. 2001 Oct 18;105(41):9396-409.
48. Chenoweth K, Van Duin A.C., Goddard W.A. ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation. *The Journal of Physical Chemistry A*. 2008 Feb 7;112(5):1040-53.
49. Senftle T.P., Hong S, Islam M.M., Kylasa S.B., Zheng Y, Shin Y.K., Junkermeier C, Engel-Herbert R, Janik M.J., Aktulga H.M., Verstraelen T. The ReaxFF reactive force-field: development, applications and future directions. *npj Computational Materials*. 2016 Mar 4;2(1):1-4.
50. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P. A novel coronavirus from patients with pneumonia in China, 2019. *New England journal of medicine*. 2020 Jan 24.
51. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature microbiology*. 2020 Apr;5(4):562-9.
52. Huang Y, Yang C, Xu X.F., Xu W, Liu S.W. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica*. 2020 Sep;41(9):1141-9.
53. Bosch BJ, Van der Zee R, De Haan CA, Rottier PJ. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *Journal of virology*. 2003 Aug 15;77(16):8801-11.
54. Vigerust D.J., Shepherd V.L. Virus glycosylation: role in virulence and immune interactions. *Trends in microbiology*. 2007 May 1;15(5):211-8.
55. Ohtsubo K, Marth J.D. Glycosylation in cellular mechanisms of health and disease. *Cell*. 2006 Sep 8;126(5):855-67.
56. Phillips M.L., Nudelman E, Gaeta F.C., Perez M, Singhal A.K., Hakomori S.I., Paulson JC. ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand, sialyl-Lex. *Science*. 1990 Nov 23;250(4984):1130-2.

57. Sperandio M, Gleissner C.A., Ley K. Glycosylation in immune cell trafficking. *Immunological reviews*. 2009 Jul;230(1):97-113.
58. Watanabe Y, Bowden T.A., Wilson I.A., Crispin M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA)-General Subjects*. 2019 Oct 1;1863(10):1480-97.
59. Banerjee N, Mukhopadhyay S. Viral glycoproteins: biological role and application in diagnosis. *Virusdisease*. 2016 Mar;27(1):1-1.
60. Bagdonaite I, Wandall H.H. Global aspects of viral glycosylation. *Glycobiology*. 2018 Jul;28(7):443-67.
61. Marshall R.D. Glycoproteins. *Annual review of biochemistry*. 1972 Jul;41(1):673-702.