UC Davis
UC Davis Electronic Theses and Dissertations

Title
Molecular Modeling for 3D Printing and Biological Applications

Permalink
https://escholarship.org/uc/item/9881w4hx

Author
Harris, Bradley Scott

Publication Date
2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Molecular Modeling for 3D Printing and Biological Applications

By

BRADLEY SCOTT HARRIS
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Roland Faller, Chair

Gang-yu Liu

Tonya Kuhl

Committee in Charge

2021

© 2021 – Bradley Scott Harris

All rights reserved.

Molecular Modeling for 3D Printing and Biological Applications

Abstract

This dissertation covers several molecular modeling efforts applied across length and time scales. These techniques are primarily applied to two major categories: 3D printing (chapters 2-4) and biological applications (chapters 5-7). These computational efforts were part of a collaborative effort to improve the experimental design process for these two applications and have corresponding experimental validation.

We first present the application of multi-scale simulation techniques to the process of controlled assembly, which is a critical pre-cursor to enabling 3D printing at the nano-scale. This controlled assembly process depends on the complex interplay between surface, solute, and solvent interactions during the evaporation process. In order to actually model the assembly that takes place during this evaporation process a computational software and methodology had to be developed to replicate evaporation of mixed solvent systems and account for potential hydrogen bonding shells. This resulting methodology was tested for different evaporation rates and scales and is presented in Chapter 2. Chapters 3 and 4 are application papers using this methodology. Chapter 3 focuses on the assembly of small linear sugar molecules analyzed using atomistic molecular dynamics simulations. This atomistic resolution allows the probing of phase transitions these sugars undergo as they assemble into various

morphological structures during evaporation. Chapter 4 focuses on the coarse-grained interactions between the phospholipid POPC, various surface types, and mixed solvents. This work investigated the specific contributions of surface type and solvent type on assembled features as a function of evaporation.

The final chapters of the work 5-7, focus on the virus that causes COVID-19, SARS-CoV-2. While no longer a part of an additive manufacturing process, these chapters remain a part of a collaborative effort to apply computational tools towards the experimental design process of a specific application, the development of a diagnostic or therapeutic tool. These chapters focus on the evaluation of SARS-CoV-2 spike protein, and specifically the receptor binding domain (RBD) and its interactions with the human angiotensin-converting enzyme 2 (hACE2) receptor. These structures were simulated using atomistic molecular dynamics, evaluated for the structural stability under different glycosylations that would derive from different synthesis methods for treatment (chapter 5). The interactions between the two during binding were evaluated using molecular force pulling simulations, with specific focus on analyzing the hydrogen bonding, electrostatic, and van der waals interactions between the two (chapter 6). This work demonstrated that they glycosylation contributes directly and indirectly to strengthening the interaction with hACE2. Finally, the structural stability of various spike truncations synthesized in Chinese hamster ovary cells for use in therapeutic / diagnostic applications were analyzed in chapter 7.

Contents

Acknowledgments

Thank you to my friends and family for the love and support they provided during the pursuit of this degree. Thank you to my advisor Roland for his mentorship, guidance, patience, insight, humor, and Pokémon Go gifts. Thank you to my advisor Gang-Yu for her humor, guidance, experimental perspectives, and physical chemistry insights, without which the punchlines for this dissertation would have been much harder to write. I thank my colleagues and collaborators: Dr. Austen Bernardi, Dr. Vincent Ustach, Dr. Tommy Harrellson, Yihan Huang, Dr. Jiali Zhang, Dr. Arpad Karsai, Dr. Umit Celik, Yunbo Zheng, Yuqi Huang, Matt Owen, Dr. John Karnes, Christy Fox, and Shiaki Minami for their council, friendship, and experimental work. My close friends and housemates, Andy Lam, Michael Bull, Noah Felvey, Matt McNulty, Jose Hernandez, Mike Meloni, Christine Smudde, Kristen Hwang, Nhu Tran, and Merve Demir, who I've leaned on for too many things to list. My collaborators at Lawrence Livermore National Lab for funding a portion of this work and advancing it through discussion. Finally thank you to professor Tonya Kuhl for her assistance as a dissertation committee member.

To my family for supporting and inspiring my pursuit of this degree.

Especially my parents Scott and Amy Harris, and my grandparents, Poppy and Gamby

Chapter 1

Introduction

Much as experimental control over temporal and spatial resolution have enabled scientific progress to reach new applications and led to the development of new fields, improvements in computational performance and parallelization have enabled more complex processes to be modeled.[1-7] These improvements allow for better physical insight into complex processes that can be difficult to elucidate experimentally. Integration of computational models into experimental workflows can improve both the understanding of results and the experimental design process.[7] Typically, these experimental systems involve processes that can't be captured by a single time or length scale, thus multiscale modeling approaches are often needed.[8]

Multiscale modeling refers to the use of multiple models that operate on different principles, length, or time scales to simultaneously describe the same system. The way these different techniques achieve these differences in resolution is often due to differences in the type of physics or assumptions that are being used in their calculations.[8,9] Multi-scale modeling is necessary for systems that don't have accurate enough macroscale models, and for systems in which the microscale models may not be efficient enough or offer too much extraneous information to justify the increased computational cost. These approaches combine techniques to achieve a balance between accuracy and computational efficiency. Multiscale

workflows are also necessary for systems that have large amounts of complexity that are best described through a combination of techniques probing different length scales. Figure 1 shows a rough illustration of the different length and time scales that various classes of simulation techniques are capable of modeling.

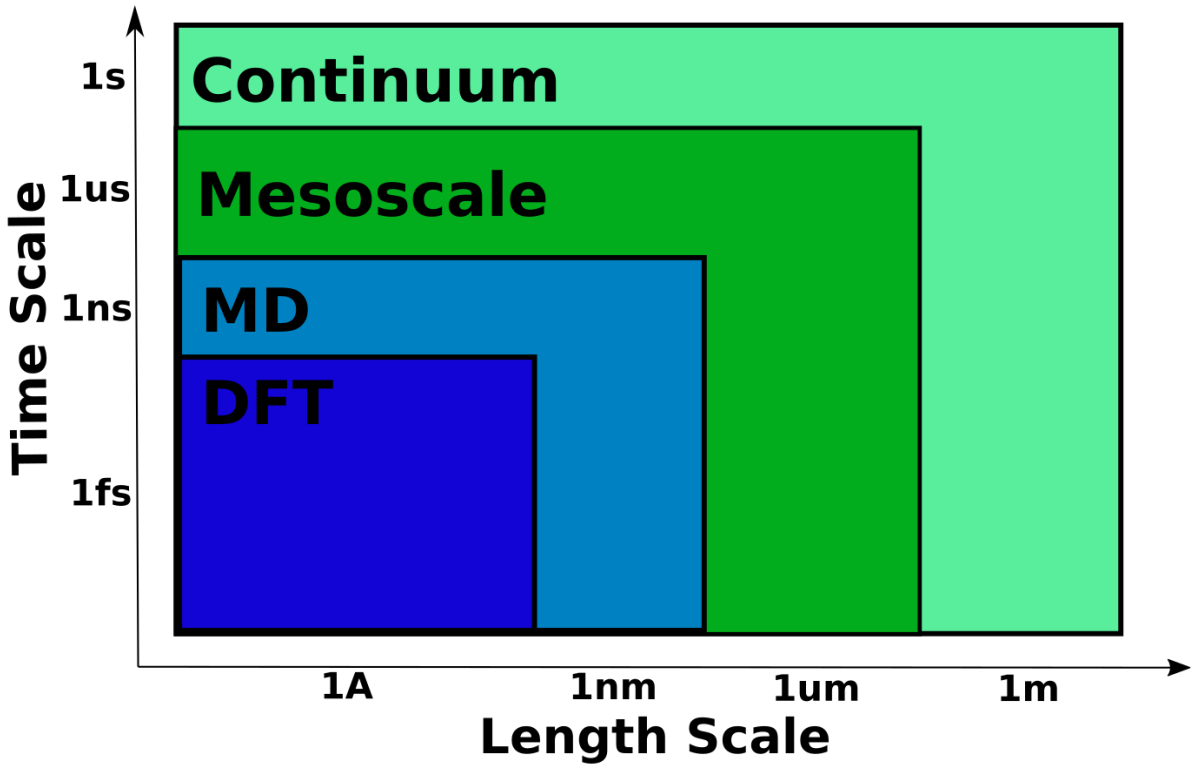


Figure 1: Multiscale modeling techniques at different time and length scales

Here it becomes clear how different techniques can be used at different scales to achieve different levels of information. For large systems or systems that require long time scales techniques based on equations of state (EOS) or empirical models are frequently applied.[10,11] These techniques are broadly considered as part of the continuum scale because they describe their system as an overall macroscopic process rather than as individual smaller components. On the opposite end of the spectrum is the field of quantum mechanics,

which describes not only the behavior of individual atoms, but also individual quantized electrons.[13-15] Techniques built on this principal are highly accurate but come with a large computational cost resulting in the ability to probe only short length and time scales. [16] Increasing in scale slightly to describe more information such as the dynamics and properties of atoms, and techniques such as molecular dynamics (MD) or Monte Carlo (MC) sampling come into play.[17-20] To sacrifice further detail in order to accomplish longer time and length scales while still describing the motion of atoms intermediate or mesoscale models such as coarse-grained molecular dynamics (CG-MD) can be used.[21,22] Several of these techniques will be described in more detail in the following subsections.

Using fluid dynamics as an example and working down in length scale, the continuum model would be the Navier-Stokes equations, which describe the motion of viscous fluids in a purely differential equation sense.[23,24] An intermediate or mesoscale model would be the lattice Boltzmann method (LBM), which describes the motion of fluids as momentum probabilities derived from the more descriptive Boltzmann distribution.[25-27] And finally at the shortest time scale the momentum of the fluids can be modeled by molecular dynamics and simulating the motion of individual atoms.[24]

The work herein will primarily use an application in controlled assembly and 3D-nanoprinting as an example by which multi-scale modeling techniques can be used to guide experimental design and gain physical insights into the behavior of complex systems. Figure 2 shows different aspects of controlled assembly and how different computational approaches could be used to model them. In descending order of scale, continuum equations based on differential equations are used to describe the quantity of fluid deposited onto surfaces,

mesoscale fluid dynamics models are utilized to investigate the flow patterns of this liquid once it is on the surface, and coarse-grained and atomistic molecular dynamics are used to model the assembly behavior of the material inside of these liquid flow environments. The work in this dissertation will primarily focus on the atomistic and coarse-grained molecular dynamics efforts for this process and molecular biology, leaving the larger scale approaches for future efforts.

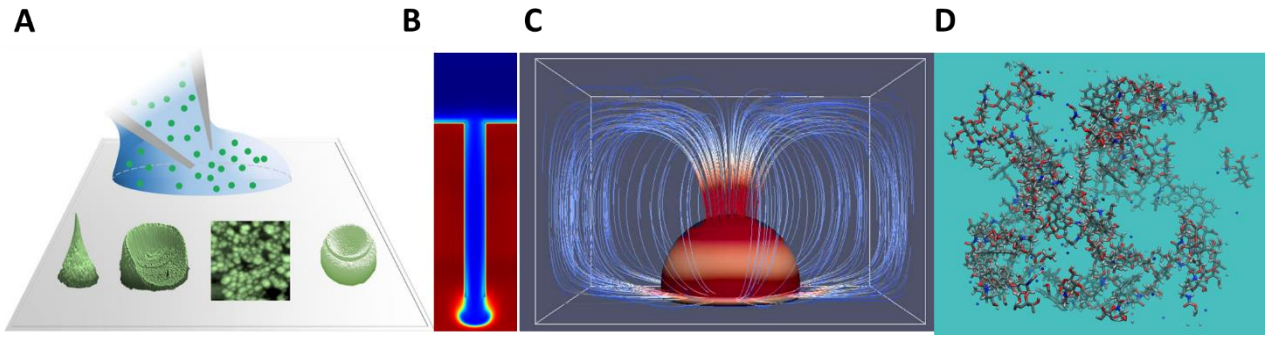


Figure 2: Example application of multi-scale modeling to an application. In this case a microfluidic AFM used for 3D-nanoprinting and controlled assembly processes. A) Cartoon of experimental fluid and solute delivery.[2] B) Mesoscale modeling of the fluid deposition. C) Mesoscale modeling of fluid evaporation on surface. D) Atomistic modeling of the assembly of material in that fluid field

1.1 Molecular dynamics

The bulk of the work in this paper was conducted using molecular dynamics (MD). MD is a computational simulation method for analyzing the movement of atoms and molecules. This technique broadly works by numerically solving equations of motion for each interacting atom, where the interactions between atoms are governed by a force field containing terms for the bonded and non-bonded electrostatic and van der waals potentials. This process can be derived starting from Hamilton's first equation[28] shown in equation 1.1.

$$\frac{d\mathbf{p}_j}{dt} = -\frac{\partial H}{\partial \mathbf{q}_j}, j = 1, \dots, N \tag{1.1}$$

Where N is the total number of particles, **p** is particle j's momentum vector, **q** is the coordinate vector, t is time and H is the Hamiltonian, which is the equivalent of the total system energy. The momentum and coordinate vectors used in equation 1.1 are comprised of three components per particle corresponding to Euclidian coordinates x, y, and z.

$$\mathbf{p}_j = < \mathrm{p}_{j,x}, \mathrm{p}_{j,y}, \mathrm{p}_{j,z} >, \mathbf{q}_j = < \mathrm{q}_{j,x}, \mathrm{q}_{j,y}, \mathrm{q}_{j,z} > \tag{1.2}$$

The Hamiltonian can be written in terms of these vectors and expanded into kinetic and potential energy components resulting in equation 1.3, where m is the particles mass and V is the particles potential energy.

$$H(\mathbf{p}_1, \dots, \mathbf{p}_N, \mathbf{q}_1, \dots, \mathbf{q}_N) = \frac{1}{2}\sum_{j=1}^{N} \frac{\left|\left|\mathbf{p}\right|\right|_j^2}{m_j} + V(\mathbf{q}_1, \dots, \mathbf{q}_N), \tag{1.3}$$

Substituting equation 1.3 into 1.1 and using the assumption that momentum can be written in terms of its position vector as **p =** (m(d**q**/dt)) results in Newton's second law, 1.4.

$$m\frac{d^2\mathbf{q}_\mathrm{j}}{dt^2} = -\frac{\partial V}{\partial \mathbf{q}_\mathbf{j}} \tag{1.4}$$

Newton's second law can be broken down in numerous ways to produce discrete equations of motion using Taylor series expansions. Several prominent discretization approaches include the Position Verlet[29], Velocity Verlet[30], and Leapfrog Verlet[24]. The work conducted here primarily utilizes the Velocity Verlet approach, which can be seen in equation 1.5 and 1.6.

$$\mathbf{q}_j^{i+1} = \mathbf{q}_j^i + \mathbf{v}_j^i \Delta t + \frac{\mathbf{F}_j^i}{2m_j}\Delta t^2, j = 1, \dots, N \tag{1.5}$$

$$\mathbf{v}_j^{i+j} = \mathbf{v}_j^i + \frac{1}{2m_j}\left(\mathbf{F}_j^i + \mathbf{F}_j^{i+1}\right)\Delta t, j = 1, \ldots, N \quad (1.6)$$

Here $\mathbf{v}_j$ is particle j's velocity vector, $\mathbf{F}_j$ is the force vector acting on particle j which can be calculated from the right hand side of Newton's second law in 1.4, and Δt is the elapsed time between integration steps. This elapsed time between integration steps is what is known as a simulation timestep. For stability purposes this timestep is typically chosen to be shorter than the timeframe of the fastest oscillation in the system which is often carbon-carbon bond bending (2fs) in atomistic systems.[31] The superscripts on all terms in equations 1.5 and 1.6 correspond to integration indices, keeping track of what integration step each term is calculated from. The positional accuracy scales with $O(\Delta t^4)$ while local velocity accuracy scales with $O(\Delta t^2)$. Higher order expansions can result in higher accuracy but require much more expensive force evaluations. These equations require an initial condition for momentum and velocity before being used to generate molecular dynamics trajectories. These structural coordinates can be generated from experimental data from X-ray diffraction (XRD)[32] or cryogenic electron microscopy (CRYO-EM) [33] Another software based option is to take a known atomic structure, and potentially refine it using iterative ab initio methods[34]. These structures are frequently solvated before simulation through the addition of water or other solvent molecules in pre-defined smaller boxes. These smaller-boxes of solvent are added to the larger box containing the molecules of interest, and any solvent molecules that would sterically clash with the existing solute is removed. The bulk of the simulation cost of these molecular dynamics techniques frequently come from the calculations involving these water molecules. After these initial coordinates are generated, they are typically refined further

through a gradient based energy minimization process. Initial velocities can be generated by sampling a Maxwell-Boltzmann distribution at the desired simulation temperature. The final component, the forces acting on each atom or molecule are typically determined as the sum of bonded, Van der Waals, and Electrostatic potentials as shown in equation 1.7. Where vector $\mathbf{r}_{ij}$ is given by equation 1.8.[24]

$$V(\mathbf{q}_1, \dots, \mathbf{q}_N) = V_{bonded} + 4\sum_{i<j}^{N} \varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{\left|\left|\mathbf{r}_{ij}\right|\right|}\right)^{12} - \left(\frac{\sigma_{ij}}{\left|\left|\mathbf{r}_{ij}\right|\right|}\right)^{6}\right) + k_0\sum_{i<j}^{N}\frac{q_i q_j}{\left|\left|\mathbf{r}_{ij}\right|\right|}, \qquad (1.7)$$

$$\mathbf{r}_{\mathrm{ij}} = \mathbf{q}_{\mathrm{j}} - \mathbf{q}_{\mathrm{i}}, \qquad (1.8)$$

The largest computational cost of molecular dynamics simulations come from the evaluation of equation 1.7, requiring $O(N^2)$ operations for the pair-wise, non-bonded particle interactions. Domain decomposition is used to help reduce the non-bonded, short range interactions to $O(N_d N_p^2)$ by dividing it into smaller more manageable sub-domains and excluding calculations outside of those domains. Here N_d is the number of sub-domains, and N^p is the number of particles within a domain plus the number of particles in the corresponding buffer shell. The allocation of these domains across simulation processors can reduce the order of this operation scaling to the number of particles squared. Several cut-offs can be used to reduce the number of computational operations required for a simulation. [35,36]

The equations above describe the dynamics for a micro canonical (NVE) system, however, molecular dynamics simulations are typically performed using isothermal (NVT) or isothermal/isobaric (NPT) states. This is due to temperature and pressure being relevant state variables for real world applications that these systems are typically attempting to recreate. The

temperature and pressure can be approximated through the addition of temperature and or pressure controlling terms on equation 1.4, adding discrete expansions to equations 1.5 and 1.6. Temperature regulation is handled by a computational thermostat, and pressure regulation by a computational barostat. Several of these thermostats and barostats have been developed, including the Nose-Hoover thermostat and barostat[37], the Parinello-Rahman barostat[38], the velocity-rescale thermostat[39], and the Berendsen thermostat and barostat[40].

In addition to the thermostats and barostats controlling for temperature and pressure, other constraints can be applied to molecular dynamics systems. These constraints include distance constraints for the movement of atomic oscillations and can be used to help stabilize the simulation timestep. Some of the constraint algorithms that have been developed include LINCS[41], SHAKE[42], RATTLE[43], and SETTLE[44]. Additionally positional restraints can be applied for the modeling of surfaces and other solids. These can be handled through methods that excludes the frozen group from the thermostat and energy evaluations or using softer positional restraints which operate by restraining molecules to a region but allowing some constrained motion to occur within those restraints.

The application of all these systems to all atom systems enables simulations reaching box sizes of low 10s of nanometers in any direction, for times on the order of 100s of nanoseconds. To achieve higher length and time scales at the cost of molecular resolution, coarse-grained (CG) molecular dynamics can be used. In these coarse-grained systems the atomic descriptions are smoothed to become beads that represent multiple atoms.[21,22] This results in a loss of chemical identity but enables simulations to reach 100s of nanometers in length and microseconds worth of time. A “semi-quantitative” CG approach using a 4-1

mapping exists in the form of the Martini force field.[45,46] This force field is “semi-quantitative” because for certain molecules, eg phospholipids, the parameterization allows it to closely match the physical behavior observed in experimental systems. The coarse-grained work in this dissertation is based on this Martini forcefield approach.

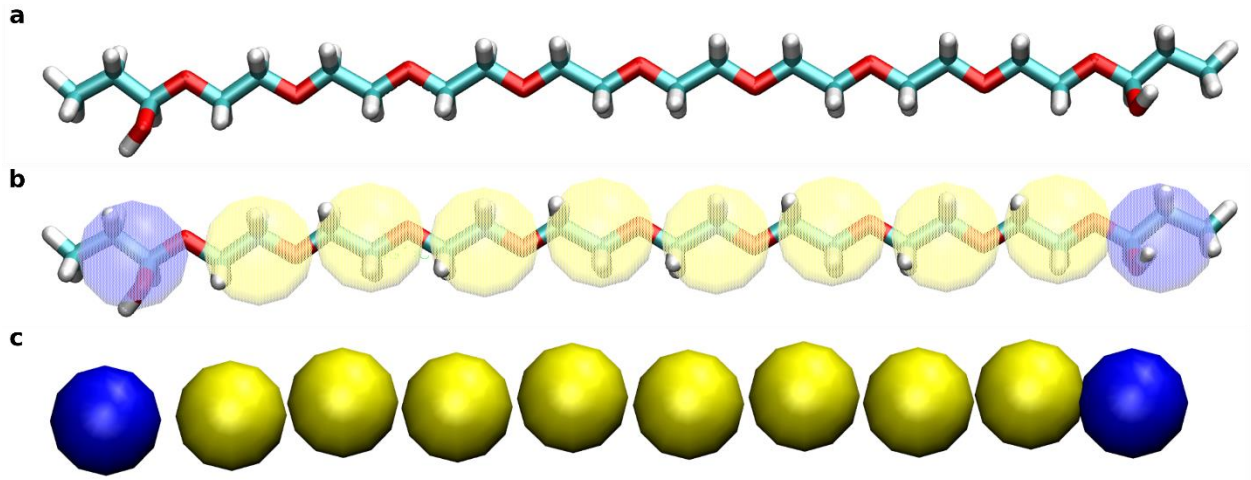


Figure 3: Atomistic to coarse-grained representations of polyethylene glycol diacrylate (PEGDA) 8mer. A) Atomistically represented PEGDA. B) Bead mapping for coarse-grained representation of PEGDA. C) Coarse-grained representation of PEGDA

Molecular dynamics is a powerful computational technique for multi-scale modeling applications due to providing highly precise detailed information. The largest drawback to the technique is the heavy computational cost associated with calculating the pair-wise interactions at every timestep. This limitation can be partially overcome through coarse-graining at the cost of chemical resolution to reach longer time and length scales of interest. As part of a broader multi-scale modeling effort that can handle aspects of the larger time and length scales these particle based simulations provide extremely valuable insight into physical behaviors and can be used to feed parameters upwards for applications in macro scale processes.

1.2 Controlled assembly

The early chapters in this dissertation focus on the notion of controlled assembly. This concept is a derivative of molecular assembly, the process by which molecules self-assemble in solution. This self-assembly behavior is typically spontaneous and occurs in solution, making it the driving force for a number of applications, some common examples are photovoltaics, medicine/cosmetics, and drug delivery.[47-50] The idea for controlled assembly stems from the notion of controlling the self-assembly behavior of the material in these solutions.[1-5] This is a very complex process depending on the interplay of surface, solvent, solute, and evaporation dynamics and interactions, providing an opportunity for multi-scale modeling to improve our physical understanding of various aspects of the process.

Experimentally controlled assembly is achieved using a microfluidic atomic force microscope (AFM). This microfluidic AFM gives spatio-temporal control over the deposition of material onto a surface of interest. The fluid environment deposited onto the surface then evaporates, acting as a “nano-reactor” environment where the material can assemble onto the surface. The control aspect in the assembly comes into play through the choice of surface, solvent, and solute. Surface type will control the type of evaporation that occurs, either a constant contact area or a constant contact angle depending on the contact angle determined by the surface’s solvent-philicity.[51-53] A schematic of this process is shown in Figure 3. The evaporation rate can be controlled by changing the solvent, using highly volatile solvents such as ethanol, highly stable solvents such as glycerol, or mixed solvent systems. The solute of interest plays a role in the way it interacts with the solvent and surface, and can be either small molecules such as sugars, or larger spherical polymers. The microfluidic AFM then provides

control over the position of the desired assembled features, and the size of the desired features by changing the dwell time, or amount of time the AFM tip spends in contact with the surface.

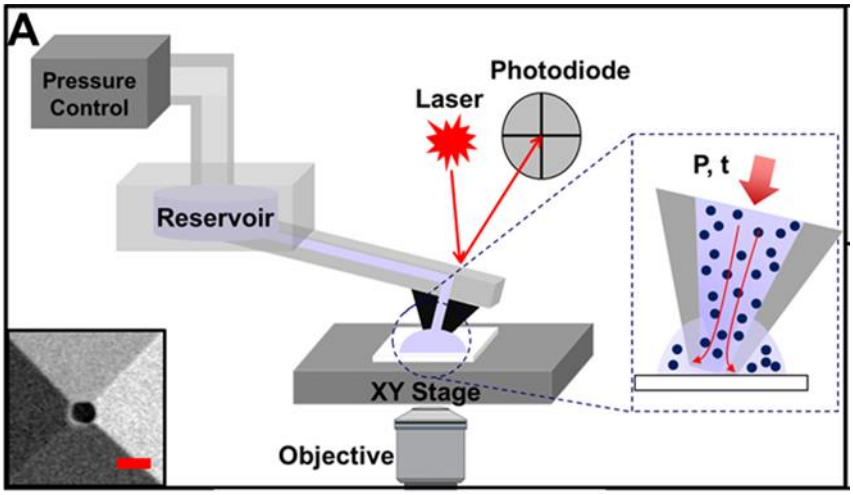


Figure 4:[2] Schematic diagram of an AFM with microfluidic delivery probe. An enlarged view of the probe-surface contact with delivery parameters is shown. The lower left insert is an SEM image of nanopipette apex with 300m aperture. Scale bar is 500nm

While making changes to these experimental conditions is fairly straightforward, each change has far reaching consequences on the physics of the assembly process. Changing surface type for instance changes not only the specific evaporation dynamics, but also contributes to differences in the assembly of the materials inside, as different surface termina will interact with the assembled solute in different ways. Changing solvent, changes the evaporation rate, but also causes new interactions with the solute, potentially forming buffer layers, or mixing with the assembled material in non-trivial ways. These effects make the final structure highly dependent on the interaction matrix between the surface, solvent, solute, and evaporation type and rate. The complexity involved in these interactions makes it a prime target for computational study, as it is much easier to systematically investigate the effects of these changes on the underlying physics using computer simulation. Additionally, because the system is taking place in extremely small volumes of liquid (~attoLiter), measuring and

determining the behavior of the material in solute during evaporation is nearly impossible experimentally. The small size also means that traditional fluid models based on macroscopic fluid approximations don't readily apply, because on the nano-scale the dynamics are much closer in scale to the movement of individual molecules.

Therefore, the purpose of the early chapters of this dissertation is to use multi-scale molecular modeling to improve our understanding of the effects of these interactions on the controlled assembly process. The work focuses on the development of new computational tools to replicate the evaporation driven assembly process in atomistic and coarse-grained molecular simulation.

1.3 SARS-CoV-2 and Glycoproteins

Towards the end of the other research conducted in this dissertation SARS-CoV-2 or COVID-19 appeared in Wuhan China, and quickly turned into a global pandemic, disrupting society on a peacetime unprecedented scale. As of November 2021, COVID-19 has infected more than 251 million people and caused over 5 million deaths.[54] The final chapters of this work are the result of applying similar molecular modeling techniques to better understanding SARS-CoV-2 and demonstrating yet another application for multi-scale modeling.

Coronaviruses are enveloped, positive-sense RNA viruses belonging to the family *Coronaviridae* and named as such because they resemble the structure of the sun. The genomic sequence for SARS-CoV-2 includes four structural proteins. The spike protein (S), the envelope, the membrane, and the nucleocapsid.[55-61] The membrane proteins maintain the viral lipid membrane structure, the envelope facilitates viral assembly and release, and the nucleocapsid

contains the viral genome responsible for infection. Finally, the S protein is responsible for the attachment to the human angiotensin-converting enzyme 2 (hACE2) receptor. The spike protein is a heavily N-linked glycosylated homotrimer with a molecular weight of around 150 kDa. The spike protein has two sub domains, the S1 receptor binding domain (RBD) that does the binding to hACE2, and the S2 portion that forms the S proteins stalk. hACE2 is a transmembrane protein expressed in the lungs, heart, kidneys, and intestine. It has been shown to be a homodimer with each monomer consisting of several domains, but with the portion responsible for binding SARS-CoV-2 being the N-terminal peptidase domain. The structure of both SARS-CoV-2 and hACE2 can be seen in figure x.

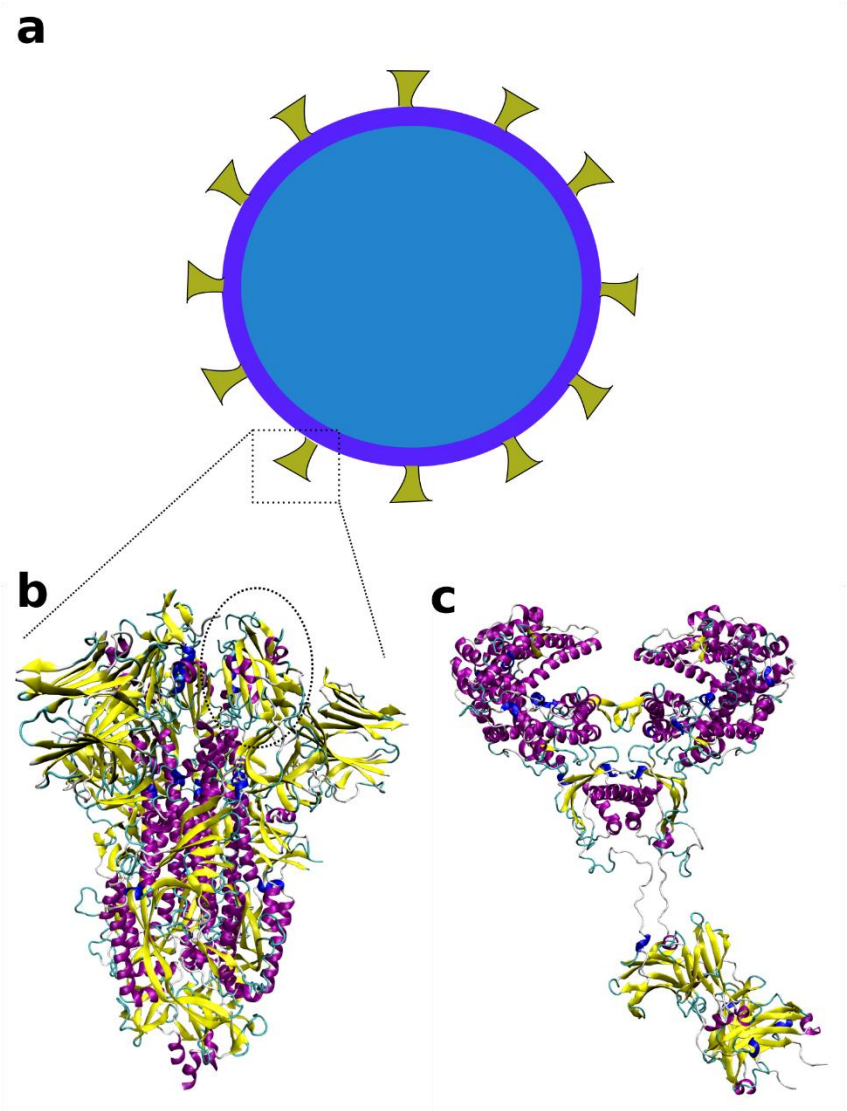


Figure 5: structure of SARS-CoV-2 and hACE2-Fc A) Cartoon representation of full SARS-CoV-2 virus protein membrane with spike protein accentuated. B) Atomistic representation of the full SARS-CoV-2 spike protein, with dashed circle on the receptor binding domain RBD. C) 2 hACE2 proteins with fusion protein Fc linker below

Both of these proteins can be categorized as glycoproteins. Glycoproteins are proteins with branched chains of sugar molecules, called glycans.[62] Viral proteins like S are often glycosylated to help evade the host immune system, modulate access to protease, and enhance cellular attachment through modification of protein structure and/or direct participation at the viral-host interface[63-69]. Additionally, glycans can significantly affect the folding and structure of glycoproteins, in turn modulating the function of the protein as a whole. Generally,

there are two classes of glycosylation: N-glycosylation and O-glycosylation. The work in this dissertation only focuses on N-glycosylation where the glycans are attached to the side chain nitrogen of asparagine (N) protein residues. These asparagines are typically followed by any other amino acid and then a serine or threonine, making them identifiable from genomic amino acid sequences.[70]

The protein synthesis and N-glycosylation occurs in the endoplasmic reticulum (ER) with further modification in the Golgi apparatus. The N-glycans are typically attached to glycoproteins in the ER and then further modified from a high mannose type to their final form by various enzymes. The glycosylation process differs between plants and humans, making glycosylation a central focus for drug design using plant-based expression.[71] The differences in the process result in different glycans depending on the species that generated the glycoprotein and can have effects on the structure and function of the synthesized proteins.

Despite the important role that glycans play in viral and protein function, there have been relatively few attempts to study the role they play in these processes.[72-73] The work in this dissertation attempts to better understand the role of glycosylation on the binding strength between the S and hACE2 proteins, as well as characterize the structure and function of hACE2 and S proteins that are synthesized for therapeutic and diagnostic purposes.

Chapter 2

GenEvaPa: A generic evaporation package for modeling evaporation in molecular dynamics simulations

This chapter is currently in submission for publication to SoftwareX (2021)

Bradley Harris,[1] Gang-yu Liu,[2] Roland Faller[1*], 2021

Abstract

This work presents a novel general tool for modeling the process of evaporation without the need for modifying existing software using Python. The tool was developed based on the MDAnalysis package, which is used to import a Molecular Dynamics trajectory. The tool then removes solvent molecules and outputs a new structure file to be used for further simulation and analysis. This process is designed to be iterated by using the resulting dynamic simulation trajectory as the input file. The evaporation is designed to randomly delete solvent molecules while preserving solvation shells around solutes. The evaporation rate can be controlled by the length of the MD simulations and the number of particles removed between dynamic simulations. Validity of the tool was tested extensively using the Gromacs suite. Advantages of this tool include its genericness, simplicity and user friendliness, as no significant modification of existing software platform or Gromacs specific tools are needed.

2.1 Motivation and Significance

Evaporation impacts the outcomes of material assembly processes, chemical reactions, crystallization and mineralization processes, physical phase transitions of matter, as well as engineering design and production. Because of the complexity and variety of evaporation processes, there are numerous ways to model the evaporation of solvents at all scales ranging from continuum to molecular [1-8]. Among them, modeling solvent evaporation and evaporation-driven solute molecular assembly are of fundamental importance towards understanding of the processes and pertaining outcomes, and guiding applications, such as 3D nanoprinting, ink jet printing, surface modification, bio-preservation etc [1, 2]. Molecular simulation techniques such as Grand canonical Monte Carlo (GCMC) and Molecular Dynamics (MD) have been utilized to model evaporation and correspondingly molecular assembly [1-4, 7]. Current GCMC approaches for assembly during evaporation typically use randomly distributed particles on lattice sites and use various models for the probability of possible particle motion directions at each site. The evaporation is handled by a Monte Carlo particle removal move, which has an acceptance criterion based on chemical potential. GCMC techniques are effective for determining equilibrium properties due to their probabilistic nature and being rooted in statistical mechanics, but Monte Carlo techniques do not mimic the real dynamics of a system as the microstates/configurations are not appearing in time order. Molecular Dynamics (MD) is used to retain this dynamics information. Using MD, an existing approach for modeling evaporation processes relies on a diffusion-based model with a removal zone for solvent molecules that diffuse into the zone [2]. This approach is likely the most realistic way to model evaporation, as it replicates the flux driven nature of evaporation. The challenge is in the

difficulty to implement generically as it is highly system dependent. Additionally, this approach is inherently limited by the diffusion rates of molecules, potentially limiting the length- and timescales that can be achieved. This same diffusion limitation also affects the ability to simulate late-stage evaporations with very small or no solvent remaining. To avoid the diffusion limitation, another MD approach was developed, i.e., randomly removing solvent molecules, e.g., using a bash script [1]. This approach benefits from being able to achieve a wide range of concentrations including the fully dried state and can be easily automated and sped up for any systems. The downside of this approach is that it does not account for hydration shells, which leads to unequal removal probabilities, and it was written specifically for their system and would require code adaptation.

To address these technical challenges, this work developed a tool to model evaporation with two primary goals: preservation of the hydration shells surrounding solute molecules and generic implementation. This led to implementation as a python wrapper to enable its use in both atomistic and coarse-grained simulations for a variety of existing MD codes, therefore making it system independent. Preserving hydration or solvation shells is scientifically significant as this more accurately models the physical process and corresponding hydrogen bonding. It also exhibits a technical advantage enabling further simulation to rapidly smooth out voids created by the removal of solvent molecules. This is done by creating those voids in places less involved in a hydrogen bonding network where solvent can diffuse more freely. The script has been tested and validated using Gromacs Versions 5.1.2016 and 2019.1 [9], thus should work for all simulation software packages that are supported by MDAnalysis [10, 11].

2.2 Software Description

GenEvaPa is an open-source python script that acts as a wrapper for MD simulation software and enables the deletion of solvent molecules outside of a certain tolerance distance from molecules of interest in order to preserve hydration shells. The script is built using the MDAnalysis python module for handling atom groups, making it compatible with any MD software suite that is compatible with MDAnalysis [11], e.g. Gromacs [9], LAMMPS [12], NAMD [13], CHARMM, etc.

2.2.1 Architecture

GenEvaPa is contained in a single python3 script that requires MDAnalysis [10, 11] as the only external package. A flow chart of the internal workings of the software is shown Figure 1. This schematic shows how the evaporation script fits into an MD workflow, with the operations of the script emphasized as a zoom in on the green oval. The script is written in an object-oriented manner with the main evaporation process implemented as a class that handles the preservation of hydration shells by calculating and comparing distances between specified solvent molecules and molecules of interest. A representative visualization of the hydration shell preservation for an atomistic system is shown in Figure 2. It sorts the solvent molecules into deleteable and non-deleteable categories based on distance from a solute and uses the python *random.sample* function to delete the desired number of solvent molecules. The distance calculation is sped up by dynamically binning the simulation box into subsections that are equal to the tolerance of the hydration shell. Periodic boundary conditions are accounted for in this binning process. The sample workflow shown in Figure 1 demonstrates how the script can be incorporated into a wider workflow. The process begins with initial structure and box

generation followed by a series of equilibration steps and followed by a molecular dynamics NPT production simulation. To study the effect of solvent removal the final frame of the production run will be taken as the initial input for GenEvaPa. This file, along with the solvent name, structure of interest, tolerance (in Angstrom), and number of solvents to remove each step are provided, and the deletion is performed. GenEvaPa performs its calculations and outputs the final coordinates without the deleted solvents, which can then be used as the input structure file for the next MD run or other purposes. This process can then be looped and automated through a bash or python script to do complete controlled solvent deletion. An example use case is shown below, and a more detailed description of the classes and methods as well as an example automated workflow and analysis scripts are available at https://github.com/bradsharris/GenEvaPa.

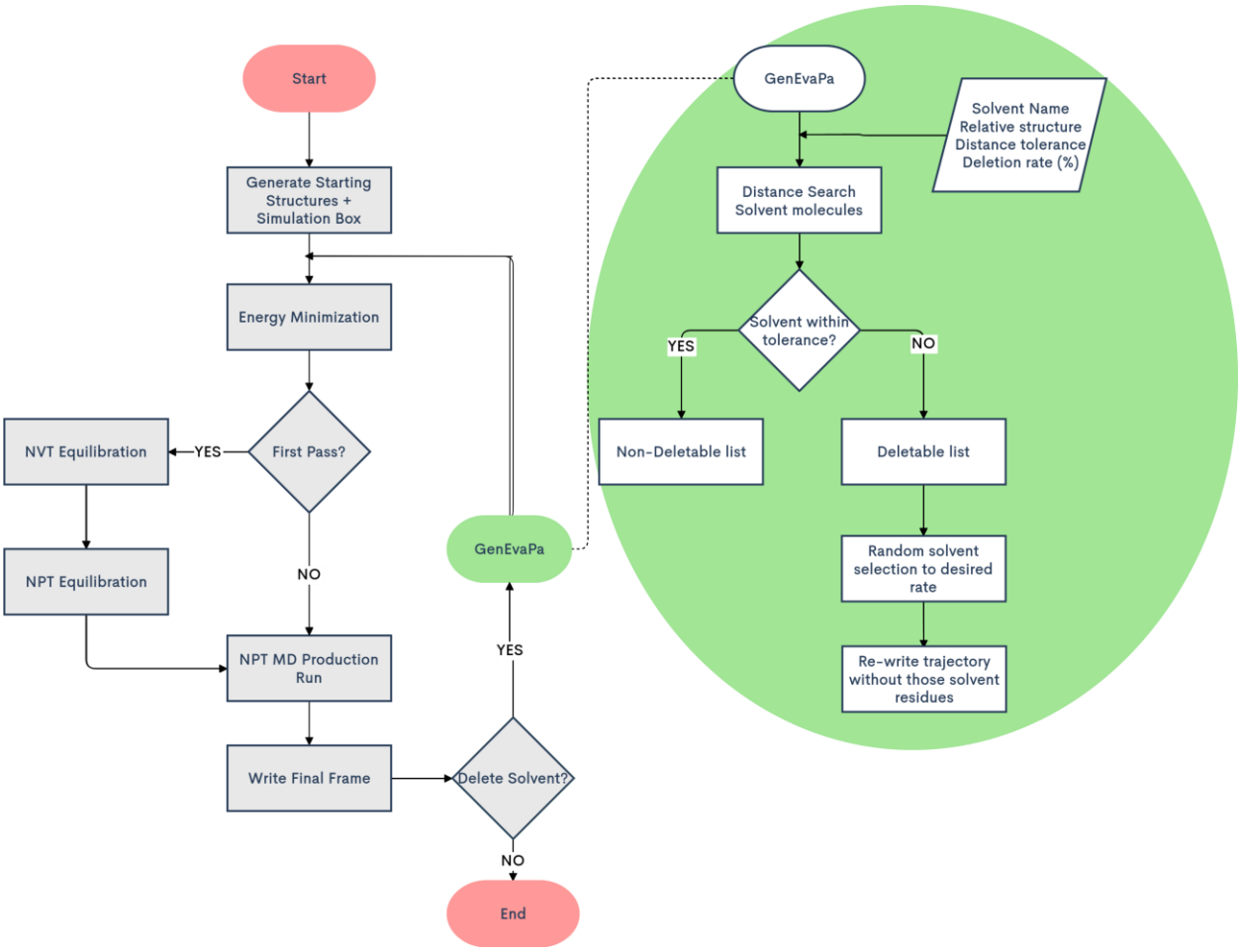


Figure 1. Schematic of GenEvaPa showing evaporation script details and sample workflow.

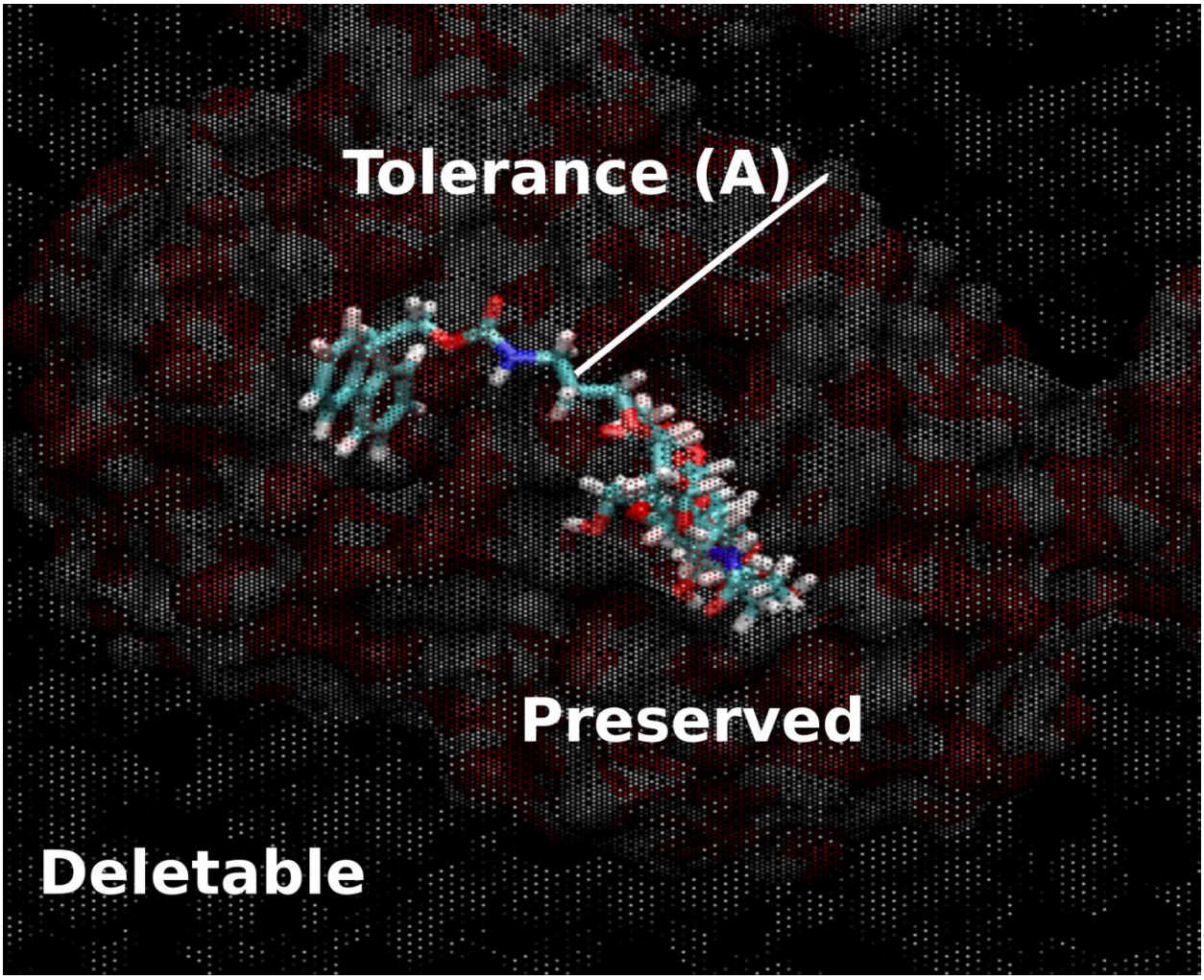


Figure 2. Visualization of hydration shell preservation during distance search. Tolerance of 10 Angstrom from the evaporation of an aqueous solution of heparosan tetrasaccharide (MW = 1.099 kD)[14].

2.3 Illustrative Examples

To assess the performance of GenEvaPa several evaporations were performed on an atomistic system using Gromacs 2019.1. These simulations were designed to test the performance of the system at various evaporation rates as well as temperatures. Further uses of the script can be seen in our published work on controlled assembly of small molecules such as heparosan tetra-saccharide.[14] The starting configurations were identical to the pre-deletion equilibrated structures from that paper. Briefly, 27 heparosan tetra-saccharide molecules were added to a

simulation box of size 17.19 nm x 17.19 nm x 17.19 nm that contained 162623 water molecules and 54 Na^+ counter ions for a total of 491514 atoms. The system was energy minimized and equilibrated in NVT for 100 ps using a velocity rescale thermostat, and then in NPT for 100 ps using a velocity rescale thermostat [15] and isotropic Parrinello-Rahman barostat.[16] The system was then simulated using NPT at 300 K and 350 K for 20 ns using the same thermostat and barostat to establish a baseline prior to solvent removal. Solvent deletion was performed using the GenEvaPa script following the workflow shown in Figure 1, and automated to perform deletions of various percentages of initial water in every loop. Deletions performed at 350 K had rates of 1%, 5%, and 10% initial water removed per loop in order to evaluate the stability of the evaporation as a function of deletion rate. The 1% deletion was also performed at 300 K and compared to 350 K to verify consistent performance at different temperatures. Each MD run following deletion during the loop was performed in Gromacs 2019.1 [9], each MD ran for 1 ns NPT simulation time, with the tolerance set to 10 Å for the first 90% of water. The tolerance was then set to 0 for the final 20 ns due to not having enough water molecules that met the original tolerance criteria to be deleted. Chosen simulation performance characteristics of energy and density are shown in Figure 3. The results show that, for these systems tested, the simulations reach equilibrated and consistent values rapidly for both temperatures and at all three deletion rates. The difference in energy and density profiles for temperature data in Figure 3A and 3B are consistent with the change in temperature, e.g., decreasing the temperature from 350 to 300 K led to a less dense system, and increases in absolute energy. The data in Figure 3C and 3D suggest that the system was stable at all three deletion rates tested, as the characteristic system parameters exhibited little rate dependence.

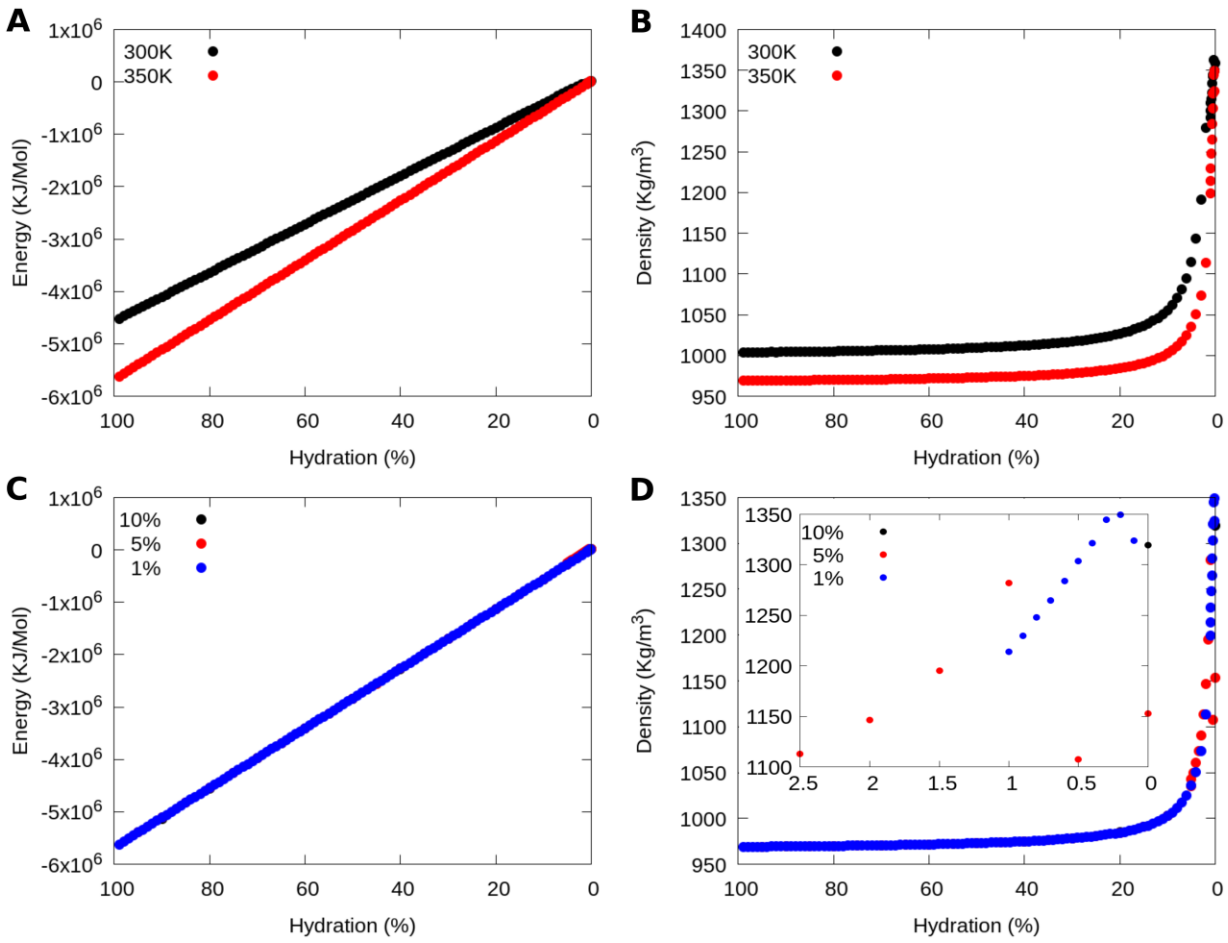


Figure 3. Simulation performance characteristics: (A) At 1% deletion rate, the total energy is plotted as a function of dehydration at 300 and 350 K. (B) at 1% deletion rate, the density is displayed as a function of dehydration. (C) Total Energy at 350 K as a function of dehydration for different deletion rates. D) Density against deletion rate as a function of dehydration, inset showing the dehydration of the final 2.5% water. Hydration levels shown as a percentage (%) relative to start.

Several properties of the heparosan tetrasaccharide were evaluated to ensure convergence, as illustrated in Figure 4. These results show that for each deletion rate the properties converge, with minor differences in the amount of interaction time available due to the solute molecules being simulated for varying total simulation times across rates, e.g., 110 ns for 1% and 22 ns for 5%. Overall, the convergence of these properties suggests that this method enables simulation of a wide range of evaporation rate, thus could accommodate various processes involving evaporation.

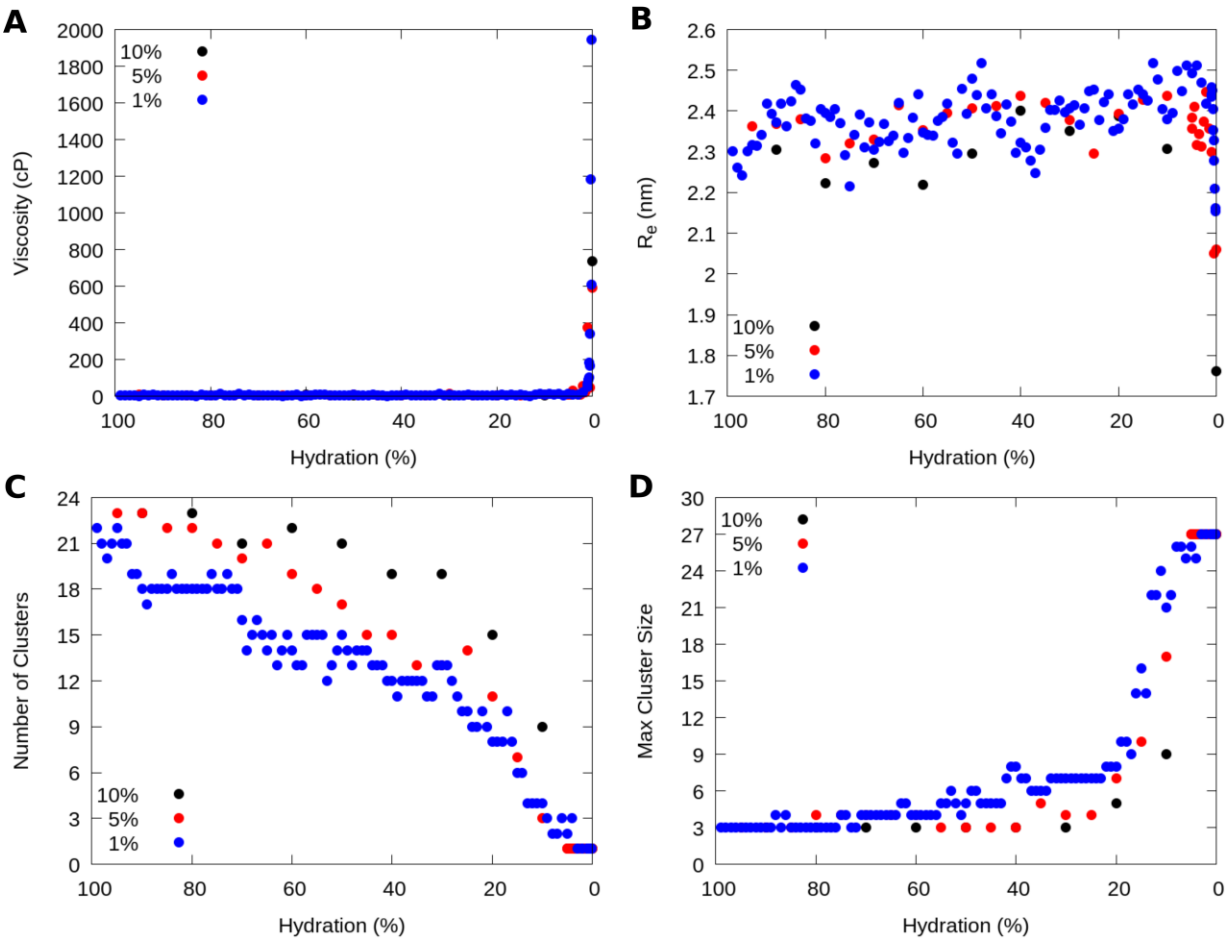


Figure 4. Properties of the assembled heparosan tetrasaccharide molecules versus evaporation driven dehydration, simulated at all three water-deletion rates. (A) Viscosity vs dehydration. (B) End-to-End distance vs dehydration. (C) Number of clusters vs dehydration. D) Maximum cluster size vs dehydration. Hydration shown as a percentage (%) relative to start.

2.4 Conclusions

A novel methodology for modeling evaporation in molecular dynamics simulation was developed and implemented as a generic python wrapper. The script has been validated for Gromacs for various temperatures and deletion rates. The key scientific advances include the preservation of hydration shells and the applicability to both atomistic and coarse-grained systems. Technical advantages include its genericness and support for a variety of existing

molecular dynamics software packages. This script has been tested for solute systems from small molecules to macromolecules, and in principle, widely supports any molecular dynamics system requiring evaporation or concentration dependence. Ongoing work is being conducted to validate the script for coarse-grained and mixed solvent systems. We anticipate its adoption and applications in many important processes involving evaporation such as chemical reactions, crystallization and mineralization processes, physical phase transitions of matter, as well as engineering design and production.

Acknowledgements

This work is supported by the National Science Foundation (CHE-1808829 and BSH was partially supported by LLNL's LDRD program, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory

2.5 References

1. Alessandri, R., et al., *Bulk Heterojunction Morphologies with Atomistic Resolution from Coarse-Grain Solvent Evaporation Simulations.* Journal of the American Chemical Society, 2017. **139**(10): p. 3697-3705.
2. Cheng, S. and G.S. Grest, *Molecular dynamics simulations of evaporation-induced nanoparticle assembly.* The Journal of Chemical Physics, 2013. **138**(6): p. 064701.
3. Crivoi, A. and F. Duan, *Three-dimensional Monte Carlo model of the coffee-ring effect in evaporating colloidal droplets.* Scientific Reports, 2014. **4**(1): p. 4310.
4. Landry, E.S., et al., *Droplet evaporation: A molecular dynamics investigation.* Journal of Applied Physics, 2007. **102**(12): p. 124301.
5. Ledesma-Aguilar, R., D. Vella, and J.M. Yeomans, *Lattice-Boltzmann simulations of droplet evaporation.* Soft Matter, 2014. **10**(41): p. 8267-8275.
6. Schlottke, J. and B. Weigand, *Direct numerical simulation of evaporating droplets.* Journal of Computational Physics, 2008. **227**(10): p. 5215-5237.
7. Yamashita, K., et al., *Grand canonical Monte Carlo and molecular dynamics simulations of capillary condensation and evaporation of water in hydrophilic mesopores.* Molecular Physics, 2017. **115**(3): p. 328-342.
8. Zhakhovskii, V.V. and S.I. Anisimov, *Molecular-dynamics simulation of evaporation of a liquid.* Journal of Experimental and Theoretical Physics, 1997. **84**(4): p. 734-745.

9. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1-2**: p. 19-25.
10. Gowers, R.J., et al., *MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations.* Proc. Of the 15th Python in Science Conf, 2016.
11. Michaud-Agrawal, N., et al., *MDAnalysis: A toolkit for the analysis of molecular dynamics simulations.* Journal of Computational Chemistry, 2011. **32**(10): p. 2319-2327.
12. Plimpton, S., *Fast Parallel Algorithms for Short-Range Molecular Dynamics.* Journal of Computational Physics, 1995. **117**(1): p. 1-19.
13. Phillips, J.C., et al., *Scalable molecular dynamics on CPU and GPU architectures with NAMD.* The Journal of Chemical Physics, 2020. **153**(4): p. 044130.
14. Zhang, J., et al., *New Means to Control Molecular Assembly.* The Journal of Physical Chemistry C, 2020. **124**(11): p. 6405-6412.
15. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* The Journal of Chemical Physics, 2007. **126**(1): p. 014101.
16. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method.* Journal of Applied Physics, 1981. **52**(12): p. 7182-7190.

Chapter 3

A New Means to Control Molecular Assembly

My contribution to the experiments and analysis of this chapter was limited to the simulation portion presented within

This chapter originally appeared in literature:

Jiali Zhang,[1] Hai Yu,[1] Bradley Harris,[2] Yunbo Zheng,[1] Umit Celik,[1] Lan Na,[1] Roland Faller,[2] Xi Chen,[1] Dominik R Haudenschild[3] Gang-yu Liu[1*] "A New Means to Control Molecular Assembly" JPC-C, 124 (11), 6405-6412, 2020

Reprinted with permission from J. Phys. Chem. C. 2020, 124, 11, 6405-6412,

https://doi.org/10.1021/acs.jpcc.9b11377

Copyright 2021 American Chemical Society

Abstract

While self-assembly of molecules is relatively well-known and frequently utilized in chemical synthesis and material science, controlled assembly of molecules represents a new concept and approach. The present work demonstrates the concept of controlled molecular assembly using a non-spherical biomolecule, heparosan tetrasaccharide (MW = 1.099 kD). The key to controlled assembly is the fact that ultra-small solution droplets exhibit different evaporation dynamics from those of larger ones. Using an independently controlled microfluidic probe in an atomic force microscope, sub-femtoliter aqueous droplets containing designed molecules produce

well-defined features with dimensions as small as tens of nanometers. The initial shape of the droplet and the concentration of solute within the droplet dictate the final assembly of molecules due to the ultrafast evaporation rate and dynamic spatial confinement of the droplets. The level of control demonstrated in this work brings us closer to programmable synthesis for chemistry and materials science which can be used to develop vehicles for drug delivery 3-D nanoprinting in additive manufacturing.

3.1 Introduction

Assembly of molecules into meso-scale structures by design still poses great challenges, despite pronounced advances in bottom-up and top-down approaches.[1, 2] Self-assembly (SA) provides a powerful means to address this challenge. Ordered structures driven by thermodynamics have been reported, including self-assembled monolayers (SAMs),[3-5] phase-separated polymers,[6-9] and lipid bilayers.[10, 11] In attempts to produce molecular assemblies by design, the concept of controlled assembly was demonstrated in our prior work via the formation of various assemblies of spherical macromolecules, such as star polymers (MW = 383 kD, Rh = 47 nm).[12] The key to controlled molecular assembly relies on an ultrafast evaporation rate and spatial confinement by small droplets (sub-fL), as illustrated in Figure 1. The rapid evaporation locks the solute molecules in place, leading to a high degree of control over the feature geometry and intrafeature molecular packing.[12]

The present work pushes the concept of controlled assembly to smaller and non-spherical molecules, e.g., heparosan tetrasaccharide (referred to as heparosan, MW = 1.099 kD), as shown in Figure 1. Heparosan is a polysaccharide consisting of disaccharide repeat units of α1–

4-linked *N*-acetyl-D-glucosamine (GlcNAc) and β1–4-linked D-glucuronic acid (GlcA).[13] It is produced as a capsule polysaccharide (CPS) by both *P. multocida* Type D and *E. coli K5*.[14] It is also the polysaccharide backbone of heparin and heparan sulfate found in mammals.[15] Assemblies of heparosan, e.g. nanogels,[16] micelles,[17-20] and nanoparticles,[21] have been used for disease diagnosis and treatment. The formation of these molecular assemblies relies primarily on self-assembly.[16-21] The size of such self-assembled units are typically in the range of 50-220 nm.[16-21] It was reported that particle size, size distribution, and morphology play important roles in their efficacy, as these factors impact cell-uptake and blood circulation time, as well as drug release.[22] Using various pre-designed ultrasmall droplets at defined solute concentration and initial droplet volume and geometry, this work demonstrates the feasibility for controlling molecular assembly of heparosan oligosaccharides. Specifically, the ability to control the assembly of heparosan nanostructures by design could significantly advance biomaterial development for drug delivery and theragnostic applications. In general, this level of control demonstrated in this work brings us closer to programmable synthesis for chemistry as well as developing vehicles for drug delivery and 3-D nanoprinting in additive manufacturing.

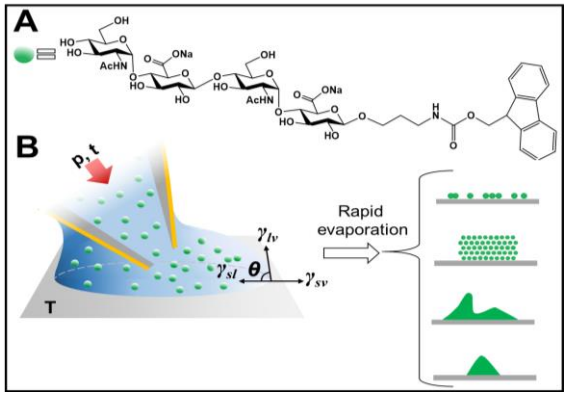


Figure 1. (A) Chemical structure of Fmoc-tagged heparosan tetrasaccharide. (B) A schematic diagram illustrating our concept of controlled assembly. An ultrasmall liquid droplet is delivered via a microfluidic probe to a surface with designed solvent-philicity. The droplet size

and shape are dictated by the delivery parameters (e.g., p, t, T and probe location), and droplet-surface interactions (θ and γ representing contact angle, and surface tensions at phase boundaries). The right column illustrates solute assemblies upon rapid solvent evaporation.

3.2 Experimental Methods

Materials. Glass slides with lateral dimensions of 75 mm × 25 mm and thicknesses of 1 mm were purchased from Fisher Scientific (Pittsburgh, PA). Glycerol (≥ 99%), sulfuric acid (95.0 – 98.0%), hydrogen peroxide (30% aqueous solution), and ammonium hydroxide (NH_4OH, 30% aqueous solution) and toluene (99.8%) were purchased from Sigma-Aldrich (St. Louis, MO). *N*-(6-aminohexyl)-aminopropyltrimethoxysilane (AAPTMS) and octadecyltrichlorosilane (OTS) were purchased from Gelest (Morrisville, PA). Ethanol (99.5%) was purchased from KPTEC (King of Prussia, PA). Ultrapure water was attained from a Milli-Q water system with resistance of 18.2 MΩ·cm at 25 ℃ (EMD Millipore, Billerica, MA). Nitrogen gas (99.999%) was purchased from Praxair, Inc. (Danbury, CT, King of Prussia, PA). AC240TS-R3 silicon cantilevers were purchased from Oxford Instruments Asylum Research (Santa Barbara, CA). All other materials were used without further treatment or modification, unless otherwise stated.

Fmoc-tagged heparosan tetrasaccharide (referred to as “heparosan” hereafter) was chemoenzymatically synthesized from chemically synthesized Fmoc-tagged monosaccharide GlcAβProNHFmoc using a sequential one-pot multienzyme (OPME) process with alternating OPME GlcNAc-activation/transfer and OPME GlcA-activation/transfer systems.[23] Briefly, disaccharide GlcNAcα1–4GlcAβProNHFmoc was enzymatically synthesized from GlcAβProNHFmoc via a one-pot four-enzyme GlcNAc-activation and transfer system containing

B. longum N-acetylhexosamine-1-kinase (BLNahK),[24] *P. multocida N*-acetylglucosamine-1-phosphate uridylyltransferase (PmGlmU),[25] *P. multocida* inorganic pyrophosphatase (PmPpA),[26] and Δ80PmHS2.[23] Trisaccharide GlcAβ1–4GlcNAcα1–4GlcAβProNHFmoc was then synthesized from the disaccharide using a one-pot four-enzyme GlcA-activation and transfer system containing *A. thaliana* glucuronokinase (AtGlcAK),[27] *B. longum* UDP-sugar pyrophosphorylase,[28] PmPpA, and Δ80PmHS2. Finally, the heparosan tetrasaccharide was readily synthesized from the trisaccharide using the OPME GlcNAc-activation/transfer system containing BLNahk, PmGlmU, PmPpA, and Δ80PmHS2.[23]

Preparation of Self-Assembled Monolayers. The goal of this step is to prepare a set of surfaces with designated solution-philicity using self-assembled monolayers (SAMs) on glass surfaces. Glass slides were cleaned following previously reported protocols. [29-35] In brief, substrates were cleaned using piranha solution by immersion for 1 h, then rinsed with copious quantities of ultrapure water. Piranha solution is a mixture of sulfuric acid and hydrogen peroxide at a (v/v) ratio of 3:1. It is highly corrosive and should be handled carefully. Then, the cleaned glass substrates were immersed in a prepared basic bath at 70 °C for 1 h. Basic bath is prepared by mixing ammonium hydroxide, hydrogen peroxide, and water at a (v/v) ratio of 5:1:1. Finally, substrates were again rinsed with copious quantities of ultrapure water and dried in nitrogen gas. The hydrophilicity of glass substrates was modified using silane chemistry, following established protocols.[29-35] To modify glass slides with amine-terminated silane AAPTMS SAMs, the clean glass slides were placed into a sealed Teflon container (100 mL) containing 200 µL of AAPTMS, then heated in an oven at 70 – 80 °C for 2 h. The substrate was then rinsed with ethanol and deionized water, sequentially, followed by drying in nitrogen gas. To prepare OTS-

modified glass slides, freshly cleaned glass slides were immersed in a solution of 5 mM OTS in toluene for 30 min, followed by rinsing with toluene and ethanol, and then dried in nitrogen gas.

Contact Angle Measurement. Contact angle data were collected for the modified substrates with a VCA Optima Contact Angle Measurement system (AST Products, Billerica, MA), following standard protocols.[12, 36-38] A 3 µL drop of designated solution was placed on surfaces using an Hamilton 700 series HPLC needle (Hamilton Co., Reno, Nev.). At least three different positions per sample were tested to assure reproducibility and accuracy. For solution of heparosan with concentration of 2×10^{-5} M in a mixed solvent of glycerol:H_2O = 5:95 (v/v), the contact angle on AAPTMS SAM and OTS SAM was 59º and 107º, respectively. Reducing concentration to of 2×10^{-6} M lead to 63 º on AAPTMS SAM and 102 º on OTS SAM. For 2×10^{-5} M aqueous solutions, the contact angle was measured to be 67º on AAPTMS SAM and 104º on OTS SAM, which are very similar to the contact angle of pure water on AAPTMS (58º)[39, 40] and OTS SAMs (110º)[41, 42], respectively.

Microfluidic Delivery of Ultrasmall Liquid Droplets. A state-of-the-art microfluidic system, FluidFM Bot, [12, 43-46] equipped with an independent hollow probe was used to dispense heparosan solution down to attoliter volume onto the designated surfaces. This system combines an AFM probe assembly, a precise x-y motorized stage, an inverted optical microscope and a pressure controller. The cantilever contains a microchannel and a reservoir and is connected to a pressure controller (-800 to 1000 mbar, 1 mbar precision) to dispense liquid inside the AFM head to guide the cantilever vertical movement. For optical view, an inverted microscope was used for imaging and facilitating stage control. A digital microscope

was installed to monitor laser position and probe movement. Microchanneled hollow probes with 300 nm square aperture at apex of the tip were utilized. The hollow probe was prefilled with designed heparosan tetrasaccharide solution, with a square pyramidal tip tilting 11° from the surface normal. The cantilevers were made of Si. Initially, 1 µL of the designated heparosan tetrasaccharide solution was filled into the probe's 2 µL reservoir. The contact force or load was set to 80 nN for all delivery reported in this work.

AFM Characterization of Heparosan Assemblies. After droplet deposition, liquid evaporates rapidly as monitored via bright field optical microscopy. To avoid any residual solvent and facilitate atomic force microscopy (AFM) characterization, the glass slides were placed in a clean ambient environment overnight unless specified. The AFM (MFP-3D, Oxford Instrument, Santa Barbara, CA) has a deflection configuration. Silicon probes (AC 240-TS, Olympus America, Central Valley, PA) with a force constant of 1.7 N/m and resonance frequency of 70 kHz. Tapping mode was utilized for imaging with damping set at 40-60%. The AFM images were analyzed using the Asylum MFP-3D software developed on an Igor Pro 6.12 platform.

3.3 Results and Discussion

Controlling Molecular Packing. Our previous work on the controlled assembly of star polymers demonstrated the packing can be controlled from random distribution to closely packed structure within monolayer-, bilayer- or multilayer-disks.[12] The disk geometry is primarily attributed to constant area evaporation of the droplets on solution-philic surfaces.[12, 47-49] In contrast to star polymers which are almost rigid with spherical shape,[12, 50] individual heparosan molecules dissolved in aqueous solutions are flexible and linear, making them "semi-flexible rods". The

critical question is whether our approach enables control over the packing of heparosan. Figure 2A reveals a heparosan disk formed upon delivery and evaporation of a 2.56 pL droplet of 2×10^{-6} M heparosan tetrasaccharide solution onto a clean aminopropyltrimethoxysilane (AAPTMS) SAM surface. The AAPTMS SAM surface is solution-philic with a contact angle of 63°. Thus, the droplet appeared to spread. Upon evaporation, a disk was formed covering 8.3 □m^2 area, i.e. the interfacial area of the initial droplet on AAPTSM SAM surface. Within the disk, randomly distributed molecular clusters were clearly visible (bright features) in Figure 2B. These clusters are relatively homogeneous in size, 35-95 nm in height. These clusters are well separated, 40-220 nm among the nearest neighbors (NN). Increasing amount of materials delivered, e.g. 4.36 pL droplets, led to larger disks, for example, a 11.2 □m^2 area disk shown in Figure 2C. Within the disk, clusters of similar size to Figure 2B are also seen, as shown in Figure 2D. These clusters exhibit heights of 12-18 nm with NN separation 18-39 nm, i.e. similar in size but more closely packed than in Figure 2A. The formation of disk containing randomly distributed solute clusters is consistent with constant contact area evaporation.[12, 47-49] The presence of clusters instead of individual heparosan molecules is attributed to molecular clustering during evaporation, based on our molecular dynamics (MD) simulation (see below). Increasing heparosan tetrasaccharide concentration and/or amount further decreases cluster separation. At 2×10^{-5} M and 0.833 pL, a heparosan disk was formed: 23.7 nm tall and 5.414 □m wide, as shown in Figure 2E. Within the disk, clusters of heparosan are packed in an over-crowding way, as seen Figure 2F, manifesting to a smoother disk surface than lower coverage cases.

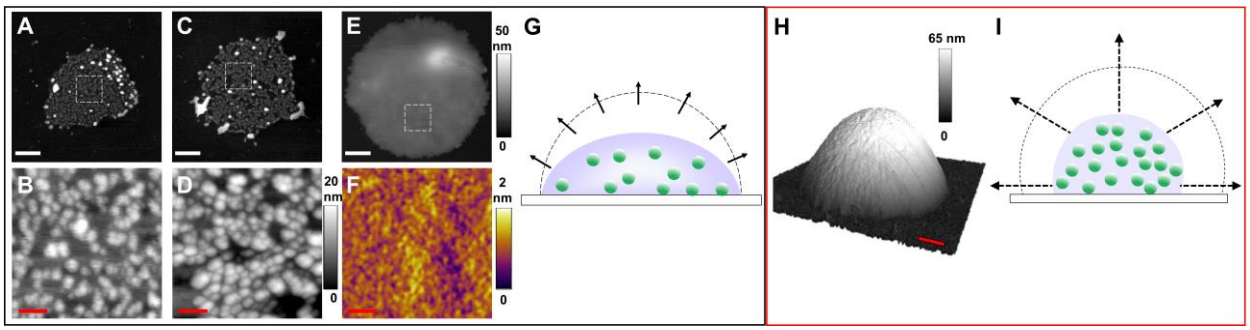


Figure 2. (A) Atomic force microscopy (AFM) topographic image of a disk of heparosan formed after delivery a droplet under 20 mbar for 4.02 s on AAPTMS/glass. Solution concentration: 2×10^{-6} M. (B) A zoomed-in view of the defined area (square frame) in (A). (C) Similar to (A) except with high pressure and longer deliver time, 100 mbar and 5 s, respectively. (D) A zoom-in image as defined in [C]. (E) Similar to (B) except higher concentration of heparosan, 2×10^{-5} M. (F) A zoom-in image of as defined in E (squared area). (G) Schematic diagram illustrates the constant contact area evaporation of a droplet on hydrophilic AAPTMS/glass surface. Broken line and array arrows represent initial droplet boundary, and evaporation direction and rate, respectively. (H) AFM topography of a mound formed on OTS/glass under the same delivery condition as (E). (I) Schematic diagram illustrates the constant contact angle evaporation for droplet on hydrophobic surface. The scale bars for white and red bars are 1 µm and 200 nm, respectively.

The molecular dynamics (MD) simulations (see details in Table S1) reveal a more detailed molecular process during evaporation. Indeed, the outcomes show a progression of heparosan from individual molecules in aqueous solution to aggregate, and then finally to a polymer melt. This is best seen quantitatively in Figure 3, suggesting the solution phase is below 0.04 M, a transition phase between 0.04-0.1 M, and a transition towards a polymer melt beginning at 0.3 M. For the y axis for Figure 3, a cluster is 1 or more sugar molecules within an angstrom of another sugar molecule. For Figure 3A this means that individual sugars appear on the y-axis as a cluster, i.e. the maximum possible number of clusters is 27. Figure 3A and 3B taken together provide insight on the relevant range of concentrations for which aggregation begins. Insets in these Figures are included to differentiate the concentrations and transition below 0.045 M. An understanding of what the sugars are doing in these regions is best seen by the end-to-end

distance in Figure 3C. At concentrations below 0.04 M, the sugars have favorable enough interactions with water that they remain flexible and independent in solution. As the sugars are pushed together as concentration increases, the individual sugar molecules extend to find each other in solution and increase favorable interactions. This mechanism explains the presence of distributed aggregates or "clusters" observed in the experimental disk features. The concentration range for this transition corresponds to approximately 0.04- 0.1 M. The maximum end-to-end distance in this region approximates the end-to-end distance obtained through ChemDraw (2.55 nm). This concentration range corresponding to chain expansion correlates to a decrease in overall number of clusters and an increase in maximum cluster size as shown in Figures 3A and 3B. Figures 4A-4C provide a more direct visualization of the transition from individual molecules to a cluster in the molecular dynamics simulations. Additional simulation analysis for this transition behavior is presented in Figures S1-S2 of Support Information (SI).

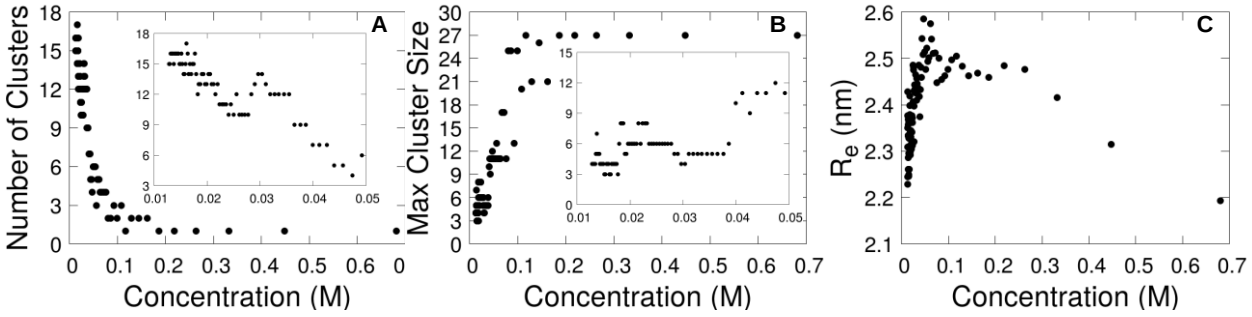


Figure 3. (A) Number of heparosan clusters during each simulation. Inlay is below 0.045 M (B) Maximum number of heparosan in a cluster at any given state during simulations during evaporation. Inlay is below 0.045 M. (C) The average end to end distance of the heparosan during each simulation.

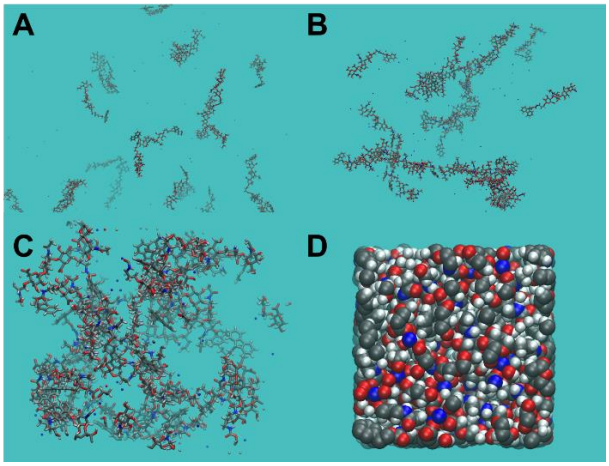


Figure 4. A ball-and-stick model displaying the molecular states from our MD simulation at (A) 0.0138 M, in a 15x15x15 nm^3 box; (B) 0.049 M in a 9.4x9.4x9.4 nm^3 box; and (C) 0.218 M in a 4.0x4.0x4.0 nm^3 box. (D) Volume filling model from our MD simulation at 1.3 M in a 3.2x3.2x3.2 nm^3 box.

In our attempt to further increase molecular packing and form a mound geometry, constant contact angle evaporation was adopted.[12, 47, 51-55] In this case, droplets of heparosan tetrasaccharide solution were delivered onto octadecyltrichlorosilane (OTS) SAM surfaces. The contact angle of the heparosan solution measured 107°, thus the droplets maintained a spherical hat geometry throughout evaporation. Evaporation of 0.113 pL of 2×10^{-5} M heparosan solution led to mound formation, as shown in Figure 2H. The mound is 1.779 µm wide at the base and 111.7 nm tall. The exterior of the mound appeared homogenous, as revealed in Figure 2H. In other words, the clusters collapsed and merged to form a single assembly. Given the spatial confinement and rapid evaporation, heparosan molecules were closely jammed within each assembly, analogous to the interdigitated state of chains among star polymers in a mound.[12] This approach enables control over the size of each mound (or number of molecules in a single assembly), by varying the heparosan concentration and droplet volume.

Our MD simulations rationalize the transition from aggregates to this polymer-melt-like state. At concentrations above 0.3 M, corresponding to the final 4% of water molecules, a transition to the melt-like state is observed. The removal of the final few percents of water corresponds to a melt-like transition shown in Figure 3C. The end-to-end distance drops from the elongated molecules observed in the transitional and aggregate state. Under spatial confined situation, such as the constant-contact-angle evaporation, the molecules begin to follow a random walk as all interactions become equal among the other heparosan, compared to the self-avoiding walk the heparosan molecules experienced with competing interactions between the heparosan and water. Figure 4D shows a visualization of this state in the simulation.

Control over the Geometry of Individual Features. As discussed in the introduction, ultra-small droplet size and rapid evaporation are key to achieving controlled molecular assembly. While previous sections addressed the molecular packing within each feature upon initial delivery of spherical hat shaped droplets, this section reports non-spherical hat shaped droplets, leading to assemblies of solutes in more complex geometries than simple disks or mounds.

Figure 5A shows 4 × 4 arrays of “volcano-like” features. Each feature was produced by dispensing 154 fL of 2×10^{-5} M heparosan tetrasaccharide solution onto an AAPTMS SAM covered glass surface. The delivery conditions: t = 1 s, p = 100 mbar with a contact force of 80 nN. Individual features can be clearly visualized as shown in Figure 5B: base diameter of 1.95-2.33 μm and with the volcanic height of 23.9-31.5 nm. The top opening measures 0.95-1.25 μm wide and 5.6-11 nm deep. The formation of this geometry is illustrated in Figure 5C. We pre-treated the exterior of the tip apex by dipping it into the solution and then letting it dry. Thus, the exterior of the tip apex becomes solution-philic, leading to symmetric climbing of liquid during delivery via capillary

interactions. The evaporation is faster than that seen in Figure 2 because the solvent here is pure water. Therefore, the final assembly adopts a similar geometry to the initial droplet. The upper "lip" was likely formed during the final separation of the tip from the droplet, when the residual amount of liquid fell atop of the partially evaporated droplet. The result shown in Figure 5A is reproducible and robust, as taller volcanos were produced by increasing the amount of material delivered, e.g., taller volcanos than the one in Figure 5A were produced (height = 42.4-87.1 nm) by increasing injection time from 1 to 3 s under delivery pressure of 400 mbar (more detailed outputs are included in Figure S3 and Table S2 in SI).

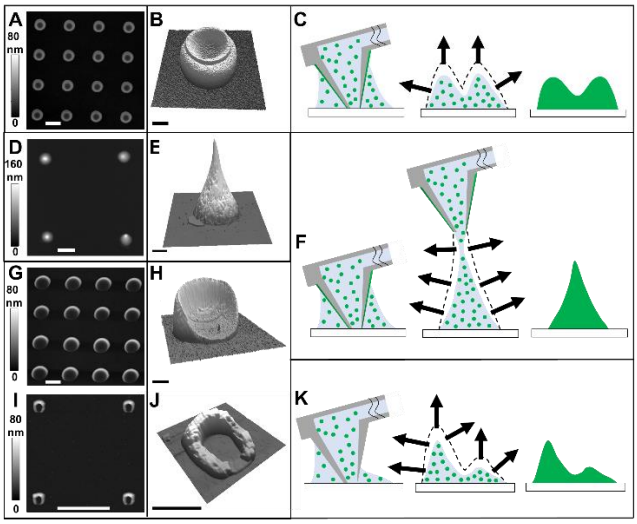


Figure 5. (A) An AFM topographic image of a 4x4 array of heparosan assemblies exhibiting volcano-like shape. (B) A 3-D display of a single feature from (A) revealing clearly the volcano-shaped geometry. (C) Schematic diagram illustrating characteristic moment of controlled assembly: during dispensing, immediately after probe withdrawal, and final assembly after solvent evaporation. Constant contact area evaporation occurred. The broken line and array arrows represent initial droplet boundary, and evaporation direction and rate, respectively. (D) AFM topographic image of a 2×2 array of heparosan assemblies with teepee-like shape. (E) A 3-D display of a single feature from (D) revealing clearly the teepee geometry. (F) Schematic diagram illustrating the assembly leading to teepee shape geometry. (G) AFM topographic image of a 4×4 array of heparosan assemblies exhibiting asymmetry. (H) A 3-D display of a single feature from (G) clearly revealing the bean-bag geometry. (K) Schematic diagram illustrates the assembly leading to asymmetric geometry. (I) AFM topographic image of a 2×2 array of heparosan assemblies exhibiting asymmetry. (J) A 3-D display of a single feature from (I). Scale bars are 2500

nm and 500 nm for white and black bars, separately. Z-scale in the 3-D displays is exaggerated to allow clear visualization of the feature geometry.

The proposed mechanism is further validated by forming “teepee” features, a 2x2 array shown in Figures 5D, under identical conditions except using a mixed solvent of glycerol:H_2O = 5:95 (v/v) and lower concentration of 2×10^{-6} M. Adding the small amount (5%) of glycerol slowed down the evaporation, allowing the solvent to merge and form capillary neck upon retreat of the tip, as illustrated in Figure 5F. The sharp tent top, clearly visible in Figure 5E, is due to the capillary neck formed at the final withdrawal. Further addition of glycerol could increase stickiness and viscosity of the solution, leading to the sharper and longer tent top, as illustrate in Figure 5F. The features in Figures 5D have a base diameter of 2.07-2.29 μm, and a height of 140-180 nm. The results were reproducible and robust. Under the same dwelling time of 0.5 s, increasing injection pressure from 500 to 900 mbar leads to increased teepee base (from 1.20 to 2.02 μm), and height (from 54 to 165 nm).

Without pre-treatment of the tip, asymmetric features, such as “bean-bag”, were formed as shown in Figure 5G (a 4x4 array). The concentration of heparosan was 2×10^{-5} M in pure water. At dwelling time of 1 s and injection pressure of 100 mbar, 176 fL of solution was delivered onto a clean and hydrophilic AAPTMS glass. As clearly shown in Figure 5H, the bean-bag measures 2.42-2.8 μm at base diameter, with high and low edges of 45.4-52.4 and 3.3-6.1 nm, respectively. We were able to tune the size of the bean-bag by varying the amount of material delivered. The formation of these asymmetric geometries are illustrated in Figure 5K. In contrast to Figure 5C, the droplet did not exhibit significant climbing up the apex exterior surfaces, instead it adopts a geometry guided by the local contact angle, as shown in Figure 5K. Since the probe tilted 11° from

the surface normal, the droplet adopted an asymmetric geometry. With the fast evaporation of water, the initial droplet shape dictated the final assembly of the solutes, leading to the bean-bag shaped features. Reducing the amount of material, further lowering the symmetry, as shown in Figure 5I, a 4x4 array of "horseshoe" features was formed. Each feature represents a heparosan assembly, after dispensing 78.4 aL of 0.01 M heparosan aqueous solution on to the surfaces and drying. The horseshoe geometry is clearly visible in Figure. 5J, with the maximum height of 95.7 nm. The full width at half maximum (FWHM) for both sides measure 137-153 and 143-161 nm with a separation of 307-359 nm. These results, collectively, demonstrate the high degree of control over the geometry of the molecular assembly by controlling the initial droplet geometry. Being stored in a clean environment under ambient conditions, these heparosan structures are stable beyond 4 months, based on time-dependent AFM imaging.

Applying Controlled Molecular Assembly to Enable 3-D Nanoprinting. With the knowledge gained above, i.e. control the assembly of heparosan via controlling the droplets dispensed, 3-D nanoprinting shall be achievable by continuous delivery of solutions following the designed trajectory. The key is maintaining consistent and homogenous assembly of heparosan throughput the printing process. In other words, the controlled molecular assembly must be maintained at each transient spot of printing. Figure 6A demonstrates that lines of heparosan can be printed. Heparosan solution (2×10^{-5} M) was delivered continuously under p = 100 mbar, at a speed of 10 μm/s. Under the delivery conditions, the transient time at each point is equivalent to a droplet deposition with dwell time of 0.4 s. Therefore, after solvent evaporation, the heparosan assembled into continuous lines with high degree of consistency and homogeneity. All 25 lines are 50 □m long, 116 nm wide and 26 nm tall, 5 of which are as shown in Figure 6A. The separation

among the lines shown in Figure 6A measures 2.15, 2.09, 2.18 and 1.87 □m, respectively, from right to left. The reproducibility was demonstrated by repeating the identical printing arrays of lines with the same geometries and size. The homogeneity and the size of the lines is consistent with the mechanism: at each point of delivery, constant area evaporation occur rapidly analogous to that in Figure 2G. The line width is tunable by the concentration and speed, e.g. slower speed leads to wider lines.

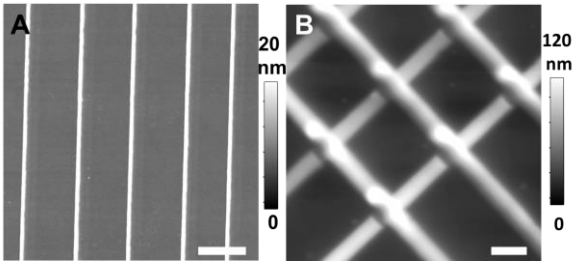


Figure 6. (A) A 15 □m x 15 □m AFM topographic image of an array of heparosan lines. (B) A 15 □m × 15 □m AFM topographic image of stacking grids consisting of heparosan. Scale bars are 2 □m.

Stacking grids, eachlayer perpendicular to another, were also printed under 200 mbar pressure, and at 2 μm/s speed. The grids covered a 15 □m x 15 □m area, with a portion shown in the AFM image in Figure 6B. Each line 39-57 nm tall and 1.01-1.09 □m wide, with periodicity or separation of 5 □m. The height of the cross section was measured to be 55-69 nm tall. The angle between top and bottom line arrays are 90.0°. The high fidelity following the design demonstrates the feasibility for 3-D nanoprinting using functional molecules such as heparosan.

3.4 Conclusions

Using heparosan tetrasaccharide as the functional solute, this work demonstrates our new concept of controlled assembly of molecules. The key to controlled assembly is the fact that

ultra-small solution droplets follow different evaporation dynamics from those of larger ones. The initial shape of the droplet and the concentration of solute within the droplet dictate the evaporation mechanism and the final assembly of molecules due to the ultrafast evaporation rate and dynamic spatial confinement of the droplets. The level of control demonstrated in this work brings us closer to programmable synthesis and assembly for chemistry and materials science. Work is in progress to explore its application in production of drug delivery vehicles, and 3-D nanoprinting in additive manufacturing.

Acknowledgements

We thank Drs. Pablo Dörig, Dario Ossola at Cytosurge, Glattbrugg, Switzerland for helpful scientific discussion, and Ms. Susan Stagner for careful proof-reading. This work is supported by Gordon and Betty Moore Foundation, the National Science Foundation (CHE-1808829), Lawrence Livermore National Laboratory LDRD project 19-ERD-012, under department of Energy Contract DE-AC52-07NA27344, UC Davis and the National Institutes of Health (NIH) Common Fund Glycoscience Program grant, U01GM125288, and 1R01AR070239 - 01A1.

3.5 References

1. Whitesides, G. M.; Grzybowski, B., Self-assembly at all scales. *Science* **2002,** 295, (5564), 2418-2421.
2. Xia, Y.; Whitesides, G. M., Soft lithography. *Annual review of materials science* **1998,** 28, (1), 153-184.
3. Bain, C. D.; Troughton, E. B.; Tao, Y. T.; Evall, J.; Whitesides, G. M.; Nuzzo, R. G., Formation of monolayer films by the spontaneous assembly of organic thiols from solution onto gold. *Journal of the American Chemical Society* **1989,** 111, (1), 321-335.
4. Sagiv, J., Organized monolayers by adsorption. 1. Formation and structure of oleophobic mixed monolayers on solid surfaces. *Journal of the American Chemical Society* **1980,** 102, (1), 92-98.
5. Nuzzo, R. G.; Allara, D. L., Adsorption of bifunctional organic disulfides on gold surfaces. *Journal of the American Chemical Society* **1983,** 105, (13), 4481-4483.

6. Hillmyer, M. A.; Bates, F. S.; Almdal, K.; Mortensen, K.; Ryan, A. J.; Fairclough, J. P. A., Complex phase behavior in solvent-free nonionic surfactants. *Science* **1996,** 271, (5251), 976-978.
7. Hayward, R. C.; Pochan, D. J., Tailored Assemblies of Block Copolymers in Solution: It Is All about the Process. *Macromolecules* **2010,** 43, (8), 3577-3584.
8. Pochan, D. J.; Chen, Z. Y.; Cui, H. G.; Hales, K.; Qi, K.; Wooley, K. L., Toroidal triblock copolymer assemblies. *Science* **2004,** 306, (5693), 94-97.
9. Bigioni, T. P.; Lin, X. M.; Nguyen, T. T.; Corwin, E. I.; Witten, T. A.; Jaeger, H. M., Kinetically driven self assembly of highly ordered nanoparticle monolayers. *Nature Materials* **2006,** 5, (4), 265-270.
10. Nagle, J. F.; Tristram-Nagle, S., Structure of lipid bilayers. *Biochimica Et Biophysica Acta-Reviews on Biomembranes* **2000,** 1469, (3), 159-195.
11. Groves, J. T.; Ulman, N.; Boxer, S. G., Micropatterning fluid lipid bilayers on solid supports. *Science* **1997,** 275, (5300), 651-653.
12. Zhang, J.; Piunova, V. A.; Liu, Y.; Tek, A.; Yang, Q.; Frommer, J.; Liu, G.-y.; Sly, J., Controlled Molecular Assembly via Dynamic Confinement of Solvent. *The Journal of Physical Chemistry Letters* **2018,** 9, (21), 6232-6237.
13. Wang, Z.; Dordick, J. S.; Linhardt, R. J., Escherichia coli K5 heparosan fermentation and improvement by genetic engineering. *Bioeng Bugs* **2011,** 2, (1), 63-67.
14. DeAngelis, P. L.; Gunay, N. S.; Toida, T.; Mao, W. J.; Linhardt, R. J., Identification of the capsular polysaccharides of Type D and F Pasteurella multocida as unmodified heparin and chondroitin, respectively. *Carbohydrate research* **2002,** 337, (17), 1547-52.
15. Sismey-Ragatz, A. E.; Green, D. E.; Otto, N. J.; Rejzek, M.; Field, R. A.; DeAngelis, P. L., Chemoenzymatic synthesis with distinct Pasteurella heparosan synthases: monodisperse polymers and unnatural structures. *The Journal of biological chemistry* **2007,** 282, (39), 28321-7.
16. Rippe, M.; Stefanello, T. F.; Kaplum, V.; Britta, E. A.; Garcia, F. P.; Poirot, R.; Companhoni, M. V. P.; Nakamura, C. V.; Szarpak-Jankowska, A.; Auzely-Velty, R., Heparosan as a potential alternative to hyaluronic acid for the design of biopolymer-based nanovectors for anticancer therapy. *Biomaterials science* **2019,** 7, (7), 2850-2860.
17. Chen, J.-X.; Zhang, M.; Liu, W.; Lu, G.-Z.; Chen, J.-H., Construction of serum resistant micelles based on heparosan for targeted cancer therapy. *Carbohydrate Polymers* **2014,** 110, 135-141.
18. Qiu, L.; Shan, X.; Long, M.; Ahmed, K. S.; Zhao, L.; Mao, J.; Zhang, H.; Sun, C.; You, C.; Lv, G.; Chen, J., Elucidation of cellular uptake and intracellular trafficking of heparosan polysaccharide-based micelles in various cancer cells. *International journal of biological macromolecules* **2019,** 130, 755-764.
19. Sun, C.; Li, X.; Du, X.; Wang, T., Redox-responsive micelles for triggered drug delivery and effective laryngopharyngeal cancer therapy. *International journal of biological macromolecules* **2018,** 112, 65-73.
20. Qiu, L.; Ge, L.; Long, M.; Mao, J.; Ahmed, K. S.; Shan, X.; Zhang, H.; Qin, L.; Lv, G.; Chen, J., Redox-responsive biocompatible nanocarriers based on novel heparosan polysaccharides for intracellular anticancer drug delivery. *Asian Journal of Pharmaceutical Sciences* **2018**.
21. Chen, J.-X.; Liu, W.; Zhang, M.; Chen, J.-H., Heparosan based negatively charged nanocarrier for rapid intracellular drug delivery. *International Journal of Pharmaceutics* **2014,** 473, (1), 493-500.

22. Danaei, M.; Dehghankhold, M.; Ataei, S.; Hasanzadeh Davarani, F.; Javanmard, R.; Dokhani, A.; Khorasani, S.; Mozafari, M. J. P., Impact of particle size and polydispersity index on the clinical applications of lipidic nanocarrier systems. **2018,** 10, (2), 57.
23. Lan Na, H. Y., Tamashree Ghosh,John B. McArthur,Thomas Asbell,Xi Chen, Engineered P. multocida heparosan synthases improve size-controlled enzymatic synthesis of longer heparosan oligosaccharides. In *The Journal of the American Chemical Society*, 2019 (Submitted).
24. Li, Y.; Yu, H.; Chen, Y.; Lau, K.; Cai, L.; Cao, H.; Tiwari, V. K.; Qu, J.; Thon, V.; Wang, P. G.; Chen, X., Substrate promiscuity of N-acetylhexosamine 1-kinases. *Molecules* **2011,** 16, (8), 6396-407.
25. Chen, Y.; Thon, V.; Li, Y.; Yu, H.; Ding, L.; Lau, K.; Qu, J.; Hie, L.; Chen, X., One-pot three-enzyme synthesis of UDP-GlcNAc derivatives. *Chem Commun (Camb)* **2011,** 47, (38), 10815-7.
26. Lau, K.; Thon, V.; Yu, H.; Ding, L.; Chen, Y.; Muthana, M. M.; Wong, D.; Huang, R.; Chen, X., Highly efficient chemoenzymatic synthesis of beta1-4-linked galactosides with promiscuous bacterial beta1-4-galactosyltransferases. *Chem Commun (Camb)* **2010,** 46, (33), 6066-8.
27. Muthana, M. M.; Qu, J.; Xue, M.; Klyuchnik, T.; Siu, A.; Li, Y.; Zhang, L.; Yu, H.; Li, L.; Wang, P. G.; Chen, X., Improved one-pot multienzyme (OPME) systems for synthesizing UDP-uronic acids and glucuronides. *Chem Commun (Camb)* **2015,** 51, (22), 4595-8.
28. Muthana, M. M.; Qu, J.; Li, Y.; Zhang, L.; Yu, H.; Ding, L.; Malekan, H.; Chen, X., Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from Bifidobacterium longum (BLUSP). *Chem Commun (Camb)* **2012,** 48, (21), 2728-30.
29. Li, J. R.; Lusker, K. L.; Yu, J. J.; Garno, J. C., Engineering the spatial selectivity of surfaces at the nanoscale using particle lithography combined with vapor deposition of organosilanes. *ACS Nano* **2009,** 3, (7), 2023-2035.
30. Li, J. R.; Garno, J. C., Elucidating the role of surface hydrolysis in preparing organosilane nanostructures via particle lithography. *Nano Letters* **2008,** 8, (7), 1916-1922.
31. Lin, W. F.; Swartz, L. A.; Li, J. R.; Liu, Y.; Liu, G. Y., Particle Lithography Enables Fabrication of Multicomponent Nanostructures. *The Journal of Physical Chemistry C* **2013,** 117, (44), 23279-23285.
32. Lin, W. F.; Li, J. R.; Liu, G. Y., Near-Field Scanning Optical Microscopy Enables Direct Observation of Moire Effects at the Nanometer Scale. *Acs Nano* **2012,** 6, (10), 9141-9149.
33. Garno, J. C.; Batteas, J. D., Nanofabrication with self-assembled monolayers by scanning probe lithography. In *Applied Scanning Probe Methods IV: Industrial Applications*, 2006; pp 105-135.
34. Tian, T.; LeJeune, Z. M.; Serem, W. K.; Yu, J. J.; Garno, J. C., Nanografting: A method for bottom-up fabrication of designed nanostructures. In *Tip-Based Nanofabrication: Fundamentals and Applications*, 2011; pp 167-205.
35. Lifang Shi, A. K., Jianli Zhao, Miquel Salmeron, Gang Yu Liu, High-Resolution Imaging of Cells Using Current Sensing Atomic Force Microscopy. *Lett. Appl. NanoBioScience* **2015,** 4, (4), 316-320.
36. Abbott, N. L.; Gorman, C. B.; Whitesides, G. M., Active control of wetting using applied electrical potentials and self-assembled monolayers. *Langmuir* **1995,** 11, (1), 16-18.

37. Bain, C. D.; Whitesides, G. M., A study by contact angle of the acid-base behavior of monolayers containing. omega.-mercaptocarboxylic acids adsorbed on gold: an example of reactive spreading. *Langmuir* **1989,** 5, (6), 1370-1378.
38. Yuan, W.; Van Ooij, W., Characterization of organofunctional silane films on zinc substrates. *Journal of Colloid Interface Science* **1997,** 185, (1), 197-209.
39. Sugimura, H.; Hayashi, K.; Saito, N.; Nakagiri, N.; Takai, O., Surface potential microscopy for organized molecular systems. *Applied Surface Science* **2002,** 188, (3-4), 403-410.
40. Hayashi, K.; Saito, N.; Sugimura, H.; Takai, O.; Nakagiri, N., Surface potential contrasts between silicon surfaces covered and uncovered with an organosilane self-assembled monolayer. *Ultramicroscopy* **2002,** 91, (1-4), 151-156.
41. Mirji, S., Octadecyltrichlorosilane adsorption kinetics on Si (100)/SiO2 surface: contact angle, AFM, FTIR and XPS analysis. *Surface Interface Analysis: An International Journal devoted to the development application of techniques for the analysis of surfaces, interfaces thin films* **2006,** 38, (3), 158-165.
42. Wang, Y.; Lieberman, M., Growth of ultrasmooth octadecyltrichlorosilane self-assembled monolayers on SiO2. *Langmuir* **2003,** 19, (4), 1159-1167.
43. Ventrici de Souza, J.; Liu, Y.; Wang, S.; Dörig, P.; Kuhl, T. L.; Frommer, J.; Liu, G.-y., Three-dimensional nanoprinting via direct delivery. *The Journal of Physical Chemistry B* **2017,** 122, (2), 956-962.
44. Deng, W. N.; Wang, S.; Ventrici de Souza, J.; Kuhl, T. L.; Liu, G.-y., New Algorithm to Enable Construction and Display of 3D Structures from Scanning Probe Microscopy Images Acquired Layer-by-Layer. *The Journal of Physical Chemistry A* **2018,** 122, (26), 5756-5763.
45. Meister, A.; Gabi, M.; Behr, P.; Studer, P.; Vörös, J. n.; Niedermann, P.; Bitterli, J.; Polesel-Maris, J.; Liley, M.; Heinzelmann, H., FluidFM: combining atomic force microscopy and nanofluidics in a universal liquid delivery system for single cell applications and beyond. *Nano letters* **2009,** 9, (6), 2501-2507.
46. Grüter, R. R.; Vörös, J.; Zambelli, T., FluidFM as a lithography tool in liquid: spatially controlled deposition of fluorescent nanoparticles. *Nanoscale* **2013,** 5, (3), 1097-1104.
47. Chhasatia, V. H.; Sun, Y., Interaction of bi-dispersed particles with contact line in an evaporating colloidal drop. *Soft Matter* **2011,** 7, (21), 10135-10143.
48. Ko, H.-Y.; Park, J.; Shin, H.; Moon, J., Rapid Self-Assembly of Monodisperse Colloidal Spheres in an Ink-Jet Printed Droplet. *Chemistry of Materials* **2004,** 16, (22), 4212-4215.
49. Park, J.; Moon, J., Control of colloidal particle deposit patterns within picoliter droplets ejected by ink-jet printing. *Langmuir* **2006,** 22, (8), 3506-3513.
50. Lee, V. Y.; Havenstrite, K.; Tjio, M.; McNeil, M.; Blau, H. M.; Miller, R. D.; Sly, J. J. A. M., Nanogel Star Polymer Architectures: A Nanoparticle Platform for Modular Programmable Macromolecular Self-Assembly, Intercellular Transport, and Dual-Mode Cargo Delivery. **2011,** 23, (39), 4509-4515.
51. Erbil, H. Y.; McHale, G.; Newton, M., Drop evaporation on solid surfaces: constant contact angle mode. *Langmuir* **2002,** 18, (7), 2636-2641.
52. Kuang, M.; Wang, J.; Bao, B.; Li, F.; Wang, L.; Jiang, L.; Song, Y., Inkjet Printing Patterned Photonic Crystal Domes for Wide Viewing-Angle Displays by Controlling the Sliding Three Phase Contact Line. *Advanced Optical Materials* **2014,** 2, (1), 34-38.
53. Kuncicky, D. M.; Velev, O. D., Surface-guided templating of particle assemblies inside drying sessile droplets. *Langmuir* **2008,** 24, (4), 1371-1380.

54. Picknett, R.; Bexon, R., The evaporation of sessile or pendant drops in still air. *J. Colloid Interface Sci* **1977,** 61, (2), 336-350.
55. Zhong, X.; Crivoi, A.; Duan, F., Sessile nanofluid droplet drying. *Adv. Colloid Interface Sci.* **2015,** 217, 13-30.

3.6 Supporting Information

3.6.1 Description of molecular dynamics simulations performed

Simulations were performed in Gromacs 5.1 using an Amber force field.[1, 2] heparosan molecules were generated by taking the Fmoc group from Amber tools and attaching to a heparosan tetrasaccharide group from the Glycam model[3] using the tool developed by Bernardi.[4] We modified some partial charges on the adjoining groups in order to correctly balance charge. The forcefield used is included in supporting file myffnonbonded.itp. 27 Fmoc-tagged heparosan tetrasaccharide molecules were placed in a periodic cubic box of dimensions 15.3 nm x 15.3 nm x 15.3 nm. The boxes were constructed such that the minimum distance between heparosan molecules was at least 1 nm in all directions. The initial NpT simulation was performed for 45 ns and data collected every 0.01 ns. This run was preceded by solvation and charge balancing, and separate 100 ps NVT and NpT equilibration runs. The solvation and charge neutralization was performed through the addition of 54 Na^+ ions. A velocity-rescale thermostat[5] was used with a reference temperature of 300 K. An isotropic Parrinello-Rahman barostat[6] was used with a reference pressure of 1 bar. Varying concentration NpT production runs were 1 ns in length each and achieved by deleting 1% of the initial water quantity (1177 water molecules deleted per simulation) from the final frame of the production simulation preceding it. The final 1% of water was deleted by removing 0.1% of the initial water (117 molecules per simulation) until no water remained. Table S1 contains the simulation information in terms of the number of

heparosan molecules, neutralizing ions, and total atoms. The box-lengths post NpT production run are listed, along with the concentrations of the heparosan calculated using the box volume. Finally, production simulation time is listed. The naming convention for simulations is as follows: FMOC, to represent the molecule of interest FMOC tagged heparosan tetrasaccharide, followed by the type of run performed. For example, the first row is the initial 45 ns warmup simulation. Following this, the md production runs with "#" (%) of water molecules deleted from the initial solvent. The letter "i" is used in row 4 to represent the range of box-lengths and concentrations for simulations FMOC_md_3_del to FMOC_md_97_del (i.e. steps 3 to 97 in the deletion sequence). The 6th row of the table shows the range of box-lengths and concentrations for the final 1% of water using the letter "j" to represent 99.1% - 99.9% initial water deletion. The deletion process was carried out as follows: care was taken to remove waters least likely to be involved in hydrogen bonding, so only water molecules at least 1 nm from the nearest heparosan residue were removed. Molecules to be deleted were added to a list based on the distance selection criteria, and then the random.sample function in Python was used to randomly select the specified number of water molecules to be deleted. For the final 4% of water, there were not enough water molecules 1 nm from a heparosan residue. The criteria was gradually relaxed until being completely turned off in order to keep the number of waters deleted constant. After each deletion, a 100 ps NVT and NpT equilibration was performed as previously described. This process was repeated until all water molecules were removed.

Table S1. Molecular Dynamics Simulation Information.

All 110 simulations performed NpT for time listed. Box-length is taken post NpT. Concentration is calculated using box-length. Simulation name defined as molecule of interest followed by type of run, warmup or production md. Number next to md in name corresponds to % of initial water deleted in that simulation. “i” corresponds to 3-97%, “j” corresponds to 99.1-99.9%. Formula to calculate the number of atoms in one of these simulations is included in the table.

Simulation	# of Heparosan	Neutralizing Ions	# Atoms	Box-Length (nm)	Concentration (mM)	time (ns)
FMOC_warmup	27	54 Na^+	353199	15.214	12.73	45
FMOC_md_1_del	27	54 Na^+	349668	15.166	12.85	1
FMOC_md_2_del	27	54 Na^+	346137	15.119	12.97	1
FMOC_md_i_del	27	54 Na^+	353199 - (3531 * i)	15.061 – 4.644	13.12 – 447.6	1
FMOC_md_98_del	27	54 Na^+	7161	4.038	680.8	1
FMOC_md_99_j_del	27	54 Na^+	7161 – (353 * j)	4.038 – 3.200	707.5 – 1165	1
FMOC_md_100_del	27	54 Na^+	3533	3.200	1248	1

3.6.2 Radial distribution function

As described in the manuscript we hypothesize based on molecular dynamics simulations that there are potentially 3 phases observed across a range of concentrations. Further evidence to support this is shown below in Figure S1. In the manuscript the images from 3 simulations in the described concentration ranges are shown. Figure S1 contains the pair-wise radial distribution function for a randomly selected heparosan molecule (number 17) with respect to the rest of the heparosan molecules in the simulations. The pair-wise radial distribution function shows the likelihood that a molecule will be encountered within a certain distance. Default Gromacs settings for the gmx rdf function[1] were used (bin size of 0.002 nm and 0 cut off distance), except exclusions were on and normalization was off. The exclusions option was included to prevent atoms in the heparosan residue used as the reference from appearing in

the data twice. The data is not normalized according to density, as is default for Gromacs, because the relative simulation densities were different. At low concentration the heparosan molecules are most likely to be found either very close to the reference molecule, the increase below 3 nm, or far away the increase above 6 nm. This behavior is consistent for a solution phase behavior. There is a broadening demonstrated at the intermediate concentration, as the heparosan molecules are expanding and interacting more strongly, and finally the broadening intensifies close to what you would expect for a melt like behavior at high concentration.

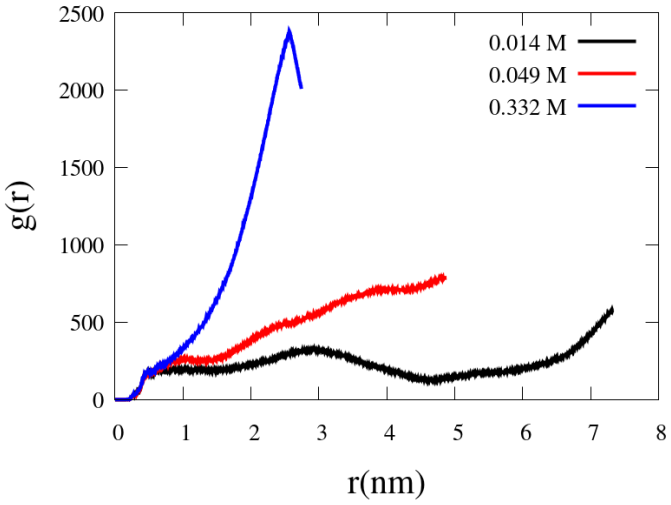


Figure S1. Pair-wise radial distribution function for heparosan molecule 17 to other heparosan molecules in simulation, with exclusions. Selected md simulations 8% deletion, 74% deletion, 96% deletion corresponding to i = 8, 74, 96 in Table S1. The same simulations were utilized for visualization shown in Figure 4 (in the main text).

3.6.3 Hydrogen bonds

Further supporting evidence for these simulated phases are the hydrogen bonding from one heparosan to another and to surrounding water molecules. The default settings for Gromacs gmx hbond were used and calculated according to the method described in the Gromacs

Manual.[1] The function determines the number and lifetime of hydrogen bonds between donors and acceptors based on a cutoff distance of 0.35 nm and a cutoff angle of 30°. -NH and -OH groups both act as hydrogen donors. The results for the number of hydrogen bonds of the two types (heparosan-heparosan and heparosan-water) are shown in Figure S2. The same phases are observed at the concentration ranges described in the paper. As described in the paper, 3 phases are observed, a solution phase below 0.045M, an elongated transition phase between 0.045M and 0.1M, and a transition towards a polymer melt beginning at 0.3M. As concentration increases, water is removed, heparosan molecules are pushed closer together, and the hydrogen bonding in the system changes. The slope for the hydrogen bonding in both plots in Figure S2 changes based on which concentration range described above it is in. The number of heparosan-heparosan hydrogen bonds increases with concentration as the system transitions from a solution to an elongated transition state as the molecules try to find one another and maximize their hydrogen bonding. Finally, the elongated state relaxes and forms a polymer melt[7] with primarily heparosan-heparosan hydrogen bonds. Hydrogen bonding between heparosan and waters follows the inverse trend as expected, as the increasing heparosan-heparosan hydrogen bonding lead to decrease of heparosan-water hydrogen bonding.

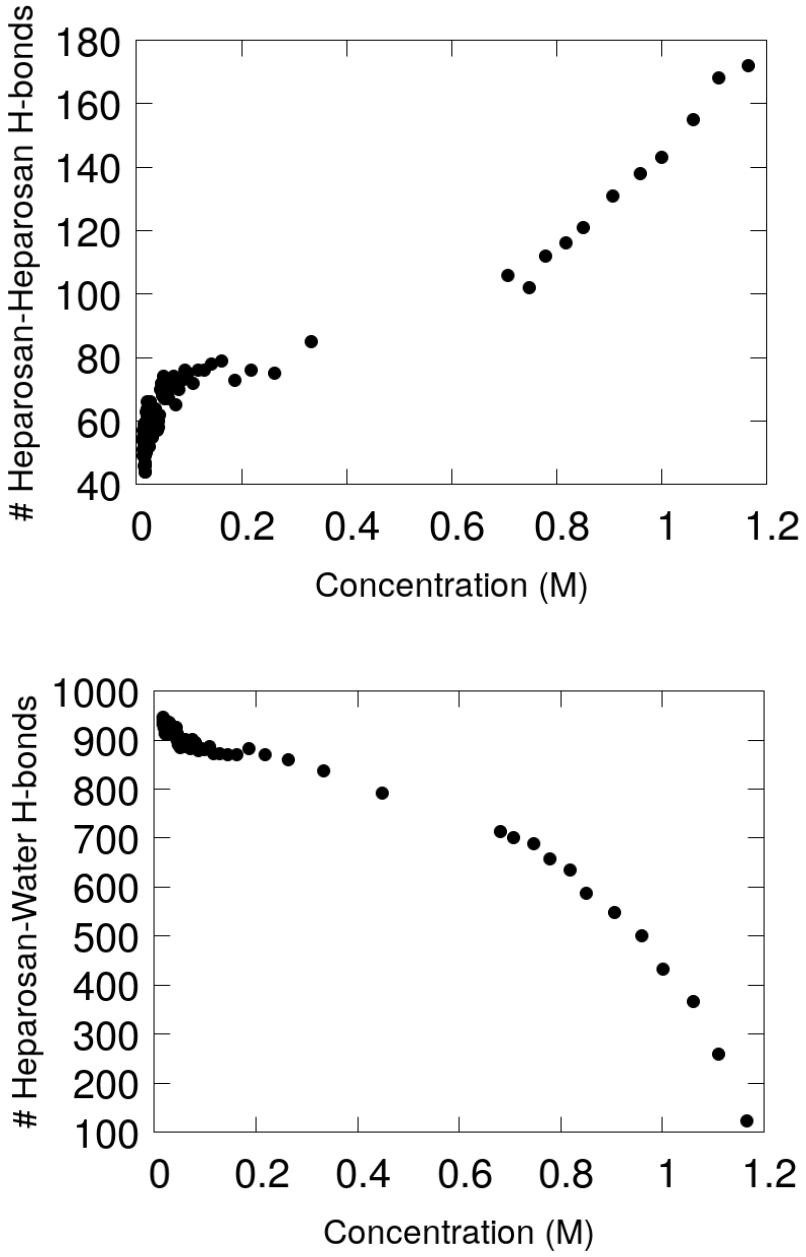


Figure S2. (A) Number of heparosan-heparosan intermolecular hydrogen bonds observed in simulation as a function of increasing heparoson concentration. (B) Number of heparosan-water intermolecular hydrogen bonds observed with the increasing heparosan concentration.

3.6.6 Asymmetric Structures Formed via Controlled Assembly

To demonstrate the robustness of forming asymmetric structure using controlled assembly, we varied the delivery conditions to produce droplets with increasing volume. As shown in Figure S3, a 3x3 arrays of heparosan droplets (2×10^{-5} M) were delivered onto an AAPTMS SAM/glass surface. The delivery time (t) and pressure (P) are indicated in Figure S3A, and the contact force was 80 nN. The dimension of each features can be measured from AFM topographic images and summarized in Table S2.

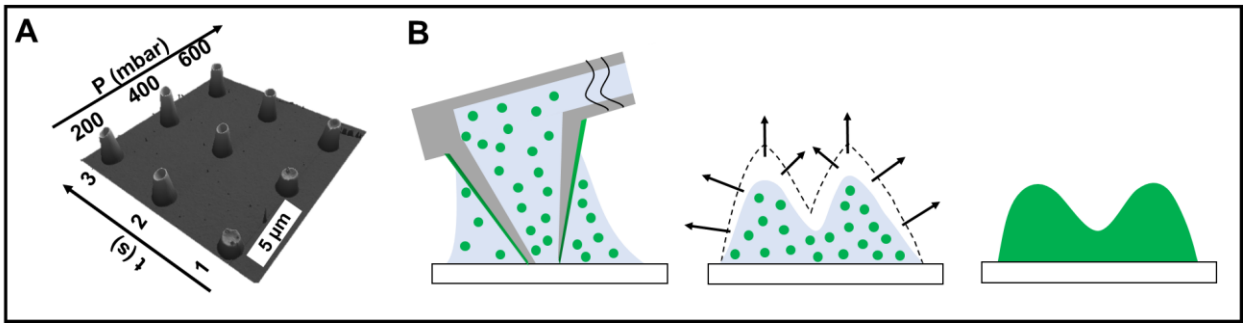


Figure S3. (A) AFM topograph of a 3x3 array of heoparosan assemblies exhibiting volcano-like shape. The t and P used to dispense each droplet are clearly indicated. (B) Schematic diagram illustrates the constant contact area evaporation of a droplet on hydrophilic AAPTMS/glass surface. Broken line and array arrows represent initial droplet boundary, and evaporation direction and rate, respectively.

Table S2. Consolidated Parameters and Metrics for the 3x3 array of heparosan assemblies exhibiting volcano-liked shape in Figure S3

Pressure (mbar)	Contact time (s)	Height (nm)		Base diameter (μm)
		Low	High	
200	1	25.9	35.1	1.845
200	2	43.9	58.9	2.191
200	3	48.1	64.3	2.075
400	1	32.1	42.4	1.956
400	2	44.4	62.3	1.899
400	3	58.6	87.1	2.083
600	1	47.3	63.2	1.978
600	2	46.6	68.3	1.957
600	3	51.6	79.9	2.014

CHAPTER 4

The impact of surface polarity on lipid assembly under spatial confinement

This work is currently in preparation for submission for publication, draft as of November 2021

Bradley Harris[1], Yuqi Huang[2], Arpad Karsai[2], Gang-yu Liu[2], Roland Faller[1*]*

[1] Department of Chemical Engineering, University of California, Davis, California, 95616, United States
[2] Department of Chemistry, University of California, Davis, California, 95616, United States

Abstract:

A collection of molecular dynamics simulations utilizing the MARTINI model were used to study the interplay between POPC and surface polarity during assembly under spatial confinement due to solvent evaporation. To the best of our knowledge this work is the first to use molecular dynamics to study the assembly of lipids during solvent evaporation as well as the first to study the effect of surface termination at variable solvent concentrations. These systems were additionally evaluated with and without glycerol, to identify its effect on structure assembly. The presented methodology provides insight into the assembly of lipids onto different surfaces and the mechanism by which glycerol stabilizes and changes those assembled structures. All surfaces resulted in the formation of lipid bilayers, with increasing homogeneity and surface mobility with increasing polarity.

4.1 Introduction

Phospholipids are a critically important aspect of biological systems due to their important role in nanoparticles such as viruses and liposomes, and for complex systems such as cells. Because of this multitude of function, they have been extensively studied experimentally and computationally. Of particular interest is the ability of these lipids to self-assemble into various forms including bilayers, vesicles, and other heterogeneous mixtures for use in identifying numerous behaviors and use in applications.[1-4] Among the reported approaches to build phospholipid bilayers and multilayer constructs, Langmuir-Blodgett, Langmuir-Schaefer, and vesicle fusion approaches are common.[5-7] While these techniques are experimentally simple to perform, the final structure depend on the complex interplay among lipid-surface interactions, intermolecular interactions among lipid molecules, surface polarity, and the local environment.[8] Thus it is difficult to control and predict the local features apriori. Using the new concept of controlled molecular assembly, our team has developed a new means to form molecular assemblies on surfaces by design.[9-10] This has previously been demonstrated for the assembly of large spherical macromolecules, as well as small non-spherical sugars. This assembly depends on the delivery of ultra-small (sub-fL) quantities of liquid with spatio-temporal control provided by a microfluidic probe. The deposition of lipids and formation of lipid structures using this technique is highly dependent on the interplay of all aspects of surface-lipid, surface-solution, and lipid solution effects that contribute to SLBs. Compared to previous SLBs which could be modeled through adsorption with an unchanging quantity of water, this controlled assembly approach is highly dependent on solvent evaporation and its interplay with surface property.

Molecular Dynamics (MD) simulations are one of the most common and powerful techniques used to analyze lipid-membrane properties directly.[11] Despite the frequency of MD lipid simulations there are comparatively few studies investigating the complexity of lipid-surface interactions. All-atom models have been used to accurately describe the confinement of hydration water and surface "lubrication" in SLB systems, but the computational cost to scale up these systems has largely prevented the study of more complex interactions.[12-14] Coarse-grained (CG) models on the other hand have been used to study more complex effects such as lipid vesicle fusion and surface roughness and inhomogeneity effects.[15-17] Additionally, the intermediate MARTINI model is extremely popular for lipid bilayer simulations due to providing a level of coarse-graining capable of replicating complex structures, while still providing enough molecular detail to compare to experimental data. The MARTINI model has been used to study the interactions between lipid bilayers and solid supports.[18] These previous studies have especially focused on the difference between the top and bottom leaflets, the effects of topology, and on the vesicle fusion to hydrophobic surfaces.[15-16,19-21] Most of these simulations have focused on one aspect of the surface-lipid, surface-solution, and lipid-solution contributions to the behavior of the corresponding experimental SLB systems. In this work we present a collection of molecular dynamics simulations utilizing the MARTINI CG model to study the interplay between POPC and hydrophobic and hydrophilic surfaces as a function of relative lipid-solvent concentration. Previous efforts to study lipid surface interactions at varying hydration levels have focused on hydrophilic surfaces with constant solvent molecules with varying surface thickness change the thickness of the lubricating water layer.[18,22] To the best of our knowledge this work represents the first to use MD to study the effect of surface polarity at

variable solvent concentrations under spatial confinement. Additionally, we evaluate the impact of evaporation rate by comparing assemblies with and without glycerol, to investigate the role glycerol plays in addition to simply slow down evaporation. This developed method is used to gain insight into the mechanisms of lipid assembly onto surfaces and the corresponding structural evolution during evaporation, as well as the mechanism by which glycerol stabilizes and changes those assembled structures.

4.2 Materials and Methods

4.2.1 CG Description

The forcefield and coarse-grained model used is the MARTINI 3 forcefield.[23] This model is based on a 4-1 mapping with 4 heavy atoms being represented as a single interactive bead, except in the case of some ring-like molecules which have a higher resolution. The model uses 4 primary CG bead types, C, N, P, Q, corresponding to nonpolar, intermediately polar, polar, and charged chemical groups respectively. The non-bonded interactions are described solely by Lennard-Jones potentials between non-charged beads, while charged beads also include Coulombic interactions. Bead sublabels are used to differentiate degrees of polarity or hydrogen donor/acceptor capabilities. All beads are the same size except for certain ring structures (S), and certain nucleotides (T). The MARTINI forcefield has been successfully used to study many biomembrane, protein, and material science problems.[24-26]

The MARTINI representations of the relevant molecular species here are shown in Figure 1. The phospholipid 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) is the lipid of interest in these simulations. POPC is represented by 12 beads, corresponding to a positively charged choline (Q1), a negatively charged phosphate(Q5), two neutrally charged glycerols (SN4a and

N4a), and 2 tails with nonpolar alkane-like beads (C1). The 2nd bead of the unsaturated first chain is represented by a more nonpolar "C4h" bead. Ethanol is the primary solvent molecule, represented with the SP1 interaction type. This corresponds to a smaller bead with a weakly polar interaction strength. Solvent glycerol is represented by the same neutrally charged 2 bead system as in POPC, with the modification being a hydrogen donor instead of acceptor (SN4d and N4d).

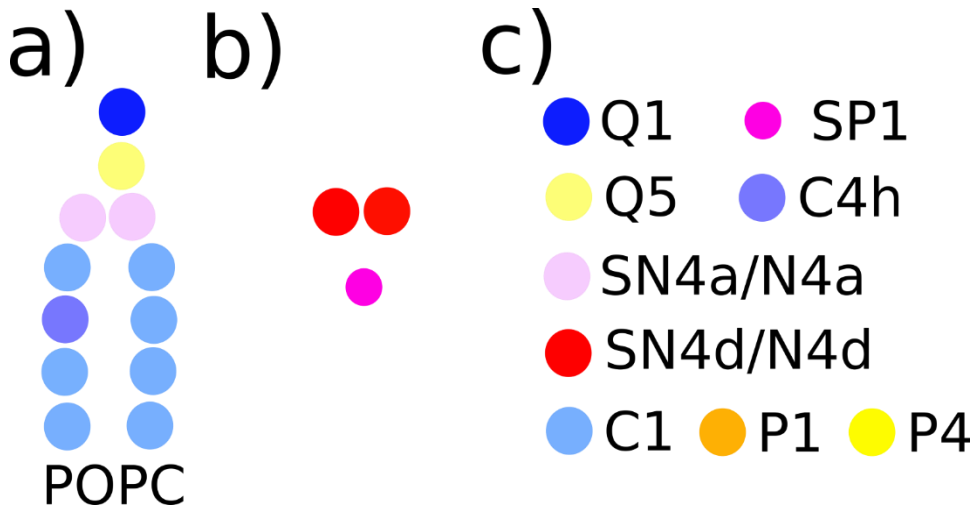


Figure 1: MARTINI 3 species representation. a) phospholipid POPC. b) Solvent molecules, glycerol (red) and ethanol (magenta). c) Bead type key

4.2.2 Configurations:

Initial surface geometry uses a regular graphene like pattern and was built according to the Martini 2 tutorial on graphene leaflets.[27] Surface shape was chosen to avoid using a perfectly flat surface with no edges, as with that level of molecular smoothness the area per lipid would be artificially set. Different surface Martini bead types were used to model a range of surface hydrophilicities. The chosen bead types were C1, P1, and P4, to represent a non-polar Carbon like surface, a weakly polar surface, and a strongly polar surface, respectively. Initial POPC lipid configurations were built using the insane.py script for bilayer stacks, and the gmx insert

command for random starting configurations.[28] Initial molecule counts are shown in table 1 and visualizations in Figure 2. Figure 2 shows configurations post 210 ns equilibration for the P4 surfaces, 2a corresponds to the 1 bilayer, 2b-c to the 6 bilayers with and without glycerol respectively, and 2d-e with random starts with and without glycerol respectively. For each surface the starting configurations studied included a single bilayer, 6 bilayer stacks with and without glycerol, and a configuration of randomly placed POPC with the same number of molecules as the 6-bilayer stack.

Table 1: Starting simulation molecule counts and starting configurations.

Table 1: Simulation system details

Surface Type	# POPC	# EOL	# GLY	# GRAP	Initial Configuration
C1	3042	176366	0	80000	1 membrane bilayer
C1	8112	114092	0	19600	6 membrane bilayer
C1	8112	97768	10000	19600	6 membrane bilayer
C1	8112	93745	0	19600	Random
C1	8112	80813	10000	19600	Random
P1	3042	176366	0	80000	1 membrane bilayer
P1	8112	114092	0	19600	6 membrane bilayer
P1	8112	97768	10000	19600	6 membrane bilayer
P1	8112	93745	0	19600	Random
P1	8112	81223	10000	19600	Random
P4	3042	176366	0	80000	1 membrane bilayer
P4	8112	114092	0	19600	6 membrane bilayer
P4	8112	97768	10000	19600	6 membrane bilayer
P4	8112	93745	0	19600	Random
P4	8112	81148	10000	19600	Random

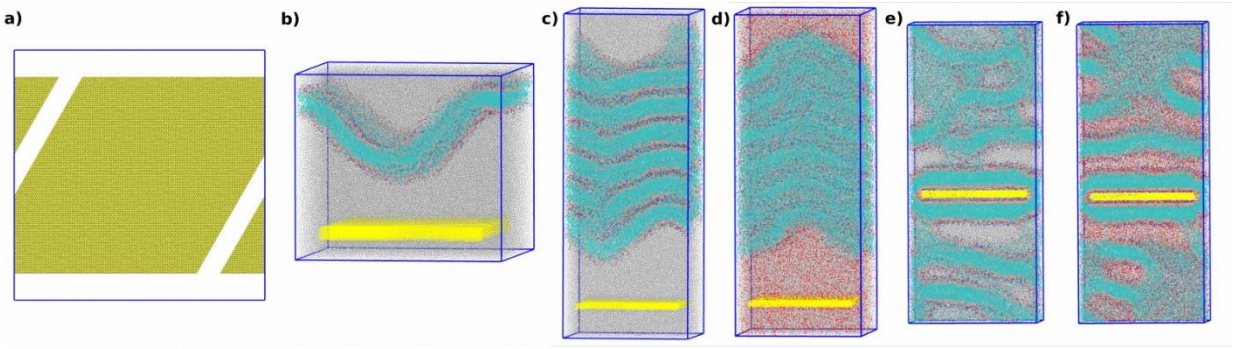


Figure 2: Representative starting configurations for P4 surface type after 210 ns equilibration before evaporation. a) surface structure b) 1 membrane bilayer start. c) 6 membrane bilayers without glycerol. d) 6 membrane bilayers with glycerol. e) random start without glycerol. f) random start with glycerol.

4.2.3 Simulation protocols:

Molecular dynamics (MD) simulations were carried out in Gromacs 2019.1 using the MARTINI 3 force field.[23,29] Following initial configuration generation, a series of minimization steps were performed. Energy minimization was performed twice, before and after solvation with ethanol. The second energy minimization was followed by short 300 picosecond NVT and NPT equilibrations. NPT and NVT were done at 300K with a Berendsen thermostat, and 1 atm Berendsen barostat for NPT.[30] Longer equilibration runs were performed NPT with a 300 K Berendsen thermostat and anisotropic 1 atm Berendsen barostat for 210 ns to ensure the starting structures were sufficiently equilibrated and had no surface interactions. Evaporation was performed using the methodology described in and used previously in our prior work.[9] Each evaporation step corresponded to the removal of 1% of the initial ethanol and was simulated for 15 ns per step using the same NPT production conditions. The final 1% of solvent was removed at a rate of 0.1% initial ethanol per step. Each evaporation required 110 simulations for full removal, making the total simulated evaporation time 1.65 μs.

Lipid density profiles were calculated using built in gromacs tools, specifically gmx density.[29] All visualizations were performed using VMD.[31] Due to curvature in the lipid

membranes during the dehydration process area per lipid calculations were performed using a Voronoi tessellation performed using the software APLVORO.[32] These Voronoi tessellations were reconstructed using python and matplotlib to show the location of the surfaces. Lipid thicknesses were calculated from density profiles and checked against the Voronoi tessellation results. Velocity profiles were calculated using the MDAnalysis streamlines package and lipid order parameters were created using the lipyphilic python package in python and plots generated using matplotlib.[33-34]

4.2.4 Experimental Methods:

4.2.5 Materials and Supplies.

Glass slides and glass coverslips were purchased from Fisher Scientific (Pittsburgh, PA). Reagents were used without futher purification. Glycerol (>99%), sulfuric acid (H2SO4, 95.0-98.0%), hydrogen peroxide (H2O2, 30% aqueous solution), ammonium hydroxide (NH4OH, 28.0-30.0% aqueous solution), and toluene (99.8%) were purchased from Sigma-Aldrich (St. Louis, MO). Octadecyltrichlorosilan (OTS) were purchased from Gelest (Morrisville, PA). Ethanol (200 Proof pure ethanol) was purchased from Koptec(King of Prussia, PA). Milli-Q water (MQ water, 18.2 MΩ·cm at 25 °C) was produced by a Milli-Q water purification system (EMD Millipore, Billerica, MA). Nitrogen gas (99.999%) were purchased from Praxair, Inc. (Danbury, CT, King of Prussia, PA). 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(7-nitro-2-1,3-benzoxadiazol-4-yl) (NBD-PE) were purchased from Avanti Lipids, Inc.(Alabaster, AL).

4.2.6 Preparation of Glass Supports.

Three different glass supports were prepared to match the simulation. Coverslips were first cleaned with ethanol and water, then plasma cleaned for five minutes with a plasma cleaner (). To prepare the OTS/SAM glass coverslips, the glass substrates were first cleaned following established protocols (DOI: 10.1021/acs.jpclett.8b02442). Briefly, the coverslips were cleaned in piranha solution for 1h. Piranha solution was prepared by mixing H_2SO_4 and H_2O_2 (v/v = 3:1). Then, the coverslips were rinsed with copious amount of MQ water. The coverslips were then soaked in a basic bath which contains NH_4OH, H_2O_2, and MQ water at a ratio of 5:1:1 (v/v) for 1h at 70℃. The clean glass coverslips were then rinsed with copious amount of MQ water and dried in nitrogen gas. The OTS/SAM coverslips were prepared by first immersing the clean coverslips in 5mM OTS solution in toluene for 3 min. Then the modified coverslips were rinsed with toluene and ethanol, and dried in nitrogen gas.

4.2.7 Fabrication of Supported Lipid Constructs.

The delivery process was carried out by using an Atomic Force Microscope (AFM) based microfluidic delivery platform FluidFM BOT (Cytosurge, Glattbrugg, Switzerland) containing an inverted optical microscope (IX-73, Olympus America, Center Valley, PA). The printing was performed using a FluidFM Nanopipette (Cytosurge, Glattbrugg, Switzerland) with a 300nm opening. To prepare the lipid solution, POPC was first dissolved in chloroform and lyophilized with 1% NBD-PE to achieve the POPC/NBD-PE mixture. The lipid mixture was then dissolved in ethanol-glycerol solvent mixture (ethanol:glycerol = 9:1) at 0.033M.

4.2.8 Characterization of Supported Lipid Constructs.

The POPC constructs were left to air dry and imaged on an AFM (MFP-3D, Oxford Instrument, Santa Barbara, CA). Silicon nitride probes (MSNL-10 E, Bruker, Camarillo, CA) were used to

characterize the geometry and size of the printed structures. Image acquisition was done using tapping mode with 60% damping. Imaging processing and display were performed using the MFP-3D software developed on the Igor Pro 6.20 platform.

4.2.9 Atomic Force Microscopy Force Measurements.

The force measurements were carried out using the MSNL-10 E probe with a typical spring constant k = 0.1 N/m. The force versus distance plots were acquired by allowing the probe to approach the lipid constructs from above the surface at a constant velocity (100nm/s) until the glass surface was encountered. During this process, the force versus distance plots were recorded and the vertical force applied to the lipid assembles was used as a direct measurement of the lateral interactions between lipid molecules (DOI: 10.1002/smll.201700147). The spring constant calibration of the AFM probe was done on the same substrate as the lipid constructs using the calibration method based on the thermal fluctuation of the cantilever (doi.org/10.1016/0956-5663(95)99227-C). All experiments were performed at 24 °C in a temperature-controlled room with stability of ± 1 °C. Force versus distance plots were displayed and analyzed using the MFP-3D software developed on the Igor Pro 6.20 platform.

4.3. Results and Discussion:

Previous efforts in studying lipid surface behavior as a function of concentration have focused on hydrophilic surfaces and used a narrow-fixed range of concentrations to try and extract trends from. To the best of our knowledge our approach is the first of its kind to sample a continuous range of concentrations during the surface assembly process. Our developed methodology and

simulations allow direct insight into the mechanisms by which experimental POPC assembles onto different surface functionalities.

4.3.1 Lipid assembly onto polar and non-polar surfaces during ethanol evaporation

Figure 3 presents the results of solvent evaporation on polar and non-polar surfaces without glycerol. The data shown is from the randomly inserted starting configuration, with 3a-e corresponding to the very polar P4 surface at different solvation levels, and 3f-j corresponding to the non-polar C1 surface at different ethanol solvation levels. Ethanol is removed from the visualization for clarity. The initial frames shown in 3a and 3f were the result of 210 ns of simulation after initial random lipid insertion. The different surface types result in different modes of attachment, resulting in different looking starting structures. The polar surface has a bilayer formed initially due to strong polar head group attraction to the polar terminated surface, the exposed tails then pack together as a typical membrane system would. The non-polar surface gets covered in a layer of lipid molecules that attach by their tails, resulting in pockets of head groups and structures that stretch perpendicular to the surface (right side of the image) in order to minimize solvent interactions with exposed lipid tails. Both systems show a buffer layer of ethanol, visualized as empty space, between the resulting lipid structures. From 75% remaining initial ethanol to 20% initial remaining ethanol shown in Figure 3B-D, the lipid features away from the surface for the very polar surface have begun to assemble into individual lipid bilayers that are gradually forced together during the dehydration process. At 20% initial solvation the lipid bilayers from the surface appear to be very uniform, with a thin layer of ethanol between them. The relative ratio of lipids to ethanol is 1 POPC: 2.81 ethanol beads at this point. From 20% to 0% of initial solvation the thin layer of ethanol is removed, and the layers begin to interpenetrate and form less ordered subdomains in the bulk. Similarly, the non-polar C1 surface

shown in Figure 3f-j shows a compacting of the initially formed lipid features into more uniform lipid bilayers with the most ordered layers in 3i, and interpenetration and mixing at 0% initial solvation Figure 3j. The non-polar surface takes longer to form uniform layers, possibly due to much less ordered initial configuration due to tail-surface interaction preference or the resulting lower surface tension. Movies showing the full evaporation and surface attachment process for every system are available in the SI.

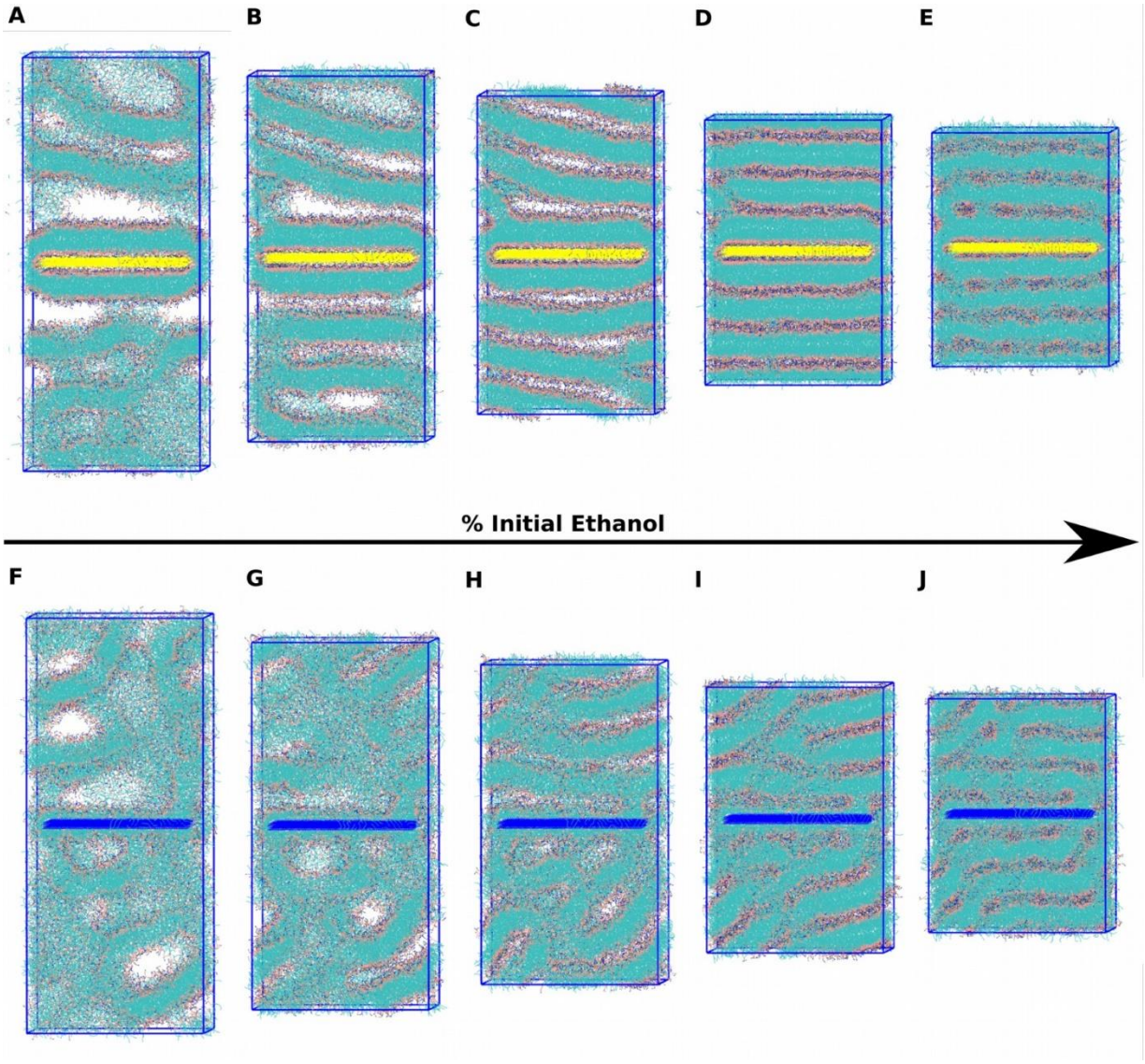


Figure 3: Molecular dynamics snapshots during solvent evaporation on polar and non-polar surfaces without glycerol, randomly inserted configuration. a) 100% solvent P4 surface. b) 75% solvent P4 surface. c) 50% solvent P4 surface. d) 20% solvent P4 surface. e) 0% solvent P4

surface. f) 100% solvent C1 surface. g) 75% solvent C1 surface. h) 50% solvent C1 surface. I) 20% solvent C1 surface. j) 0% solvent C1 surface.

While the random starting configurations are most comparable to our experimental deposition process, the stacked membrane bilayers are more comparable to Langmuir Blodget or Langmuir-Shafer deposition methods for SLBs. When analyzing the stacked membrane bilayers starting configurations shown in Figure S_ we find that the process of lipid adsorption to a surface during the dehydration process is comparable to those observed in existing simulations under constant hydration conditions. This similarity is clearly seen by the presence of the lubrication effect, or a hydration layer forming between the surface and the lipids, the gradual sliding effect of the lipids onto the surface, and the inversion of the lipid membrane on the non-polar surface.

4.3.2 Effect of glycerol on lipid assembly during evaporation

Experimentally, glycerol is known to stabilize and preserve lipid solutions.[35] In our experiments glycerol is added to slow down the evaporation rates to improve control of the liquid environment being deposited. To investigate the effect this glycerol addition plays on the formation of these lipid structures we added 10000 glycerol molecules to the simulations shown in Figure 3. Snapshots are shown in Figure 4 at the same ethanol solvation levels shown previously. Glycerol is not removed during simulated evaporation due to its much lower volatility than ethanol, and to study the effect of remaining glycerol on the final structures. Figure 4a-e corresponds to the highly polar P4 surface, and Figure 4f-j to the non-polar C1 surface. Ethanol is once again removed from the visualization for clarity. The first half of ethanol removal is very similar for the strongly polar surface when compared to the system without glycerol (cf. Figure 3a-c). The final 50% of ethanol removal shows deviation in the formation of final structures. There is clearly more separation between the layers attached directly to the surface and the bulk in Figure 4d and 4e. Additionally, glycerol shows a stabilizing effect by

adding a buffer layer between POPC layers resulting in more uniform bilayer stacks beginning in Figure 4c and continuing through 4e. Bilayer dissipation due to increased lipid interaction is not present close to the surface as a result of this effect. Similar trends are observed for the non-polar C1 surface. Initially the surface shows lipid tail surface interactions with pockets of headgroups, and a less ordered structure above the surface buffered by solvent. At 75% of initial ethanol the glycerol is beginning to stabilize the lipid structures away from the surface with clear separation beginning to form between layers. At 50% initial ethanol solvation the more ordered layers from the solution have begun to form layers diagonally from the surface, likely between the pockets of headgroups to minimize tail-solvent exposure. At 20% ethanol solvation the glycerol has begun to separate the layers above the surface in addition to the layer closest to the surface. Finally at 0% ethanol solvation the glycerol has aided the formation of bilayers across the entire system and prevented interpenetration, with glycerol isolating layers and the pockets of headgroups attached to the surface. The movies for these systems are also available in the SI.

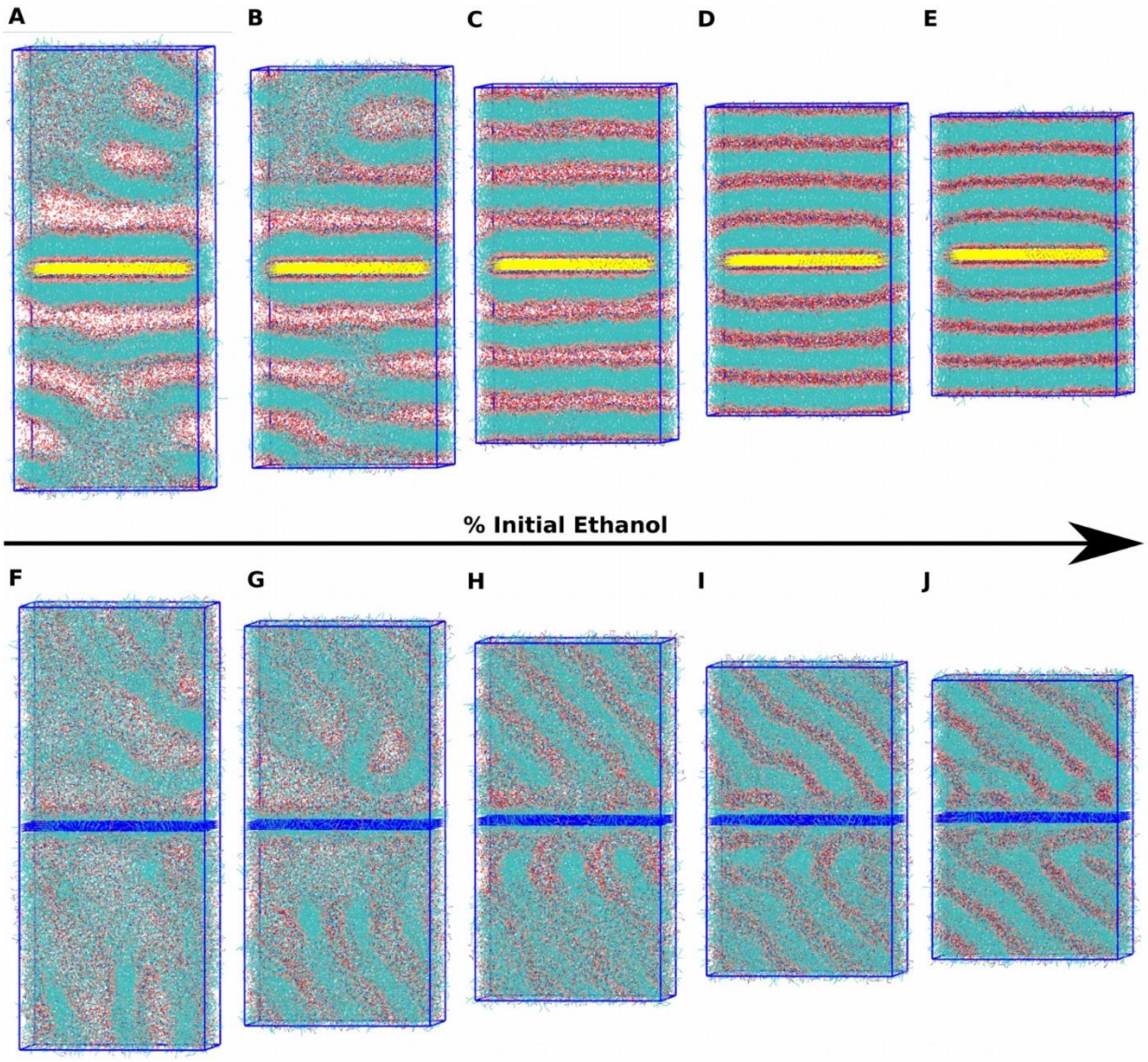


Figure 4: Molecular dynamics snapshots for solvent evaporation on polar and non-polar surfaces with glycerol, 6 bilayers starting configuration. a) 100% solvent P4 surface. b) 75% solvent P4 surface. c) 50% solvent P4 surface. d) 20% solvent P4 surface. e) 0% solvent P4 surface. f) 100% solvent C1 surface. g) 75% solvent C1 surface. h) 50% solvent C1 surface. I) 20% solvent C1 surface. j) 0% solvent C1 surface.

The flexibility of our methodology also enabled us to study the effect that glycerol plays on this assembly process. Glycerol clearly provides a larger buffer layer compared to the ethanol layer. This larger layer is observed at both the surface and the layers between the lipid layers. These glycerol buffer layers are clearly visible in Figure 4. For the systems that started as stacked bilayers the penetration depth of the glycerol layers extends several layers but not to the final

layer Figure S_. This effect is comparable to the effects observed with the introduction of cryosolvents on lipid structures, whereby a crosslinked network forms on the surface of the headgroups. Our simulations show that these crosslinked networks isolate and stabilize individual bilayers in dehydrated structures.

4.3.3 Quantification of resulting lipid structures

In order to quantify these observations lipid density profiles were calculated and shown in Figure 5. Figure 5a-b corresponds to the systems without glycerol and random start at 20% and 0% solvation respectively and Figure 5c-d to the system with glycerol and random start at 20% solvation and 0% solvation respectively. Grey lines correspond to the density profile of the stacked membrane away from the surface at full solvation without and with glycerol as a control. Black lines correspond to the very polar P4 surface, blue lines to the non-polar C1 surface. Figure 5a shows that lipid stacks formed on the surface closely align with those freely floating in solution. The C1 surface shows maximum peaks with an offset from the center, due to attachment of tails to the surface. The maxima of these peaks appear to roughly correspond to similar thickness as the control group, with a similar offset for each peak. The non-polar peaks, however, are broader, suggesting the presence of lipid head groups in locations not corresponding to perfect bilayer stacks. Comparing against the visualization in Figure 3i, these likely correspond to the bilayer stacks not forming at an angle to the surface rather than parallel. The fully dry structures shown in Figure 5b demonstrate a thinning of the membrane compared to the control profile. The peaks for both surfaces become more uniform, with the same tail-surface related offset for the C1 surface, and slightly broader peaks for the C1 surface related to the slight angle of the structure. The effect of the addition of glycerol is shown in Figure 5c-d. For the polar surface, the overall trends are similar, almost matching the control in 5c, and

thinning in 5d. There appears to be a general broadening of the peaks for all systems in the presence of glycerol, and a slight offset from the control group in 5c for the very polar surface. The assembled structures on the C1 surface show a dramatically different profile in the presence of glycerol. In both 5c and 5d there are two distinct peaks on either side of the surface at 0, with the same offset corresponding to tail-surface attachment. Beyond the two peaks however there is an even distribution of phosphate beads away from the surface, making it difficult to distinguish between membrane bilayers in the z-direction. Comparison to the visualization in Figure 4i-j suggests this is due to the angle the membrane bilayers have formed onto the surface. Density profiles for the starting bilayer stack and other configurations that are not shown here are available in the supplementary information figure S__.

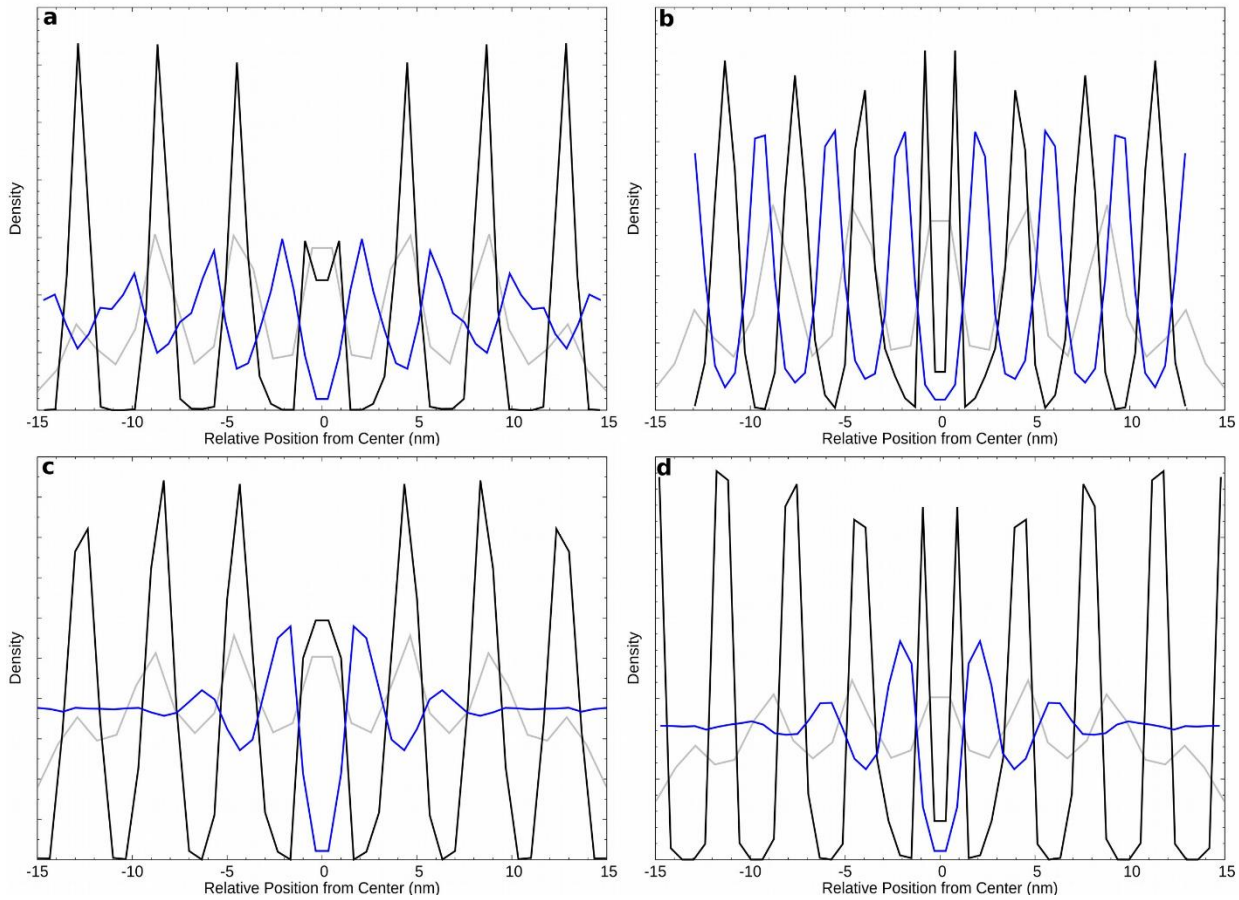


Figure 5: Lipid density profiles normalized in the z direction and centered around the center of mass of Phosphate head group beads. a) No glycerol 20% of initial ethanol from random start. b) No glycerol 0% of initial ethanol from random. c) 20% of initial ethanol with glycerol from random start. d) 0% of initial ethanol with glycerol from random start. Grey line corresponds to control bilayer stack with or without glycerol, black lines to very polar P4 surface, blue lines to non-polar C1 surface.

Thickness values from Figure 5 were calculated and shown in Table 2 along with lateral diffusion coefficients for these states. Table 2 clearly shows that the membranes at 20% align with the initial away from the surface membrane thickness and lateral diffusion behavior. As these structures are further dehydrated the membranes break up resulting in interdigitation as shown previously, this results in a decreased membrane head-head thickness. Interestingly this effect is shown to also occur in the presence of glycerol. Lateral diffusion slows down with dehydration, as the membranes become more locked in and less fluid. The presence of glycerol

increases the fluidity of these membranes resulting in a less dramatic slowdown in lateral diffusion.

Table 2: Thickness and lateral diffusion coefficients from random starting configurations with and without glycerol

Parameter	Initial	C1		C1 Glycerol		P4		P4 Glycerol	
		20% EtOH	0% EtOH	20% EtOH	0% EtOH	20% EtOH	0% EtOH	20% EtOH	0% EtOH
Thickness (nm)	4.14	4.20	3.69	4.65	3.63	4.18	3.70	4.07	3.47
Lateral D (1e-5 cm/s^2)	0.037	0.034	0.013	0.034	0.211	0.037	0.013	0.049	0.031

4.3.4 Mechanisms associated with lipid assembly during dehydration

To study the mechanistic behavior of POPC assembly during simulated evaporation the potential energy, surface tension, and lipid order parameter was evaluated. The potential energies as a function of percentage remaining ethanol are shown in Figure 6. Figure 6A and 6B highlight the differences due to starting configuration for simulations without and with glycerol respectively. Figure 6C and 6D show the effect that glycerol plays for each starting configuration, bilayer stack and random starting configurations respectively. Only the most polar and non-polar surfaces are shown. The polar P1 surface details are available as SI figure S_. For starting configuration comparisons in 6A-B, the systems that start as 6 stacked bilayers have lower initial potential energy before the final structures converge in the final 20% of ethanol solvation. This is expected as the membrane bilayers will be in a more energetically favorable state compared to

the distributed mixture. The convergence of all systems suggests that the starting configuration affects the mechanism associated with surface attachment, but not the energetic favorability of the final structure when all ethanol is removed. At slightly higher ethanol hydration there remains a small difference in potential energy caused by starting configuration, likely related to the distribution of solvent in the final structures.

Evaluation of the effect glycerol has on lipid assembly during evaporation is shown in Figures 6C-D. These plots show clearly that glycerol reduced the potential energy values when compared to their counterparts without, again supporting the notion that glycerol stabilizes the lipid assemblies. In Figure 6C, the dark blue dots corresponding to the C1 surface with the bilayer stack starting configuration show a slight dip around 82% initial remaining ethanol. For non-polar surfaces it is known that lipid micelles and membranes go through a flip when attaching to the surface. Visual inspection from evaporation movies shown in S_ reveal that this dip likely corresponds to this flip and surface attachment, resulting in a slight decrease in potential energy. The presence of glycerol appears to mask the changes in potential energy during the evaporation process, with no visible dips, and what appears to be a constantly increasing slope compared to the changing plots without glycerol. For all systems the non-polar C1 surface has higher potential energy values as compared to the highly polar surface, demonstrating a less energetic favorability.

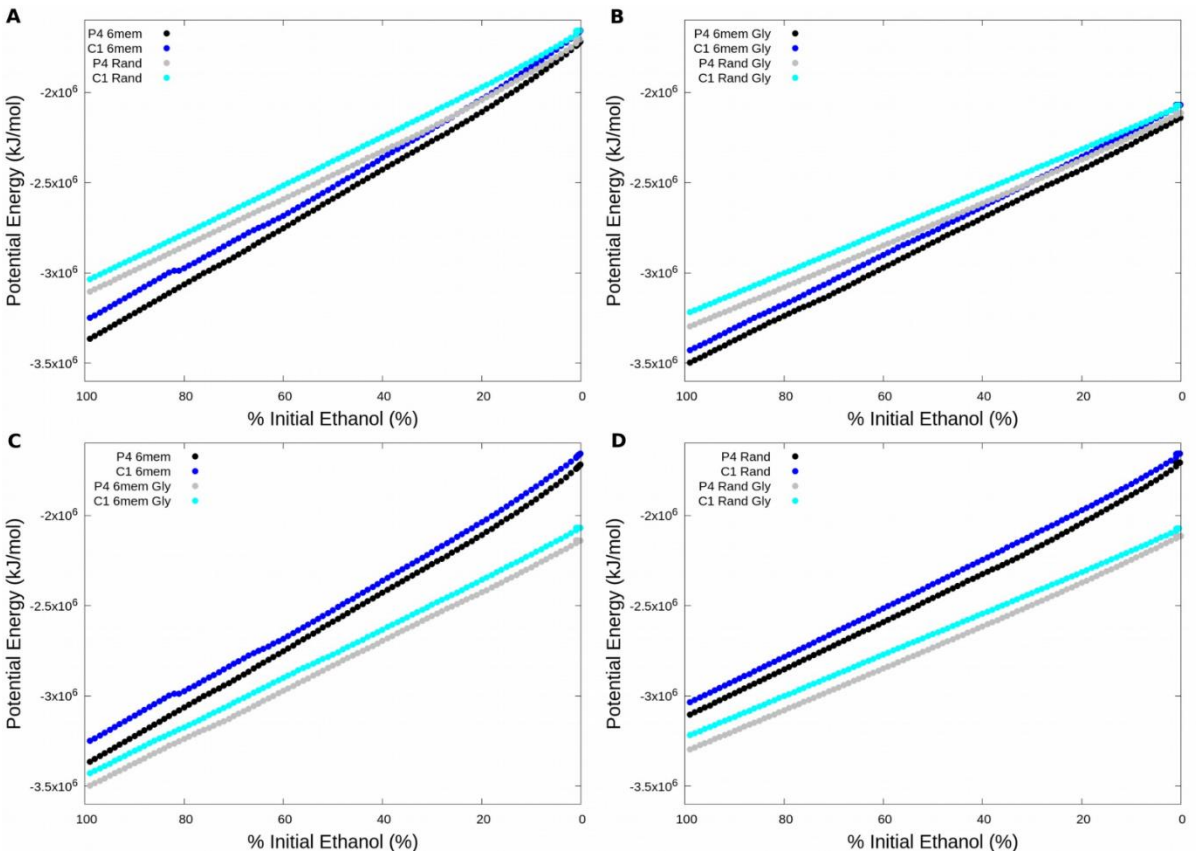


Figure 6: Potential Energies as a function of % of remaining initial ethanol. A) Bilayer stack vs random without glycerol. B) Bilayer stack vs random with glycerol. C) Bilayer stack with and without glycerol. D) Random configurations with and without glycerol.

To further investigate the contributions of starting configuration and glycerol on evaporation mechanism the surface tensions as a function of ethanol concentrations are shown in Figure 7. Figures 7A-D correspond to the same comparisons as in Figure 6, starting configurations without, and with glycerol, and differences caused by the glycerol for the membrane and random starting configurations respectively. Beginning with the starting configuration comparison, clearly the polar surfaces have a higher surface tension for all states but the stacked bilayer starting configuration with glycerol. This is likely due to the differences in starting location of the glycerol. The random starting configuration was more evenly distributed, while the 6 bilayers trapped more glycerol at the surface and was not able to diffuse and distribute evenly across the

structure, resulting in differences in surface tension as you move away from the surface. The presence of glycerol seemingly results in more structured membranes in the random start when the glycerol is more evenly dispersed in the solution. The non-polar surfaces all converged with low surface tension for all configurations and solutions. Non-polar stacked bilayer without glycerol shows an increase in surface tension beginning around 82% remaining initial ethanol, providing further evidence the transition was related to the initial process of lipid inversion onto the surface, eventually transitioning back to the final values for the random configuration.

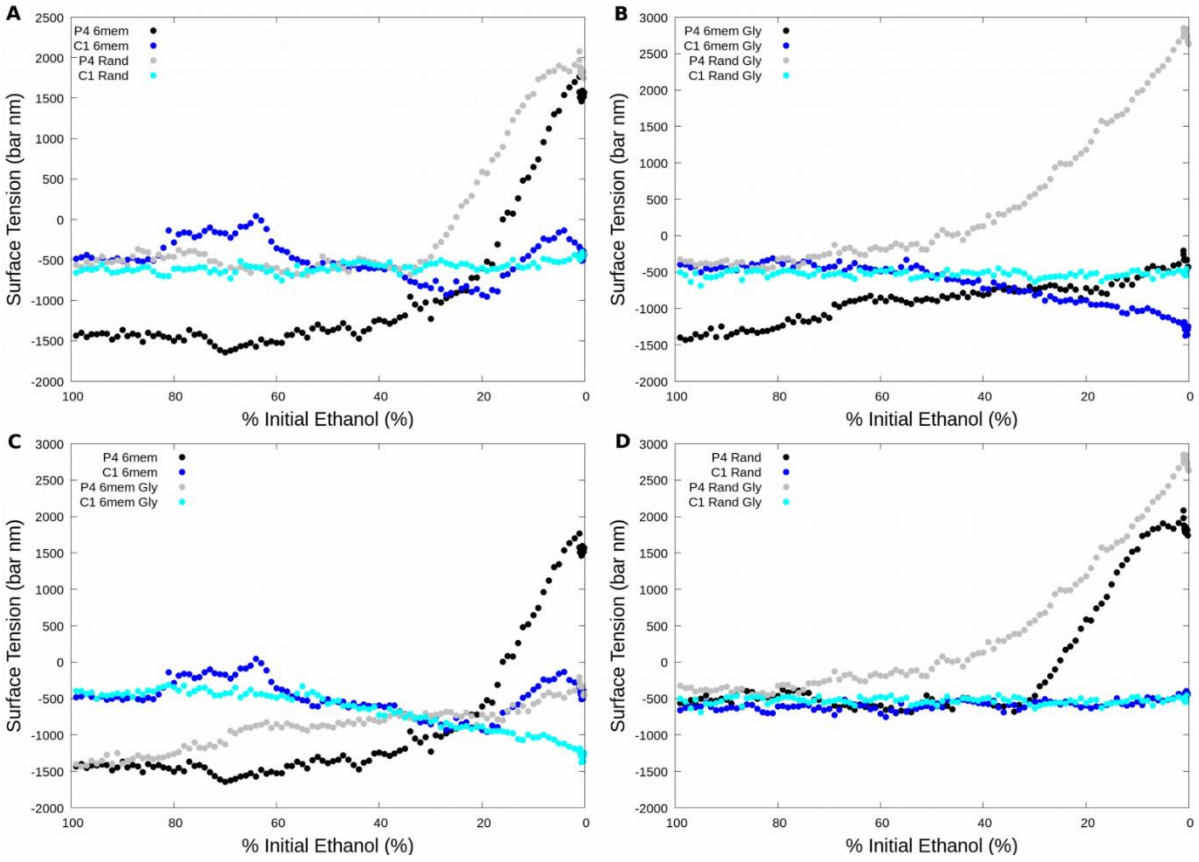


Figure 7: Surface tension as a function of % of remaining initial ethanol. A) Bilayer stack vs random without glycerol. B) Bilayer stack vs random with glycerol. C) Bilayer stack with and without glycerol. D) Random configurations with and without glycerol.

Finally, the coarse-grained lipid tail order parameter was evaluated for the same systems and is shown in Figure 8. The coarse-grained lipid order parameter is calculated as shown in equation 1.

$$Scc = \frac{< 3\cos^2\theta >}{2} \quad (1)$$

The systems that start as bilayers show a high tail order due to having an ordered bilayer starting structure. This structure breaks up during evaporation resulting in a drop of lipid order between 90% and 40% solvation. From 40% solvation the structure gains order until peaking around 20% solvation. After 20% solvation the structure starts to disintegrate as the tails can interact with the layers that were separated by ethanol. The randomly generated structures show a linear increase in tail order parameter peaking around 35% and then decreasing without the presence of glycerol and plateauing in its presence. The polar surfaces in the random starting configurations show a higher tail order parameter compared to the nonpolar surface due to the presence of an ordered bilayer at the surface due to the strong head group interactions. This suggests the presence of ethanol stabilizes the structures that are present and prevents lipid interdigitation. The addition of glycerol appears to stabilize the layers that are formed on the surface as or more strongly than the thin ethanol layer remaining at 20% for all configurations.

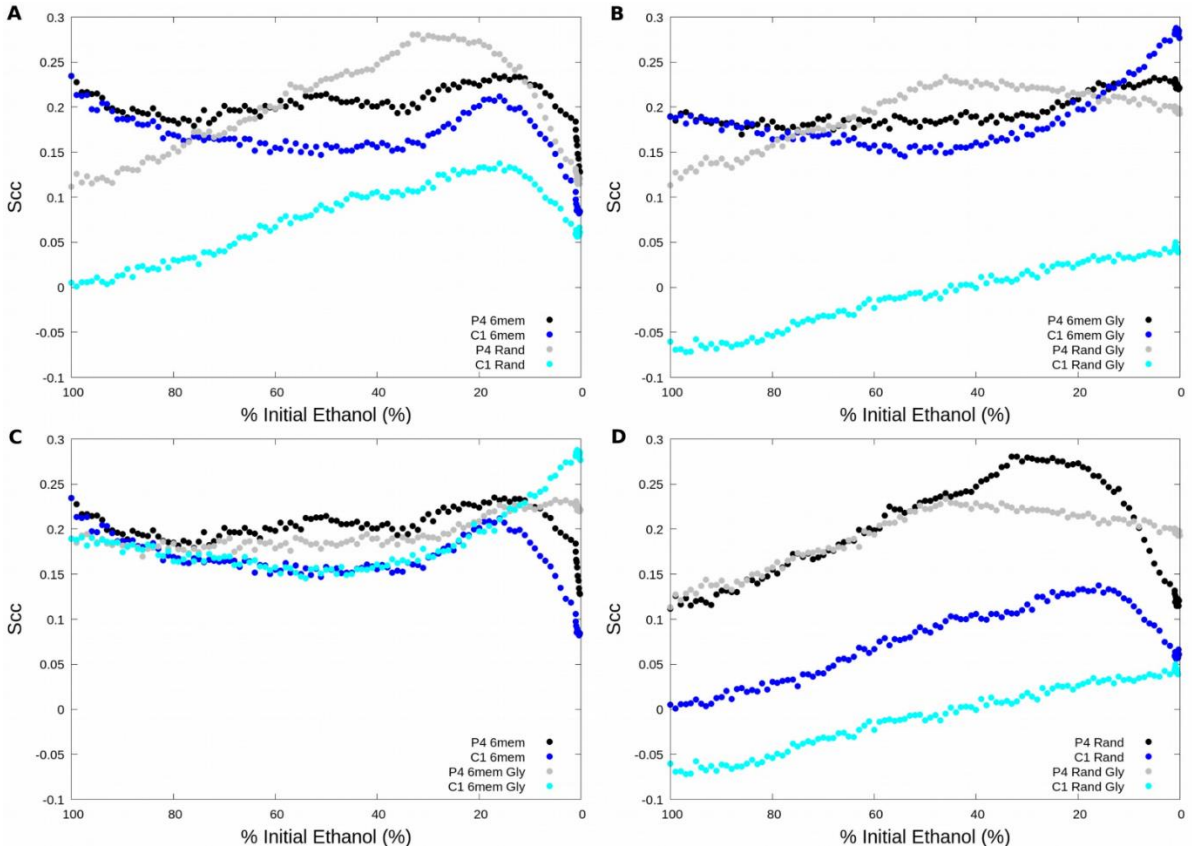


Figure 8: Coarse-grained lipid tail order parameter as a function of % of remaining initial ethanol. A) Bilayer stack vs random without glycerol. B) Bilayer stack vs random with glycerol. C) Bilayer stack with and without glycerol. D) Random configurations with and without glycerol.

Lipid assembly from the random starting configuration on C1 surface show lower tail order parameter during the entire process with and without glycerol. Correlating this with the visualizations in Figures 4f-j and 3f-j suggests the difference in the final states as being related to the off parallel angle the structures eventually form at. At full solvation the system is expected to be less ordered given the lack of a bilayer formed on the surface like is formed on the polar surface. Clearly, the drying process causes the lipids above the surface to be less ordered for the C1 surface compared to the P4 surface. This suggests that the local structure at the surface affects the assembly and lipid behavior of the surrounding region.

4.3.5 Effects of surface termination on assembled lipid structure

Due to curvature and breakup and reformation of lipid membranes area per lipid calculations through standard approaches were difficult. To enable the calculation of area per lipid Voronoi tessellations were performed on the 1 bilayer starting structures for each surface at 100%, 20%, and 0% solvation (cf. Figure 9). Figures 9a-c correspond to the P4 surface, 9e-f to P1 surface, and 9g-I to the C1 surface. The starting structures at 100% solvation all look similar and show a band of higher area per lipid due to starting curvature in the lipid membrane as seen in Figure 2a. This membrane then gradually attaches to the surface and forms pockets that resemble liposomes. The remnant of one of these structures can be seen in 9e in the center of the surface. The polar surfaces behave very similarly, though at 20% solvation the presence of this ring has dissipated for the strongly polar surface suggesting that the interaction strength affects the speed at which the lipids adhere to the surface. At 0% solvation the polar surfaces show multiple phases, a closely packed structure in the edges and gaps between the surface, and the layer formed on the top of the structure. This contrasts with the behavior of the non-polar C1 surface that adheres to the surface through tail groups, forming pockets of head groups with extremely high area per lipid as shown in Figure 9h. At 0% ethanol solvation the non-polar surface exhibits a very different local phase as compared to the polar surfaces. This suggests that the head groups packing to the surface result in a highly predictable layer with tight packing in the edges, while the non-polar surface results in islands of head groups with some having high area available and some with low area available. Interestingly this pattern does not follow the surface boundary as closely as in the case of the polar surfaces.

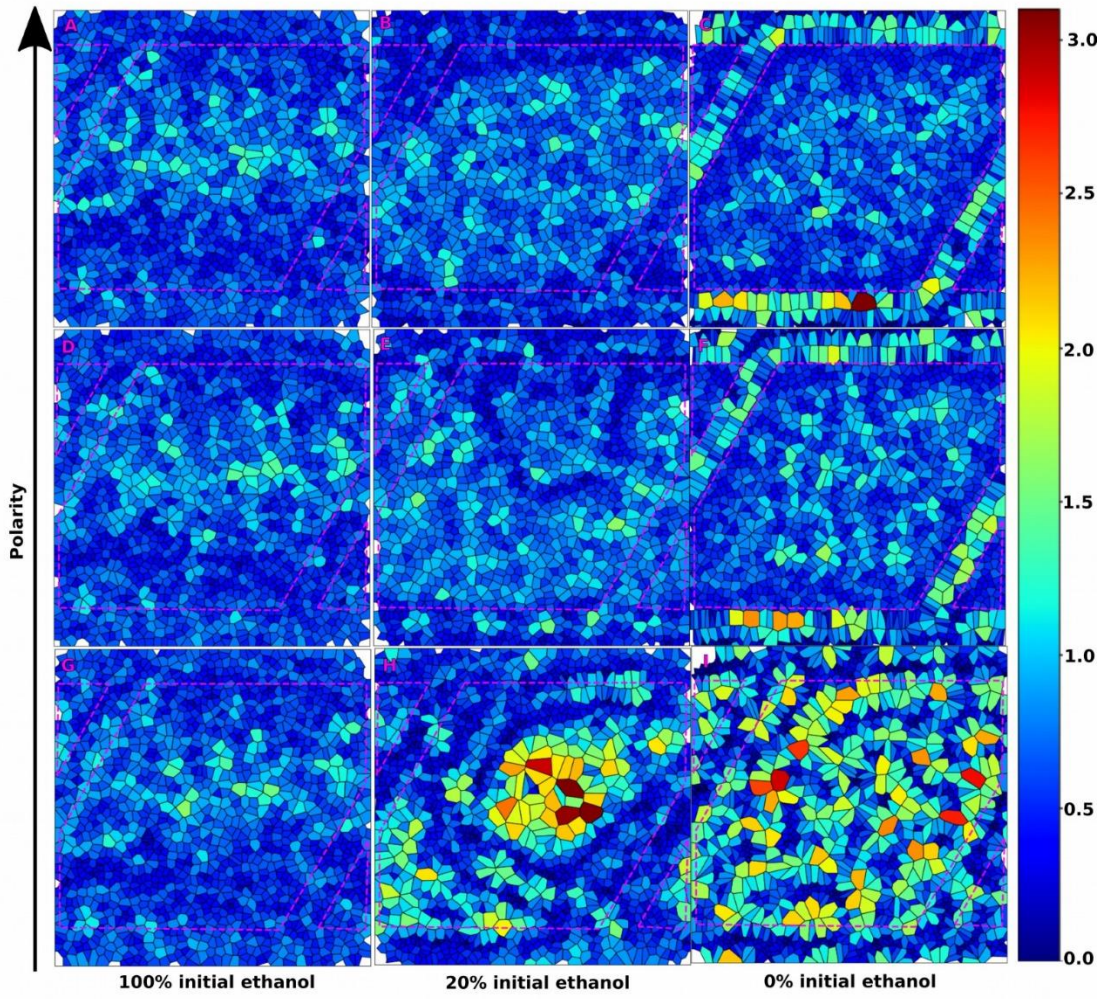


Figure 9: Voronoi tessellations of area per lipid. Cells correspond to PO4 beads in lipid head group normalized to a 2d plane, generated from 1 bilayer starting configuration. a-c) P4 surface at 100%, 20%, and 0% ethanol solvation respectively. d-f) P1 surface at 100%, 20%, and 0% ethanol solvation respectively. g-I) C1 surface at 100%, 20% and 0% ethanol solvation respectively

These Voronoi tessellations were also used to calculate and visualize the thickness of the layer directly above the surface. To inspect the mobility of these lipids the streamlines function in MDAnalysis was used to calculate the velocities of the lipid headgroups along the surface.[33] The results of these calculations are shown in Figure 10. Figure 10a and d correspond to the non-polar C1 surface, 10b and e to polar P1 surface, and 10c and f to strongly polar P4 surface. The homogeneity of the polar surfaces is clearly demonstrated by the thickness profiles in 10b and

10c, while the heterogeneity of the non-polar surface in 10a due to the distinct lipid domains is clear. Interestingly the mobility of the polar head groups is much higher for the polar surfaces in 10e and 10f compared to the non-polar surface in 10d. This is manifested in the number of streamlines as well as the sustained high velocity levels. The direction of the headgroup motion associated with the streamlines suggests mobility across the entire leaflet on the surface, while the headgroups on the non-polar surface get confined to locally ordered subdomains.

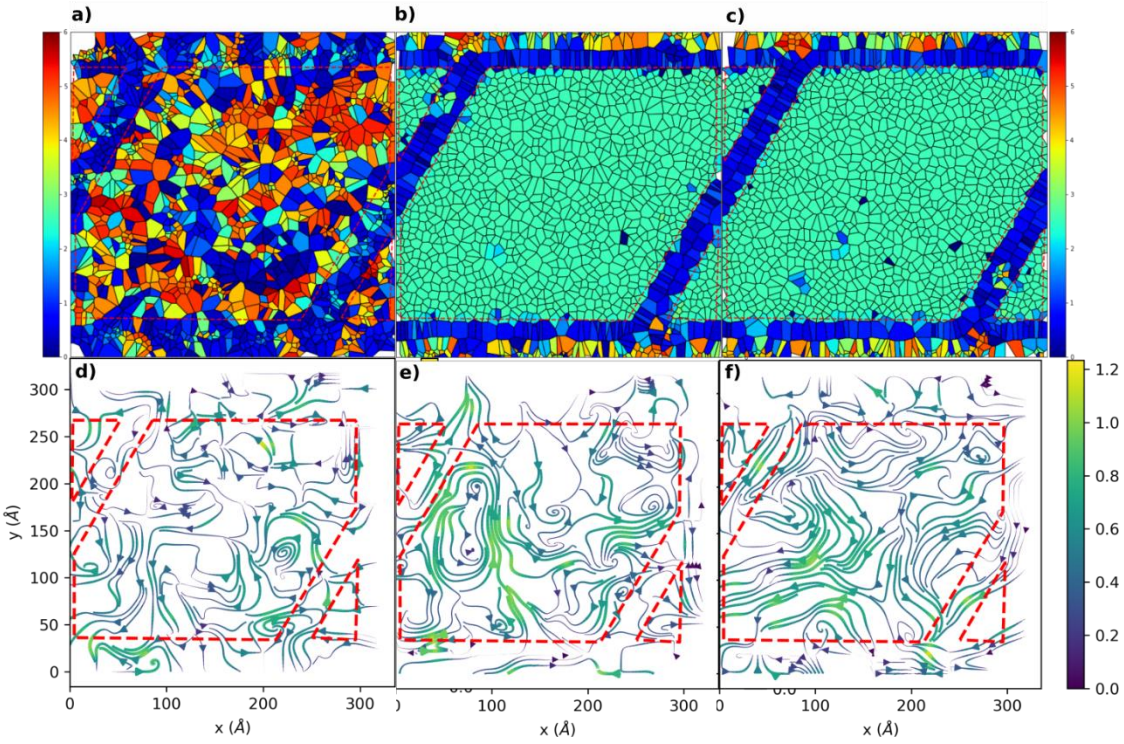


Figure 10: Lipid thickness and surface velocity plots. Lipid thickness applied to Voronoi tessellation plots. Velocity streamlines calculated at the surface for the final 15ns of simulation time. All structures at 0% ethanol solvation from 1 bilayer starting configuration. a) C1 surface thickness. b) P1 surface thickness c) P4 surface thickness. d) C1 velocity profile. e) P1 velocity profile. f) P4 velocity profile

Surface functionality clearly plays a role in the resulting structures from this dehydration process. Area per lipid and Voronoi tessellation analysis reveals that the surface type affects the homogeneity of the layers that are attached to the surface. Polar surfaces result in an attachment

of the head groups to the surface, with stronger polarity resulting in a stronger attachment. The non-polar surface type results in a more distributed attachment of the tails, which forms islands of the head groups that appear less structured on the Voronoi tessellations. Velocity analysis of these structures reveals a difference in the dynamics of these head groups that are attached to the surface. On a polar surface the lipids are still mobile, and capable of moving around on the surface with velocity pattern that traverses the entire surface. The non-polar surface results in much less surface velocity, with motion confined, likely due to the stronger binding of the tail groups to the surface.

4.3.6 Experimentally assembled POPC structures appear to be composed of stacked lipid bilayers

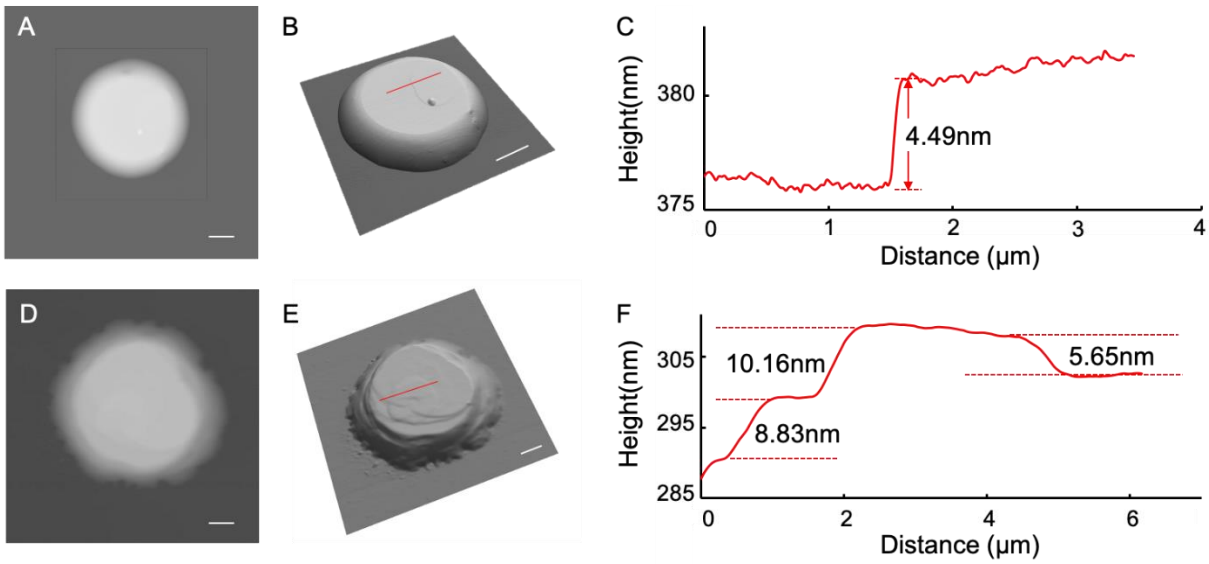


Figure 11: AFM topographic images exhibit step-by-step feature and the measured step thickness matched the POPC bilayer thickness. A) AFM topographic image of a POPC assembly. POPC/NBD-PE was dissolved in ethanol-glycerol mixture (ethanol: glycerol = 9:1) at 0.033 M and delivered to plasma cleaned glass. Scale bar = 2 μm. B) 3D display of (A). Scale bar = 2 μm. C) Height profile from the cursor indicated in (B). D) AFM topographic image of a POPC construct exhibiting the layer-by-layer geometry. POPC/NBD-PE was dissolved in ethanol-glycerol mixture (ethanol: glycerol = 9:1) at 0.033 M and delivered to octadecyltrichlorosilane (OTS) SAM/glass. Scale bar = 2 μm. E) 3D display of (D). Scale bar = 2 μm. F) Height profile from the cursor indicated in (D).

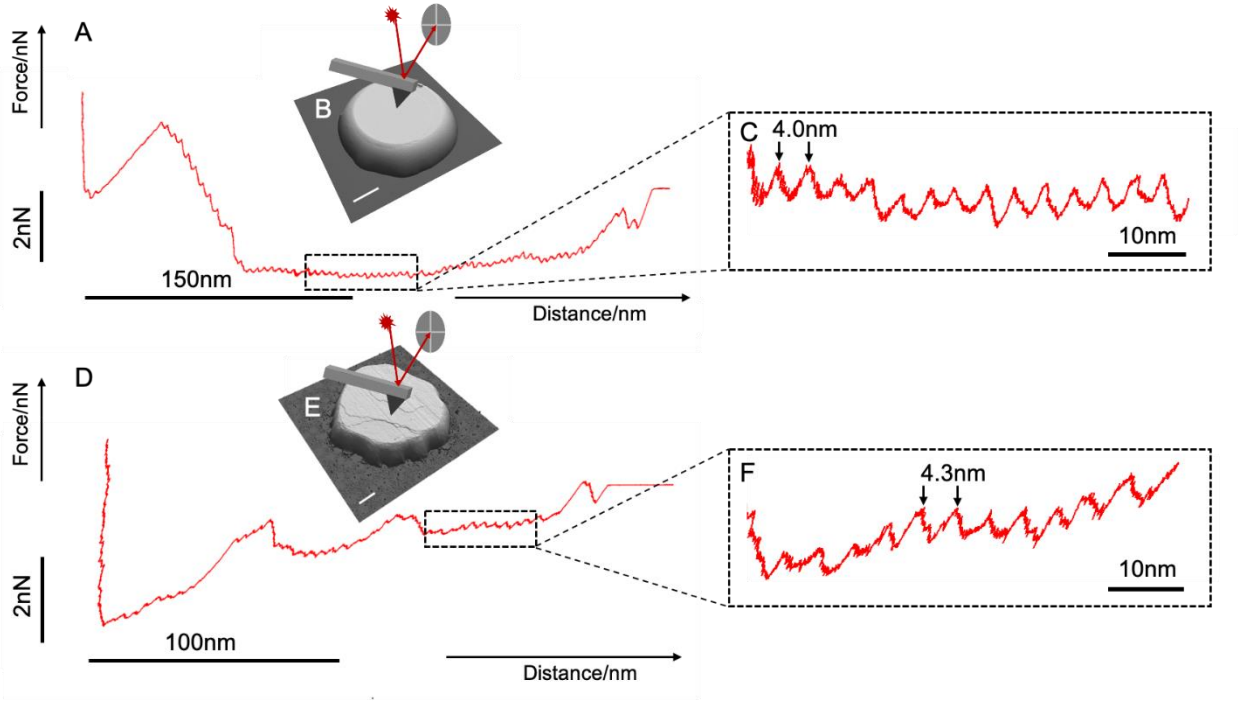


Figure 12: AFM force indentation experiment exhibit the saw-tooth pattern which corresponds to the rupture event of POPC bilayers. A) The force versus distance plot for a POPC assembly on plasma-cleaned glass. POPC/NBD-PE was dissolved in ethanol-glycerol mixture (ethanol: glycerol = 9:1) at 0.033 M and delivered to plasma cleaned glass. B) A 3D display of the POPC assembly on plasma-cleaned glass. Scale bar = 2 μm. C) Zoom-in image of the force versus distance plot shows bottom-to-bottom distance = 4.0 nm. D) The force versus distance plot for a POPC assembly on OTS/SAM glass. POPC/NBD-PE was dissolved in ethanol-glycerol mixture (ethanol: glycerol = 9:1) at 0.033 M and delivered to octadecyltrichlorosilane (OTS) SAM/glass. E) A 3D display of the POPC assembly on OTS/SAM glass. Scale bar = 2 μm. F) Zoom-in image of the force versus distance plot shows bottom-to-bottom distance = 4.3 nm.

Assembly from ultrasmall solution droplets follows a different dynamic from that of larger scales.[10] In the experimental section, subfemtoliter aqueous droplets containing POPC molecules were delivered onto glass substrate with different surface modification using FluidFM to

compare with the simulation results. AFM images were taken to provide in situ topographic information of the POPC constructs.

The POPC assembly on plasma-cleaned glass exhibited the “gumdrop” geometry with a base diameter of 9.5 μm and a height of 382 nm. Topographic image (Figure 11A) showed the flat smooth top and a circular base. Terraced structures were observed from the 3D display and the height profiles showed a layer thickness of 4.49 nm, which matches the thickness of a POPC bilayer (Figure 11B-C). The POPC assembly on OTS/SAM showed a similar geometry with more terraced structures on the top (Figure 11D-E). Lipid tendrils were present at the edge of the POPC assembly around the bottom of the base, which suggests that the OTS/SAM surface leads to tail group bonding and lower velocity, either disrupting line tension or leading to nucleation sites as these lipid islands form. The height profile (Figure 11F) showed step-by-step feature with the step thickness measured to be 5.65 nm, 8.83 nm, and 10.16 nm, which matches the thickness of POPC bilayer or integer multiples of POPC bilayers thickness. The overall POPC assembly on OTS/SAM glass has a height of 295 nm and a base diameter of 14 μm. The flatter and more spread-out geometry of the POPC assembly on the OTS/SAM glass can also be due to the lower fluidity of the tail group bonding on the non-polar surface.

To get a better understanding of the lipid structure of the bulk of the POPC assembles, force indentation experiments were performed on lipid features on both surfaces. The force versus distance plots from puncturing both the POPC assemblies on plasma-cleaned and OTS/SAM glass exhibited the saw-tooth pattern, with each saw-tooth pattern corresponding to the rupture event of a POPC bilayer (Figure 12A, D). For the POPC assembly on the plasma-cleaned glass, each rupture event occurs at an average layer thickness = 3.96nm ± 0.32nm (n

=61). The POPC assembly on OTS/SAM glass showed an average layer thickness = 4.36nm ± 0.51nm (n =27).

4.4. Conclusions

A collection of molecular dynamics simulations utilizing the MARTINI CG model were used to study the interplay between POPC and hydrophobic and hydrophilic surfaces as a function of relative lipid-solvent concentration. To the best of our knowledge this work is the first to use MD to study the assembly of lipids during solvent evaporation, and the first to study the effect of surface termination at variable solvent concentrations. Additionally, these systems were evaluated with and without glycerol to investigate the role it plays in the assembly of lipid structures during the drying process. This developed method was used to gain insight into the mechanisms of lipid assembly onto surfaces and the corresponding structural transformation, as well as the mechanism by which glycerol stabilizes and changes those assembled structures. Surface functionality was shown to affect the local structure of lipid assemblies at the surface, with increasing homogeneity at the surface with polarity. The changed functionality at the surface was also shown to clearly affect the mobility of the layer directly interacting with the surface, this could potentially affect the final assembled structures through the creation of potential nuclear sites, or a disruption of line tension. The different terminations determined whether the interaction was head-surface or tail-surface, resulting in uniform parallel bilayers in the case of head-surface interactions, and angled bilayers on tail-surface favored surfaces. Despite the differences in surface attachment and packing angle, all structures still resulted in the formation of bilayers in the bulk. Glycerol was shown to strengthen the formation and stability of bilayers due to the crosslinked network formed between bilayers, especially as the systems were

dehydrated. This stabilizing effect during dehydration suggests glycerol could act as a sugar would in the water replacement hypothesis. On non-polar terminated surfaces the glycerol stabilized the angle the lipid structure formed at. The systematic nature of the modeled evaporation process allowed for the reconstruction of assembly during evaporation. For the stacked bilayer configuration that most closely resembles the Langmuir-Blodgett and Langmuir-Schafer processes, we observed previously shown effects such as lubrication and layer inversion for the initial attachment. The starting configuration of the simulations was shown to play a role in the entrapment of glycerol locally to the surface. This entrapment formed parallel bilayers in the case of non-polar surfaces, unlike those assembled through random starting configurations. These simulated structures were validated experimentally through the delivery of ultra-small droplets of ethanol and glycerol containing POPC. These experimentally generated structures showed the formation of bilayers in the bulk for all surfaces, and the presence of defects at the edge of non-polar surfaces, suggesting a competing effect with line tension or nucleation caused by tail group interactions. These experimentally generated structures showed the formation of bilayers in the bulk for both plasma-cleaned glass (polar surface) and OTS/SAM glass (non-polar surface), which is in agreement with the simulation result in Figure 4D and Figure 4I. The presence of defects at the edge of the POPC assembly on the non-polar surface suggests a competing effect with line tension or nucleation caused by tail group interactions.

4.5 References:

1. Goetz, R.; Gompper, G.; Lipowsky, R. Mobility and Elasticity of Self-Assembled Membranes. *Phys. Rev. Lett*. **1999**, 82, 221-224.
2. Brannigan, G.; Phillipds, P. F.; Brown, F. L. H. Flexible Lipid Bilayers in Implicit Solvent *Phys. Rev. E*. **2005**, 72, 011915
3. Marrink, S. J.; Mark, A. E Molecular dynamics Simulation of the Formation, Structure, and Dynamics of Small Phospholipid Vesicles. *J. Am. Chem. Soc.* **2003**, 125, 15233-15242
4. Schindler, T.; Kroner, D.; Steinhauser, M. O. On the dynamics of molecular self-assembly and the structural analysis of bilayer membranes using coarse-grained molecular dynamics simulations. *Biochim. Biophys. Acta, Biomembr.* **2016**, 1858, 1955-1963
5. Roberts, G. Langmuir-blodgett films. Springer Science and Business Media, **2013**
6. Liu, J.; Conboy, J. C.; Structure of a gel phase lipid bilayer prepared by the Langmuir-blodgett/Langmuir-Schaefer method characterized by sum-frequency vibrational spectroscopy *Langmuir*, **2005**, 21, 20, 9091-9097
7. Lind, T.K.; Cardenas, M.; Understanding the formation of supported lipid bilayers via vesicle fusion—a case that exemplifies the need for the complementary method approach. *Biointerphases*, **2016**, 11, 2, 010801
8. Jackman, J.; Cho, N.J.; Supported lipid bilayer formation: beyond vesicle fusion, *Langmuir*, **2020**, 36, 1387-1400
9. Zhang. J.; Yu, H.; Harris, B.S.; Zheng, Y.; Celik, U.; Na, L.; Faller, R.; Chen, X.; Haudenschild, D.R.; Liu, G.Y.; New means to control molecular assembly *J. Phys. Chem. C.* **2020**, 124, 11, 6405-6412
10. Zhang. J.; Piunova, V.A.; Liu, Y.; Tek, A.; Yang, Q.; Frommer, J.; Liu, G.Y.; Sly, J.; Controlled molecular assembly via dynamic confinement of solvent *J. Phys. Chem. Lett.* **2018**, 9, 21, 6232-6237
11. Saiz, L.; Bandyopadhyay, S.; Klein, M.; Towards an understanding of complex biological membranes from atomistic molecular dynamics simulations *Biosci Rep*, **2002**, 22, 2, 151-173
12. Ceccarelli, M.; Marchi, M.; Molecular dynamics simulation of POPC at low hydration near the liquid crystal phase transition. *Biochemie*, **1998**, 80, 5-6, 415-419
13. Kanduc, M.; Schneck, E.; Netz, R.R.; Hydration interaction between phospholipid membranes: insight into different measurement ensembles from atomistic molecular dynamics simulations. *Langmuir* **2013**, 29, 29, 9126-9137
14. Schneemilch M.; Quirke N.; Free energy of adsorption of supported lipid bilayers from molecular dynamics simulations. *Chem. Phys. Lett.* **2016**, 664, 1, 199-204
15. Mannelli, I.; Sagues, F.; Pruneri, V.; Reigada, R.; Lipid vesicle interaction with hydrophobic surfaces: a coarse-grained molecular dynamics study. *Langmuir*, **2016**, 32, 48, 12632-12640
16. Bennun, S.V.; hoopse, M.I.; Xing, C.; Faller, R.; Coarse-grained modeling of lipids. *Chemistry and physics of lipids,* **2009***,* 159, 2, 59-66

17. Poursoroush, A.; Sperotto, M.M. Laradji, M.; Phase behavior of supported lipid bilayers: A systematic study by coarse-grained molecular dynamics simulations. *The Journal of Chemical Physics,* **2017**, 146, 15, 154902
18. Benedetti, F.; Fu, L.; Thalmann, F.; Charitat, T.; Rubin, A.; Loison, C.; Coarse-grain simulations of solid supported lipid bilayers with varying hydration levels. *J. Phys. Chem. B.* **2020**, 124, 38, 8287-8298
19. Xing, C.; Faller, R.; Coarse-grained simulations of supported and unsupported lipid monolayers. *Soft Matter,* **2009**, 5, 22, 4526-4530.
20. Hoiles, W.; Gupta, R.; Cornell, B.; Cranfield, C.; Krishnamurthy, V.; The effect of tethers on artificial cell membranes: A coarse-grained molecular dynamics study. *PloS one*, **2016**, 11, 10, e0162790
21. Hoopes, M.; Deserno, M.; Longo, M.; Faller, R.; Coarse-grained modeling of interactions of lipid bilayers with supports. *The Journal of chemical physics,* **2008**, 129, 17, p11B602
22. Koster, K.; Webb, M.; Bryant, G.; Lynch D.; Interactions between soluble sugars and POPC (1-palmitoyl-2-oleoylphosphatidylcholine) during dehydration: vitrification of sugars alters the phase behavior of the phospholipid. *Biochemica et biophysica Acta (BBA) Biomembranes*, **1994**, 1193, 1, 143-150
23. Souza, P.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics, *Nature Methods*, **2021**, 18, 382-388
24. Marrink, S.; Risselada, J.; Yefimov, S.; Tieleman, P.; de Vries, A.; The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations, *J. Phys. Chem. B,* **2007**, 111, 27, 7812-7824
25. Monticelli, L.; Kandasamy, S.; Periole, X.; Larson, R.; Tieleman, P.; Marrink, S.; The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory and Comput*. **2008**, 4, 5, 819-834
26. Rossi, G.; Monticelli L,; Puisto, SR.; Vattulainen I,; Ala-Nissila T.; Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case. *Soft Matter*. **2011**, 7, 2, 698-708
27. Piskorz, T.; Gobbo, C.; Marrink, S.; Feyter, S.; de Vries, A.; van Esch, J.; Nucleation Mechanisms of Self-Assembled Physisorbed Monolayers on Graphite *J. Phys. Chem C.* **2019**, 123, 28, 17510-17520
28. Wassenaar T.A.; Ingolfsson, H.I.; Bockmann, R.A.; Tieleman, D.P.; Marrink, S.J.; Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *JCTC*, **2015**, 11,2144-2155
29. Lindahl, A.; Hess, V.D.S.; Gromacs 2019 Source Code, 10.5281/zenodo.2424363
30. Berendesen, H.J.C.; Postma, J.; van Gunsteren, W.F.; DiNola, A.; Haak, J.; Molecular Dynamics with coupling to an external bath. *J. Chem. Phys*. **1984**, 81
31. Humphrey, W.; Dalke, A.; Schulten, K.; VMD – Visual Molecular Dynamics *J. Molec. Graphics,* **1996**, 14, 33-38
32. Lukat, G.; Kruger, J.; Sommer, B.; APL@Voro: A Voronoi-Based Membrane Analysis Tool for GROMACS Trajectories. *J. Chem. Inf. Model*. **2013**, 52, 11, 1908-2925
33. Michaud-Agrawal, N.; Denning, E.J.; Woolf, T.B.; Beckstein, O.; MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem*. **2011**, 32, 2319-2327

34. Smith, P.; Lorenz, C.; LiPyphilic: A Python Toolkit for the analysis of lipid membrane simulations, *J. Chem. Theory Comput*. **2021**, 17, 9, 5907-5919
35. Malajczuk, C.; Hughes, Z.; Mancera, R.; Molecular dynamics simulations of the interactions of DMSO, mono- and polyhydroxylated cryosolvents with a hydrated phospholipid bilayer. *BBA-Biomembranes*, **2013**, 1828, 9, 2041-2055

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 limited to the simulation of AF^{M8} and AF^{M8}/SpFr

This work previously appeared in literature as

Austen Bernardi [a], Yihan Huang [b], Bradley Harris [a], Yongao Xiong [a], Somen Nandi[a,c], Karen A. McDonald[a,c], Roland Faller [a*] "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) e0237395, 2020

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 modeling 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.

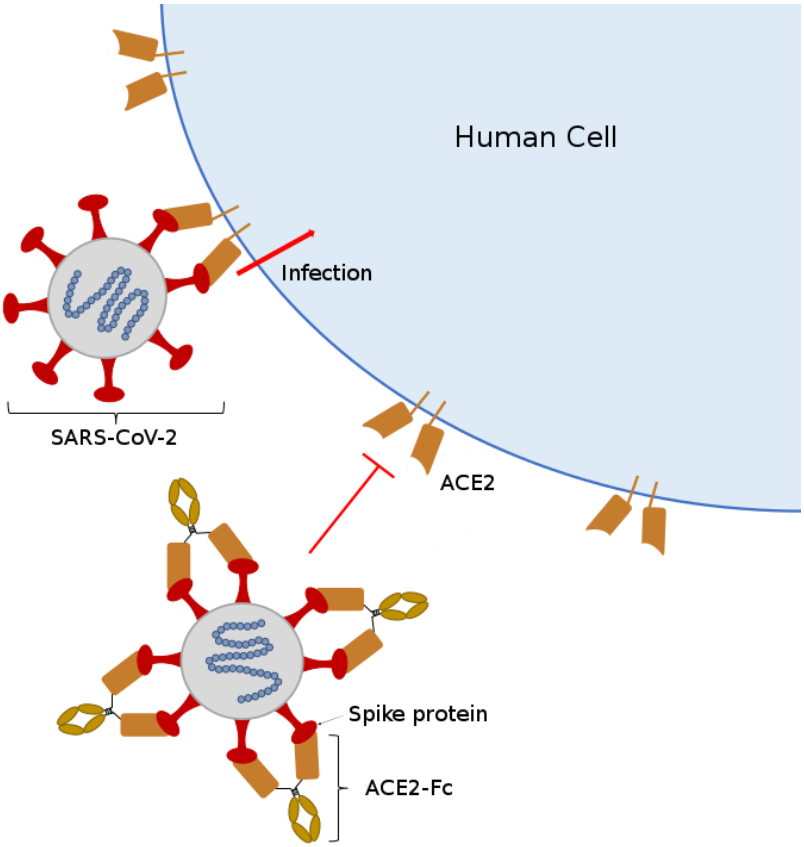


Figure 1. Proposed strategy for SARS-CoV-2 neutralization by ACE2-Fc immunoadhesin. ACE2-Fc binds to the spike (S) protein on the virus and blocks binding to the human cellular receptor ACE2, preventing cellular entrance of SARS-CoV-2.

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 glycosylation (6-8), 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 cell (9), the recombinant protein, as well as production (10, 11) and purification methods (12). 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 domains (15). The neutralization strategy

behind ACE2-Fc is shown in Figure 1 (15). 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 experiments (22). 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 conformation (13). 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 available (23-26) among them recently a proposed glycosylated model of the Spike protein (27). Very recently a short simulation of the Spike protein with glycosylation has been published which is enabling longer studies (28). 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 proteins (30-32). Also, molecular docking studies have been performed with the older related SARS-CoV-1 virus protein implicated in the SARS epidemic in the early 2000s (33). 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

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 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).

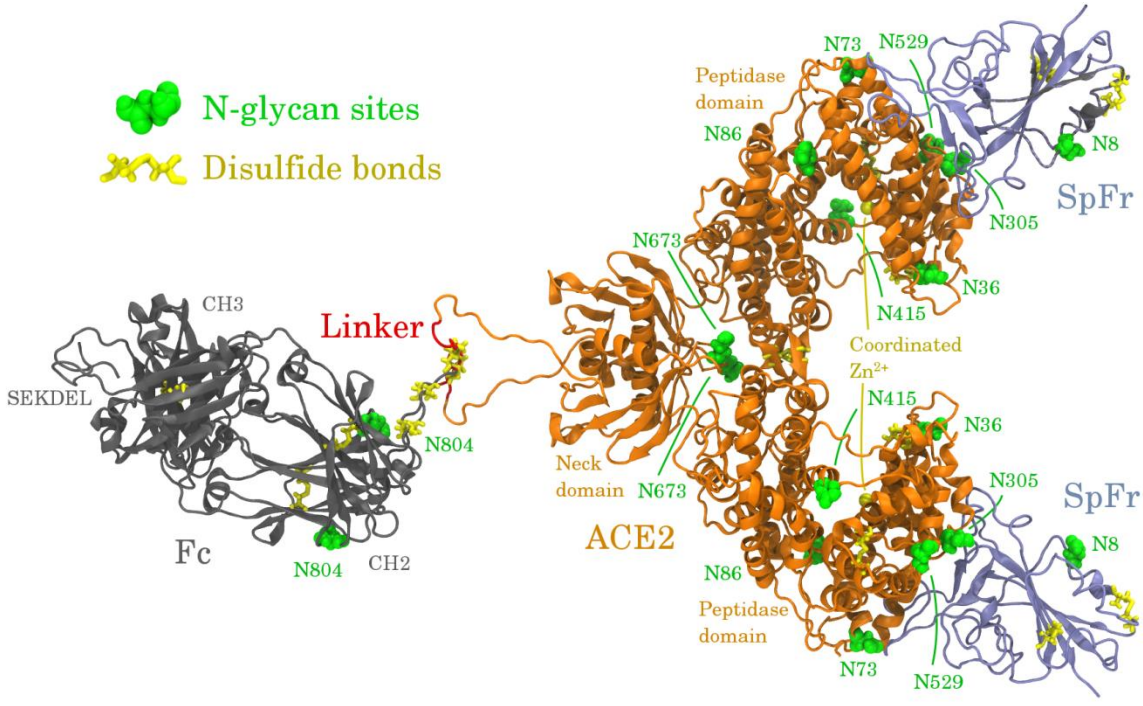


Figure 2. Infographic of the ACE2-Fc variant 1 homodimer bound to two SpFr.

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 study (35), and the SpFr is glycosylated with AnaF^{6} (36). Figure 3 shows these glycans using the Consortium of Functional Glycomics nomenclature.

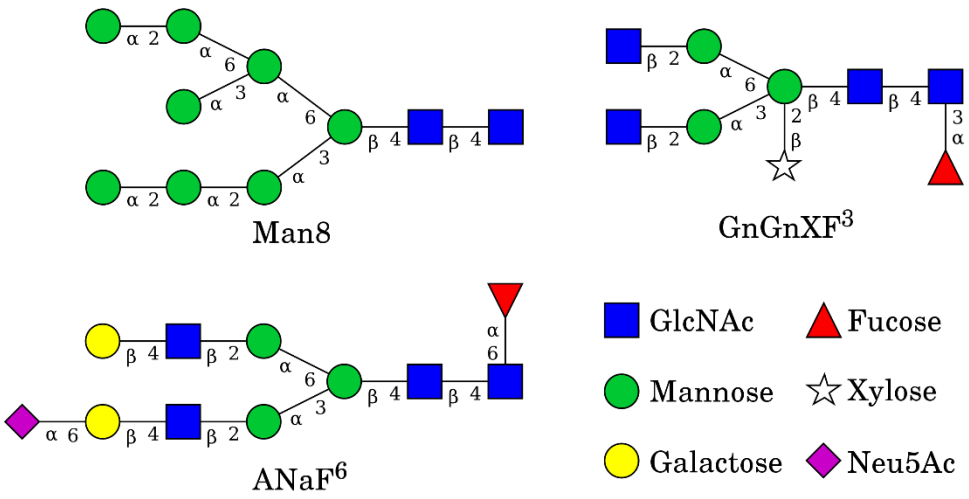


Figure 3. Glycans used in the simulated systems. All structures were built using GlycanBuilder (37).

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 GnGnXF^{3} 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 AnaF[6]. Table 1 summarizes the simulated systems.

Table 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-B^0AT1 and ACE2-B^0AT1/SpFr structures, corresponding to PDB codes 6M18 and 6M17, respectively (38). The ACE2 and ACE2/SpFr domains were taken from these structures and fused to the Fc domain (PDB 3SGJ) (39). The Zn^{2+} 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 1R42 (40). Histidine protonation states for each system were determined using Reduce (41), and are summarized in Table ST2.

Simulation Procedure

The simulation procedure includes the following steps:

1. Fuse ACE2 with Fc to ACE2-Fc using Modeller (42)
2. Model Zn^{2+} and coordinating residues with MCPB.py (43)

3. Attach glycans using glycam.org (44)
4. Merge structures from 2. And 3. Using github.com/austenb28/MCPB_Glycam_merge (45)
5. Generate topology files using AmberTools (43)
6. Convert topology files to Gromacs format using Acpype (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 Zn^{2+} 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 AmberFF14SB (48)forcefield for protein atoms, the Glycam06-j (49) forcefield for glycan atoms, and the SPC/E water model (50). Steps 8 through 14 are performed using the Gromacs suite (51-53). 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 thermostat (54) and Parinello-Rahman barostat (55) using time constants of 0.1 ps and 2 ps, respectively. All water bonds are constrained with SETTLE (56); all other bonds are constrained with LINCS (57). A 1 nm cutoff was used for short-range nonbonded interactions. Particle Mesh Ewald was used to model long-range electrostatics (58). 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 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 protein (28). 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 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 proteins (13).

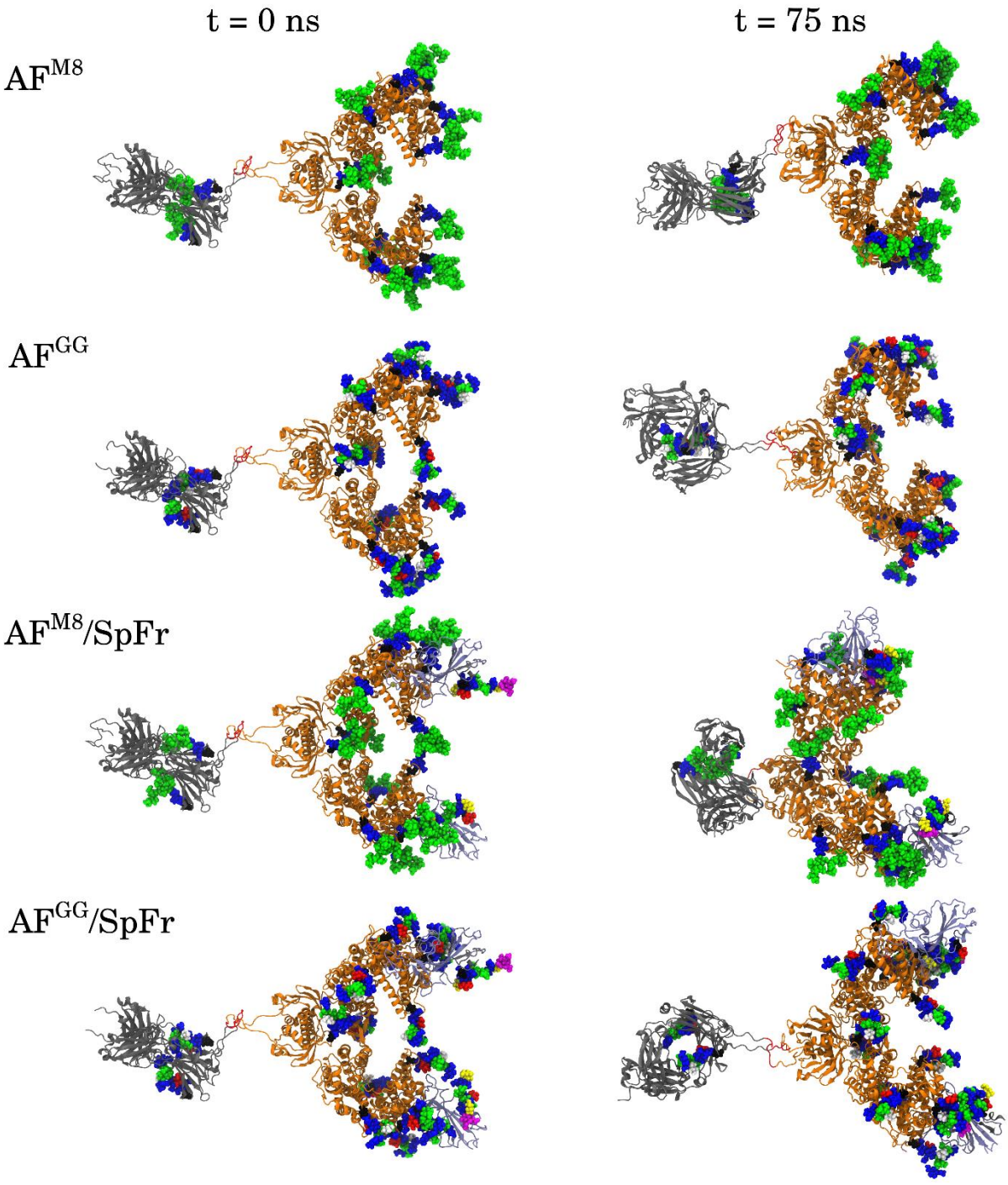


Figure 3: Initial (left) and 75 ns simulated (right) configurations of all systems.

To quantitatively assess structural stability, the root mean square deviation (RMSD) of the ordered domains of ACE2-Fc are shown in Figure 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 $GnGnXF^3$ 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.

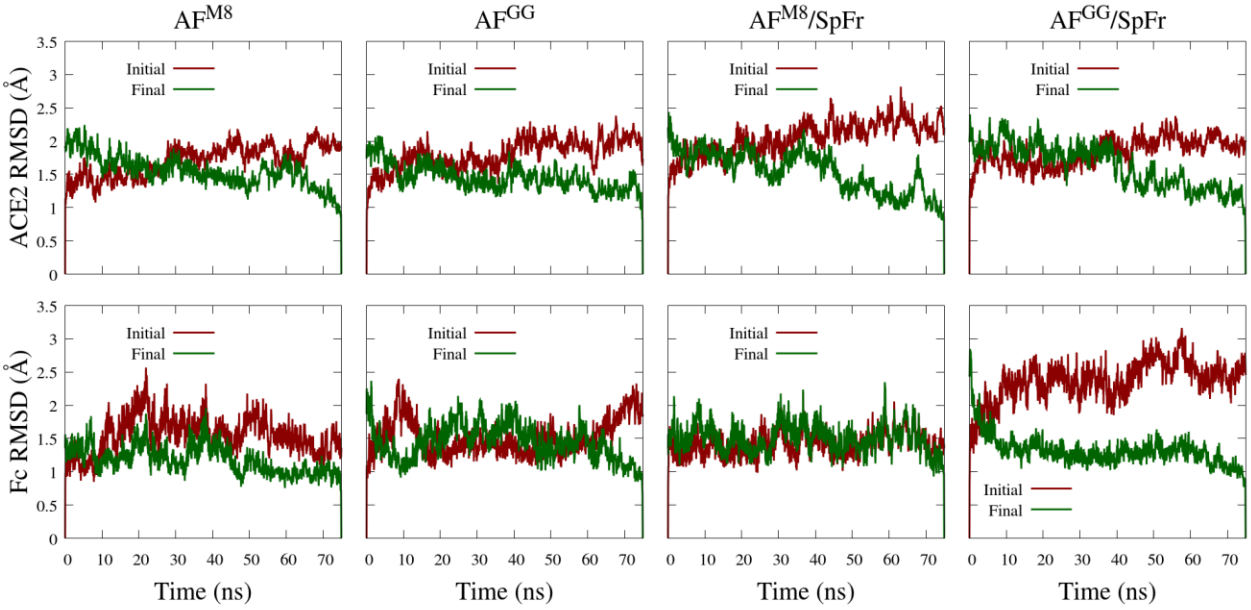


Figure 4. Backbone RMSD profiles of ACE2 (top) and Fc (bottom) ordered domains referenced to initial and final simulation configurations. ACE2: residues 4-707. Fc: residues 745-950. (see SI for sequences)

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

5.5 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, Veesler 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, Liau 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 orientation11Edited 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. Parinello 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.

5.6 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 - QSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGD -  50
 51 - KWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSEDKSKRLN - 100
101 - TILNTMSTIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNERLWAWES - 150
151 - WRSEVGKQLRPLYEEYVVLKNEMARANHYEDYGDYWRGDYEVNGVDGYDY - 200
201 - SRGQLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYISPIGCLPAHLL - 250
251 - GDMWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFVS - 300
301 - VGLPNMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKGDFRILMCTKVTMD - 350
351 - DFLTAHHEMGHIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPKH - 400
401 - LKSIGLLSPDFQEDNETEINFLLKQALTIVGTLPFTYMLEKWRWMVFKGE - 450
451 - IPKDQWMKKWWEMKREIVGVVEPVPHDETYCDPASLFHVSNDYSFIRYYT - 500
501 - RTLYQFQFQEALCQAAKHEGPLHKCDISNSTEAGQKLFNMLRLGKSEPWT - 550
551 - LALENVVGAKNMNVRPLLNYFEPLFTWLKDQNKNSFVGWSTDWSPYADQS - 600
601 - IKVRISLKSALGDKAYEWNDNEMYLFRSSVAYAMRQYFLKVKNQMILFGE - 650
651 - EDVRVANLKPRISFNFFVTAPKNVSDIIPRTEVEKAIRMSRSRINDAFRL - 700
701 - NDNSLEFLGIQPTLGPPNQPPVSSSERKCCVECPPCPAPELLGGPSVFLF - 750
751 - PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRE - 800
801 - EQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQP - 850
851 - REPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKT - 900
901 - TPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSL - 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 Zn^{2+} 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 - QSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGD -  50
 51 - KWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSEDKSKRLN - 100
101 - TILNTMSTIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNERLWAWES - 150
151 - WRSEVGKQLRPLYEEYVVLKNEMARANHYEDYGDYWRGDYEVNGVDGYDY - 200
201 - SRGQLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYISPIGCLPAHLL - 250
251 - GDMWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFVS - 300
301 - VGLPNMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKGDFRILMCTKVTMD - 350
351 - DFLTAHNEMGNIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPKH - 400
401 - LKSIGLLSPDFQEDNETEINFLLKQALTIVGTLPFTYMLEKWRWMVFKGE - 450
451 - IPKDQWMKKWWEMKREIVGVVEPVPHDETYCDPASLFHVSNDYSFIRYYT - 500
501 - RTLYQFQFQEALCQAAKHEGPLHKCDISNSTEAGQKLFNMLRLGKSEPWT - 550
551 - LALENVVGAKNMNVRPLLNYFEPLFTWLKDQNKNSFVGWSTDWSPYADQS - 600
601 - IKVRISLKSALGDKAYEWNDNEMYLFRSSVAYAMRQYFLKVKNQMILFGE - 650
651 - EDVRVANLKPRISFNFFVTAPKNVSDIIPRTEVEKAIRMSRSRINDAFRL - 700
701 - NDNSLEFLGIQPTLGPPNQPPVSSSERKCCVECPPCPAPELLGGPSVFLF - 750
751 - PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRE - 800
801 - EQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQP - 850
851 - REPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKT - 900
901 - TPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSL - 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 - CPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPT -  50
 51 - KLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIA - 100
101 - WNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEG - 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				

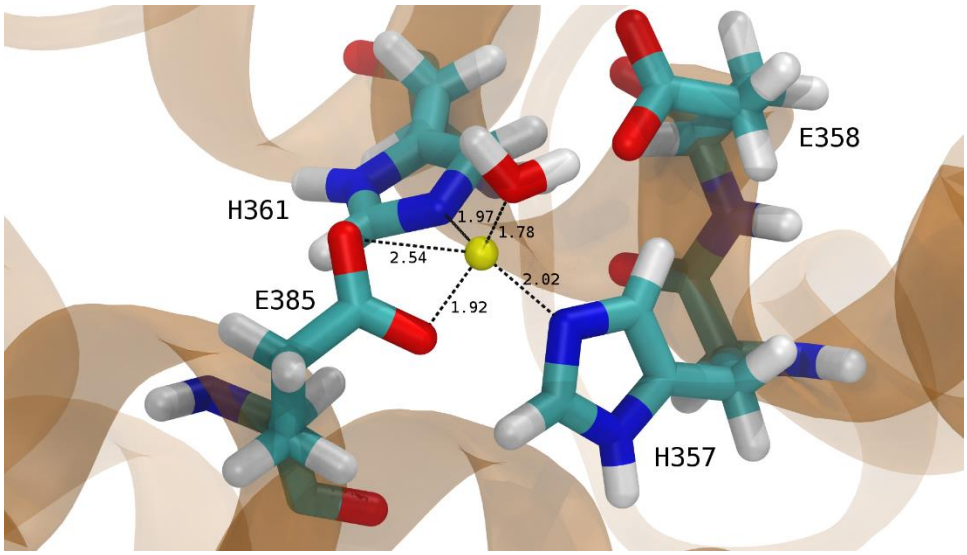


Figure SF1. Coordinating residues of the Zn^{2+} (yellow) active site of variant 1. Coordinating bonds with their distances in angstroms are labeled.

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

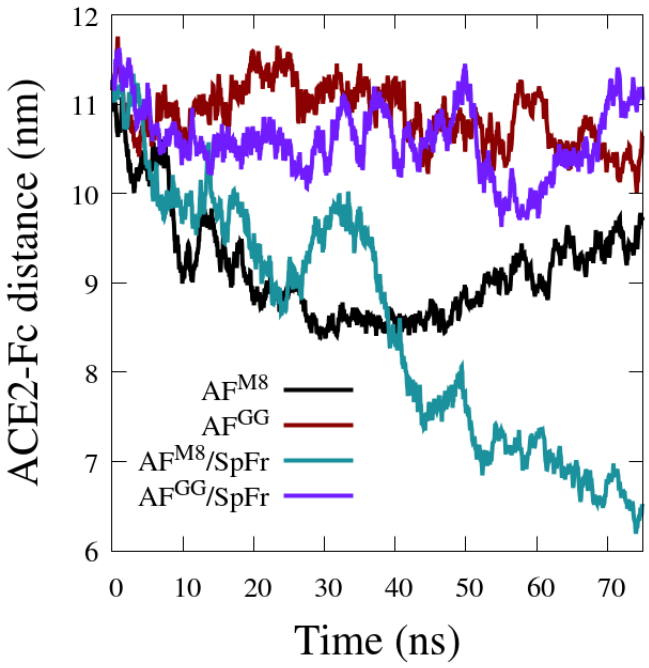


Figure SF2. Center of mass distance between the ordered domains of ACE2 and Fc. ACE2: residues 4-707. Fc: residues 745-950.

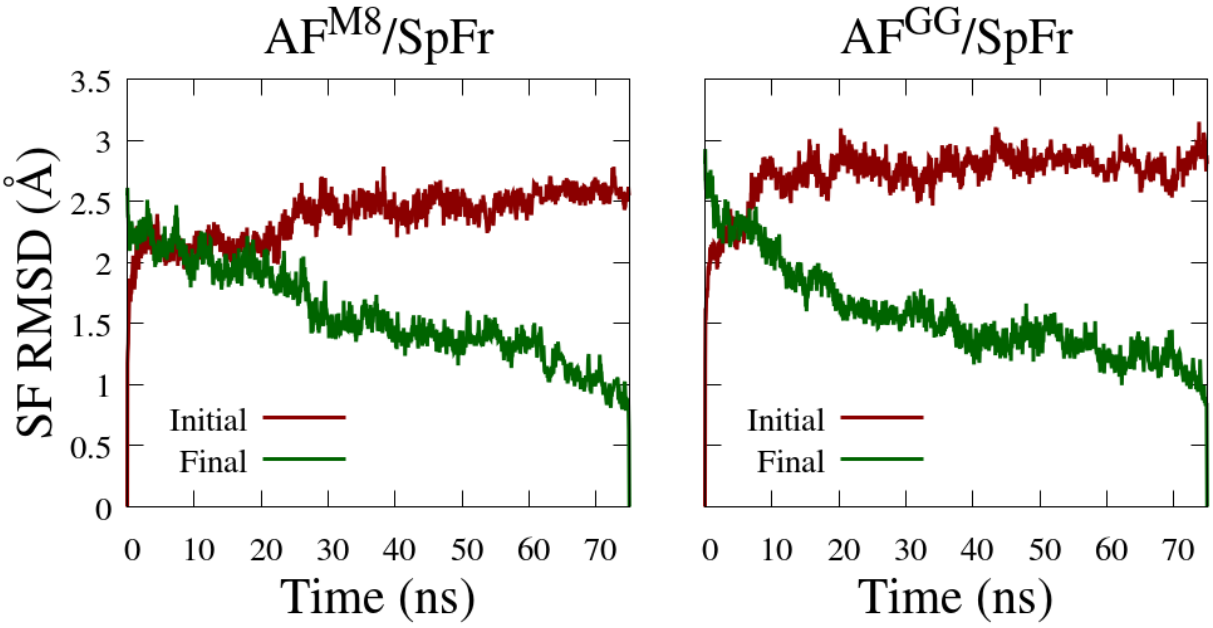


Figure SF3. Backbone RMSD profiles of the SpFr referenced from initial and final simulation configurations.

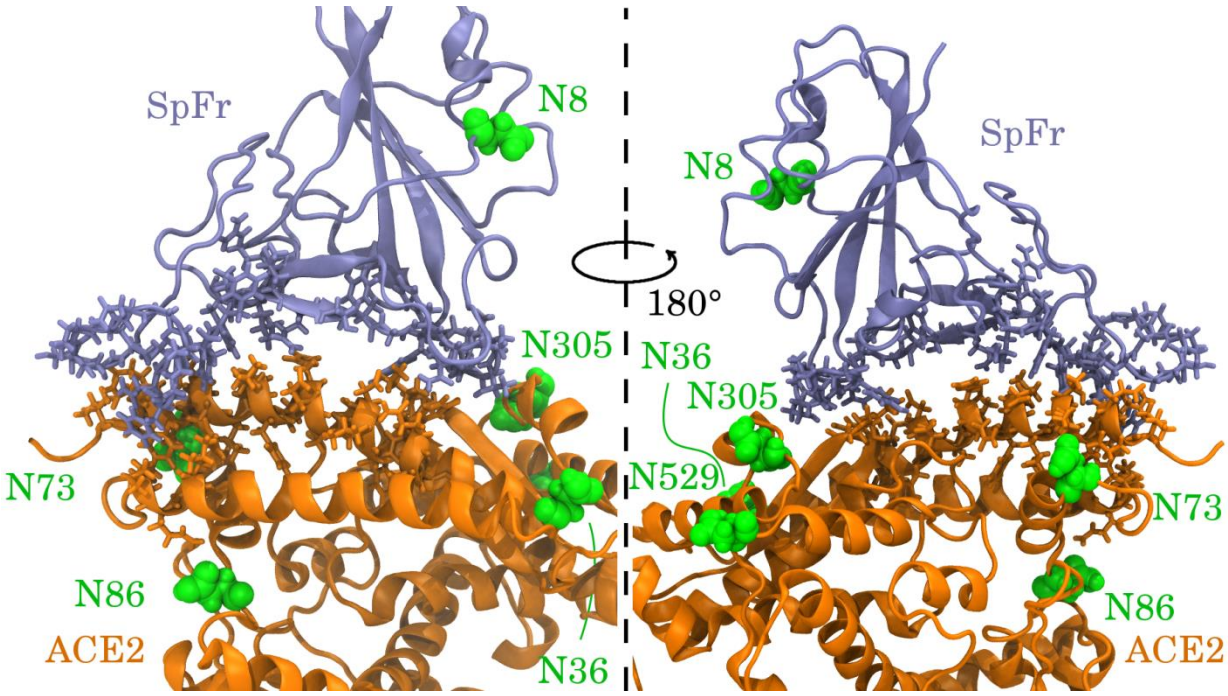


Figure SF4. Front (left) and back (right) zoom views of the interface between the ACE2 and SpFr domains. Interfacial residues are shown in licorice.

Chapter 6

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

My contribution was the simulation of A1 fragments, Yihan the A2.
Analysis and writing were split equally

Yihan Huang[1X], Bradley S. Harris[2X], Shiaki A. Minami[2], Seongwon Jung[2], Priya S. Shah[2,3], Somen Nandi[2,4], Karen A. McDonald[2,4], Roland Faller[2,*]

[1]Department of Materials Science, UC Davis; [2]Department of Chemical Engineering, UC Davis; [3]Department of Microbiology and Molecular Genetics, UC Davis; [4]Global HealthShare Initiative, UC Davis

[X] equal first authors
* corresponding author: rfaller@ucdavis.edu

This work appears 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), doi: https://doi.org/10.1016/j.bpj.2021.12.002

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 computationally (1-7). 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 interface (8-14). Furthermore, many mammalian viruses use glycans on cell-surface glycoproteins or glycolipids as receptors (15). 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 CryoEM (16). 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 sites (17). Only 16 N-linked glycosylation sites were observed in a cryo-EM map of S produced in HEK293F cells (18). A study by Watanabe *et al*. (2020) determined site-specific glycoform analysis of full-length trimeric S protein made in HEK293F cells (16). In another study of S glycosylation patterns including O glycosylation were determined (19). In a similar vein, it has recently been argued that glycosylation can have influences post-vaccination and for vaccine resistance (20). 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 account (23-26). 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 human proteins 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-tagging (37), Variant 1 was fully glycosylated with MAN8 glycans. Variant 2 was fully glycosylated with GlcNAc2XylFucMan3GlcNAc2 ($GnGnXF^3$) that is a standard plant glycoform, and the S protein fragment was glycosylated with $AnaF^6$ (2). Figure 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.

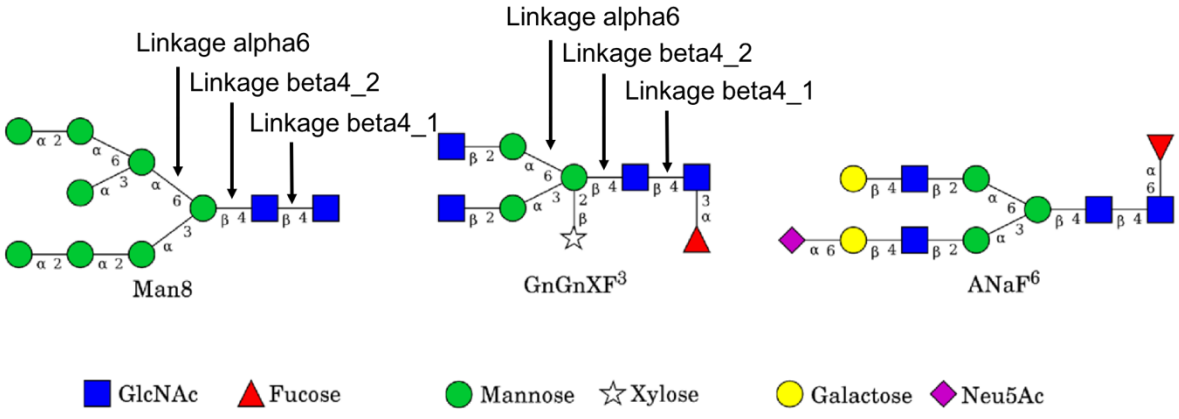


Figure 1: Glycans used in the simulations, adapted from previous work (2), with linkages of interest in MAN8 and GnGnXF^3 glycans for dynamic analysis.

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 proteins (38). The starting atomic coordinates for all pulling systems were taken from the final 75 ns configurations of our previous paper (2). 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 6M17 (39)). These structures were fused to the Fc domain (PDB 3SGJ (40)). The Zn^{2+} coordinating residues and water were taken from structure PDB 1R42 (41) in the case of Variant 1 ACE2. Variant 2 has 2 mutations that prevent Zn^{2+} 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 Zn^{2+} 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 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 AmberTools (44) as described previously (2). Briefly, the molecules were trimmed and glycans were removed, then Man8 glycans were reattached to the truncated Variant 1 of ACE2, $GnGnXF^3$ to the truncated Variant 2 of ACE2 and $AnaF^6$ to the SpFr in both variants using Glycam.org (45). The coordinating Zn^{2+} was reattached to truncated and glycosylated Variant 1 using MCPB.py (46). 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 S-S bridges are retained in our simulations.

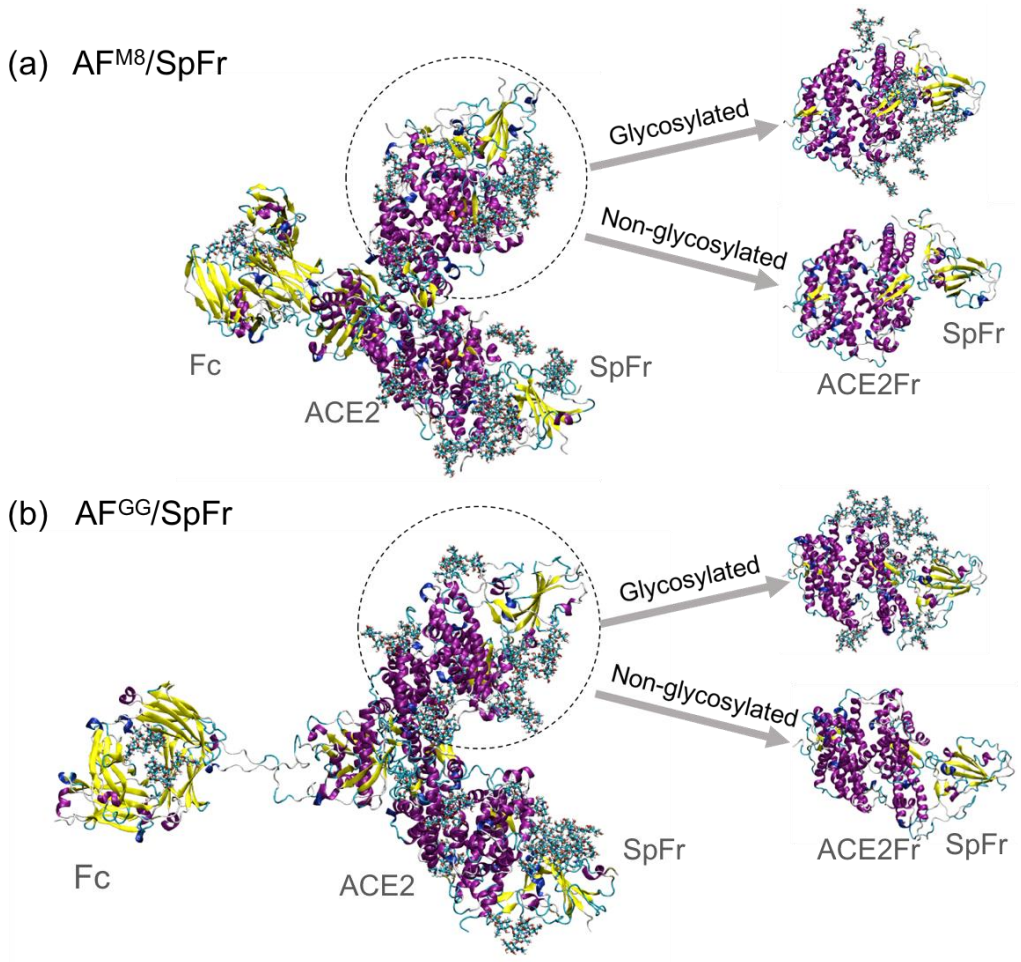


Figure 2: Schematics of generating the different initial structures. The full recombinantly made Ace2-FC systems on the left are used in the BLI experiments for determining binding affinities to the RBD, whereas the four truncated systems on the right containing only a fragment of ACE2 are modeled in 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 thermostat (47) and Parrinello-Rahman barostat (48). All water bonds were constrained with SETTLE (49), and all other bonds were constrained with LINCS (50). Box expansion, solvation, and equilibration were performed using the Gromacs suite version 2019.1 (51).

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/nm^2, a Nose-Hoover thermostat (52) 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 2019.1 (51). 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 3. Systems were pulled over a distance of 8 nm until full separation (no interaction) was achieved (see Figure 3).

Hydrogen bonds were analyzed using the built-in Gromacs bond command (51) 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 matplotlib (53) 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 command (54).

6.2.2 Experiments:

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 µL of Laemmli sample buffer (Bio-Rad) and 2 µL β-mercaptoethanol (Bio-Rad) were added to 30 µL 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).

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 Filtering (55) 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:

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

6.3 Results:

Figure 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 3a and 3b the aglycosylated structures return to baseline at roughly 2.5 nm of pulling distance. Importantly the glycosylated structures in Figure 3a and 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.

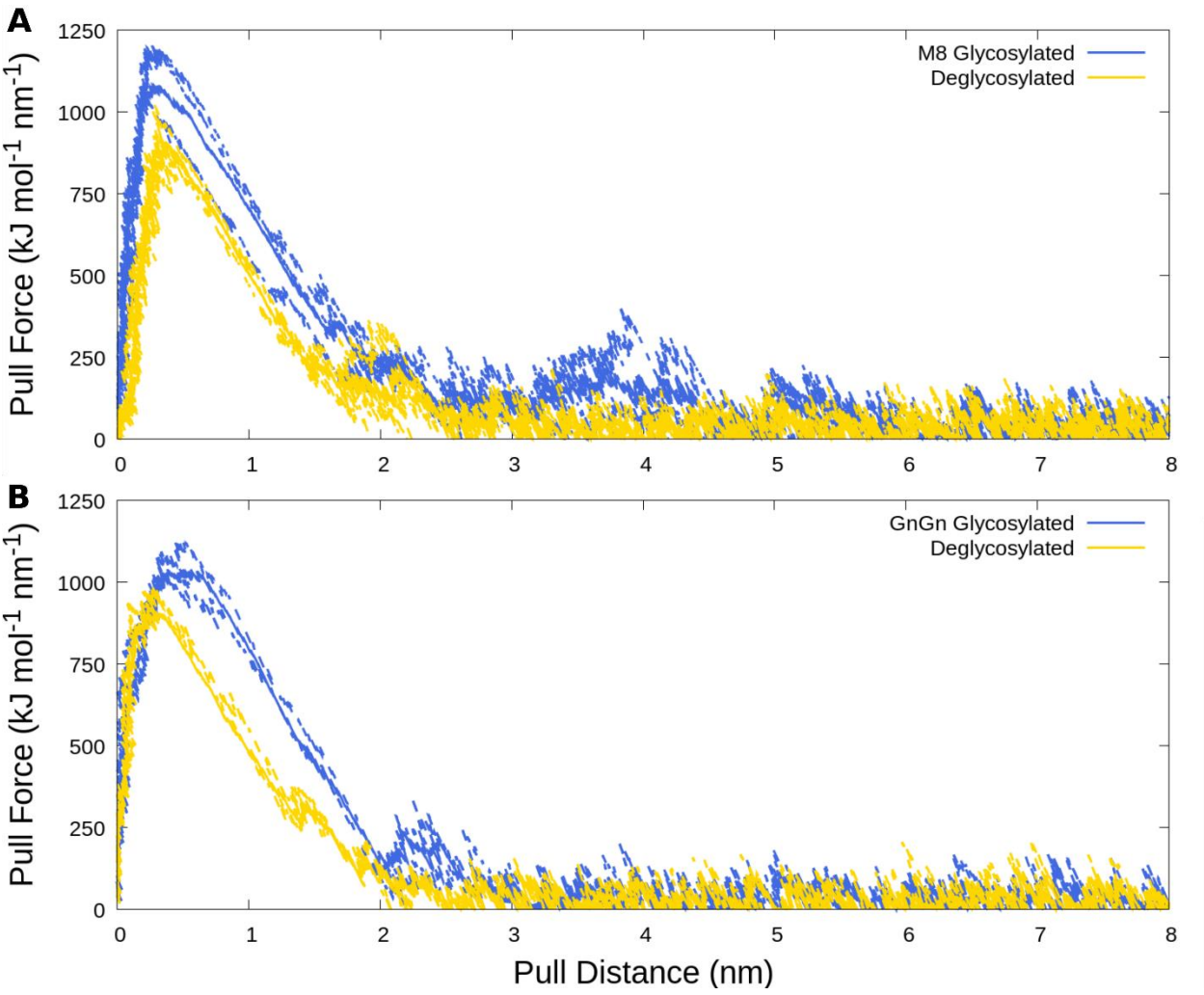


Figure 3: Traces of pull force versus pull distance. A) Man8 glycosylated A1Fr^{M8}/SpFr and aglycosylated structure A1Fr/SpFr B) GnGnXF glycosylated A2Fr^{GG}/SpFr. andaglycosylated Structure A2Fr/SpFr. Blue lines correspond to glycosylated structures, gold to deglycosylated. Dashed lines are individual replicas, solid lines are averages.

To further characterize the extension of binding interactions, Figure 4 shows hydrogen bonding interaction maps between the ACE2 and RBD proteins. Figure 4a and 4c are for A1Fr^{M8}/SpFr and A2Fr^{GG}/SpFr respectively while 4b and 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 4a) the predominant interactions involve glycans directly while for $A2Fr^{GG}$/SpFr (Figure 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 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 4b and 4d express maps of similar protein-protein interactions, though the A2Fr/SpFr shown in Figure 4d contains many more interactions as indicated by the increased number of rows.

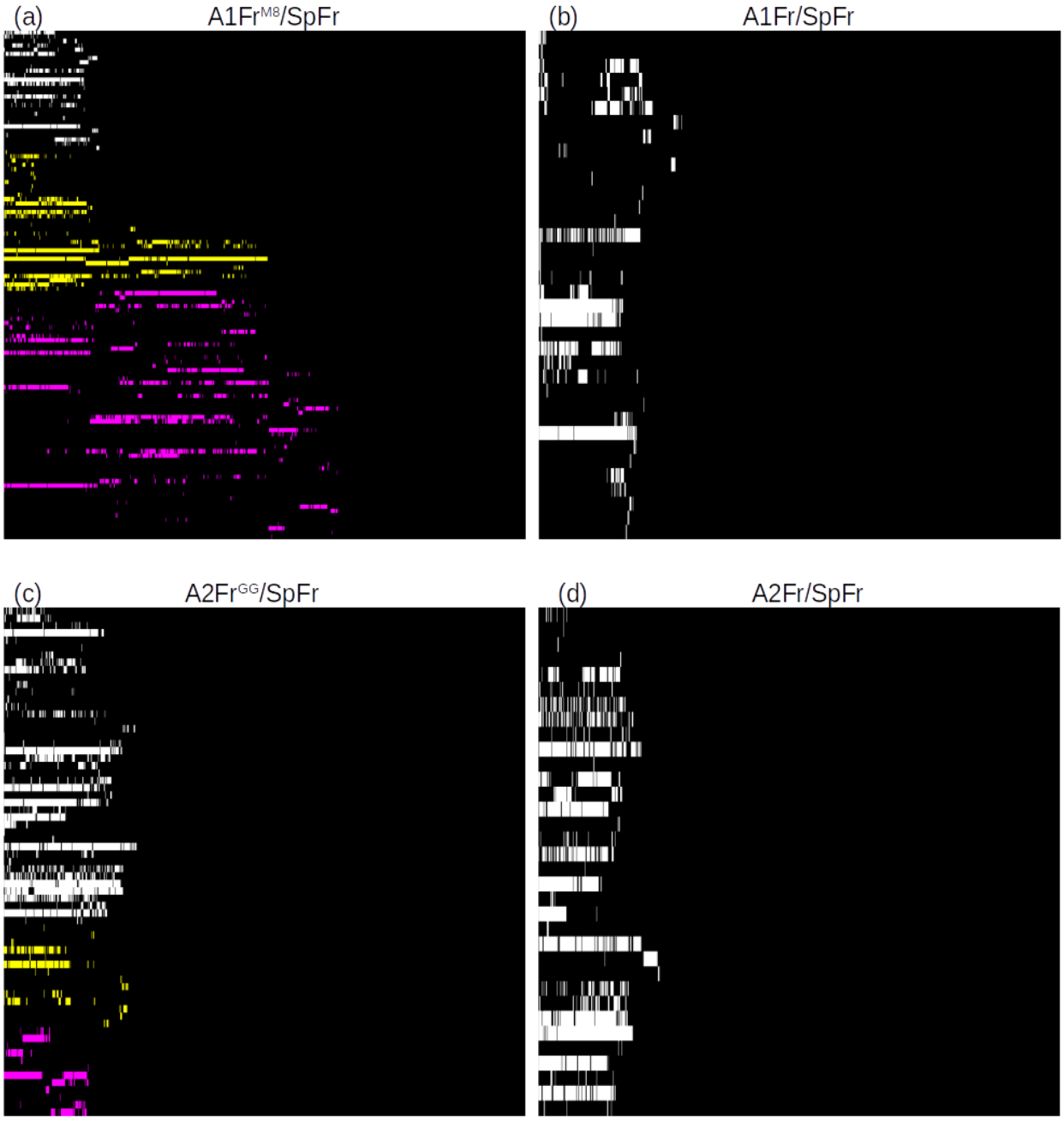


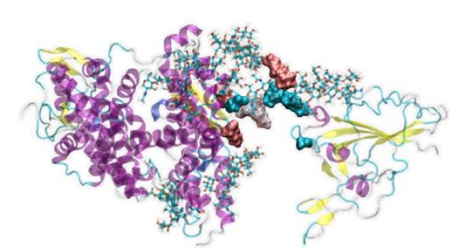
Figure 4: Hydrogen bond interactions vs simulation time. A) Man8 Glycosylated A1Fr^{M8}/SpFr. B) Aglycosylated Structure A1Fr/SpFr. C) GnGnXF^{3} Glycosylated A2Fr^{GG}/SpFr. D) Aglycosylated Structure A2Fr/SpFr. Colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.

Figure 5 shows the configurations where RBD with and without ANaF^{6} 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 nomenclature (56). 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 shows how the different structures of MAN8 and GnGnXF^3 affect the hydrogen bonding regime. Although MAN8 and GnGnXF^3 have similar size (223 atoms vs 222 atoms), their shapes are very different. MAN8 is relatively flatter comparing to GnGnXF^3, making it bend less flexibly. Therefore, when MAN8 is close to AnaF^6, they interact in a side-by-side fashion, whereas when GnGnXF^3 is close to AnaF^6, they interact in a head-to-head fashion, forming less hydrogen bonds than the MAN8/GnGnXF^3 pair.

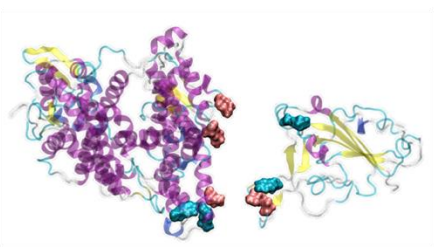
(a)



Glycosylated ACE2_Man8 and RBD Top 5 Hbonds

Donor Acceptor Pair (Donor : Acceptor)	Total Occupancy (%)
RBD-gly_6LB789O2:ACE2-gly_VMA815O2	24.6
ACE2_NLN488ND2:RBD-gly_0SA790O5N	20.2
RBD-gly_0SA790O4:ACE2_MET489O	19.8
RBD_GLY167N:ACE2_LYS519O	18.3
RBD-gly_4YB788O6:ACE2-gly_VMB814O2	18.1

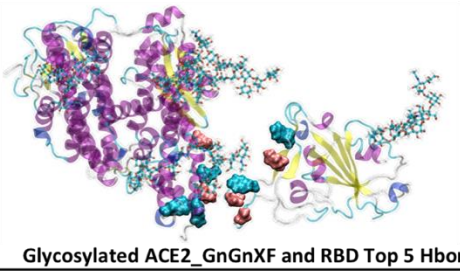
(b)



NonGlycosylated ACE2_MAN8 and RBD Top 5 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

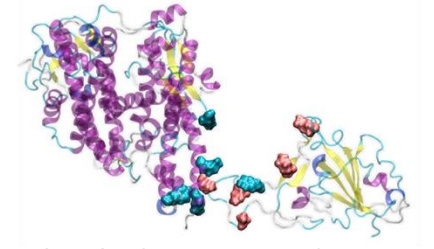
(c)



Glycosylated ACE2_GnGnXF and RBD Top 5 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

(d)



NonGlycosylated ACE2_GnGnXF and RBD Top 5 Hbonds

Donor Acceptor Pair (Donor : Acceptor)	Total Occupancy (%)
ACE2_TYR249OH:RBD_ASN152OD1	19.8
ACE2_GLN190NE2:RBD_ALA140O	19.2
RBD_TYR154OH:ACE2_THR193OG1	18.7
ACE2_LYS197NZ:RBD_GLN158OE1	16.8
ACE2_LYS519NZ:RBD_GLN163OE1	14.6

Figure 5: Figure 5: Top 5 hydrogen bond donor:acceptor pairs and occupancy. A) Man8 Glycosylated A1Fr^{M8}/SpFr. B) Aglycosylated Structure A1Fr/SpFr. C) GnGnXF^{3} 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.

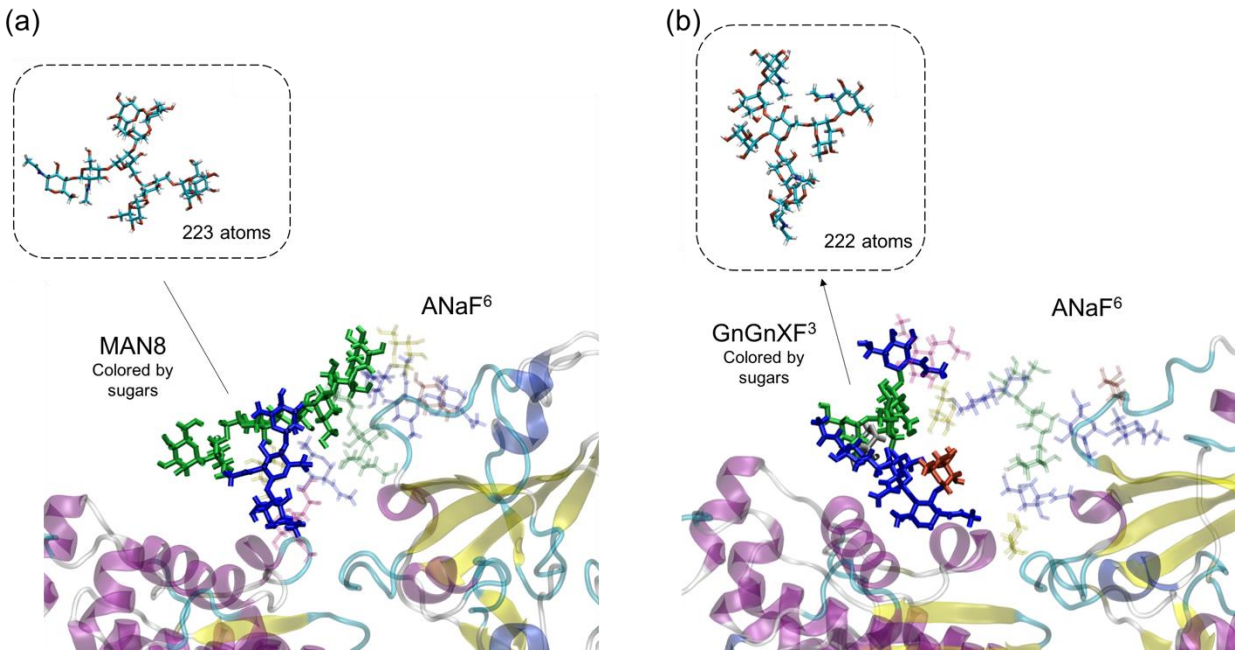


Figure 6: Different structures and hydrogen bonding regimes of MAN8 and GnGnXF^3 when interacting with AnaF^6 on RBD. A) MAN8 that interacts with AnaF^6. B) GnGnXF^3 that interacts with AnaF^6. Inserts: shape and size of the MAN8 and GnGnXF^3 without bending towards AnaF^6. The glycans attached to proteins 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 GnGnXF^3, 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 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 ANaF^6 on RBD. To improve statistics, trajectories from the previous 75 ns runs (2) were

used for the ACF analysis. Figure 7 shows the angle and dihedral motions for both MAN8 and $GnGnXF^3$ 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 $ANaF^6$ 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 $GnGnXF^3$, ACF_{Angle} and $ACF_{Dihedral}$ of MAN8 decrease either at similar rate or slower than those of $GnGnXF^3$ 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 $GnGnXF^3$ for the angle and dihedral motions at linkage beta4_1, beta4_2, and alpha6. The glycans at N488 are the ones interacting with $ANaF^6$ on RBD. All angle motions and dihedral motions of MAN8 at N488 are less flexible than for $GnGnXF^3$, which further proves that side-by-side hydrogen bonding fashion with $ANaF^6$ is favored by MAN8 resulting in more hydrogen bonds between glycans before pulling, whereas a head-to-head arrangement is favored by $GnGnXF^3$ 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 ANaF^6 on the other end.

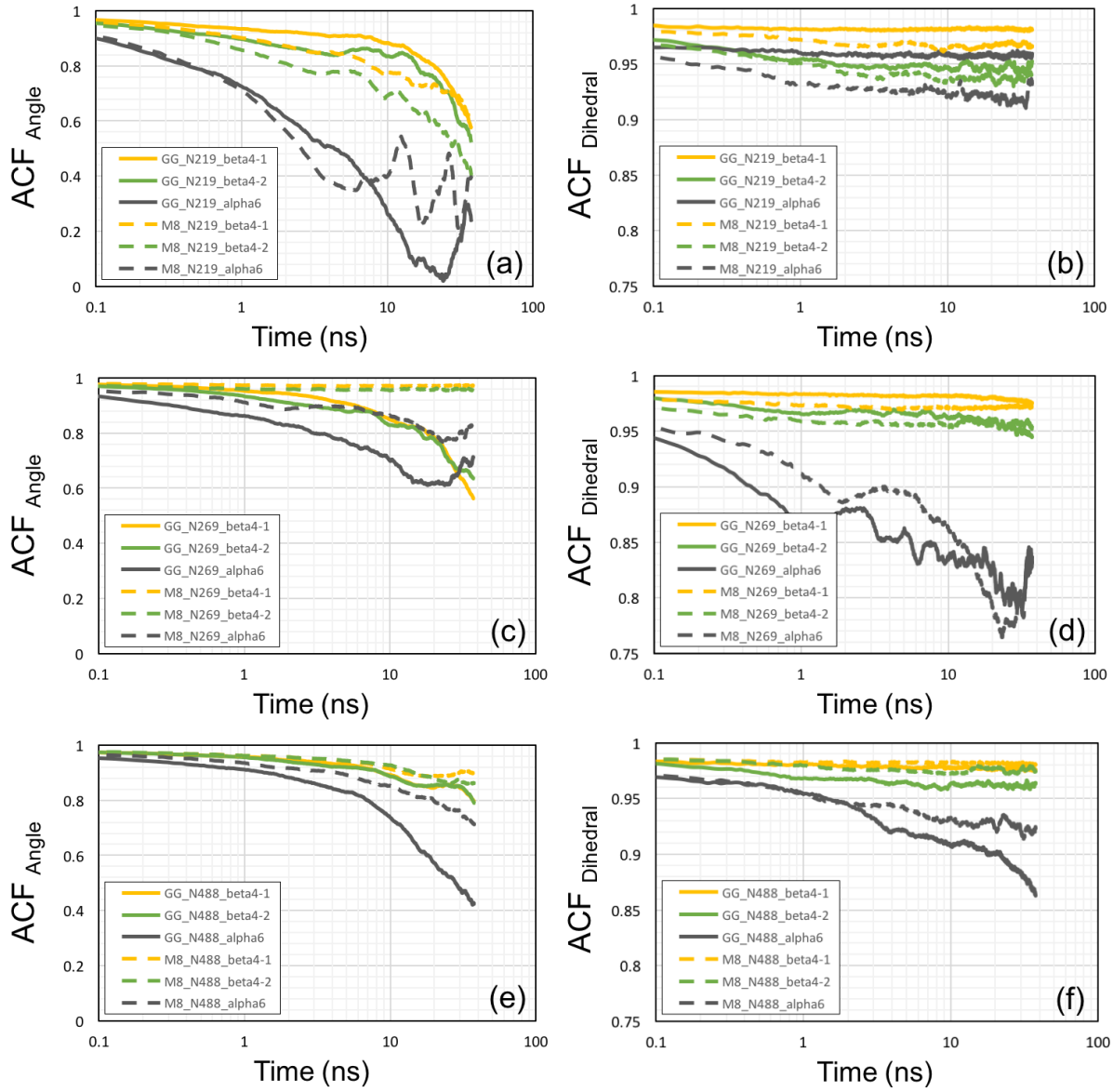


Figure 7: Autocorrelation function analysis of angles and dihedrals at linkage beta4_1, beta4_2, and alpha6 for MAN8 and GnGnXF^3 at ACE2 fragment glycosylation sites in semi-log lots. Glycans at N219 (a-b) and at N269 (c-d) shows typical behaviors, and glycan at N488 (e-f) are the ones directly interacting with AnaF^6 on RBD. Dashed lines are the dynamic motions of MAN8, and solid lines are the dynamic motions of GnGnXF^3.

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 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.

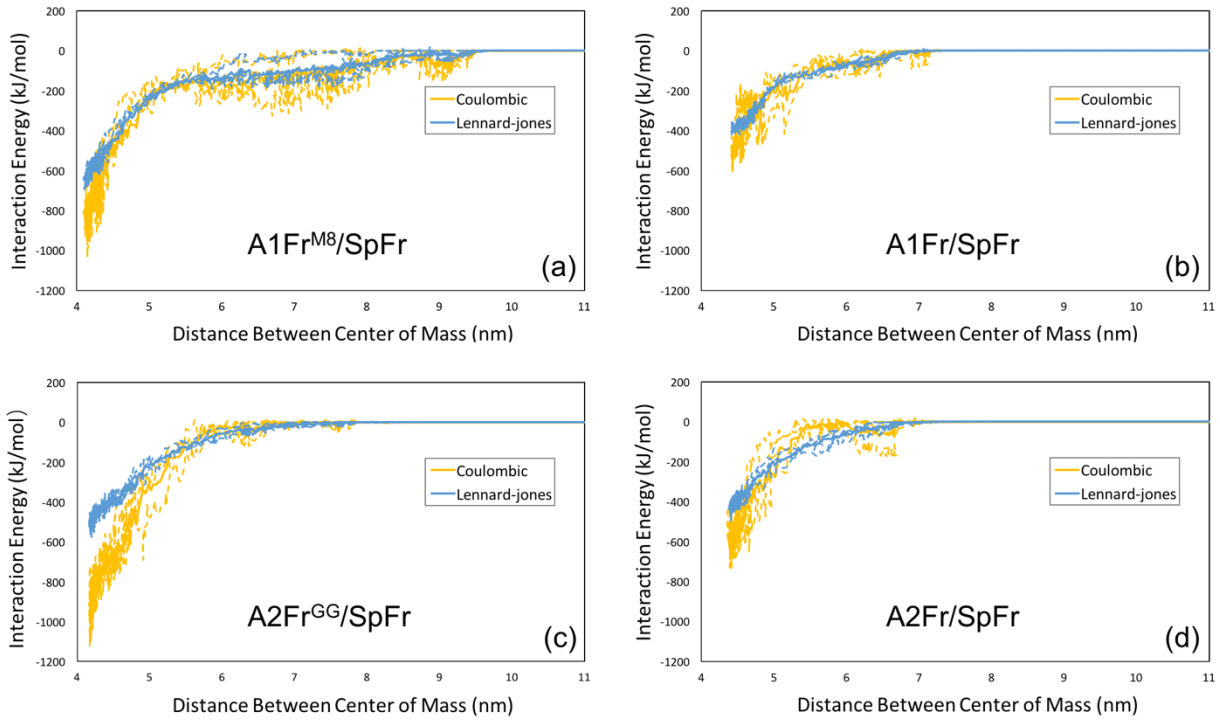


Figure 8: Lennard Jones and electrostatic energies. a) Man8 Glycosylated A1Fr^{M8}/SpFr. b) Aglycosylated Structure A1Fr/SpFr. c) GnGnXF Glycosylated A2Fr^{GG}/SpFr. d) Aglycosylated Structure A2Fr/SpFr.

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 publication (2) to determine the dominant motions of the RBD. Results of the PCA are presented in Figure 9 and S12-S15. Figures 9a and 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 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 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).

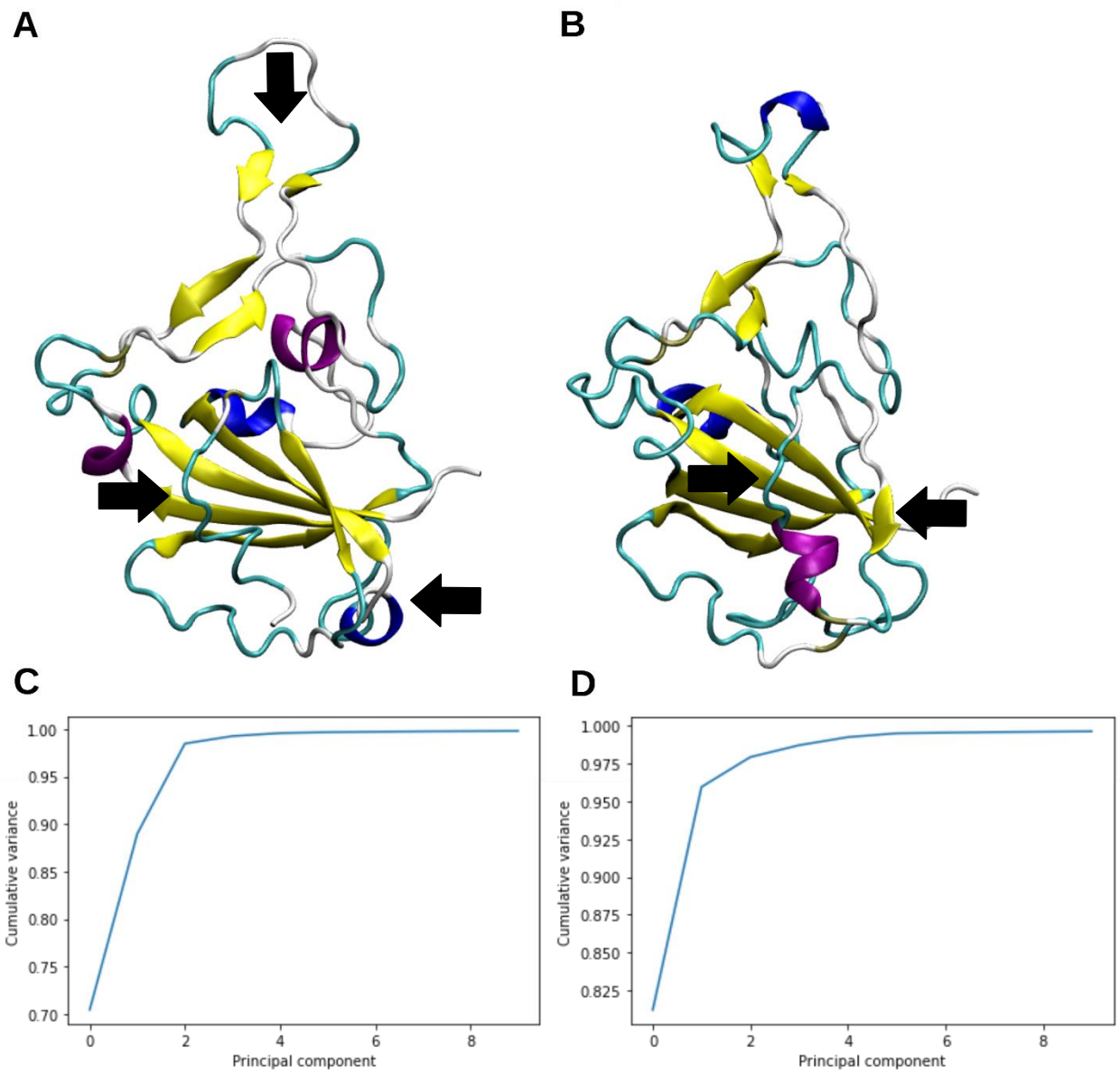


Figure 9: First Principal Component (PC1) projected motion and cumulative variance. A) PC1 projected motion for A1Fr^{M8}/SpFr. B) PC1 projected motion for A2Fr^{GG}/SpFr C) Principal component cumulative variance A1Fr^{M8}/SpFr. D) Principal component cumulative variance A2Fr^{GG}/SpFr. Arrows indicate 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 binding (57). The response is measured as a shift in wavelength in units of nm. Figure 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 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 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.

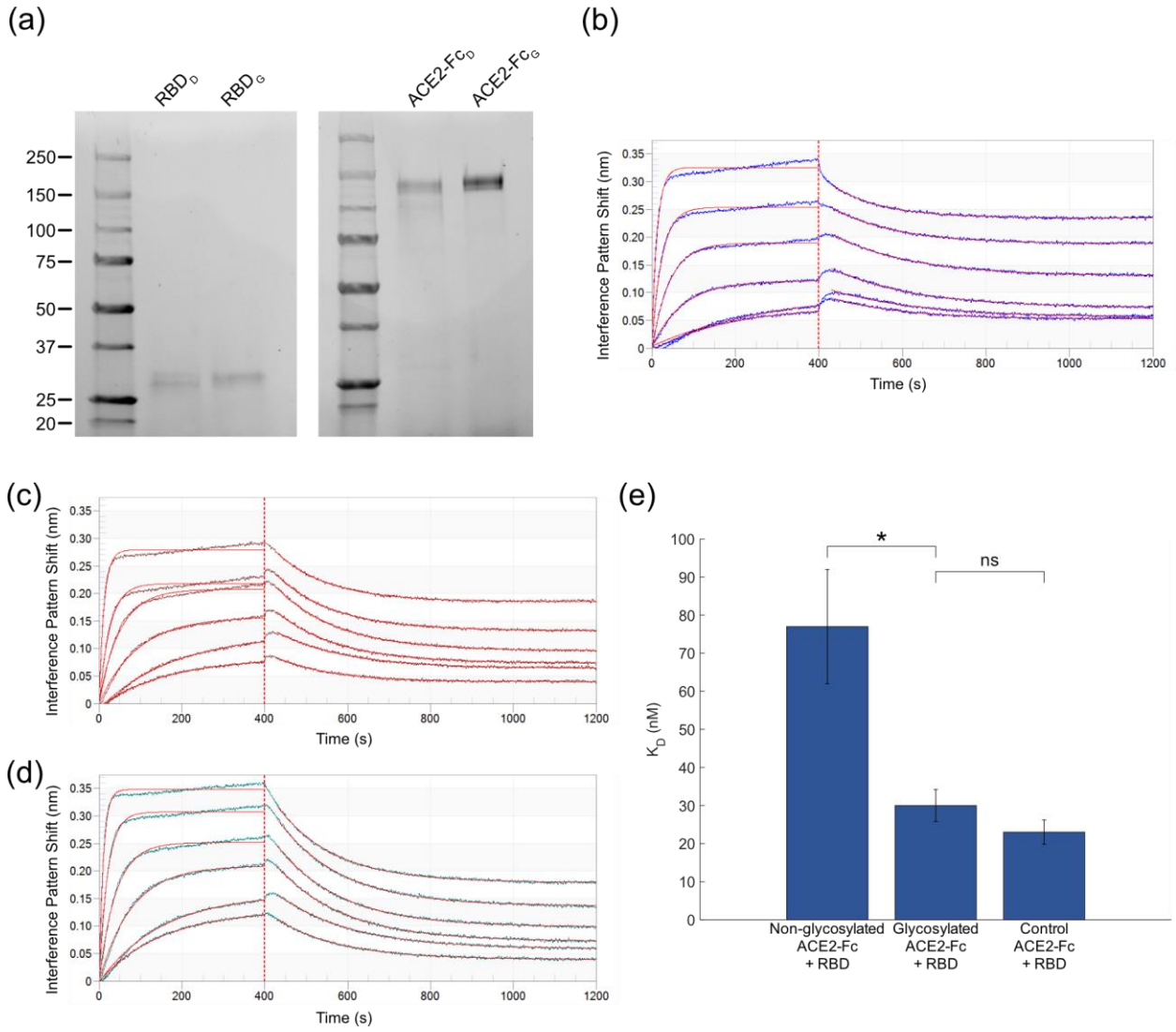


Figure 10. Figure 10. Biolayer Interferometry on glycosylated and deglycosylated ACE2-Fc and RBD. (a) SDS-PAGE on ACE2-Fc and RBD with and without PNGase F treatment. A total of 1 µg of protein is loaded onto each lane. Subscript D indicates deglycosylated proteins, and subscript G indicates glycosylated 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 Discussion:

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 infectivity (58-60). 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 pentasaccharide (32). 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 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 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 $GnGnXF^3$ 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 $AnaF^6$, 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 $GnGnXF^3$, especially for the MAN8 and $GnGnXF^3$ glycans that directly interact with $AnaF^6$, proving that MAN8 is more constrained by the hydrogen bonds between MAN8 and $AnaF^6$. The hydrogen bond map of $A2Fr^{GG}$/SpFr in Figure 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 3) as well as the interaction energies vs center of mass distances (Figure 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 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 loop (32). 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 Zn^{2+} on the stability of ACE2. Figures 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 Fig. 10 b,d.

6.5 Conclusion:

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.

Funding:

YH and RF were partially supported by the National Science Foundation under grant no. CBET 1911267. SAM and SJ were partially supported by a COVID-19 Research Accelerator Funding Track award by the UC Davis Office of Research (https://covid19research.ucdavis.edu/tags/craft). 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. KAM and SN were partially supported by NASA Space Technology Research (award number NNX17AJ31G) and by the Translational Research Institute through NASA (grant number NNX16AO69A).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration (NASA) or the Translational Research Institute for Space Health (TRISH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

6.6 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, doi: https://doi.org/10.1002/jmv.25726, https://doi.org/10.1002/jmv.25726 (https://doi.org/10.1002/jmv.25726).
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, doi: 10.1371/journal.pone.0237295, https://doi.org/10.1371/journal.pone.0237295.
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, doi: 10.1126/science.abc0870, https://science.sciencemag.org/content/sci/369/6508/1261.full.pdf.
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, doi: 10.1101/2021.01.13.426558, https://www.biorxiv.org/content/biorxiv/early/2021/01/13/2021.01.13.426558.full.pdf.
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, doi: 10.1073/pnas.2100425118, https://www.pnas.org/content/pnas/118/19/e2100425118.full.pdf.
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 Modeling*. 101:107716, doi: https://doi.org/10.1016/j.jmgm.2020.107716, https://www.sciencedirect.com/science/article/pii/S1093326320305052.
7. Davidson, A. M., J. Wysocki, and D. Batlle. 2020. Interaction of SARS-CoV-2 and Other Coronavirus With ACE (Angiotensin-Converting Enzyme)-2 as Their Main Receptor. *Hypertension*. 76(5):1339-1349, doi: doi:10.1161/HYPERTENSIONAHA.120.15256, https://www.ahajournals.org/doi/abs/10.1161/HYPERTENSIONAHA.120.15256
8. Vigerust, D. J., and V. L. Shepherd. 2007. Virus glycosylation: role in virulence and immune interactions. *Trends in Microbiology*. 15(5):211-218, doi: 10.1016/j.tim.2007.03.003, https://doi.org/10.1016/j.tim.2007.03.003.
9. Ohtsubo, K., and J. D. Marth. 2006. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*. 126(5):855-867, doi: 10.1016/j.cell.2006.08.019, https://doi.org/10.1016/j.cell.2006.08.019.

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, doi: 10.1126/science.1701274, https://science.sciencemag.org/content/sci/250/4984/1130.full.pdf.
11. Sperandio, M., C. A. Gleissner, and K. Ley. 2009. Glycosylation in immune cell trafficking. *Immunological Reviews*. 230(1):97-113, doi: https://doi.org/10.1111/j.1600-065X.2009.00795.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-065X.2009.00795.x.
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, doi: https://doi.org/10.1016/j.bbagen.2019.05.012, https://www.sciencedirect.com/science/article/pii/S0304416519301333.
13. Banerjee, N., and S. Mukhopadhyay. 2016. Viral glycoproteins: biological role and application in diagnosis. *VirusDisease*. 27(1):1-11, doi: 10.1007/s13337-015-0293-5, https://doi.org/10.1007/s13337-015-0293-5.
14. Bagdonaite, I., and H. H. Wandall. 2018. Global aspects of viral glycosylation. *Glycobiology*. 28(7):443-467, doi: 10.1093/glycob/cwy021, https://doi.org/10.1093/glycob/cwy021.
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, doi: https://doi.org/10.1016/j.coviro.2019.01.004, https://www.sciencedirect.com/science/article/pii/S1879625718301494.
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, doi: 10.1126/science.abb9983, https://science.sciencemag.org/content/sci/early/2020/05/01/science.abb9983.full.pdf.
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, doi: 10.1038/s41591-020-0820-9, https://doi.org/10.1038/s41591-020-0820-9.
18. Walls, A. C., Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 181(2):281-292.e286, doi: 10.1016/j.cell.2020.02.058, https://doi.org/10.1016/j.cell.2020.02.058.
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, doi: 10.1021/acs.analchem.0c03173, https://doi.org/10.1021/acs.analchem.0c03173.
20. Fernández, A. 2021. Glycosylation of SARS-CoV-2 Steers Evolutionary Outcomes in the Postvaccination Phase. *ACS Pharmacology & Translational Science*. 4(1):410-412, doi: 10.1021/acsptsci.1c00015, https://doi.org/10.1021/acsptsci.1c00015.
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, doi: https://doi.org/10.1016/j.jmb.2020.166762, https://www.sciencedirect.com/science/article/pii/S0022283620306872.

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, https://www.mdpi.com/1999-4915/13/5/927.
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, doi: 10.1101/2021.02.15.431212, https://www.biorxiv.org/content/biorxiv/early/2021/05/17/2021.02.15.431212.full.pdf.
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, doi: 10.1021/acs.jpcb.0c04553, https://doi.org/10.1021/acs.jpcb.0c04553.
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, doi: 10.1021/acs.jctc.0c01144, https://doi.org/10.1021/acs.jctc.0c01144.
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, doi: 10.1021/acscentsci.0c01056, https://doi.org/10.1021/acscentsci.0c01056.
27. Vuorio, J., I. Vattulainen, and H. Martinez-Seara. 2017. Atomistic fingerprint of hyaluronan–CD44 binding. *PLOS Computational Biology*. 13(7):e1005663, doi: 10.1371/journal.pcbi.1005663, https://doi.org/10.1371/journal.pcbi.1005663.
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, doi: 10.1038/s41598-021-84569-z, https://doi.org/10.1038/s41598-021-84569-z.
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, doi: 10.1371/journal.pone.0187994, https://doi.org/10.1371/journal.pone.0187994.
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), doi: 10.3389/fpls.2019.00768, https://www.frontiersin.org/article/10.3389/fpls.2019.00768.
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, doi: 10.1038/srep08926, https://doi.org/10.1038/srep08926.

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, doi: https://doi.org/10.1016/j.bpj.2021.02.007, https://www.sciencedirect.com/science/article/pii/S0006349521001417.
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, doi: 10.1038/s41586-020-2179-y, https://doi.org/10.1038/s41586-020-2179-y.
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, doi: 10.1038/s41467-020-18319-6, https://doi.org/10.1038/s41467-020-18319-6.
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, doi: 10.1016/j.tibtech.2005.09.003, https://doi.org/10.1016/j.tibtech.2005.09.003.
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, doi: 10.1371/journal.pone.0054836, https://doi.org/10.1371/journal.pone.0054836.
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), doi: 10.3389/fpls.2014.00473, https://www.frontiersin.org/article/10.3389/fpls.2014.00473 (Mini Review).
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. Deuflhard, 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, doi: 10.1126/science.abb2762, https://science.sciencemag.org/content/sci/367/6485/1444.full.pdf.
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, doi: 10.1073/pnas.1108455108, https://www.pnas.org/content/pnas/108/31/12669.full.pdf.
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, doi: 10.1074/jbc.M311191200, https://doi.org/10.1074/jbc.M311191200.

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, doi: https://doi.org/10.1016/j.jinorgbio.2019.110955, https://www.sciencedirect.com/science/article/pii/S0162013419305549.
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, doi: 10.1016/j.tibs.2006.12.007, https://doi.org/10.1016/j.tibs.2006.12.007.
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. http://glycam.org
46. Bernardi, A. 2020. https://github.com/austenb28/MCPB_Glycam_merge
47. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*. 126(1):014101, doi: doi:http://dx.doi.org/10.1063/1.2408420, http://scitation.aip.org/content/aip/journal/jcp/126/1/10.1063/1.2408420.
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, http://dx.doi.org/10.1002/jcc.540130805.
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, doi: https://doi.org/10.1016/j.softx.2015.06.001, http://www.sciencedirect.com/science/article/pii/S2352711015000059.
52. Nosé, S. 1984. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*. 81(1):511-519, doi: doi:http://dx.doi.org/10.1063/1.447334, http://scitation.aip.org/content/aip/journal/jcp/81/1/10.1063/1.447334.
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, doi: https://doi.org/10.1016/j.cpc.2021.107931, https://www.sciencedirect.com/science/article/pii/S0010465521000631.

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, doi: 10.1021/ac60214a047, https://doi.org/10.1021/ac60214a047.
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, doi: https://doi.org/10.1016/j.ab.2008.03.035, https://www.sciencedirect.com/science/article/pii/S0003269708001437.
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, https://doi.org/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, doi: 10.1038/s41467-021-21118-2, https://doi.org/10.1038/s41467-021-21118-2.
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, doi: 10.1038/s41422-021-00496-8, https://doi.org/10.1038/s41422-021-00496-8.
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, doi: doi:10.1128/JVI.79.19.12311-12320.2005, https://journals.asm.org/doi/abs/10.1128/JVI.79.19.12311-12320.2005.

6.7 Supporting Information

6.7.1 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.

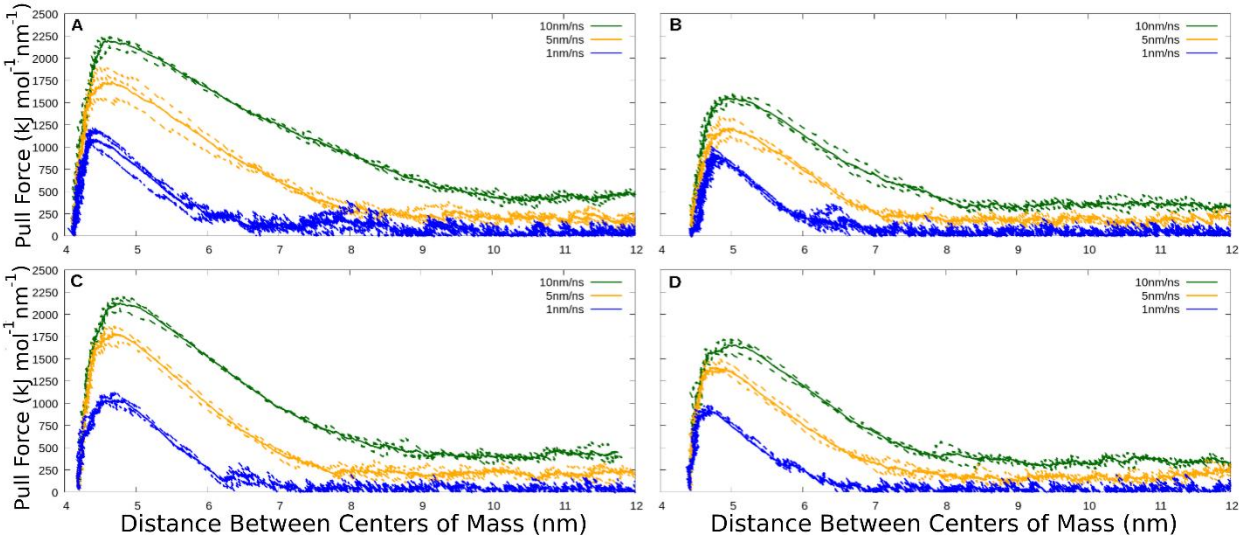


Figure S1: Traces of pull force versus distance. A) Man8 Glycosylated A1Fr^{M8}/SpFr. B) Non-glycosylated Structure A1Fr/SpFr. C) GnGnXF Glycosylated A2Fr^{GG}/SpFr. D) Non-glycosylated Structure A2Fr/SpFr. Different colors indicate different pull rate. Dashed lines are individual replicas, solid lines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

6.7.2 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 dataframe 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.

6.7.3 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.

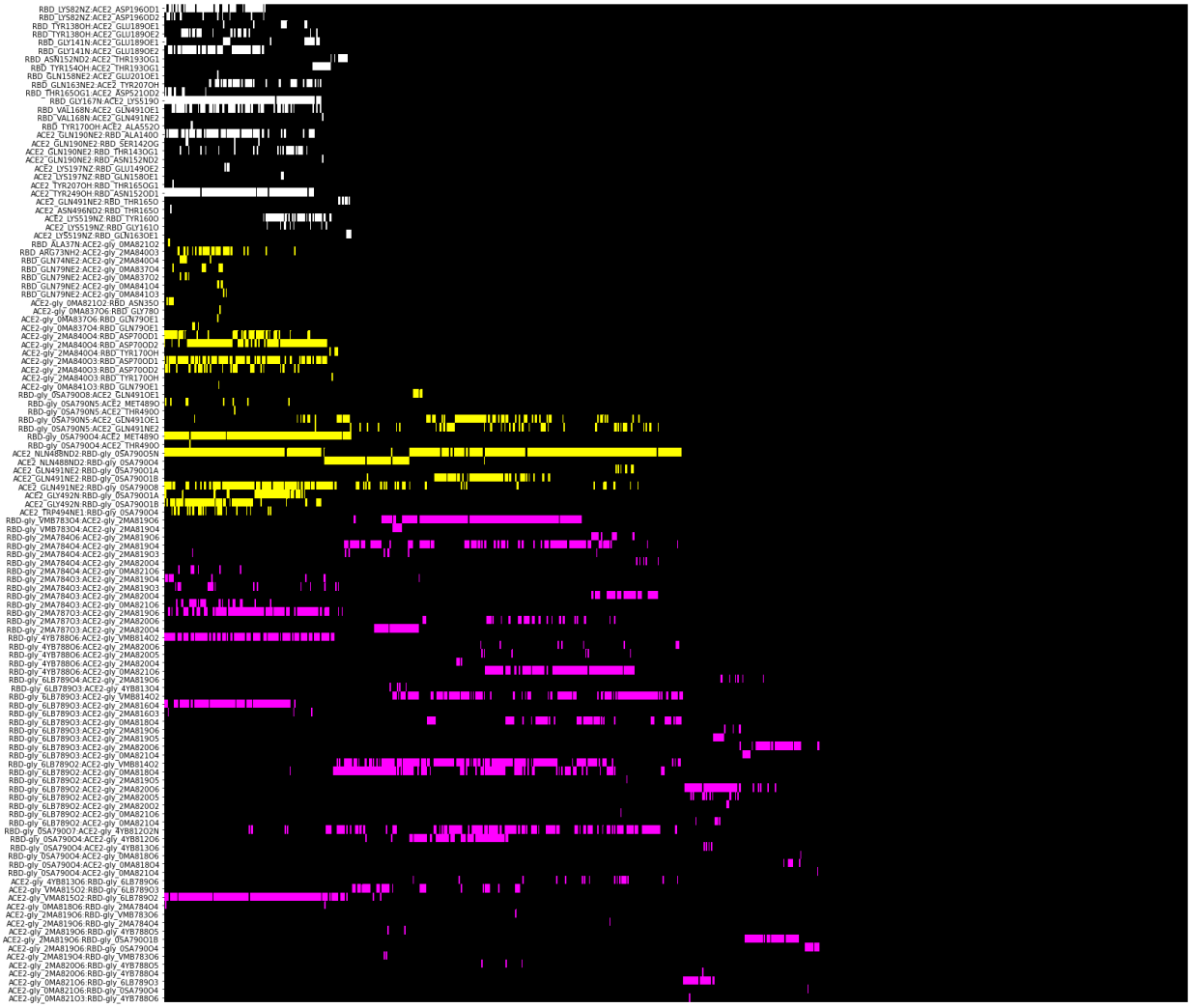


Figure S2. Hydrogen bonding donor:acceptor pairs vs simulated time for A1Fr^{M8}/SpFr. 1 ns / nm pulling speed. Colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan

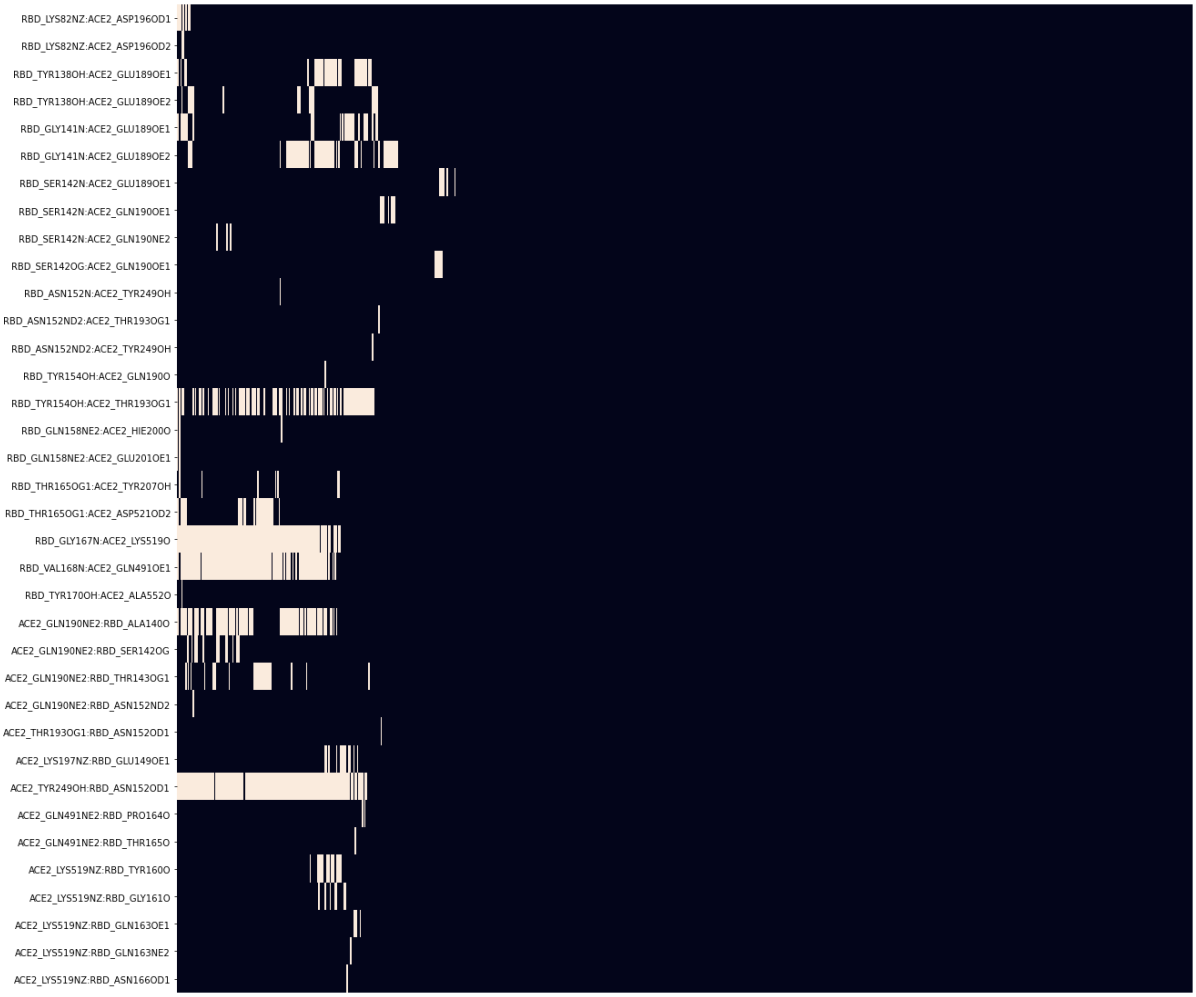


Figure S3. Hydrogen bonding donor:acceptor pairs vs simulated time for A1Fr/SpFr. 1 ns / nm pulling speed. Colors indicate interaction type: White: protein-protein

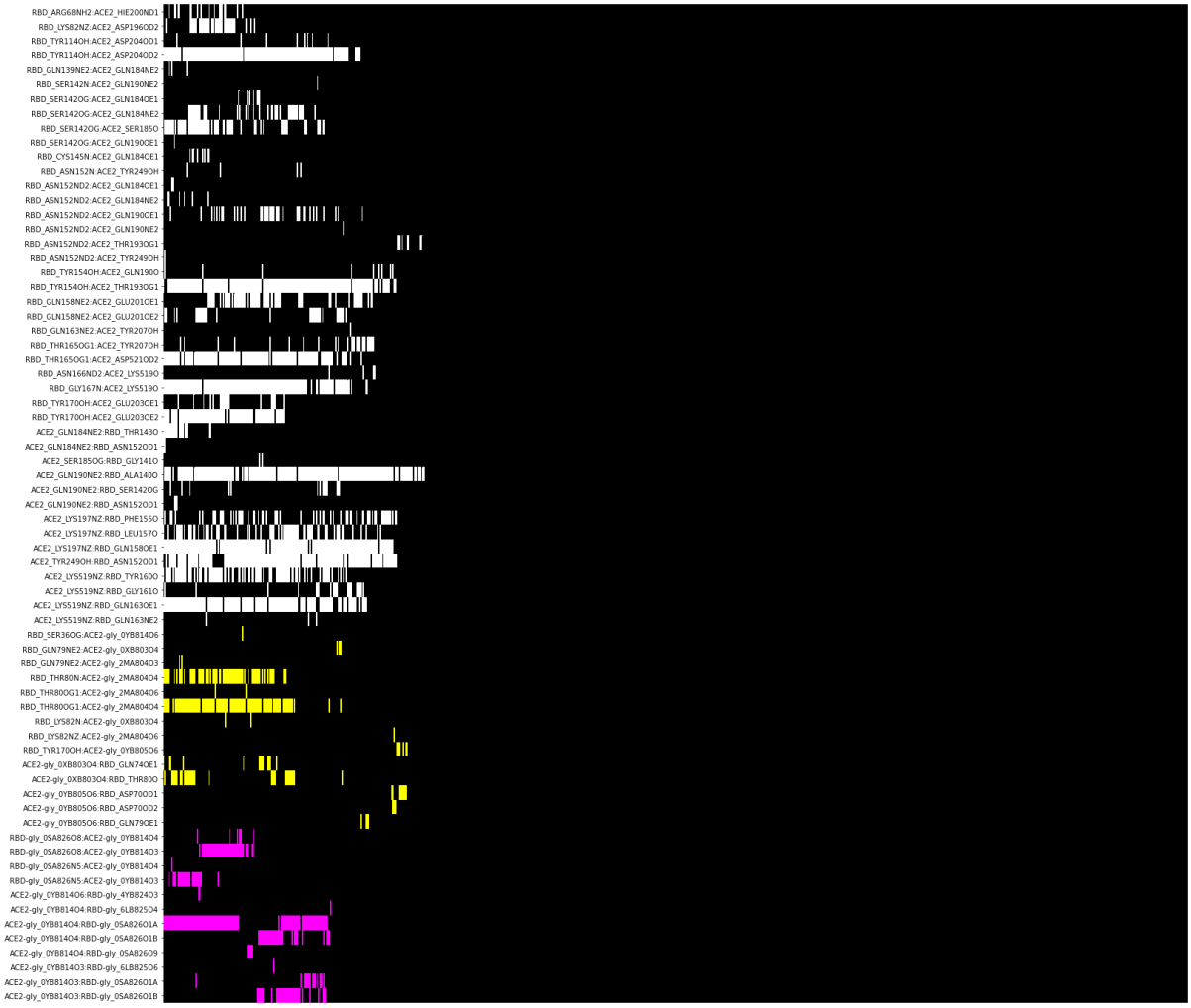


Figure S4. Hydrogen bonding donor:acceptor pairs vs simulated time for A2Fr^{GG}/SpFr. 1 ns / nm pulling speed. Colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan

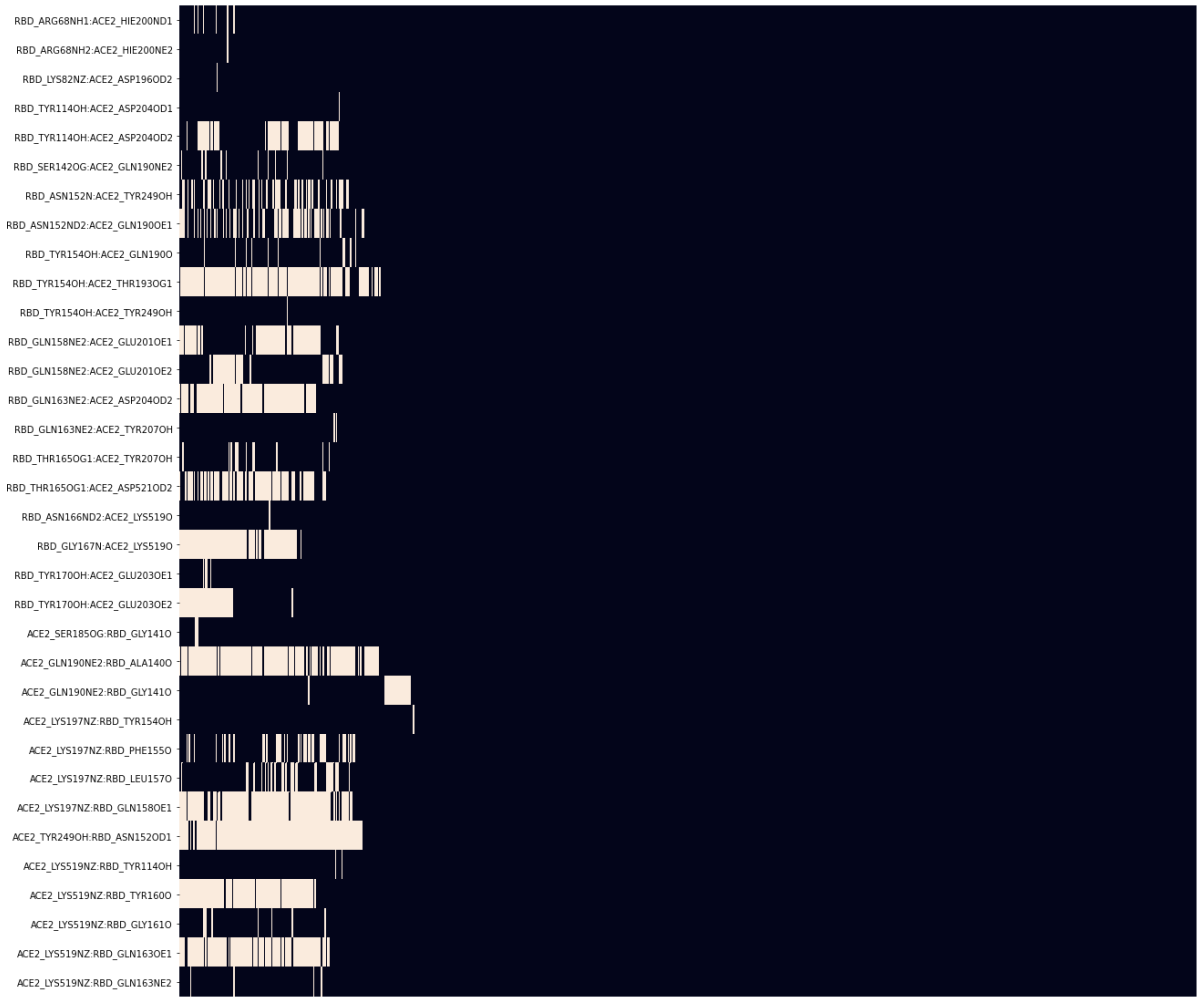
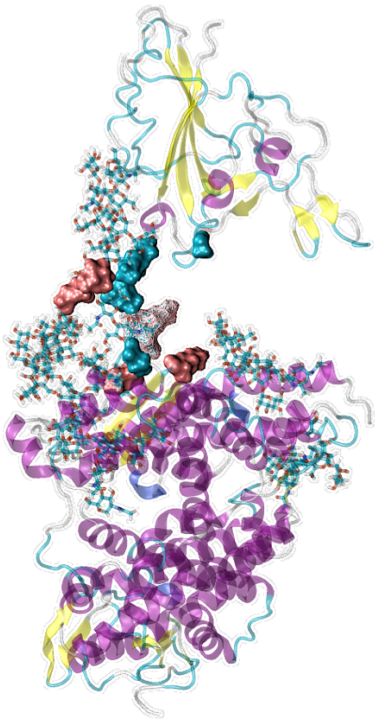


Figure S5. Hydrogen bonding donor:acceptor pairs vs simulated time for A2Fr/SpFr. 1 ns / nm pulling speed. Colors indicate interaction type: White: protein-protein

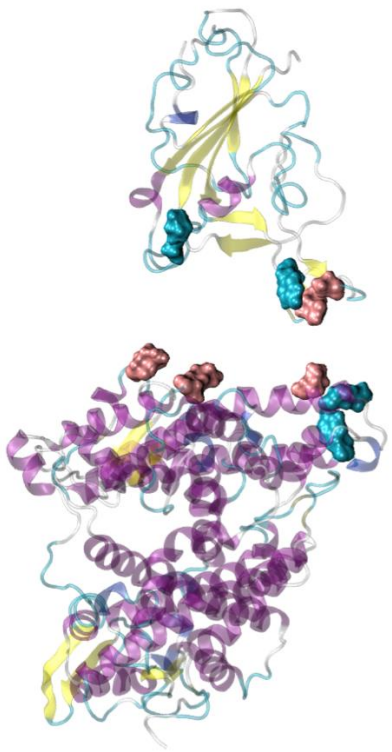
6.7.4 Hydrogen Bond Occupancy

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 AnaF[6] 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.



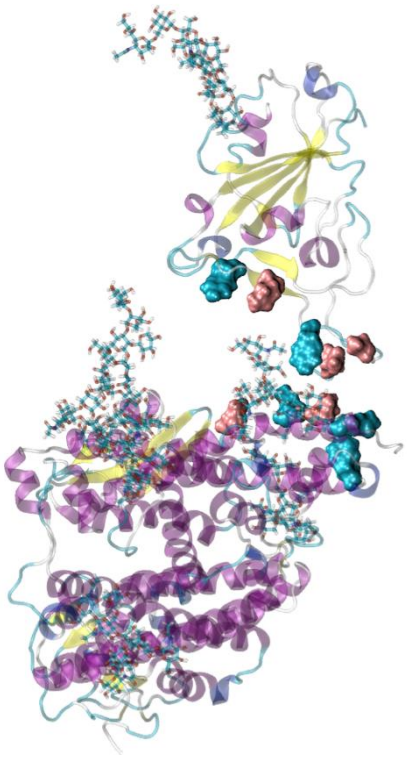
Glycosylated ACE2_Man8 and RBD Top 25 Hbonds	
Donor Acceptor Pair (Donor : Acceptor)	**Total Occupancy (%)**
RBD-gly_6LB789O2:ACE2-gly_VMA815O2	24.6
ACE2_NLN488ND2:RBD-gly_0SA790O5N	20.2
RBD-gly_0SA790O4:ACE2_MET489O	19.8
RBD_GLY167N:ACE2_LYS519O	18.3
RBD-gly_4YB788O6:ACE2-gly_VMB814O2	18.1
RBD-gly_6LB789O3:ACE2_gly_2MA816O4	17.9
ACE2_TYR249OH:RBD_ASN152OD1	17.9
ACE2-gly_2MA840O4:RBD_ASP70OD1	15.6
RBD-gly_2MA787O3:ACE2-gly_2MA819O6	15.4
ACE2_GLN190NE2:RBD_ALA140O	14.2
ACE2_GLY492N:RBD-gly_0SA790O1A	13.8
ACE2_GLY491NE2:RBD-gly_0SA790O8	13.0
RBD_VAL168N:ACE2_GLN491OE1	11.3
RBD-gly_2MA784O3:ACE2-gly_2MA819O3	11.0
ACE2-gly_2MA840O3:RBD_ASP70OD1	11.0
ACE2_TRP494NE1:RBD-gly_0SA790O4	8.7
RBD_GLY141N:ACE2_GLU189OE1	8.1
ACE2-gly_VMB814O2:RBD-gly_6LB789O2	7.3
ACE2_GLY492N:RBD-gly_0SA790O1B	7.2
RBD_TYP138OH:ACE2_GLU189OE2	6.3
ACE2-gly_2MA840O3:RBD_ASP70OD2	5.4
RBD_GLY141N:ACE2_GLU189OE2	4.7
RBD_LYS82NZ:ACE2_ASP196OD1	4.3
ACE2-gly_2MA819O3:RBD-gly_6LB789O2	4.3
RBD_TYR154OH:ACE2_THR193OG1	4.0

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.



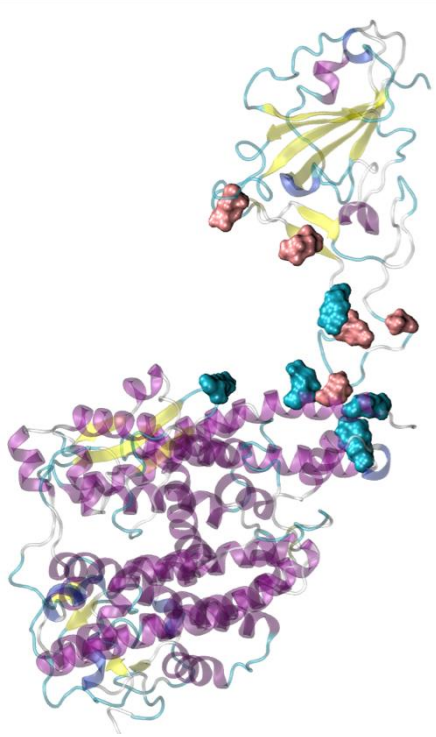
NonGlycosylated ACE2_MAN8 and RBD Top 25 Hbonds	
Donor Acceptor Pair (Donor : Acceptor)	**Total Occupancy (%)**
ACE2_TYR249OH:RBD_ASN152OD1	18.0
RBD_GLY167N:ACE2_LYS519O	15.6
RBD_VAL168N:ACE2_GLN491OE1	14.2
RBD_TYR154OH:ACE2_THR193OG1	11.2
ACE2_GLN190NE2:RBD_ALA140O	10.5
RBD_GLY141N:ACE2_GLU189OE2	7.4
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 GnGnXF^{3} glycosylated A2Fr^{GG}/SpFr. Table colors indicate interaction type: White: protein-protein, Yellow: protein-glycan, Magenta: glycan-glycan.



NonGlycosylated ACE2_GnGnXF and RBD Top 25 Hbonds	
Donor Acceptor Pair (Donor : Acceptor)	**Total Occupancy (%)**
ACE2_TYR249OH:RBD_ASN152OD1	19.8
ACE2_GLN190NE2:RBD_ALA140O	19.2
RBD_TYR154OH:ACE2_THR193OG1	18.7
ACE2_LYS197NZ:RBD_GLN158OE1	16.8
ACE2_LYS519NZ:RBD_GLN163OE1	14.6
ACE2_LYS519NZ:RBD_TYR160O	14.5
RBD_GLN163NE2:ACE2_ASP204OD2	14.0
RBD_GLY167N:ACE2_LYS519O	12.5
RBD_THR165OG1:ACE2_ASP521OD2	11.2
RBD_GLN158NE2:ACE2_GLU201OE1	9.7
RBD_TYR114OH:ACE2_ASP204OD2	8.6
RBD_ASN152ND2:ACE2_GLN190OE1	8.6
RBD_ASN152N:ACE2_TYR249OH	6.7
RBD_TYR170OH:ACE2_GLU203OE2	6.2
ACE2_LYS197NZ:RBD_PHE155O	5.6
RBD_GLN158NE2:ACE2_GLU201OE2	5.2
ACE2_LYS197NZ:RBD_LEU157O	4.4
ACE2_GLN190NE2:RBD_GLY141O	3.1
RBD_THR165OG1:ACE2_TYR207OH	1.7
RBD_TYR154OH:ACE2_GLN190O	1.5
RBD_SER142OG:ACE2_GLN190NE2	1.4
ACE2_LYS519NZ:RBD_GLY161O	1.1
RBD_ARG68NH1:ACE2_HIE200ND1	0.8
RBD_TYR170OH:ACE2_GLU203OE1	0.6
ACE2_LYS519NZ:RBD_GLN163NE2	0.6

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.

6.7.5 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.

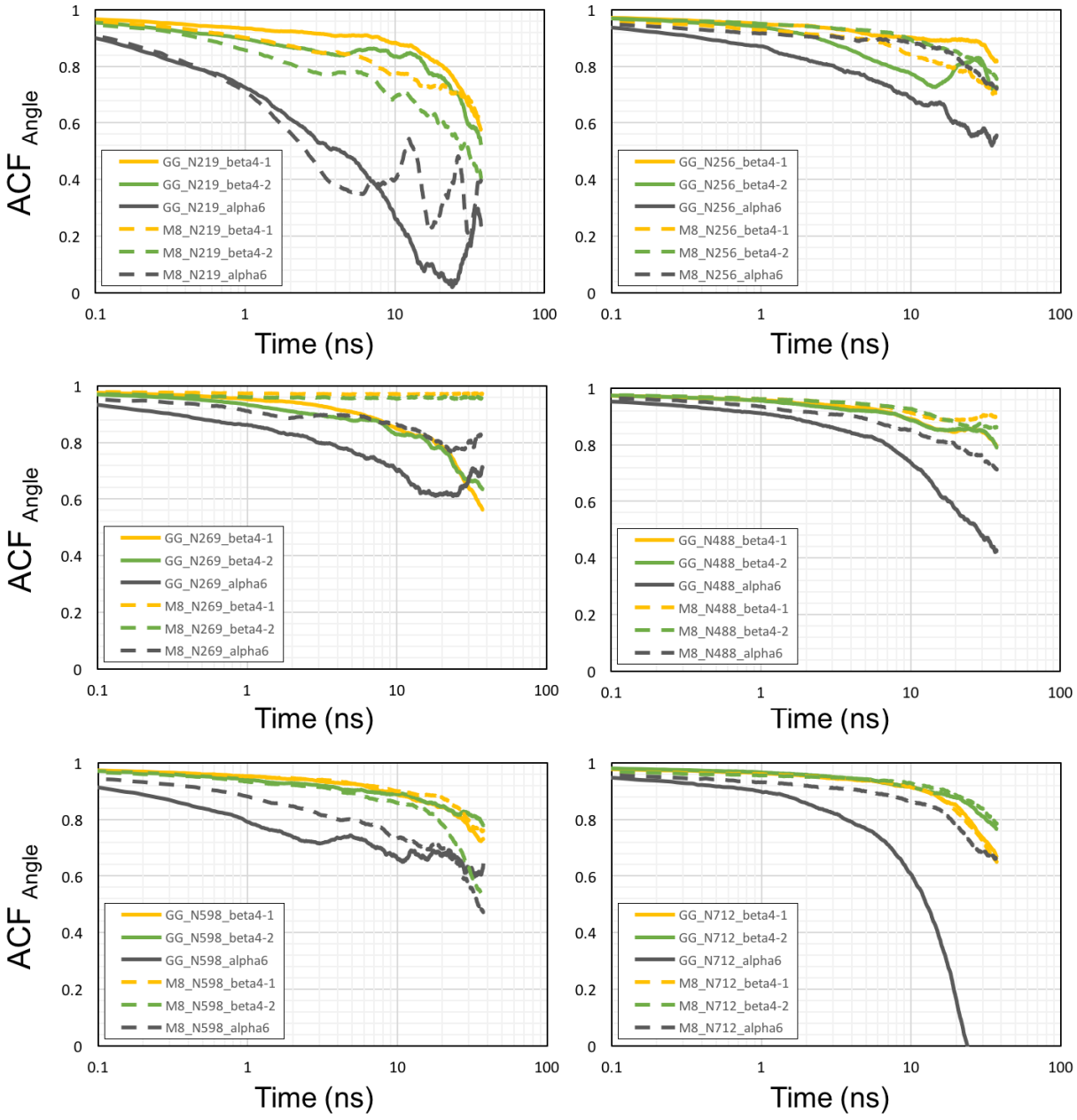


Figure S10. Autocorrelation function analysis of angles at linkage beta4_1, beta4_2, and alpha6 of MAN8 and GnGnXF^3 at ACE2 fragment glycosylation sites in semi-log plots. Dashed lines are the dynamic motions of MAN8, and solid lines are the dynamic motions of GnGnXF^3.

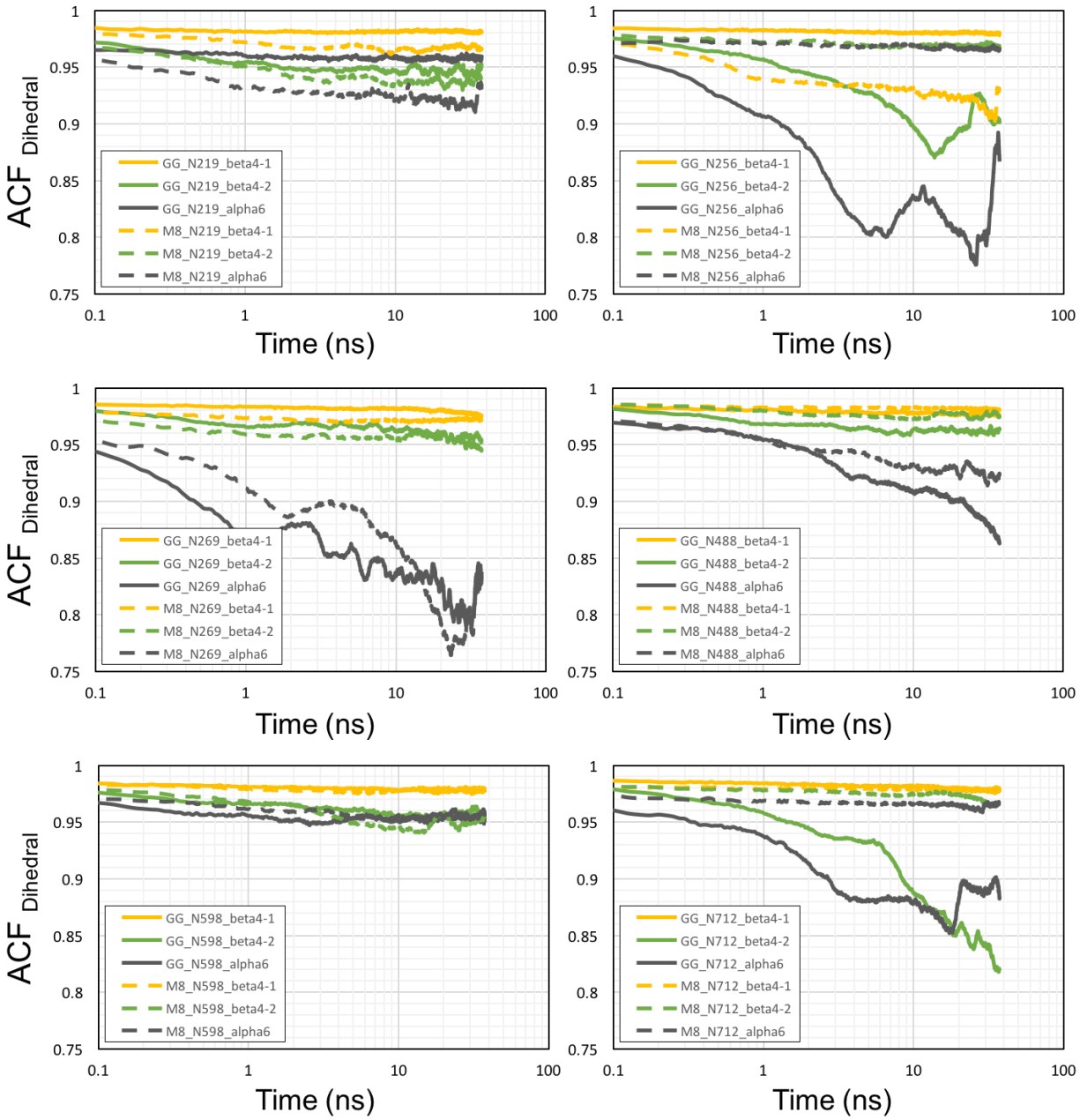


Figure S11. Autocorrelation function analysis of dihedrals at linkage beta4_1, beta4_2, and alpha6 of MAN8 and GnGnXF^3 at ACE2 fragment glycosylation sites in semi-log plots. Dashed lines are the dynamic motions of MAN8, and solid lines are the dynamic motions of GnGnXF^3.

6.7.6 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.

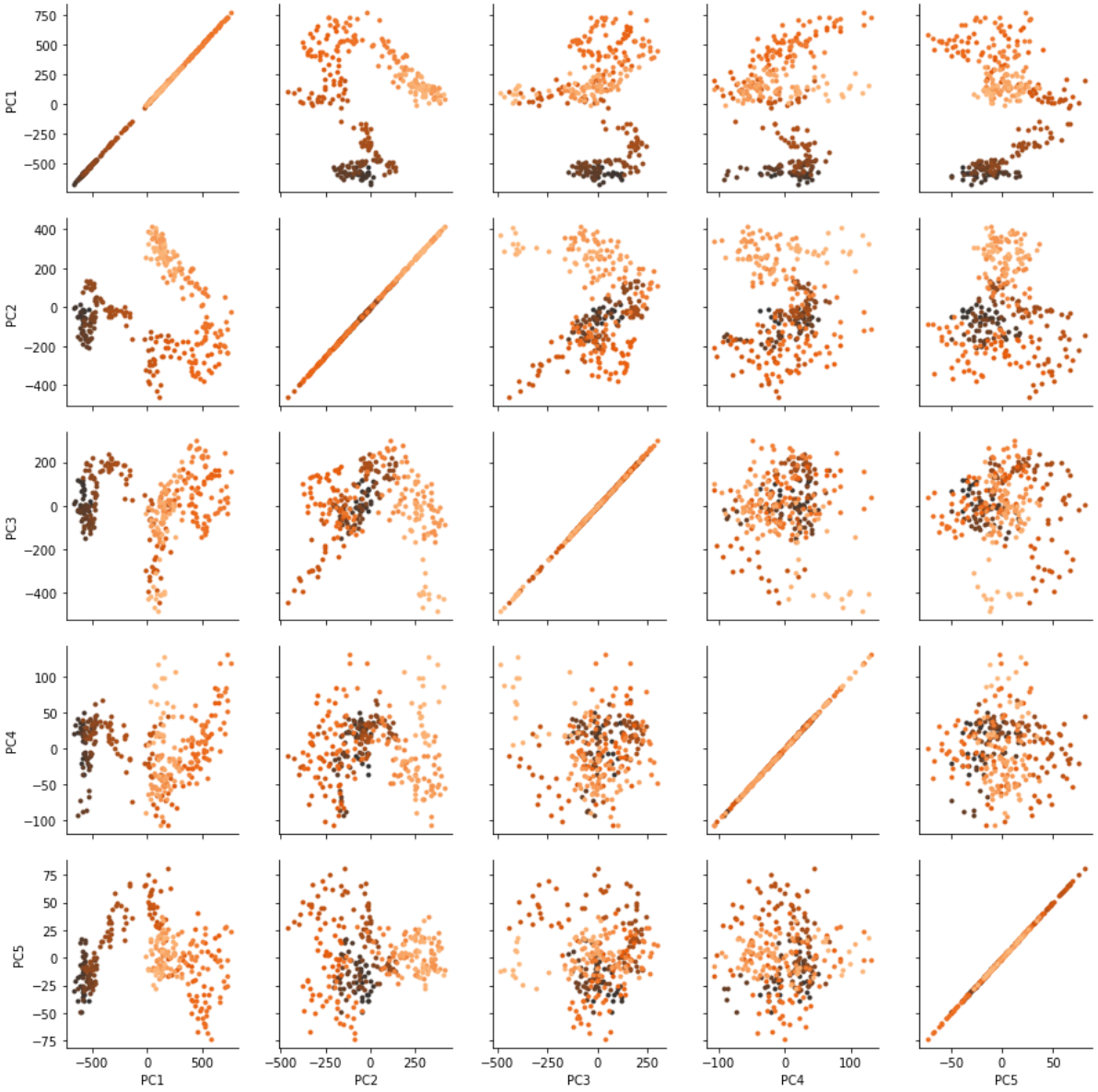


Figure S12. Principal component pair-wise interaction map for A1 variant system. First 5 principal components are shown. Color corresponds to time and goes black to dark orange.

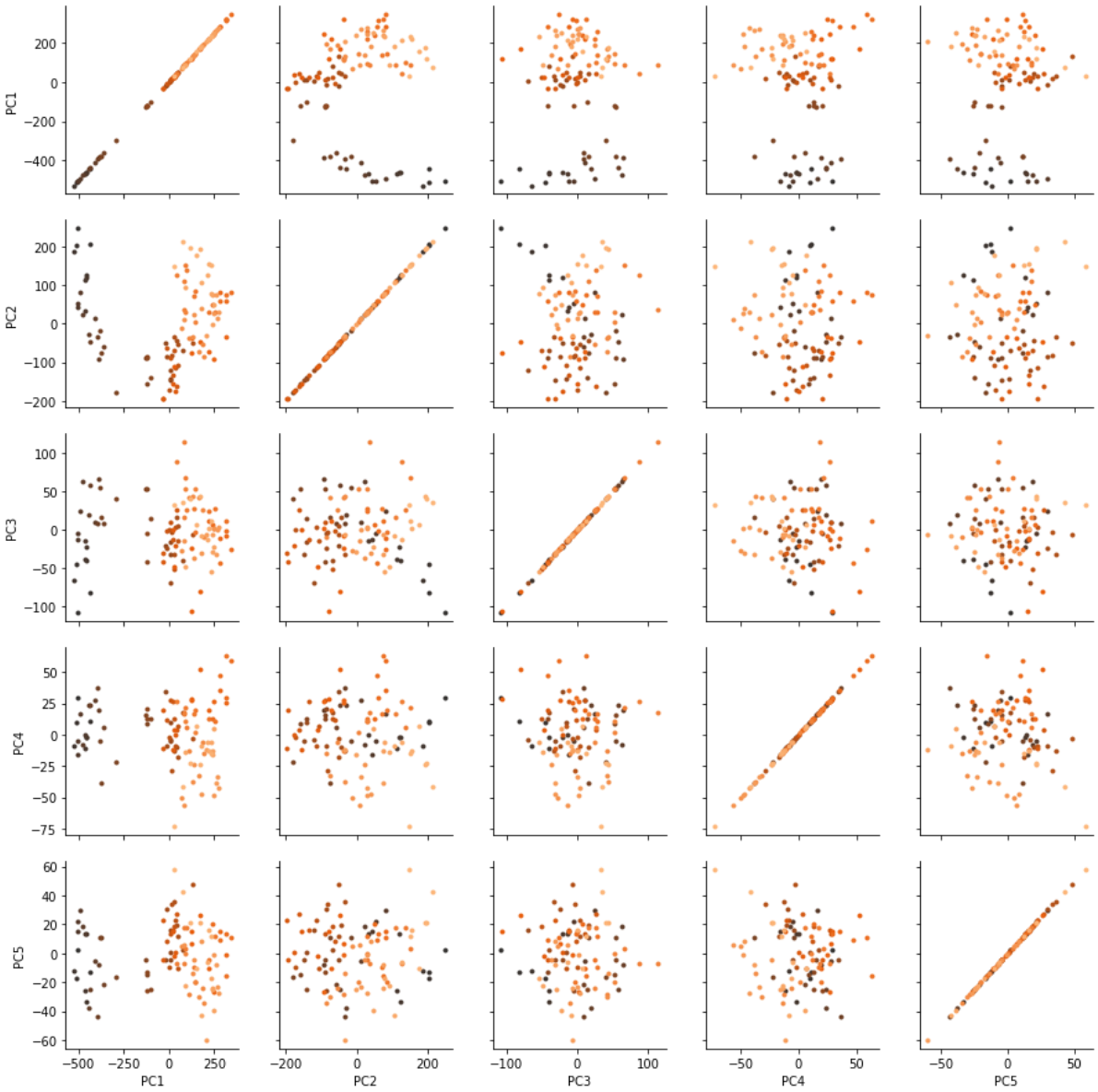


Figure S13. Principal component pair-wise interaction map for A2 variant system. First 5 principal components are shown. Color corresponds to time and goes black to dark orange.

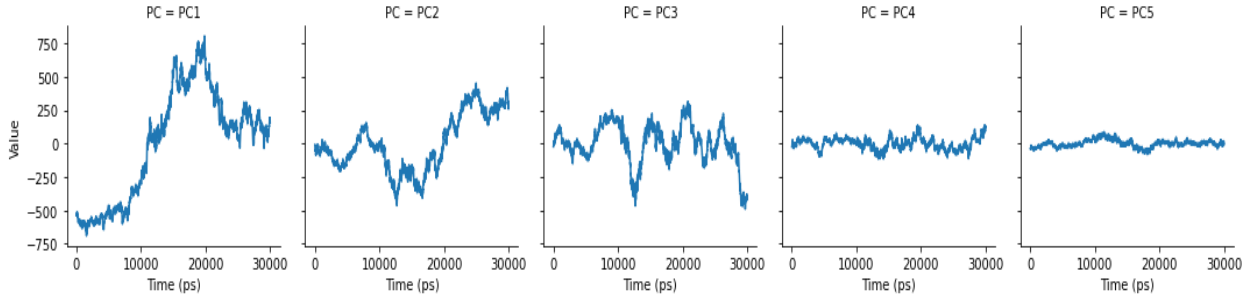


Figure S14. Principal component vs time for A1 variant system.

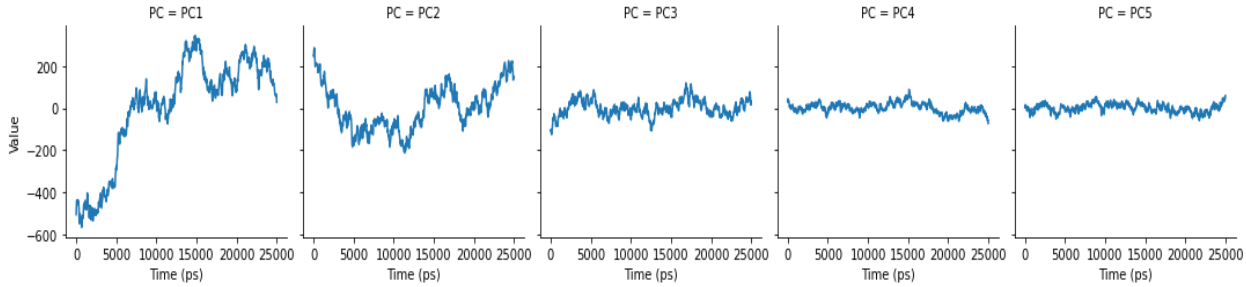


Figure S15. Principal component vs time for A2 variant system.

Chapter 7

Production of novel Spike fragments in Chinese hamster ovary cells

My contribution to this work was the simulation and analysis of and related to spike fragments T1, T3, T3 modified, and RBD with HIS tag with associated writing and editing.

This work is currently in preparation for submission to biotechnology and bioengineering (2021)

(preparation draft as of 11/22/2021)

Shiaki A. Minami[1], Seongwon Jung[1], Yihan Huang[2,#], Bradley S. Harris[1,#], Matthew W. Kenaston[3], Roland Faller[1], Somen Nandi[1,4], Karen A. McDonald[1,4], Priya S. Shah[1,3,*]

1 Department of Chemical Engineering
2 Department of Materials Science and Engineering
3 Department of Microbiology & Molecular Genetics
4 Global HealthShare Initiative
University of California, Davis
# These authors contributed equally
* Correspondence: prsshah@ucdavis.edu

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 first improve 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 3 days post-transfection. 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 is higher than our RBD titers. Evaluating purified proteins for binding to monoclonal and polyclonal antibodies raised against full-length Spike, we find T1 has the same sensitivity as Spike against the monoclonal antibody and even outperforms Spike for the polyclonal antibody. These results suggest T1 is a promising Spike alternative for use in clinical 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 XXX infections and XXX deaths globally since November 2019. Major aspects of containing this global pandemic are surveillance (large-scale and rapid asymptomatic testing) and herd immunity (an 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, 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, respectively (Huang et al., 2020). The receptor-binding domain (RBD) of Spike lies within the S1 domain (Fig. 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 (Heath et al., 2021).

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 5-30mg, with one report of 150 mg/L (Amanat et al., 2020; Hsieh et al., 2020; Stuible et al., 2021). Low expression of Spike precludes rapid, cost-effective production. A possible alternative is to express the Spike receptor binding domain (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 (Amanat et al., 2020). 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 (Smaoui & Yahyaoui, 2021; Starr et al., 2020). Rational structure-based approaches have also been used to improve stability of full-length Spike (Hsieh et al., 2020). Nonetheless, identifying sequence-independent methods to increase expression is essential for matching Spike in existing variants as closely as possible, as well as determining Spike fragments that have high expression and antibody binding in diagnostic applications.

In this work, we transiently express Spike and RBD in Chinese Hamster Ovary (CHO) cells, an industrial workhorse for mammalian glycoprotein production. 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 protein than the other and has higher binding to anti-Spike antibodies. Taken together, these alternative proteins 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 (Amanat et al., 2020). 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. 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 digestion of pCAGGS-RBD with XbaI and XhoI, followed by Gibson Assembly. 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 PredictProtein (Yachdav et al., 2014). Starting and ending residues were selected to avoid interrupting major secondary structures (Kelley et al., 2015; Meng et al., 2006; Pettersen et al., 2021).

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% CO_2, 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 x 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 proteins (Esposito et al., 2020). 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 Supplementary 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 described (Xiong et al., 2019). 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), which were developed against full-length Spike. 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. Plates were loaded with 100 µL/well goat anti-hIgG, HRP secondary

antibody at 1:4000 in PBS containing 1% BSA. Plates were developed. 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 nm (Ernst & Zor, 2010), 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 stain-free SDS-PAGE gel. Bands were extracted and directly submitted for proteomic analysis. Searches were conducted against the known sequences of T1 and T4, and alignments were performed using Multiple Alignment using Fast Fourier Transform (Katoh et al., 2019).

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 (Micsonai et al., 2015). 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 (Frishman & Argos, 1995).

7.2.9 Simulations:

Starting configurations for molecular dynamic 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 modeller (Webb & Sali, 2016), which modifies amino acid sequences of proteins. The newly HIS tagged structures were prepared for and had glycans attached using Glycam (Woods Group. GLYCAM Web Athens, GA: Complex Carbohydrate Research Center, University of Georgia; 2005–2020 [Available from: http://glycam.org]). RBD and it's N-glycosylation sites had the glycoform FA2 attached, this is true for the RBD portion of all truncations. T1 contained no additional N-glycosylation sites, T3 contained an additional FA3 glycoform, T4 contained an additional M5 glycoform. (Glycoforms used can be seen in (SI? Figure? attached pdf) Amber ff14SB and Glycam06 forcefields (Kirschner et al., 2008; Maier et al., 2015) were used and generated using acpype.py following the method shown previously (Bernardi et al., 2017, 2019). Simulations were conducted using the Gromacs 2019.1 suite with similar energy minimization procedure as in previous simulations (Abraham et al., 2015; Pronk et al., 2013; Van Der Spoel et al., 2005) including ones involving glycosylated RBD (Cite Bernardi 2017 and the BPJ paper/preprint). 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 (Fig. 1A). We also replaced the previously tested secretion signal with an alternative secretion signal to determine whether it affects expression and secretion of Spike and RBD (Spike-1, Spike-2, RBD-1, and RBD-2, Fig. 1B).

Relative amounts of protein in the supernatant and cell lysate were determined by western blot 5 days post-transfection (Fig. 1B and 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 (Fig. 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 (Fig. 1E-H). Spike concentration in the supernatant increased steadily until 7 days post-transfection, after which time it remained stable (Fig. 1E and 1F). In comparison, RBD

concentration in the supernatant peaked at day 3, then decreased (Fig. 1 G and 1H). From these results, Spike and RBD harvests were determined to be 7 and 3 days post-transfection, respectively.

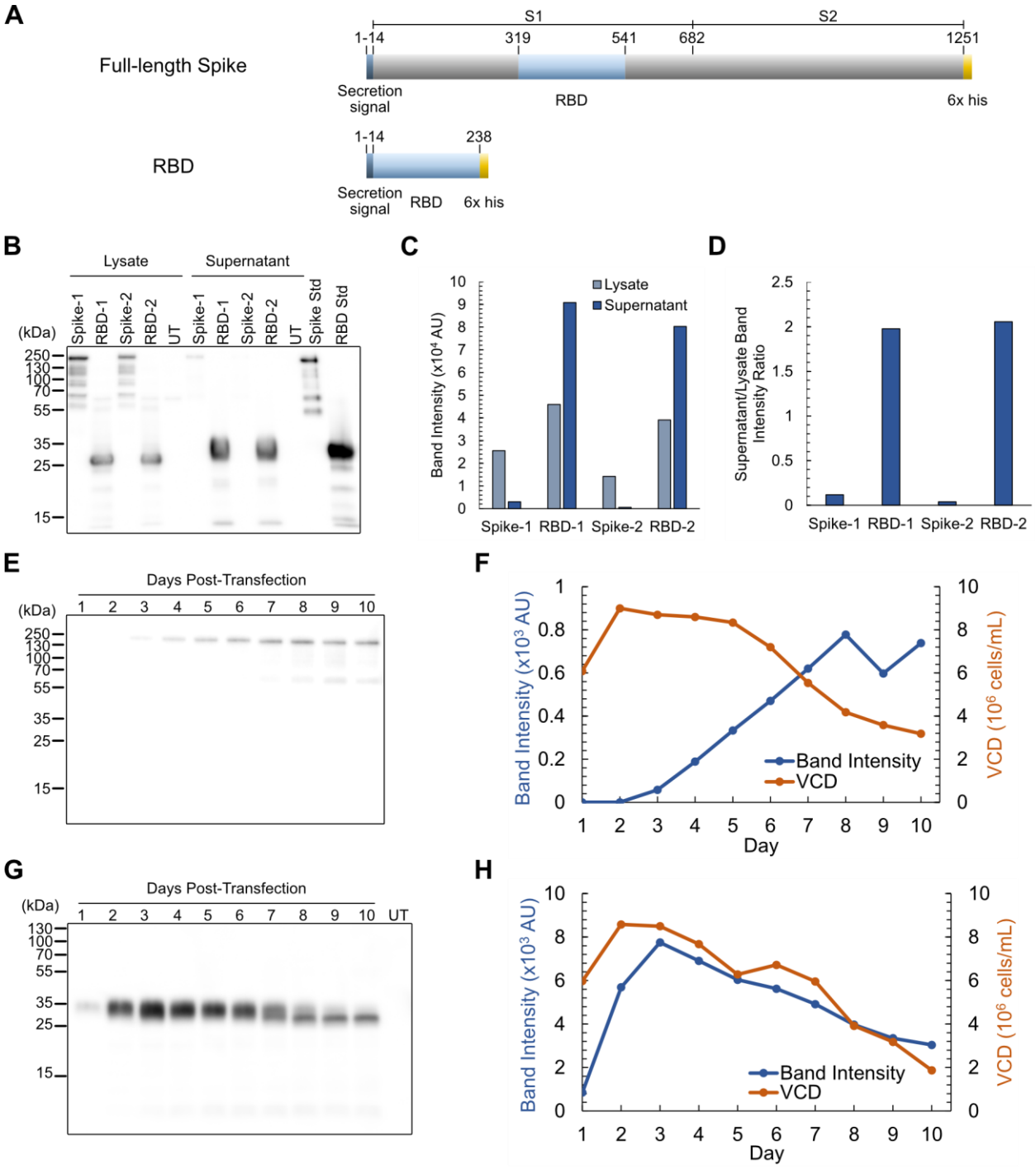


Figure 1. Transient Spike and RBD production in CHO cells. (**A**) Construct diagram of full-length Spike (1257 aa) and RBD (244 aa) constructs. Residues are labeled starting from the beginning of the secretion signal. (**B**) Western blot and (**C**) densitometry comparing two secretion signals for Spike and RBD. (**D**) Ratio of band intensities of supernatants and lysates. (**E**) Western blot and (**F**) densitometry on Spike expression timecourse. (**G**) Western blot and (**H**) densitometry on RBD expression timecourse. Abbreviations: aa (amino acids); untransfected (UT); standard (std); viable cell density (VCD).

To produce high 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 (Fig. S1). The crude supernatants were purified through FPLC (Fig. 2 and S2). For purification of Spike, SDS-PAGE revealed bands in elution fractions E3, E4, and E5 (Fig. 2A), which were confirmed by western blot to include immunoreactive bands consistent with full-length Spike (Fig. 2B). For RBD, SDS-PAGE (Fig. 2C) and western blot (Fig. 2D) showed bands in elution fractions E3, E4, and E7. FPLC-purified samples were dialyzed using PBS and concentrated through spin columns. E3 fractions of both proteins were used for subsequent experiments.

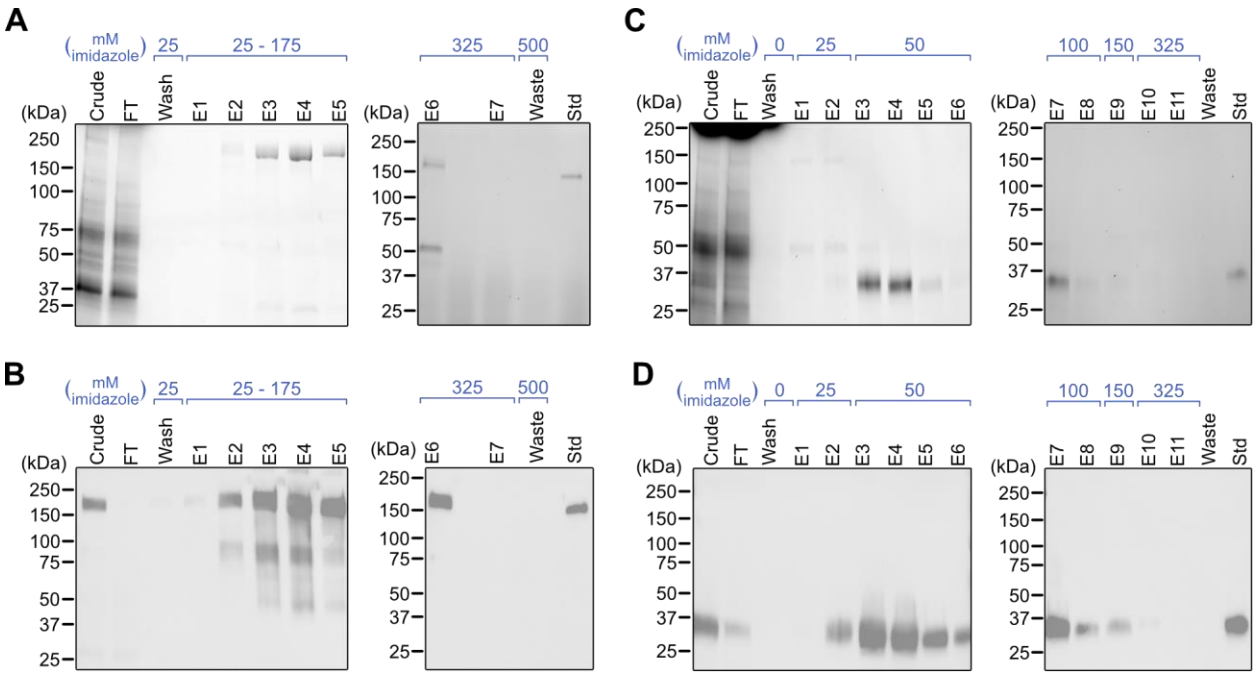


Figure 2. Purification of Spike and RBD.
(**A**) SDS-PAGE and (**B**) western blot on Spike fractions. (**C**) SDS-PAGE and (**D**) western blot on RBD fractions. FT samples are collected during sample loading onto the column. FT and wash samples are pooled from multiple fractions at equal volumes. Abbreviations: flow-through (FT); elution (E); standard (std).

7.3.2 Novel truncations to improve protein titers

Given the difficulty in expressing full-length Spike but its higher sensitivity in serological assays (Amanat et al.), we sought to determine whether other truncations that are larger than the RBD could be highly expressed and maintain high sensitivity. We developed eight truncations, T1-T8, by adding increments of approximately 50 amino acids on the N- or C-terminal ends of RBD (Fig. 3A). 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 (Fig. 3B). T1 and T4 had the highest expression, as well as the highest supernatant/lysate signals (Fig. 3C and 3D). In particular, T1 had even higher expression and relative secretion than RBD did. Given that S1 has previously been studied (Ren et al., 2020), T1 and T4 were selected for scaleup and purification. To determine optimal harvest dates for T1 and T4, cells were transfected in 25 mL of culture media and a sample of the supernatant was taken every 24 hours for 10 days. For T1, western blot and densitometry (Fig. 3E and 3F) showed that expression peaks 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 (Fig 3G and 3H). Therefore, T1 and T4 harvest dates were determined to be 3 and 4 days post-transfection, respectively.

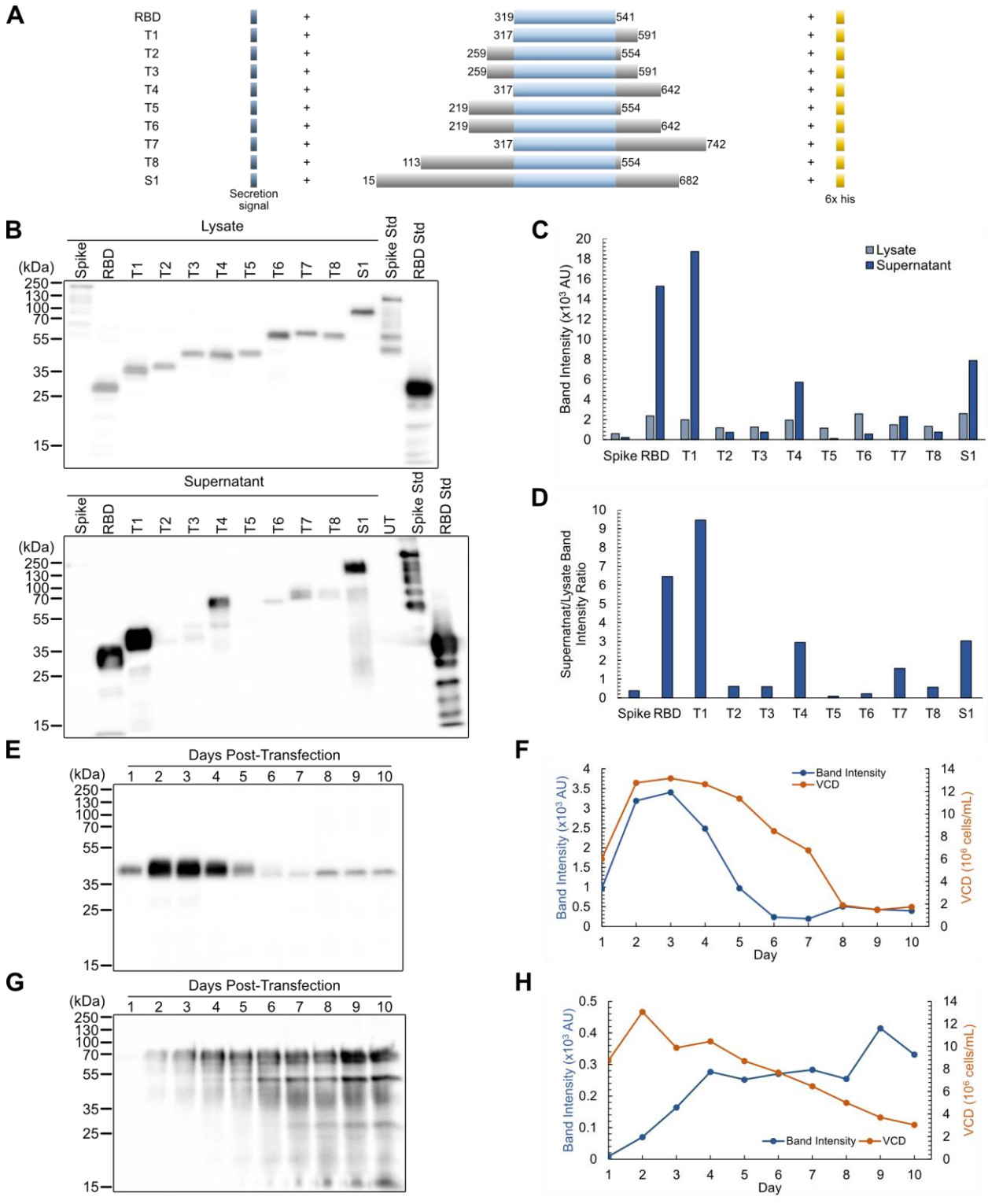


Figure 3. Expression of Spike truncations. (**A**) Construct diagram of Spike truncations. Residue numbers indicate the region in full-length Spike, including the secretion tag, in which the truncation resides. (**B**) Western blot and (**C**) densitometry on truncations for lysates and supernatants. (**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 150 mL cultures and samples were purified through FPLC (Fig. 4 and S3). SDS-PAGE revealed two bands for T1, one of which was determined to be T1 through western blot (Fig. 4A and 4B). To remove the impurity, T1 was dialyzed and repurified. For T4, three bands were detected- one band at 70 kDa, one band at 50 kDa, and one band at slightly below 50 kDa (Fig. 4C). Western blot detected bands in both the 70 kDa and 50 kDa regions (Fig. 4D), indicating two forms of T4. The impurity was removed through dialysis and repurification. Pure T1 and T4 were obtained in this manner and further prepared for further characterization through dialysis and spin column concentration. The bands detected for T1 and T4 through western blot were cut out of an SDS-PAGE gel and submitted for proteomic analysis. Coverage for T1 and T4 were very high, and the two bands of T4 on the SDS-PAGE gel suggest different post-translational modifications (Fig. S4). Quantifying crude titers, we found that T1 and T4 expressed at 130 mg/L and 73 mg/L, respectively, which are higher than our RBD titers (Fig. S5). For further characterization and binding assays, concentrations of purified samples were determined using Sandwich ELISAs and SDS-PAGE densitometry (Fig. S6).

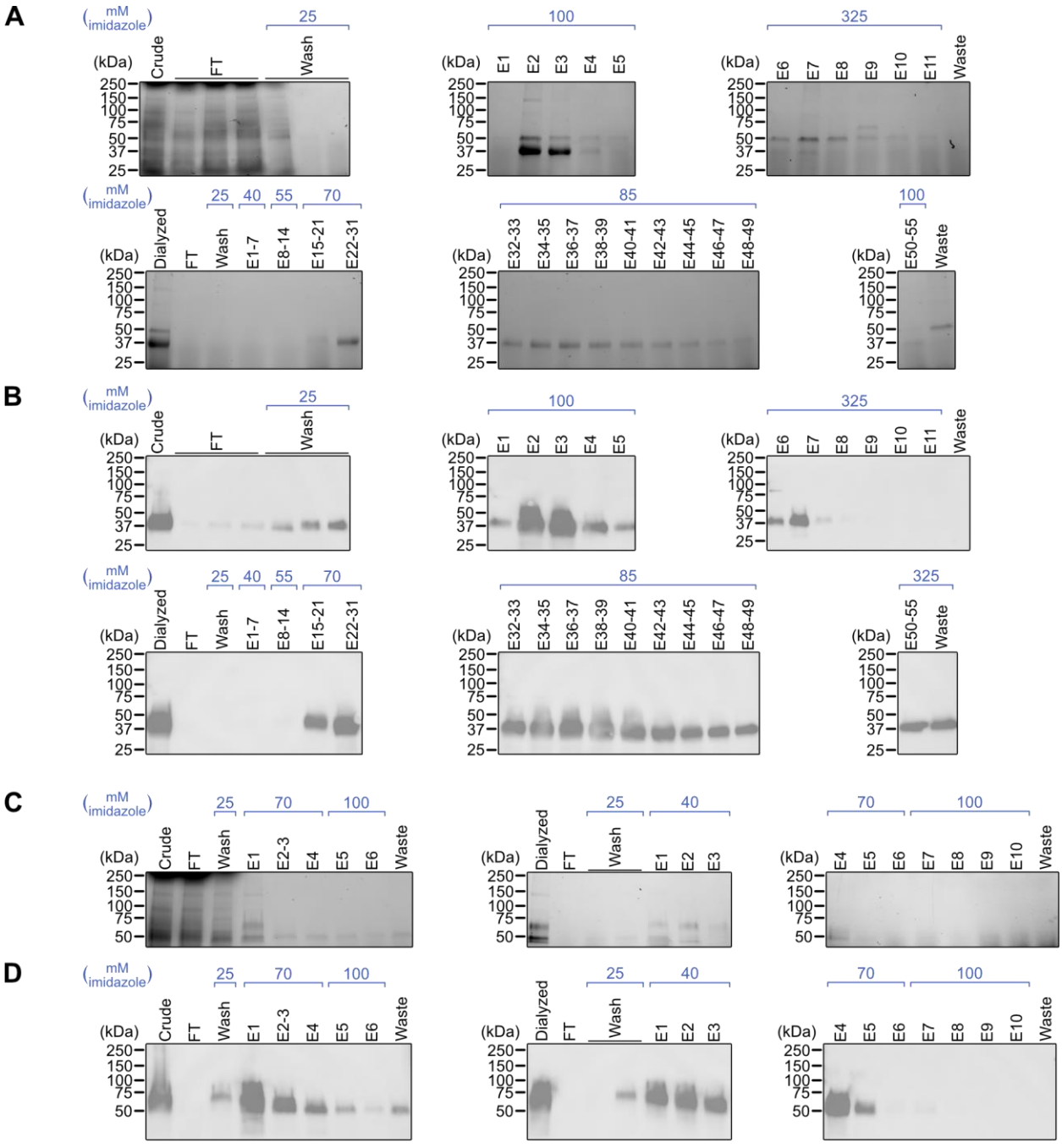


Figure 4. Purification of T1 and T4. (**A**) SDS-PAGE and (**B**) western blot on T1 fractions from crude purification and repurification. (**C**) SDS-PAGE and (**D**) western blot on T4 fractions from crude purification and repurification. FT and wash samples are pooled from multiple fractions at equal 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 Spike (Yuan et al., 2020). Serial dilutions of CR3022 were bound to Spike, RBD, T1, and T4 (Fig. 5A). Binding sensitivities were compared by taking the areas under the curves (Fig. 5B). To compare with another recombinant source of Spike, Sf9 insect Spike was also used in the assays.

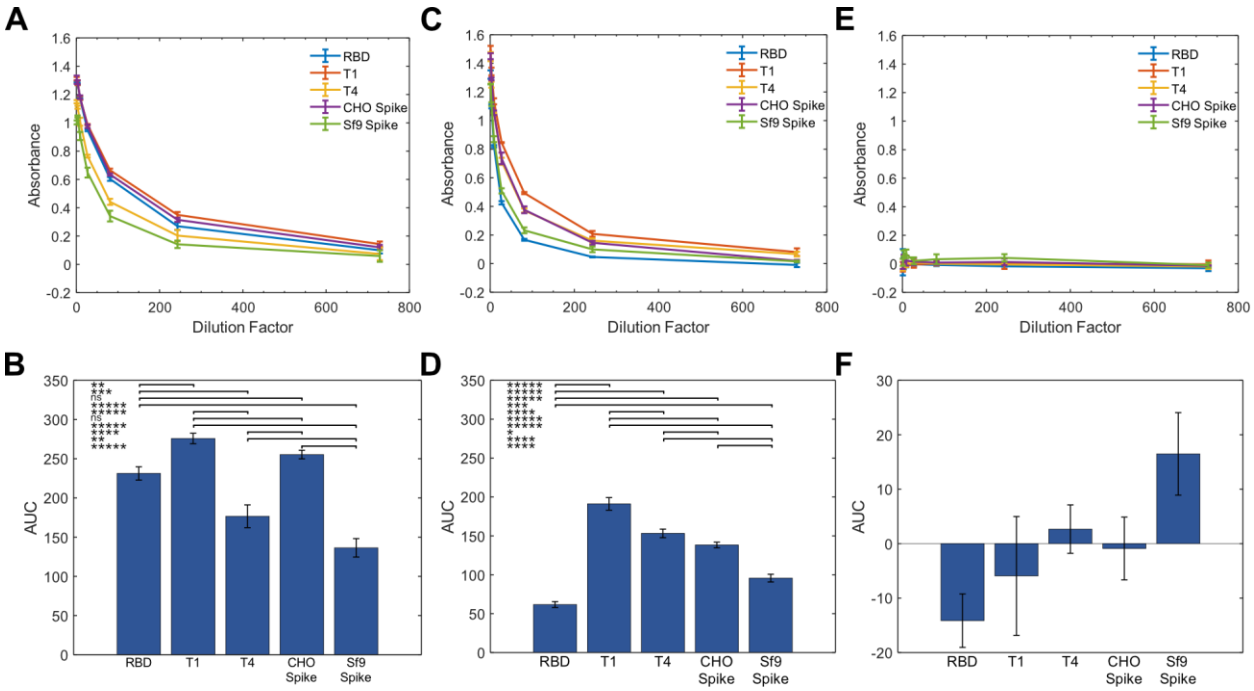


Figure 5. Binding assays of Spike truncations with anti-Spike antibodies.
(**A**) Absorbance as a function of dilution factor of CR3022. (**B**) AUC calculated from (A). (**C**) Absorbance as a function of dilution factor of an anti-Spike polyclonal antibody. (**D**) AUC calculated from (C). (**E**) Absorbance against dilution factor of a rabbit normal IgG antibody. (**F**) AUC calculated from (E). Antigens were loaded at 100 ng/well. Error bars represent ± SD of technical triplicates. p-values were calculated using a one-way ANOVA followed by Tukey's Test. * indicates $p < 5x10^{-2}$, ** $p < 5x10^{-3}$, *** $p < 5x10^{-4}$, **** $p < 5x10^{-5}$, and ***** $p < 5x10^{-6}$. Abbreviations: area under curve (AUC); ns (not significant).

CHO-expressed Spike has higher binding to CR3022 than Sf9-expressed Spike. This may be due to differences in folding or glycosylation between the insect and CHO 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 developed against full-length Spike were tested (Fig. 5C and 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 Sf9-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 (Fig. 5E and 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 molecular dynamics (MD). Snapshots of simulated structures of RBD, T1, and T4 at 0 ns and 100 ns simulation times are shown (Fig. 6A-D). The RBD portion of all structures remained stable in this duration. 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 (Fig. 6E and 6F). These RMSD plots show deviation relative to initial structures. The RMSD plots provide further evidence that the RBD sub-domains 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.

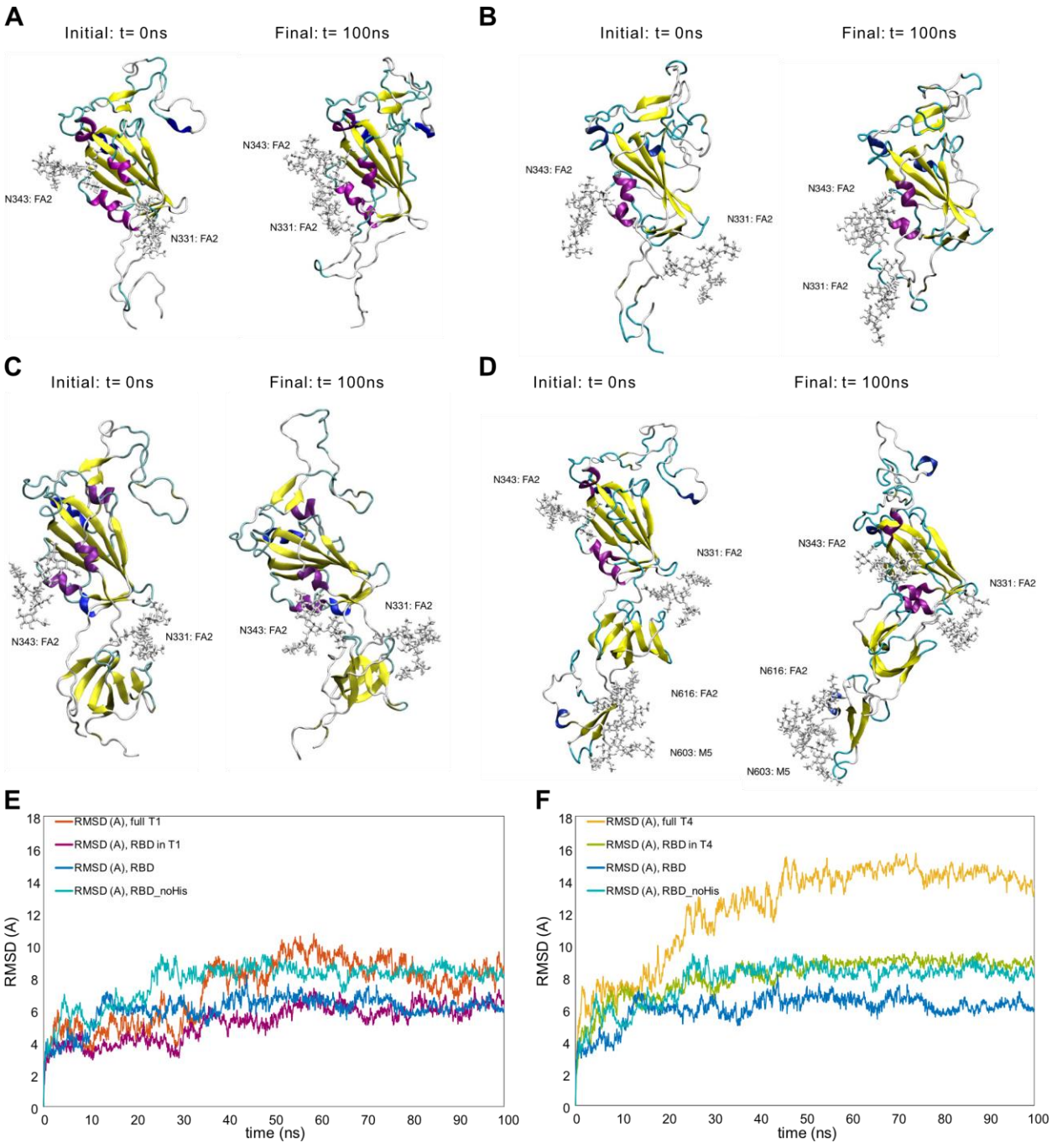


Figure 6. Molecular dynamics structural stability snapshots and analysis. MD snapshots are visualized for (**A**) RBD, (**B**) RBD without the 6x His tag, (**C**) T1, and (**D**) T4 at 0 ns and 100 ns.

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.

We hypothesize that some truncations that did not express well may also have had improper or unstable folding, which could lead to higher degradation or retention in the cells. To explore this idea, we compared T1 and T3, which only differ by ~50 amino acids at the N-terminal end but had vastly different expression (Fig. 3B and Fig. 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 (Fig. S7). Visualization of T3 revealed that the difference in RMSD was due to the additional FA3 glycan self-binding to its own RBD, which could contribute to the lack of expression. Removal of the self-binding glycan resulted in RMSD and secondary structure behavior that matched more closely to T1 (Figure S??). 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). Δε values were obtained, which is a measure of the difference in absorbance of left- and right-circularly polarized light. Using the BeStSel server, Δε as a function of wavelength was analyzed to predict the secondary structure compositions. The distributions of observed secondary structures were similar for most proteins (Fig. 7A and 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, Walls et al., 2020). 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.

CD-derived structural information was also compared with MD secondary structures for RBD, T1, and T4 determined using DSSP (DOI: 10.1002/bip.360221211) (Fig. 7B-D). DSSP was performed using the built in timeline function in VMD (https://doi.org/10.1016/0263-7855(96)00018-5) . Truncated 6VXX structures containing the relevant residues for each truncation were also included, which represent structural composition had truncating Spike not resulted in any structural changes. For all three proteins, 6VXX and MD structures have high similarity, with CD-derived structures having lower turn content. Overall, MD and CD results suggest that T1 and T4 retain accurate RBD structure, which is consistent with their high sensitivities in the ELISAs against anti-Spike antibodies.

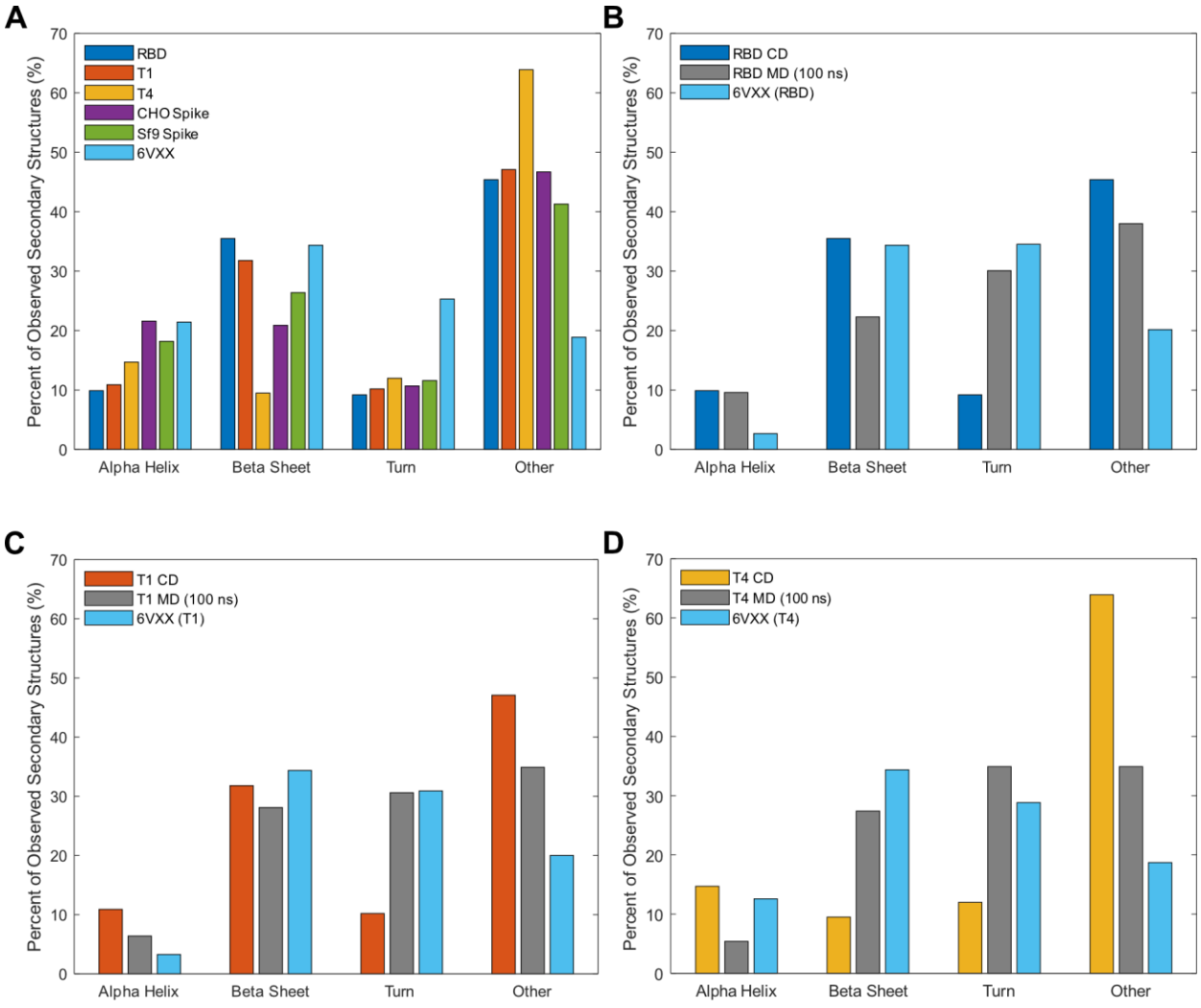


Figure 7. Structural composition of Spike and truncations produced in CHO cells. (**A**) Overall secondary structure compositions of CHO-derived proteins and 6VXX. (**B**) Comparison of RBD MD to CD data and the RBD region of 6VXX. (**C**) Comparison of T1 MD to CD data and the T1 region of 6VXX. (**D**) 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.

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 mutations (Hsieh et al., 2020), comparison of different cell lines (Stuible et al., 2021), and optimization of production conditions, such as temperature shifts (Johari et al., 2021). 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 (Fig. 3C and 3D). This suggests that residues downstream from the C-terminus of the T1 region may be contributing to decreased titers. Comparing T1 to T2 and T3, residues upstream of T1 also appeared 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 (Fig. 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.

In the ELISA sensitivity assays for CR3022 and the PAb, CHO-expressed Spike has higher AUCs for both antibodies compared to Sf9 Spike (Fig. 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 patterns (Esko & Stanley, 2015). We also found that CHO Spike produces higher signal than RBD when probed with the PAb, consistent with results from serological assays (Amanat, 2020). This was also expected because polyclonal antibodies target multiple epitopes, and full-length Spike may contain more binding epitopes than does 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. Visualization of binding through cryo-electron microscopy and analysis of binding kinetics and thermodynamics through methods such as biolayer interferometry and steered MD may elucidate the reason for their high affinities.

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.

Disclosures

All authors are listed as inventors on a record of invention describing T1 for research and clinical applications.

7.6 References

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. https://doi.org/10.1016/j.softx.2015.06.001

Amanat, F., Stadlbauer, D., Strohmeier, S., Nguyen, T. H. O., Chromikova, V., McMahon, M., Jiang, K., Arunkumar, G. A., Jurczyszak, D., Polanco, J., Bermudez-Gonzalez, M., Kleiner, G., Aydillo, T., Miorin, L., Fierer, D. S., Lugo, L. A., Kojic, E. M., Stoever, 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. https://doi.org/10.1038/s41591-020-0913-5

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. https://doi.org/10.1016/j.softx.2019.100241

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. https://doi.org/10.1371/journal.pone.0187994

Ernst, O., & Zor, T. (2010). Linearization of the Bradford Protein Assay. *Journal of Visualized Experiments : JoVE*, *38*, 1918. https://doi.org/10.3791/1918

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. http://www.ncbi.nlm.nih.gov/books/NBK453088/
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. https://doi.org/10.1016/j.pep.2020.105686
Frishman, D., & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, *23*(4), 566–579. https://doi.org/10.1002/prot.340230412
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*, *0*(0), null. https://doi.org/10.1056/NEJMoa2107659
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, N.y.)*, eabd0826. https://doi.org/10.1126/science.abd0826
Huang, Y., Yang, C., Xu, X., Xu, W., & Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, *41*(9), 1141–1149. https://doi.org/10.1038/s41401-020-0485-4
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. https://doi.org/10.1002/bit.27615
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. https://doi.org/10.1093/bib/bbx108
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. https://doi.org/10.1038/nprot.2015.053
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. *Journal of Computational Chemistry*, *29*(4), 622–655. https://doi.org/10.1002/jcc.20820
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. https://doi.org/10.1021/acs.jctc.5b00255
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. https://doi.org/10.1186/1471-2105-7-339
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. https://doi.org/10.1073/pnas.1500851112
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. https://doi.org/10.1002/pro.3943

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. https://doi.org/10.1093/bioinformatics/btt055

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. https://doi.org/10.1016/j.vaccine.2020.06.066

Smaoui, M. R., & Yahyaoui, H. (2021). Unraveling the stability landscape of mutations in the SARS-CoV-2 receptor-binding domain. *Scientific Reports*, *11*(1), 9166. https://doi.org/10.1038/s41598-021-88696-5

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., Veesler, 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. https://doi.org/10.1016/j.cell.2020.08.012

Stuible, 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. https://doi.org/10.1016/j.jbiotec.2020.12.005

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. https://doi.org/10.1002/jcc.20291

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. https://doi.org/10.1016/j.cell.2020.02.058

Webb, B., & Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*, *54*(1), 5.6.1-5.6.37. https://doi.org/10.1002/cpbi.3

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. https://doi.org/10.3389/fpls.2019.00768

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. https://doi.org/10.1093/nar/gku366

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. https://doi.org/10.1126/science.abb7269

7.7 Supplemental information

Table S1. Protein sequences of CHO-derived Spike truncations. Signal sequences are highlighted in gray, and potential N-linked glycosylation sites are highlighted in purple. 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%)	MFVFLVLLPLVSSQNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCV ADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYN YKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCN GVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCV NFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSHHHHHH
T2 (25.1%)	MFVFLVLLPLVSSQTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVAD YSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYK LPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGV EGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNF NFNGLTGTGVLTEHHHHHH
T3 (28.1%)	MFVFLVLLPLVSSQTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVAD YSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYK LPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGV EGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNF NFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSHHHHHH

T4 (27.5%)	MFVFLVLLPLVSSQNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCV ADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYN YKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCN GVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCV NFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVIT PGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVHHHHHH
T5 (28.3%)	MFVFLVLLPLVSSQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAY YVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESI VRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTK LNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKV GGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTEHHHHHH
T6 (35.3%)	MFVFLVLLPLVSSQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAY YVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESI VRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTK LNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKV GGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQ QFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPV AIHADQLTPTWRVYSTGSNVHHHHHH
T7	MFVFLVLLPLVSSQNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCV ADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYN

	YKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCN GVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCV NFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVIT PGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHV NNSYECDIPIGAGICASYQTQTNSPASVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTIS VTTEILPVSMTKTSVDCTMYICGDHHHHHH
T8 (35.5%)	MFVFLVLLPLVSSQKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEF RVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLP QGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLK YNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFG EVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYAD SFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFR KSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFEL LHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTEHHHHHH
S1 (54.7%)	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLM DLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQ TLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSET KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC

	NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVI TPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH VNNSYECDIPIGAGICASYQTQTNSPAHHHHHH

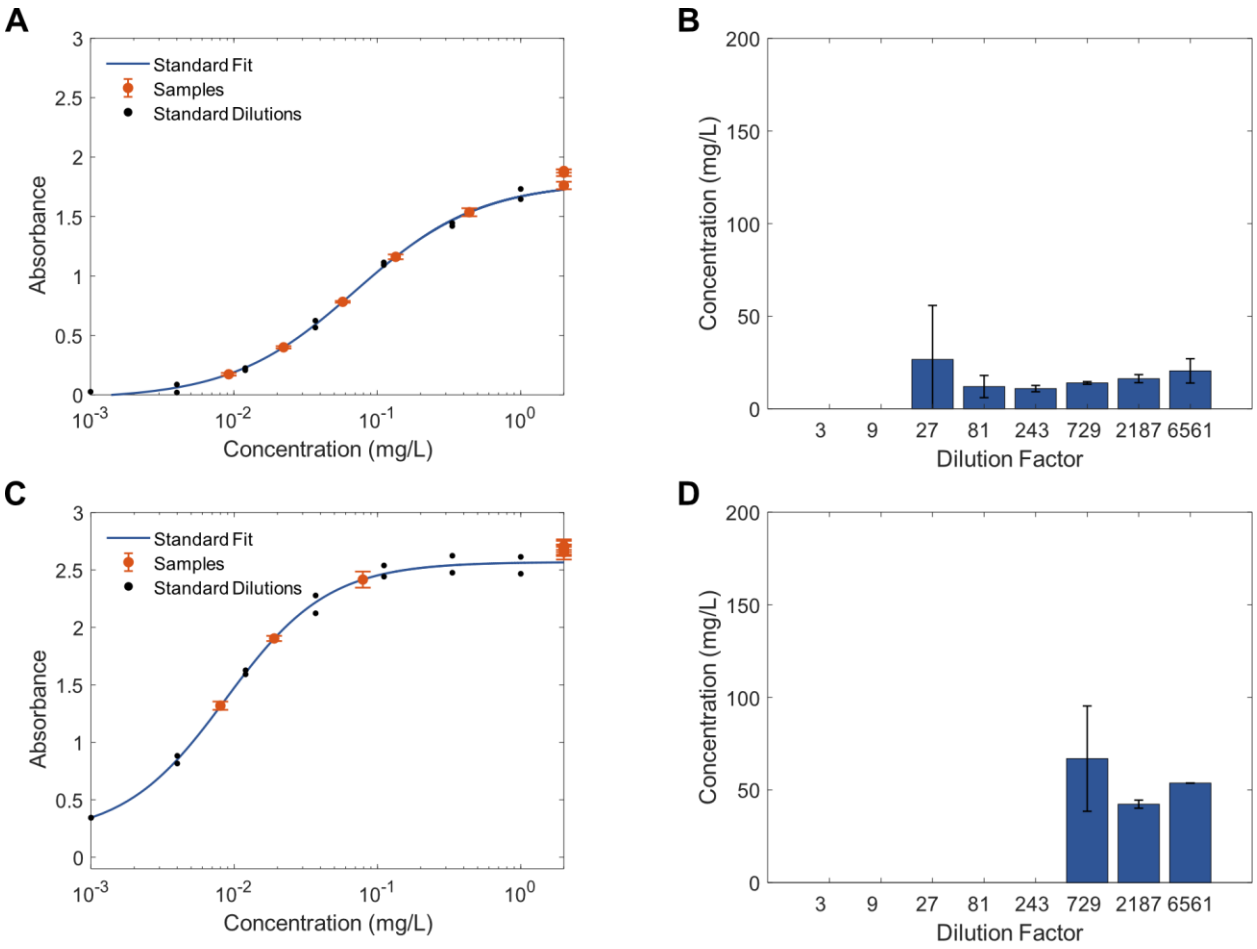


Figure S1. Measurement of crude titers for Spike and RBD via sandwich ELISA. (**A**) Dilutions of crude Spike plotted against the Spike standard curve. (**B**) Back-calculated concentrations of Spike crude titers. (**C**) Dilutions of crude RBD plotted against the RBD standard curve. (**D**) Back-calculated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

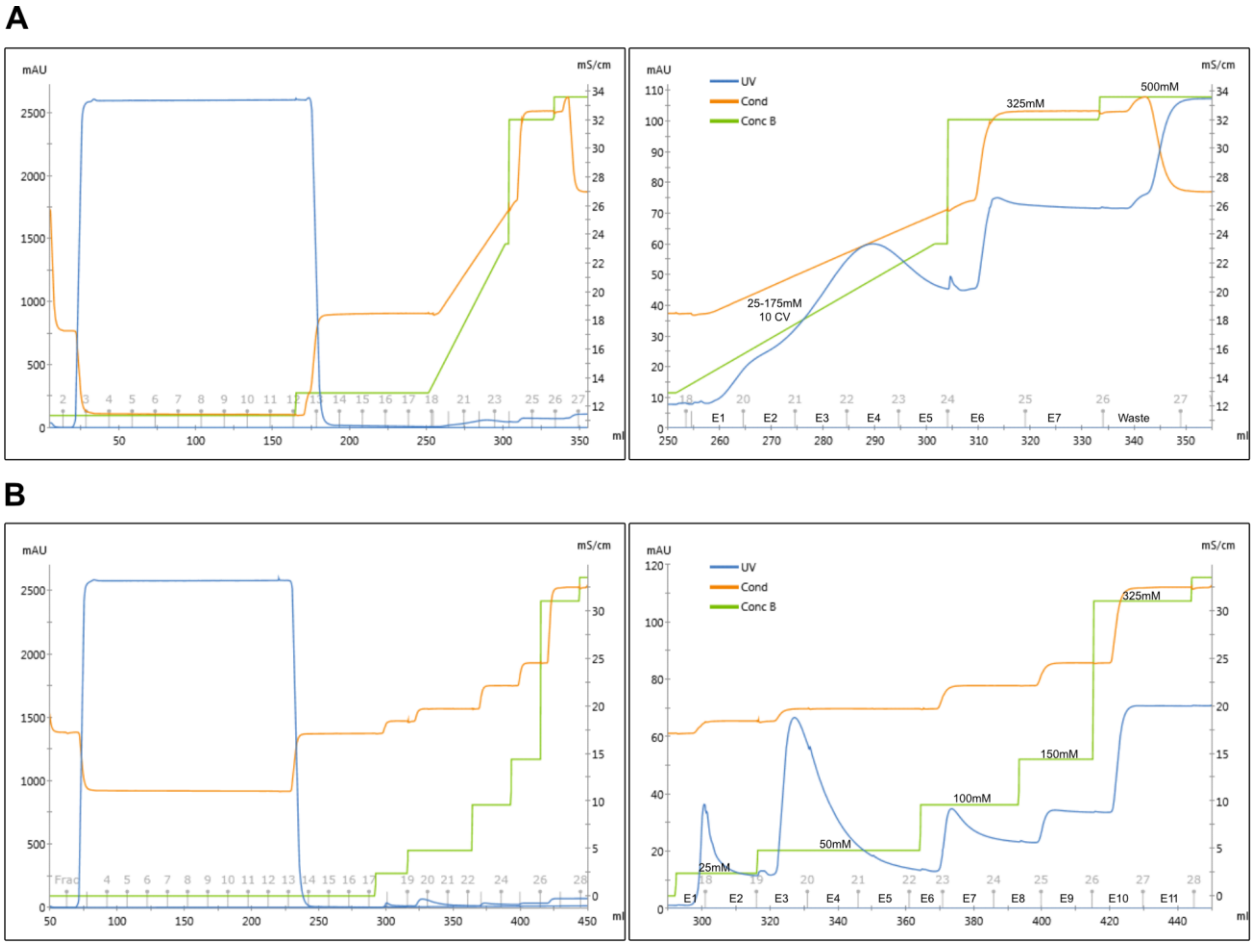


Figure S2. Spike and RBD chromatograms plotted over volume flowed through the system. (**A**) Chromatogram of entire Spike purification (left) and of zoomed in elution fractions (right). 150 mL of sample was loaded and washed with PBS containing 25 mM imidazole. A continuous gradient was applied from 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.

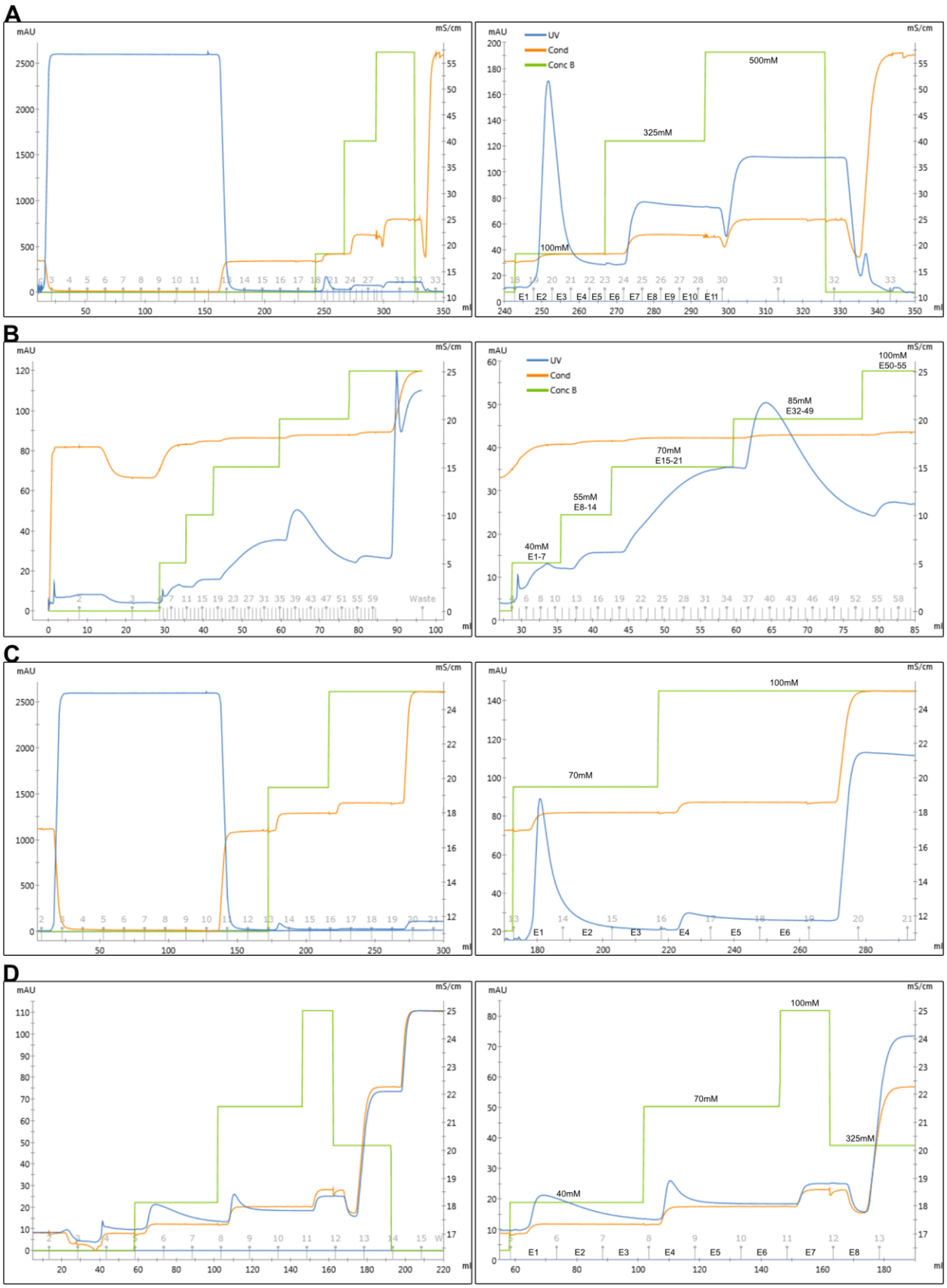

A
mAU
mS/cm
UV
Cond
Conc B
100mM
325mM
500mM
E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11
B
mAU
mS/cm
Waste
UV
Cond
Conc B
40mM
E1-7
55mM
E8-14
70mM
E15-21
85mM
E32-49
100mM
E50-55
C
mAU
mS/cm
70mM
100mM
E1 E2 E3 E4 E5 E6
D
mAU
mS/cm
40mM
70mM
100mM
325mM
E1 E2 E3 E4 E5 E6 E7 E8

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      NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFST
T1_Coverage     -FRVQPTESIVR-----------------FASVYAWNRKRISNCVADYSVLYNSASFST
                 **********                 ****************************

Ref_T1_Seq      FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAW
T1_Coverage     FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAW
                ************************************************************

Ref_T1_Seq      NSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYG
T1_Coverage     NSNNLDSKVGGNYNYLYR---KSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYG
                *****************   ************************************

Ref_T1_Seq      FQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN
T1_Coverage     FQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN
                ************************************************************

Ref_T1_Seq      KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSHHHHHH
T1_Coverage     KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSHHHHHH
                ****************************************
```

B

```
Ref_T4_Seq      NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFST
T4_Top_Coverage --RVQPTESIVR-----------------FASVYAWNRKRISNCVADYSVLYNSASFST
                  **********                 ****************************

Ref_T4_Seq      FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAW
T4_Top_Coverage FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAW
                ************************************************************

Ref_T4_Seq      NSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYG
T4_Top_Coverage NSNNLDSKVGGNYNYLYR---KSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYG
                *****************   ************************************

Ref_T4_Seq      FQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN
T4_Top_Coverage FQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN
                ************************************************************

Ref_T4_Seq      KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVN
T4_Top_Coverage KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCS------------------------
                **********************************

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

C

```
Ref_T4_Seq      NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFST
T4_Bottom_Cover -FRVQPTESIVR-----------------FASVYAWNRKRISNCVADYSVLYNSASFST
                 **********                 ****************************

Ref_T4_Seq      FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAW
T4_Bottom_Cover FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAW
                ************************************************************

Ref_T4_Seq      NSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYG
T4_Bottom_Cover NSNNLDSKVGGNYNYLYR---KSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYG
                *****************   ************************************

Ref_T4_Seq      FQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN
T4_Bottom_Cover FQPTNGVGYQPYRVVVLSFELLHAPATVCGPKK--------CVNFNFNGLTGTGVLTESN
                ********************************        ******************

Ref_T4_Seq      KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVN
T4_Bottom_Cover KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCS------------------------
                **********************************

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.

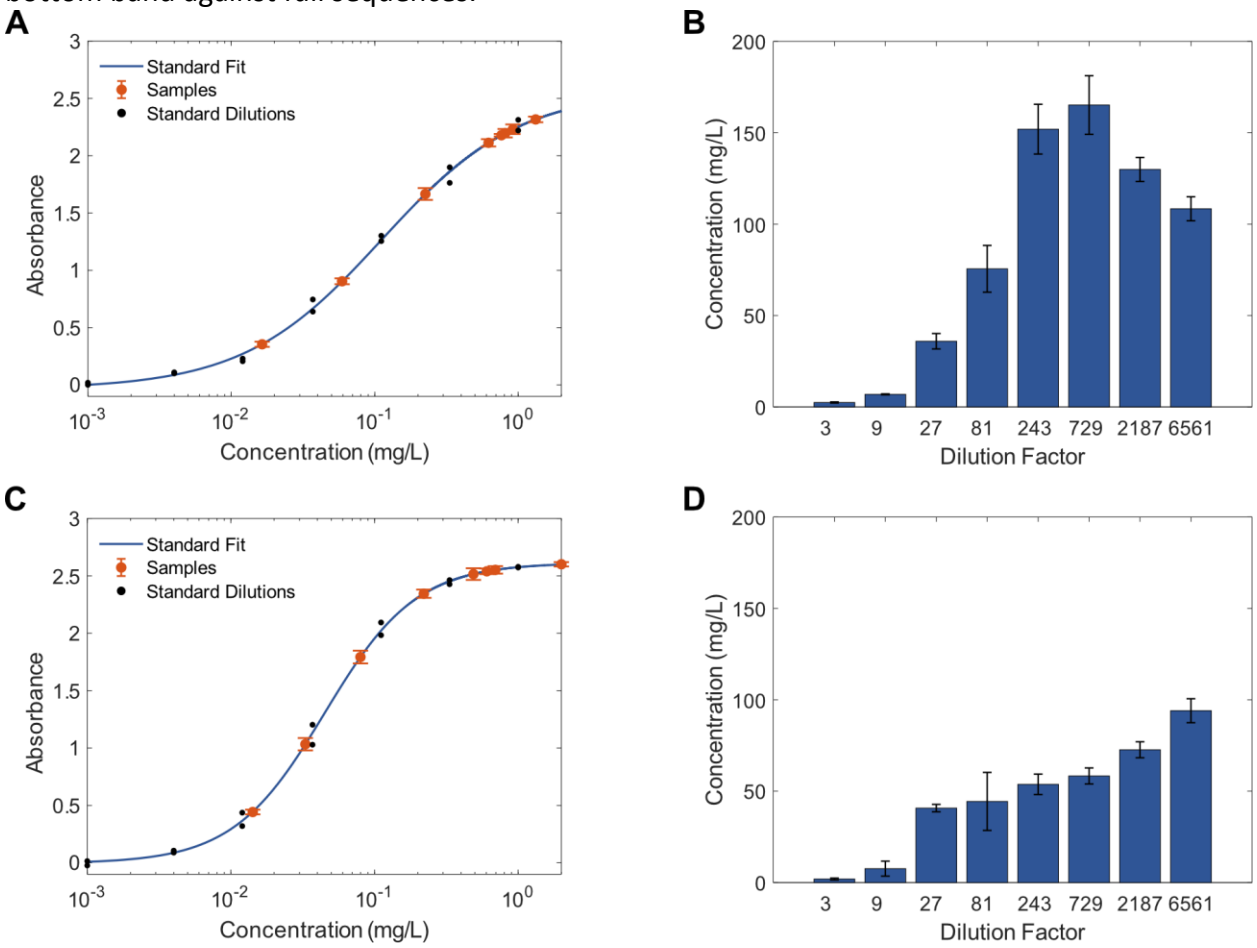


Figure S5. Measurement of crude titers for T1 and T4 via sandwich ELISA. (**A**) Dilutions of crude T1 plotted against the T1 standard curve. (**B**) Back-calculated concentrations of T1 crude titers. (**C**) Dilutions of crude RBD plotted against the RBD standard curve. (**D**) Back-calculated concentrations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

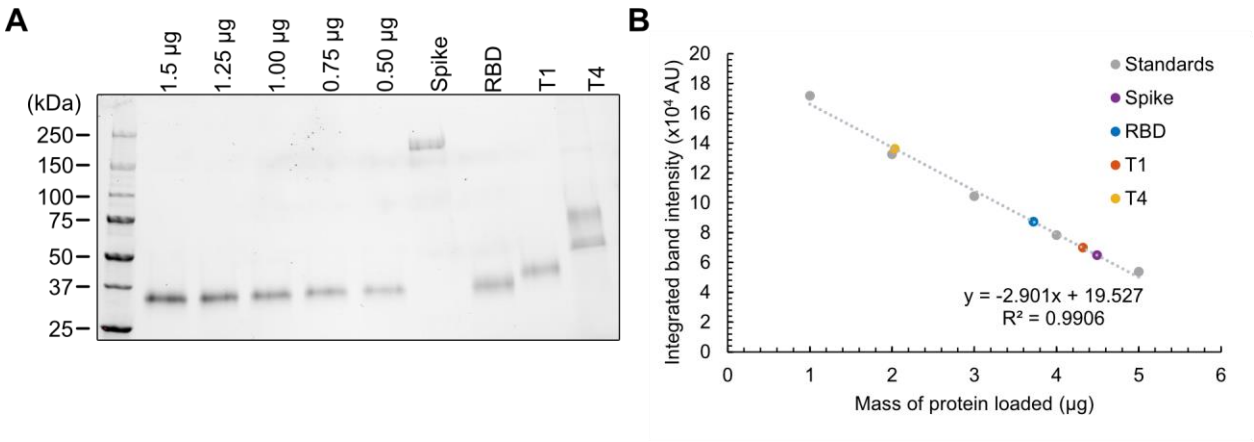


Figure S6. Quantification of purified proteins. (**A**) SDS-PAGE and (**B**) quantification for purified proteins and serial dilutions of RBD obtained from BEI Resources. A standard curve was prepared using a linear fit to serial dilutions of the standard protein.

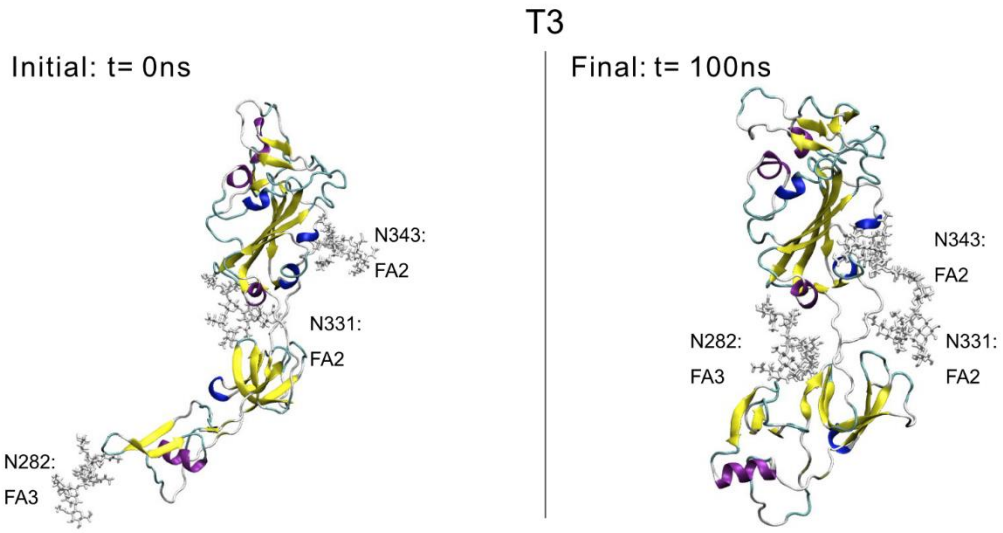


Figure S7. MD-simulated structures of T3. RMSD plots in the same format as T1 and T4.

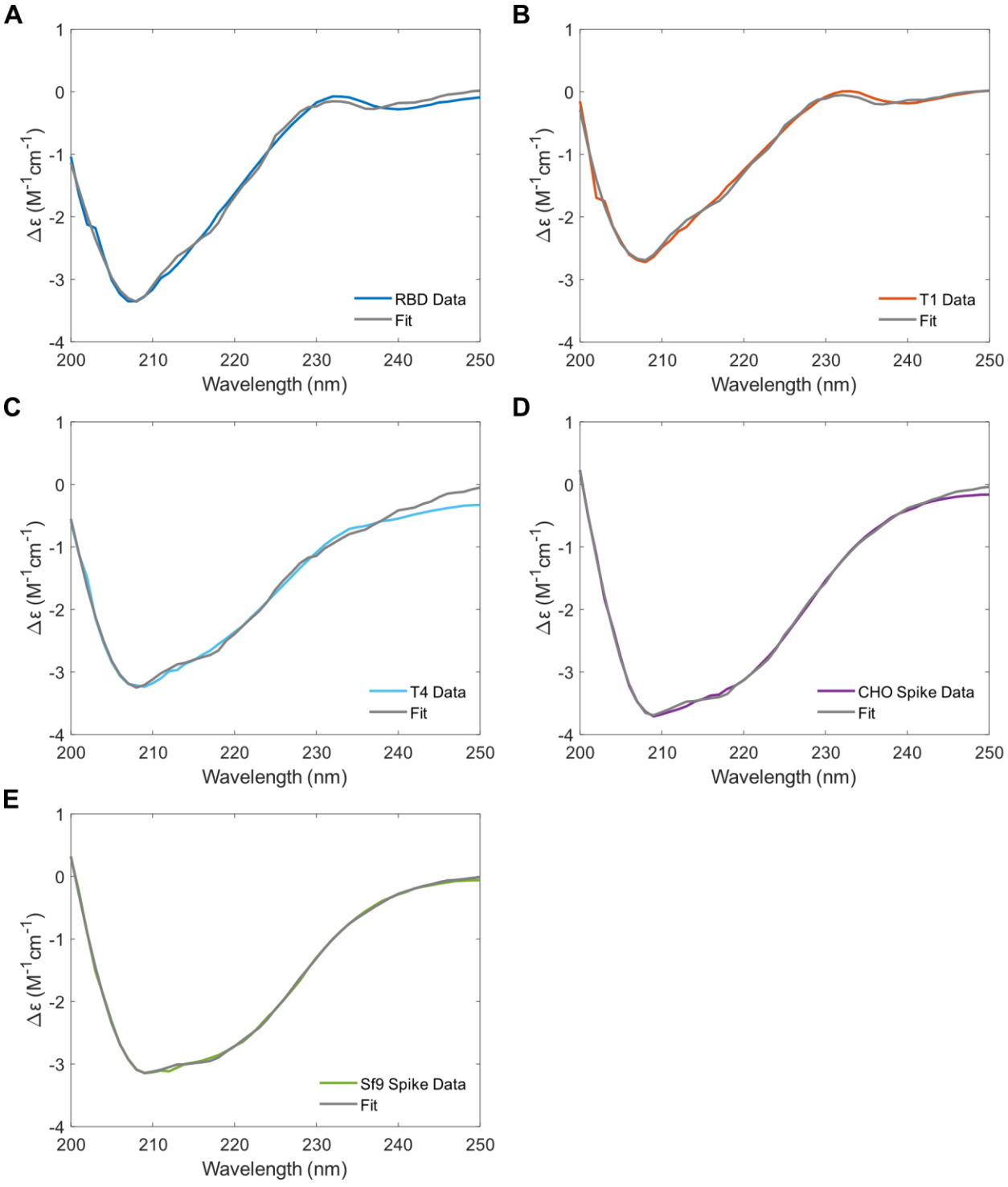


Figure S8. Raw spectral data on proteins analyzed via circular dichroism. Δε is plotted against wavelength for (**A**) RBD, (**B**) T1, (**C**) T4, (**D**) CHO Spike, and (**E**) Sf9 Spike.

Appendix 1: Analysis of Significance of extra Amino Acids on T1

The following is my analysis of the significance of 2 additional amino acids we added for potential stability to the T1 structure in the previous paper. This work may be incorporated into a future manuscript

Analysis:

To analyze the stability and contribution to overall stability of the first few amino acids (NFRVQPTGSI) and HIS tag, Root mean squared deviation (RMSD), root mean squared fluctuation (RMSF), secondary structure evolution with time, and pair distance calculations were performed. Visualizations were performed using VMD. RMSD calculations were performed using the Gromacs suite, the calculation was performed for both the first 2 amino acids of interest, ASN, and PHE, and the first 10 amino acids, with both the initial and final structures as references. The contribution of these Amino Acids are shown next to the overall T1 RMSD for reference. RMSF calculations were performed using the Gromacs suite's RMS command, with the residue flag on to treat the amino acid as an entity instead of as individual atoms. Secondary structures evolution was calculated using VMD's timeline command, counts of individual points from the corresponding heatmap were performed using a custom python script. Minimum and Maximum pair distances were performed using Gromacs pairdist command with the HIS tag as reference comparing the first 2 and first 5 amino acids.

Results and Discussion:

Visual inspection of the first few amino acids at the C? terminus and the HIS tag at the N? terminus of T1 from the original T1 simulation was performed and is shown in Figure 1. The first 3 amino acids and the start and end of the HIS tag were labeled using the VMD label tool. Zoomed in snapshots are shown from the starting 0 ns configuration, and the final frame after 100 ns of molecular dynamics simulation. The 0 ns configuration appears to show minimal interactions between the two termina, while after 100ns of configuration the termina appear to be aligned parallel to each other suggesting an increase in interactions over the course of the simulation. It is important to note that the secondary structure assignment for visualization was performed using the final frame (100ns), and the dark yellow structure represents an isolated bridge, which is more ordered than the white coil. These increasing interactions suggest the first few amino acids may contribute to the stability of the termina.

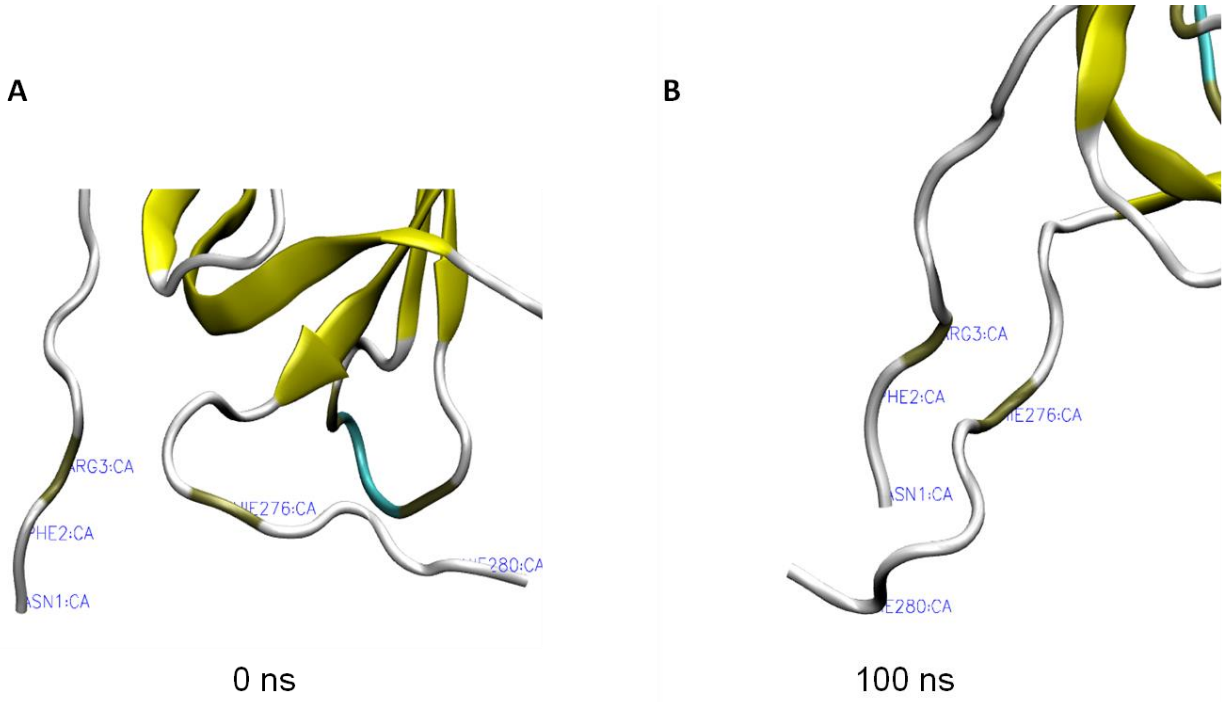


Figure 1: Visualization of T1 C and N termina. A) snapshot taken at initial frame 0 ns simulation B) snapshot taken at final frame 100 ns simulation

To determine the contribution of these amino acids on the stability of the structure, RMSD calculations were performed for the first 2 amino acids and the first 10 amino acids. Figure 2 shows the RMSD contribution for the first 2 and first 10 amino acids compared to their initial 2A and final 2B structures shown above. This is compared to the overall T1 RMSD shown in Figure 2C. The first two amino acids show low deviation and an equilibrated structure (blue). Expanding this analysis to the first 10 amino acids shows a greater deviation, with some evolution towards a more stable structure seen by the reduction in fluctuation over time in 2A.

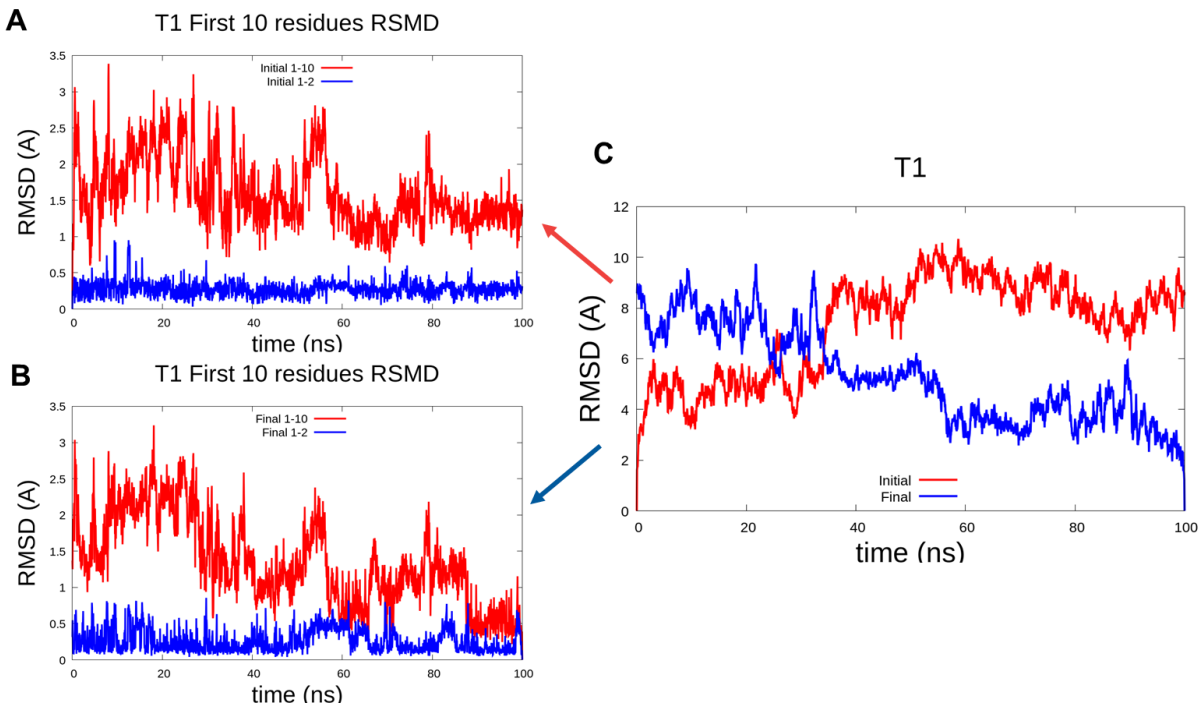


Figure 2: RMSD calculations A) First few amino acids compared to initial structure B) First few amino acids compared to final structure C) Overall T1 compared to initial and final structures. Red corresponds to first 10 AA in A and B, blue to first 2 AA in A and B. Red and blue correspond to initial and final structure in C.

RMS fluctuations were calculated for all amino acids in T1 and the first 68 amino acids in T3 (The amino acids that are additional to T1). Fluctuations were determined based on the deviation from the average position across the full 100ns simulation. Figure 3 shows the results of this RMSF calculation, 3A shows T1 and the first 10 amino acids and 3B shows T3, with the same 10 amino acids (59-68) labeled. The results suggest high fluctuation for the first 2 amino acids, which is to be expected as they are at the end of the structure and can more freely move around. Interestingly the fluctuations for end of T3 are greater than those for T1, suggesting that the additional AA are not as stable as they extend out.

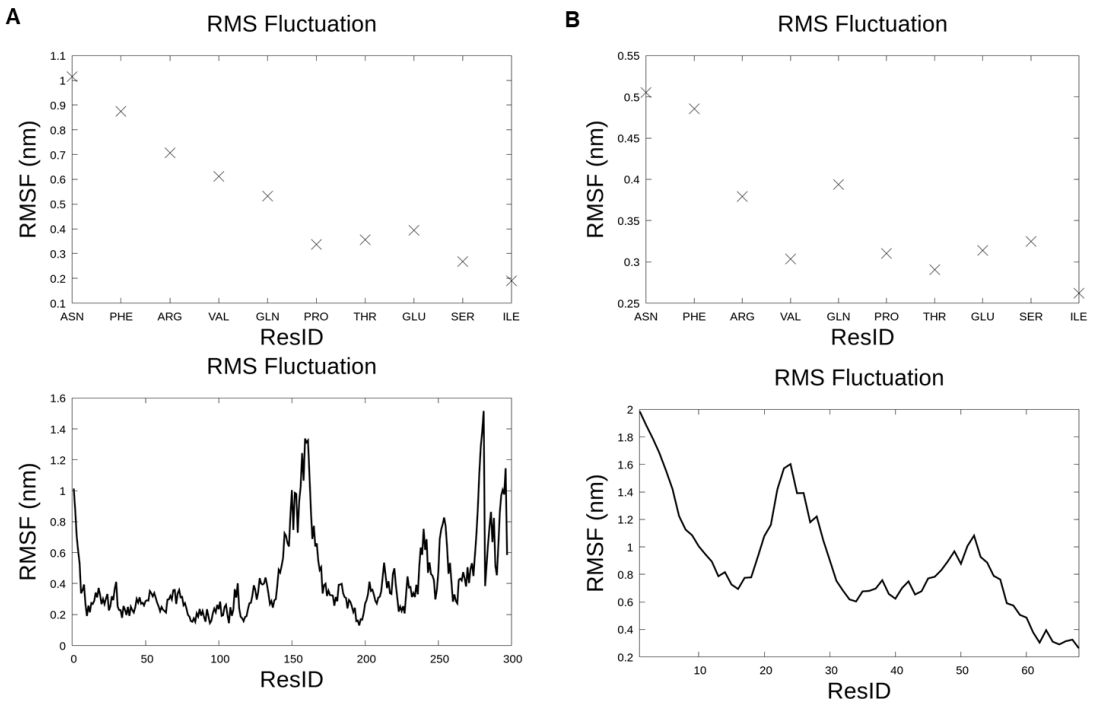
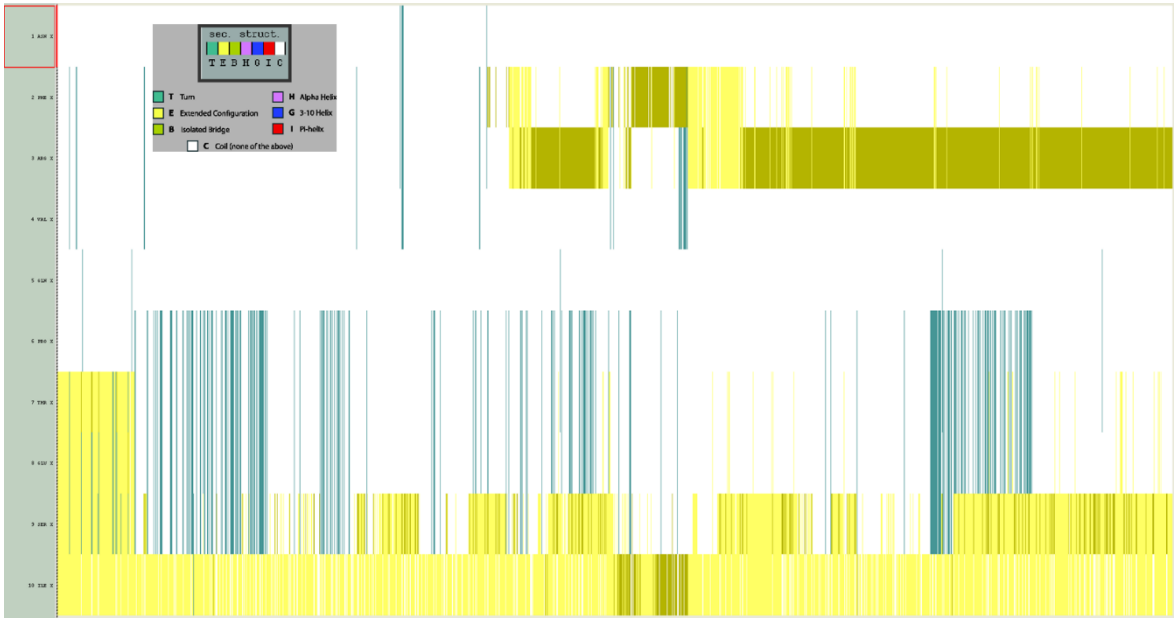


Figure 3: RMSF for T1 and T3. A) T1 and first 10 AA, B) T3 and same 10 AA (59-68)

To determine the stability of the secondary structures a secondary structure timeline analysis was performed and shown as a heatmap in Figure 4. The 2nd and 3rd amino acids show an evolution from a coil to a more structured extended configuration and isolated bridge as seen



in visualization. This suggests these 2 additional amino acids increase the potential interaction area between the two termina. Counts of each type for each of these first 10 amino acids are shown in Figure 5.

Figure 4: Secondary structure timeline analysis.

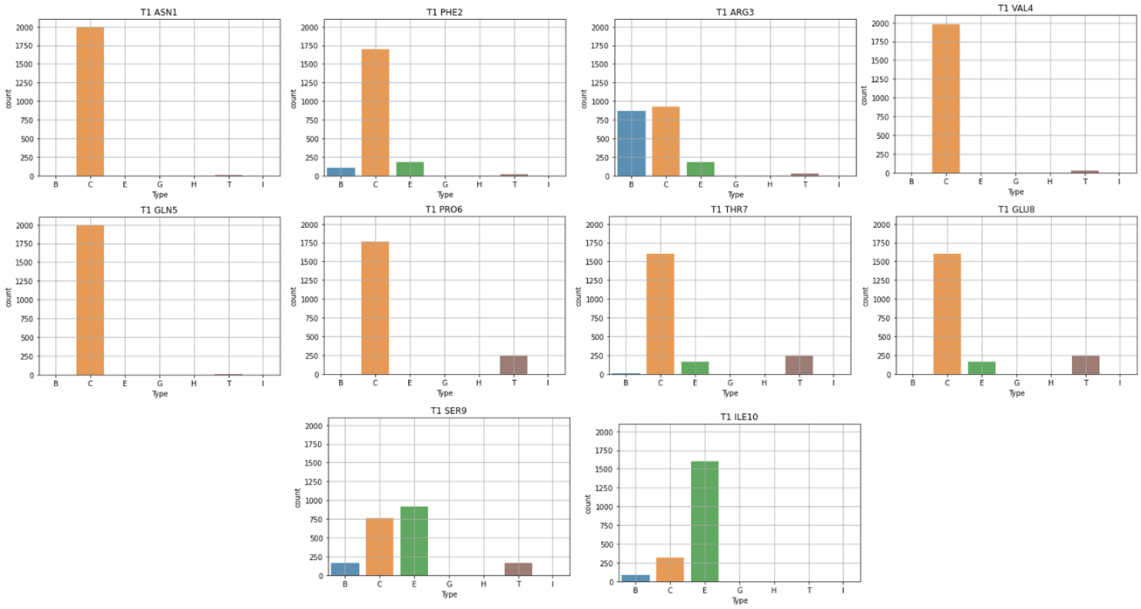


Figure 5: Secondary structure counts from timeline analysis

Pair distance calculations were performed between the first 2 and first 5 amino acids and the HIS tag and results are shown in Figure 6. 6A shows the minimum distance between these residues, 6B shows the maximum distance between these residues, the calculation takes the minimum or maximum distance between any of the atoms on any of the residues and plots them all together. The minimum distance plot suggests that both the first 2 and the first 5 amino acids start interacting more strongly with the HIS tag, as seen by the decrease in overall distance as the simulation progressed. The same trend is seen in the maximum distance plots, where the decrease in maximum distance also suggests that the HIS tags and those residues are

getting closer together. When taken with the visualization and the secondary structure it suggests that the two termina interact, and that the number of amino acids on the ends of either termina influence each other.

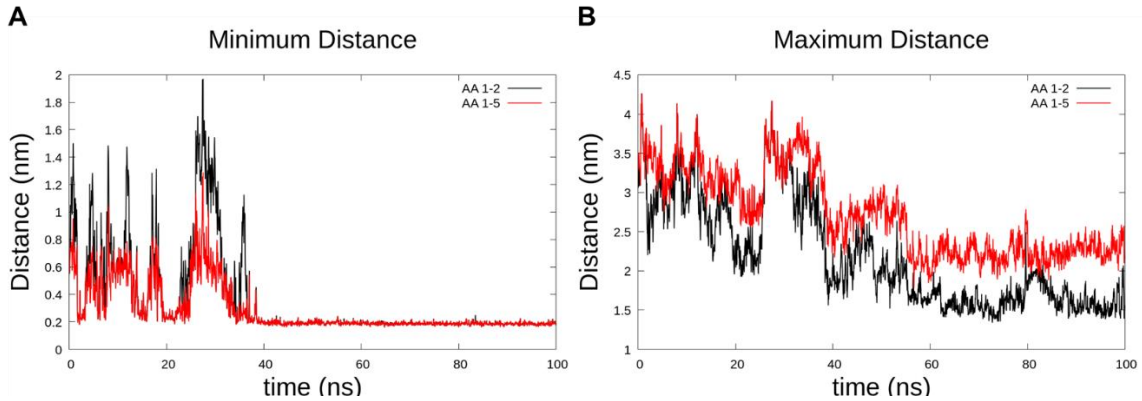


Figure 6: Pair distance calculations, HIS tag as reference structure, Black curves AA 1-2 and red AA 1-5. A) Minimum distance between any atom in reference and AAs. B) Maximum distance between any atom in reference and AAs.

Chapter 8

Summary and Outlook

This work has covered the development and application of molecular modeling workflows for the improvement of experimental design using 3D printing and biological applications, including lipid and protein structure prediction as examples. It began with the development of a computational software and methodology to replicate the evaporation of mixed solvent systems while accounting for potential hydrogen bonding shells. This software and workflow were developed to better understand the evaporation of ultra-small droplets and the controlled assembly of solute particles inside them. This evaporation process is of core importance to the idea of controlled molecular assembly, which is a critical precursor for enabling 3D printing at the nano-scale. The process is dependent on the resulting solute, solvent, and surface interactions, and improving our understanding of the physics for this process is crucial to improving experimental workflows. To improve our understanding of these processes the developed workflow was applied to two biologically relevant systems at different length scales. First, an atomistically represented sugar molecule was used to probe phase transitions as these sugars assembled into various morphological structures. Secondly, coarse-grained phospholipids in mixed solvents deposited onto surfaces with different functionalities were modelled. The increased length and time scales of the coarse-grained approach allowed for the study of solute, solvent, and surface interactions during the evaporation process at the

same time. Both computational approaches were validated experimentally and incorporated into the experimental design process to improve future experimental efforts.

The second half of this work discussed the application of molecular simulation on SARS-CoV-2 spike protein and human angiotensin converting enzyme 2 (hACE2) receptor, specifically for the development of a diagnostic or therapeutic tool. This work focused on the evaluation of structural stability of the hACE2 receptor and spike receptor binding domain (RBD) with different glycosylation patterns that would derive from different synthesis methodologies. The corresponding interactions of the RBD and ACE2 were analyzed using molecular force pulling simulations, specifically focusing on the effect glycosylation pattern played on the hydrogen bonding, electrostatic, and van der waals interactions between the two. Finally, the structural stability of various spike truncations with the prospect of improved expression and response were analyzed.

As demonstrated, molecular dynamics is an immensely powerful technique for the applications shown here, with key advantages being the knowledge of position, velocity, and fore at any given time step. This information can be used to reconstruct many experimentally difficult to determine properties related to structure, dynamics, reaction pathways, and thermodynamics of a system. While the strengths of this technique are clear, it is important to acknowledge the potential weaknesses and limitations of the technique, as well as look at potential options for filling in gaps or for future improvements.

First, molecular dynamics and any computational technique are inherently limited by computational resources. In the case of molecular dynamics, the simulation time required is heavily dependent on the number of particles involved. For atomistic simulations with a

standard 2 fs time step it is virtually impossible at this juncture to capture all behaviors that would be present. Using protein simulations as an example, protein folding happens on the order of microseconds to seconds, while atomistic simulations in this work were limited to 75-100ns, several orders of magnitude less.[74] These 100ns runs took roughly 2 weeks for the large protein systems like ACE2-Fc and RBD proteins shown in chapter 5. The evaporation simulations shown in chapter 3 required roughly 3 weeks of continuous simulation time for a single full evaporation run.

Coarse graining, like the work shown in chapter 4, is one option for improving achievable simulation time. While this results in a decrease in resolution it also results in an increase in simulation time step which enables the capacity for longer and larger simulations, on the order of microseconds or 100s of nm. These approaches have their own downsides, in addition to still not being able to reach time and length scales required for certain applications, coarse-graining removes chemical specificity. The type of parameterization for a coarse-grained force field also makes a tremendous impact and must be tailored for the specific application of interest. A classic example of parameterization decisions is center of mass vs center of geometry for a coarse-grained bead to represent. The center of mass approach is often better for speed when representing multiple heavy atoms and gives functionally the same level of chemical detail as center of geometry when the mapping is 4:1 or greater. Center of geometry approaches on the other hand give much more accurate results as you reduce the level of coarse-graining, and become useful for applications dependent on replicating specific interactions. For the coarse-grained work shown in chapter 4, we used the MARTINI 3

parameterization [46], which has been specifically parameterized to give good accuracy for common lipids, including POPC, as compared to experiment.

For both atomistic and coarse-grained approaches, the bulk of the computational resources are spent calculating the forces of the surrounding solvent. Options for improvement in achievable time and length scales using molecular dynamics include a shift to GPU parallelization, as GPU performance has outpaced the typical CPU performance for most benchmarks. Currently, most existing workflows are optimized for CPU based performance, and the number of codes capable of supporting GPUs has slowly been catching up. Another option for improved resolution is a method that removes the water from the calculations, and a growing approach in the literature and field is the usage of an implicit solvent method like the lattice-boltzmann method along with molecular dynamics.

It is worth mentioning a protein specific limitation with molecular dynamics, the lack of readily available protein structures. This is due to experimental difficulties in getting good protein crystals, a field in which improvements have continuously been made. From a computational perspective, recent advancements in the development of machine learning predicted protein structures, like those of AlphaFold[75], present the opportunity for future workflows optimizing previously undiscovered proteins using AlphaFold to predict an initial structure for molecular dynamics.

Ultimately, no single simulation technique is accurate or powerful enough to replicate all of the complex behaviors of all real experimental systems. This is why the application of multiple simulation techniques at multiple scales to determine the behaviors of different aspects of these complex systems is vital for improved physics understanding. This also shows

why experimental validation, and integration with computational workflows is crucial for improving the experimental design process in a meaningful way. There are a number of advances in the areas of scientific computing that show the growing viability and necessity of these integrations, including improvements in computing architecture, structure prediction, and workflow optimization.

Bibliography

1. Jiali Zhang, Victoria A Piunova, Yang Liu, Andy Tek, Qingbo Yang, Jane Frommer, Gang-yu Liu Joseph Sly. Controlled Molecular Assembly Via Dynamic Confinement of Solvent. Journal of Physical Chemistry Letters, 9(21): 6232-6237, 2018

2. Jiali Zhang, Hai Yu, Bradley Harris, Yunbo Zheng, Umit Celik, Lan Na, Roland Faller, Xi Chen, Dominik R. Haudenschild, and Gang-yu liu. New Means to Control Molecular Assembly. JPC C, 11(124): 6405-6412, 2020.

3. Viral Chhasatia and Ying Sun. Interaction of bi-dispersed particles with contact line in an evaporating colloidal drop, Soft Matter, 7(21): 10135-10143, 2011

4. Hwa-Young Ko, Hyunjung Shin, and Jooho Moon. Rapid Self-Assembly of Monodisperse Colloidal Spheres in an Ink-Jet Printed Droplet. Chemistry of Materials, 16(22): 4212-4215, 2004

5. Jungho Park and Jooho Moon. Control of Colloidal Particle Deposit Patterns within Picoliter Droplets Ejected by Ink-Jet Printing. Langmuir, 22 (8): 3506-3513, 2006

6. Kathleen E Hamilton, Catherine D. Schuman, Steven R. Young, Ryan S. Bennink, Neena Imam, Travis S. Humble. Accelerating Scientific Computing in the Post-Moore's Era, ACM Transactions on Parallel Computing 7(1): 1-31, 2020

7. P.J Denning, D.E comer, D. Gries, M.C Mulder, A. Tucker, A.J. Turner, P.R. Young, Computing as a discipline. Computer 22(2): 63-70, 1989

8. E Weinan, Principles of Multiscale Modeling. Cambridge University Press, 2011

9. Mark F Horsttemeyer, Multiscale Modeling: A Review, Practical Aspects of Computational Chemistry, 2009

10. Joachim Gross and Gabriele Sadowski, Application of the Perturbed-Chain SAFT Equation of State to Associating Systems, Ind. Eng. Chem. Res. 41(22):5510-5515, 2002

11. Eleuterio F. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics, Springer Publishing, 1997

12. J.E Moyal Quantum mechanics as a statistical theory, Mathematical Proceedings of the Cambridge Philosophical Society, 45(1): 99-124, 1949

13. David J Griffiths and Darrel F Schroeter, Introduction to Quantum Mechanics (3rd ed.), Cambridge University Press, 2018

14. P Geerlings, F. De Proft, W. Langenaeker, Conceptual Density Functional Theory, Chem. Rev, 103(5):1793-1874, 2003

15. Eberhard Engel, Reiner M. Dreizler, Density Functional Theory, Springer Publishing, 2013

16. Aron J Cohen, Paula Mori-Sanchez, Weitao Yang, Challenges for Density Functional Theory, Chem. Rev. 112(1), 289-320, 2012

17. Dennis C Rapaport, The Art of Molecular Dynamics Simulation, Cambridge University Press, 2004

18. Martin Karplus and Gregory A. Petsko, Molecular dynamics simulations in biology, Nature, 347: 631-639, 1990

19. J.M. Hammersley and D.C Handscome, Monte Carlo Methods, Fletcher & Son Ltd 1964

20. Reuven Y. Rubinstein and Dirk P. Kroese, Simulation and the Monte Carlo Method (3rd ed.) Wiley Publishing, 2016

21. Soumil Y. Joshi and Sanket A. Deshmukh, A review of advancements in coarse-grained molecular dynamics simulations, Molecular Simulation, 47(10): 786-803, 2020

22. Robert E. Rudd, Jeremey Q Broughton, Coarse-grained molecular dynamics and the atomic limit of finite elements, Physical Review B. 58(10), 1998

23. Roger Temam, Navier-Stokes Equations: Theory and Numerical Analysis, AMS Chelsea Publishing, 1984

24. Michael P Allen and Dominic J Tildesley. Computer simulation of liquids. Oxford university press, 2017

25. Sauro Succi, The Lattice Boltzmann Equation: for Fluid Dynamics and Beyond, Oxford Science Publications, 2001

26. Xiaoyi He and Li-Shi Luo, Theory of the lattice Boltzmann method: From the Boltzmann equation to the lattice Boltzmann equation, Phys. Rev. E, 56, 1997

27. Shiyi Chen and Gary D. Doolen, Lattice Boltzmann Method for Fluid Flows, Rev. Fluid Mech, 30:329-64, 1998

28. Louis N Hand and Janet D Finch. Analytical mechanics. Cambridge University Press, 1998.

29. Loup Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. Physical review, 159(1):98, 1967

30. William C Swope, Hans C Andersen, Peter H Berens, and Kent R Wilson. 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, 76(1):637–649, 1982.

31. Jong Cheol Jo and Byeong Cheol Kim, Determination of Proper Time Step for Molecular Dynamics Simulation, Bulletin of the Korean Chemical Society, 21(4):410-424, 2000

32. Jan Drenth. Principles of protein X-ray crystallography. Springer Science & Business Media, 2007.

33. Jacqueline LS Milne, Mario J Borgnia, Alberto Bartesaghi, Erin EH Tran, Lesley A Earl, David M Schauder, Jeffrey Lengyel, Jason Pierson, Ardan Patwardhan, and Sriram Subramaniam. Cryo-electron microscopy–a primer for the non-microscopist. The FEBS journal, 280(1):28–45, 2013

34. Anastassis Perrakis, Richard Morris, and Victor S Lamzin. Automated protein model building combined with iterative structure refinement. Nature structural & molecular biology, 6(5):458, 1999.

35. Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. The Journal of Chemical Physics, 98 (12):10089–10092, 1993.

36. Roger W Hockney and James W Eastwood. Computer simulation using particles. crc Press, 1988.

37. Denis J Evans and Brad Lee Holian. The nose–hoover thermostat. The Journal of chemical physics, 83(8):4069–4074, 1985.

38. M Parrinello and A Rahman. Crystal structure and pair potentials: A molecular-dynamics study. Physical Review Letters, 45(14):1196, 1980.

39. Sulltan Darvesh, David A Hopkins, and Changiz Geula. Neurobiology of butyrylcholinesterase. Nature Reviews Neuroscience, 4(2):131–138, 2003.

40. H. J. C. Berendsen, J. P.M Postma, W.F van Gunsteren, A. DiNola, and J.R Haak. Molecular Dynamics with coupling to an external bath. J. Chem. Phys, 81, 1984

41. Berk Hess. P-lincs: A parallel linear constraint solver for molecular simulation. Journal of Chemical Theory and Computation, 4(1):116–122, 2008.

42. Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of computational physics, 23(3):327– 341, 1977.

43. Hans C Andersen. Rattle: A velocity version of the shake algorithm for molecular dynamics calculations. Journal of Computational Physics, 52(1):24–34, 1983.

44. Shuichi Miyamoto and Peter A Kollman. Settle: an analytical version of the shake and rattle algorithm for rigid water models. Journal of Computational Chemistry, 13(8):952–962, 1992.

45. Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations, J. Phys. Chem. B, 111(27):7812-7824, 2007

46. Paulo C. T. Souza, Riccardo Allesandri, … Siewert J. Marrink, Martini 3: a general purpose force field for coarse-grained molecular dynamics, Nature Methods, 18:382-388, 2021

47. M. Mastrangeli, S. Abbasi, C. Varel, C. Van hoof, J-P Celis, K.F Bohringer. Self-assembly from milli- to nanoscales: methods and applications. Journal of Micromechanics and Microengineering, 19(8), 2009

48. Janos H. Fendler, Chemical Self-assembly for Electronic Applications. Chem. Mater. 13(10):3196-3210, 2001

49. P. Alexandridis, B. Lindman. Ampiphilic Block Copolymers: Self-Assembly and Applications, Elsevier Science, 2000

50. Youssef Habibi, Lucian A. Lucia, Orlando J. Rojas. Cellulose Nanocrystals: Chemistry, Self-Assembly, and Applications. Chem. Rev. 110(6),3479-3500, 2010

51. H. Yildririm Erbil, G. McHale, M.I Newton. Drop evaporation on solid surfaces: Constant contact angle mode. Langmuir, 18(7):2636-2641, 2002

52. R. Picknett, R. Bexon. The Evaporation of sessile or pendant drops in still air. Journal of Colloid and Interface Science, 61(2):336-50, 1977

53. Xin Zhong, Alexandru Crivoi, and Fei Duan. Sessile nanofluid droplet drying. Advances in Colloid and Interface Science, 217():13-30, 2015

54. Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". *Published online at OurWorldInData.org.*

55. 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

56. Austen Bernardi, Yihan Huang, Bradley Harris, Yongao Xiong, Somen Nandi, Karen A. McDonald, and Roland Faller. 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, 2020

57. Chan, K. K., D. Dorosky, P. Sharma, S. A. Abbasi, J. M. Dye, D. M. Kranz, A. S. Herbert, and E. Procko. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. Science. 369(6508):1261-1265, 2020

58. Nelson, G., O. Buzko, P. Spilman, K. Niazi, S. Rabizadeh, and P. Soon-Shiong. 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.2001.2013.426558, 2021

59. Mehdipour, A. R., and G. Hummer. Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. Proceedings of the National Academy of Sciences. 118(19):e2100425118, 2021

60. Prajapat, M., N. Shekhar, P. Sarma, P. Avti, S. Singh, H. Kaur, A. Bhattacharyya, S. Kumar, S. Sharma, A. Prakash, and B. Medhi. 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 Modeling. 2020

61. Davidson, A. M., J. Wysocki, and D. Batlle. Interaction of SARS-CoV-2 and Other Coronavirus With ACE (Angiotensin-Converting Enzyme)-2 as Their Main Receptor. Hypertension. 76(5):1339-1349, 2020

62. William Lennarz. The biochemistry of glycoproteins and proteoglycans. Springer Science & Business Media, 2012.

63. Vigerust, D. J., and V. L. Shepherd. Virus glycosylation: role in virulence and immune interactions. Trends in Microbiology. 15(5):211-218, 2007

64. Ohtsubo, K., and J. D. Marth. Glycosylation in Cellular Mechanisms of Health and Disease. Cell. 126(5):855-867, 2006

65. Phillips, M., E. Nudelman, F. Gaeta, M. Perez, A. Singhal, S. Hakomori, and J. Paulson. ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand, sialyl-Lex. Science. 250(4984):1130-1132, 1990

66. Sperandio, M., C. A. Gleissner, and K. Ley. Glycosylation in immune cell trafficking. Immunological Reviews. 230(1):97-113, 1990

67. Watanabe, Y., T. A. Bowden, I. A. Wilson, and M. Crispin. Exploitation of glycosylation in enveloped virus pathobiology. Biochimica et Biophysica Acta (BBA) - General Subjects. 1863(10):1480-1497, 2019

68. Banerjee, N., and S. Mukhopadhyay. Viral glycoproteins: biological role and application in diagnosis. VirusDisease. 27(1):1-11, 2016

69. Bagdonaite, I., and H. H. Wandall. 2018. Global aspects of viral glycosylation. Glycobiology. 28(7):443-467, 2018

70. RD Marshall. Glycoproteins. Annual review of biochemistry, 41(1):673–702, 1972

71. Richard Strasser, Plant Protein glycosylation, Glycobiology, 26(9):926-939, 2016

72. Allen, J. D., Y. Watanabe, H. Chawla, M. L. Newby, and M. Crispin. Subtle Influence of ACE2 Glycan Processing on SARS-CoV-2 Recognition. Journal of Molecular Biology. 433(4):166762, 2021

73. Nguyen, K., S. Chakraborty, R. A. Mansbach, B. Korber, and S. Gnanakaran. Exploring the Role of Glycans in the Interaction of SARS-CoV-2 RBD and Human Receptor ACE2. Viruses. 13(5):927, 2021

74. Ken A Dill and Justin L MacCallum. The protein-folding problem, 50 years on. Science, 338(6110):1042–1046, 2012

75. John Jumper et al. Highly accurate protein structure prediction with AlphaFold, Nature 596:583-589, 2021.