UNIVERSITY OF CALIFORNIA SAN DIEGO

Computational Microscopy of Glycoproteins: from Viruses to T Cells

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Biochemistry and Molecular Biophysics

by

Mia Aurora Rosenfeld

Committee in charge:

Professor Rommie E. Amaro, Chair
Professor Elizabeth A. Komives, Co-Chair
Professor Jack A. Gilbert
Professor Judy E. Kim
Professor Andrew McCammon
Professor Susan S. Taylor

2023

The dissertation of Mia Aurora Rosenfeld is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To Prof. Martin H. Rosenfeld, PhD. I stand on the shoulders of giants.

EPIGRAPH

You came to take us

All things go, all things go

To recreate us

All things grow, all things grow

We had our mindset

All things know, all things know

You had to find it

All things go, all things go

Sufjan Stevens, *Chicago* (2005)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

ACKNOWLEDGEMENTS

I'd like to begin by recounting the philosophy of Shabbat, the Jewish day of rest. Every Friday, we come together as a community and marvel at the tiny, seemingly stochastic events that lead all of us to that very location at that very moment in time in the history of this very universe. Me being here, writing the acknowledgements section of my doctoral dissertation, I can't help but marvel that this is indeed the sum of many, many small decisions and experiences that brought me here–and a direct result of the influence of so many incredible people along my personal and scientific journey.

I'd like to begin by acknowledging the importance of community and community support during grad school. At many moments during my doctoral journey, I was deeply depressed, isolated, and wanted to quit. The science is always cool and fun, but sometimes (correction: a lot of the time) that isn't enough. My saving grace has always been my incredible communities, specifically: my queer community (Emily Mercado, Gabi Kirsch, DC Queers et al, Aaron McLeod, Victor Wong, Rachel Hall, and Jovan Dragelj), my Jewish community (Eleanor Castracane, Nicole Kennedy, Danielle Harvey, Klara Vesela, UCSD Hillel Graduate Student community, and Onward Israel/Career Passport), and my student government community (Giulia Corno, Ross Turner, Angela Garais, Patiemma Rubio, and the rest of UCSD GPSA).

To those who planted the first seeds of curiosity in me: everyone at Glendale Montessori who turned me from a tiny, unknowing, 7-year-old blob into a real, functioning, curious human, especially Hal Tyler, Teri Jover, Bob Layport, Nina O'Brien, and Marta Homo-Tyler; my truly incredible teachers at Immaculate Heart middle & high school, especially Carla Trujillo and Maya Diakoff for making the initial spark in my soul bright for math and chemistry, and Stephanie Wong for teaching me the beauty of acceptance and compassion through the lens of spirituality; and Professor Michelle Scholefield at Santa Monica College, encouraging me that I was capable of actually doing Chemistry by simply being a wonderful and engaging teacher, and unwittingly convincing me to get a degree in Chemistry after finishing community college (Dr. Scholefield, you're probably not reading this, but if you are – now you know! ;)). Thank you.

Neverending thank you to my first ever PIs, Professors Antje and Paulo Almeida at my undergraduate alma mater University of North Carolina Wilmington. Antje saw something in me that I didn't know existed: I came to her and asked if I could wash glassware in her lab, and she replied with "how about we set you up with a project instead?" The trajectory-changing experience of working in her lab for ~3 years in undergrad turned me from a bright-eyed pre-med student into a bright-eyed biophysicist-in-training, forever shifting my path from medicine to research. An additional thank you to my NIH mentors, Dr. Alex Kelly and Dr. Ana Robles, who let me join their labs over the span of two summers during my undergraduate education, giving me the confidence to solidify my decision to pursue a career in science (maybe even at NIH one day!).

Unfathomably large thank you to Dr. Fiona Kearns and Dr. Lorenzo Casalino for being the best postdoctoral mentors I could have ever asked for. When I switched from experimental biophysics to computational biophysics, they took me under their wing, even though I knew approximately… ~nothing. They were (and still are!) incomprehensibly generous with their expertise, time, and support. In a million years, I couldn't have dreamed up better mentors and friends than Fiona and Lorenzo. They have been a shimmering, brilliant bright spot of my PhD. Thank you, Fiona and Renzo. I am so, so lucky.

On a similar note, I'd like to thank Constanza Torres-Paris and Dr. Maira Rivera for teaching me nearly everything I know about protein purification. Long hours in the lab were turned from hellish to absolutely, positively delightful because of Cony and Maira. And to Cony: thank you for being a confidant about life, science, and everything in between when I needed it most.

To my two labs that I call home, the Amaro Lab and the Komives Lab: thank you for always making doing science feel like a warm hug, especially Abby Dommer, Emilia Pecoro de Barros, Nick Wauer, Shirley Ahn, Sarah Kohaneck, Xandra Nuqui, Patiemma Rubio, Ozlem Demir, Mo Shehata, Teri Simas, Hannah Baughman, Calvin Lin, Riley Peacock, and Wei Chen.

Huge shout out to my experimental collaborators who were willing to take chances on lofty joint projects with me and my AmaroLab coworkers, especially Profs. Stacy Malaker and Ronit Freeman. Stacy and Ronit have served as shining examples of strong women in science for me, with their willingness to combine

techniques that span scales and their openness to give advice, whether it's science-related or life-related. Additionally, thank you to all of my collaborators on the Gordon Bell 2022 team for your hard work and dedication to doing groundbreaking and absolutely massively-scaled science. Pushing the boundaries of what is possible with this team was truly exhilarating and reminded me why I love doing science.

Thank you to Sydney Cromer, whose friendship emboldened me to really become me.

Thank you to Gary & Sara for friendship, Trader Sam's cocktails, and Disneyland trips when I was in the depths of PhD despair. And thank you, internet, for bringing these incredible humans into my life.

My family, to whom one "thank you" isn't enough: to Becky and Paul, for being my second set of east coast parents and for always rooting for me; to Bubby, Pops, Donna, Barry, Clare, Lauren, Jonathan, Cara, and Allie, for always being a soft, familiar, and loving place for me to land each Spring; to Adam and Frances, for teaching endless curiosity by example, and for watching me grow up and helping me embrace who I truly am along the way; to my mom, who has fiercely and unconditionally loved me through every single part of my personal and scientific journey, even if my science doesn't always make complete sense to her; and last but not least, to my dad, my absolute best friend in the entire world, who instilled a spirit of curiosity and a love for learning in me from birth, solidified in all of the little moments: going over flashcards at the diner in middle school, ideating in your car during rides through Forest Lawn in high school, and talking through my first ever research project at Killer Cafe in college.

To my committee members, Professor Susan S. Taylor, Professor J. Andrew McCammon, Professor Judy E. Kim, and Professor Jack A. Gilbert: thank you for seeing me through this journey and for encouraging me to always strive for greatness & rigor in my research.

Finally, I'd like to acknowledge my two spectacular PhD advisors, Prof. Rommie Amaro and Prof. Elizabeth (Betsy) Komives. I spent the first two years of my PhD working in Betsy's experimental biophysics lab. The first two years of a PhD are canonically the most personally and scientifically difficult, and there were many moments I nearly gave up. Betsy unwittingly convinced me not to quit just by being herself: merely

being in her midst to experience the way she fosters community and unconditionally supports her students gave me the strength to not give up. If Betsy believed in me, I knew I could do it.

Things inevitably shifted during the pandemic, and Rommie gave me the opportunity to live with my mom in North Carolina, work completely remotely, and switch to being entirely computational. If I'm being honest, I never thought I'd enjoy being a computational biophysicist – after all, I had been wearing a lab coat and micropipetting clear buffer from one test tube to another since I was 19 in undergrad. However, upon immersing myself in Rommie's entirely computational SARS-CoV-2 spike glycoprotein projects, it dawned on me: "Oh. This is it. This is what I've been looking for. This is what I want to do for the rest of my life."

Once things slowed down (marginally) with the SARS-CoV-2 projects, me and my mentor/work partner/dear friend, Fiona Keans, got really intrigued by mucins. Yep, you heard that right: Mucins. Mucus. Boogers. This class of bio-macromolecule are everywhere in our bodies and are involved in a diverse array of functions, but are incredibly understudied, making them (in my opinion) one of the remaining great frontiers of biological science. Fiona and I worked a few weeks under the table trying to put together our first pass of an all-atom, computational model of a respiratory mucin – something that had never, ever been done before. We brought it to Rommie to show it to her and to ask if we could continue to pursue studying this class of molecule, and she immediately said yes. This is emblematic of Rommie: she values innovation, diversity of thought, creativity, and pushing the boundaries of what's possible. Rommie Amaro is not only my doctoral mentor, but also one of my greatest scientific idols. I hope to continue the forward march of science with this same philosophy of doing the un-doable, always with Rommie and Klaus in mind. I can't believe my luck when I think back on my fairytale PhD with Rommie, all of which having been enabled by her always advocating for me, supplying me with resources to try whatever "pie-in-the-sky" thing I proposed, and allowing me to learn what kind of scientist I wanted to be through leading by example.

Two years of my PhD was generously funded by the UCSD Interfaces Graduate Training Program (NIH T32 EB009380).

Chapter 2, in full, is a modified reprint of the material as it appears in “Chang-ki Oh , Tomohiro Nakamura, Nathan Beutler, Xu Zhang, Juan Piña-Crespo, Maria Talantova, Swagata Ghatak, Dorit Trudler, Lauren N. Carnevale , Scott R. McKercher, Malina A. Bakowski, Jolene K. Diedrich, Amanda J. Roberts, Ashley K. Woods, Victor Chi, Anil K. Gupta, Mia A. Rosenfeld , Fiona L. Kearns, Lorenzo Casalino , Namir Shaabani, Hejun Liu , Ian A. Wilson , Rommie E. Amaro, Dennis R. Burton , John R. Yates III, Cyrus Becker , Thomas F. Rogers, Arnab K. Chatterjee , Stuart A. Lipton. *Targeted protein S-nitrosylation of ACE2 inhibits SARS-CoV-2 infection.* Nature Chemical Biology 19, 275–283 (2023).” The dissertation author was a primary co-investigator of all computational findings in this work, and a co-author of this work.

Chapter 3, in full, is a modified reprint of the materials as it appears in “Kim SH†, Kearns FL†, Rosenfeld MA†, Casalino L, Papanikolas MJ, Simmerling C, Amaro RE, Freeman R. *GlycoGrip: Cell Surface-Inspired Universal Sensor for Betacoronaviruses.* ACS Cent Sci. 2022 Jan 26;8(1):22-42.” The dissertation author was a primary co-investigator and co-author of this work.

Chapter 4, in full, is a modified reprint of the materials as it has been accepted for publication in Cell Reports Physical Science as “Kim SH†, Kearns FL†, Rosenfeld MA†, Votapka L, Casalino L, Papanikolas MJ, Amaro RE, Freeman R. Positively bound: *SARS-CoV-2 Evolved Variants Optimize Binding to Cellular Glycocalyx.* Cell Reports Physical Science. 2023. (accepted).” The dissertation author was a primary co-investigator and co-author of this work.

Chapter 5, in full, is a modified reprint of the materials as it appears in “Dommer, A., † Casalino, L.†, and Kearns, F. †; Rosenfeld, M.; Wauer, N.; Ahn, S-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A.; Trifan, A.; Brace, A.; Sztain, T.; Clyde, A.; Ma, H.; Chennubhotla, C.; Lee, H.; Turilli, M.; Khalid, S.; Tamayo-Mendoza, T.; Welborn, M.; Christensen, A.; Smith, D.; Qiao, Z.; Sirumalla, S.; O’Connor, M.; Manby, F.; Anandkumar, A.; Hardy, D.; Phillips, J.; Stern, A.; Romero, J.; Clark, D.; Dorrell, M.; Maiden, T.; Huang, L.; McCalpin, J.; Woods, C.; Grey, A.; Williams, M.; Barker, B.; Rajapaksha, H.; Pitts, R.; Gibbs, T.; Stone, J.; Zuckerman, D.; Mulholland, A.; Miller, T.; Jha, S.; Ramanathan, A.; Chong, L.; Amaro, R. *#COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol*. Int J High Perform

Comput Appl. 2023 Jan;37(1):28-44." The dissertation author was a co-investigator and co-author of this work. **Notably, this work was selected as a Finalist for the 2021 ACM Gordon Bell Special Prize for COVID-19 Research.**

Chapter 6, in full, is a modified reprint of the materials as it has been submitted for publication in Science as "Joann Chongsaritsinsuk‡, Alexandra D. Steigmeyer‡, Keira E. Mahoney‡, Mia A. Rosenfeld‡, Taryn M. Lucas, Deniz Ince, Fiona L. Kearns, Alexandria S. Battison, Marie A. Hollenhorst, D. Judy Shon, Katherine H. Tiemeyer, Victor Attah, Catherine Kwon, Carolyn R. Bertozzi, Michael J. Ferracane, Rommie E. Amaro, Stacy A. Malaker. Glycoproteomic landscape and structural dynamics of TIM family immune checkpoints enabled by mucinase SmE. 2023." The dissertation author was a primary co-investigator and co-author of this work.

VITA

2012	Valedictorian, Polk County Early College
2012-2014	Associates in Science, Santa Monica College
2014-2018	Bachelor of Science in Chemistry Bachelor of Arts in Mathematics University of North Carolina Wilmington
2018-2020	Teaching Assistant, University of California San Diego
2018-2023	Doctor of Philosophy in Chemistry University of California San Diego

PUBLICATIONS

Özlem Demir†, Emilia P Barros†, Tavina L Offutt†, **Mia A. Rosenfeld**†, Rommie E Amaro†, An integrated view of p53 dynamics, function, and reactivation, Current Opinion in Structural Biology, Volume 67, 2021, Pages 187-194, ISSN 0959-440X, doi.org/10.1016/j.sbi.2020.11.005.

Kim SH†, Kearns FL†, **Rosenfeld MA†**, Casalino L, Papanikolas MJ, Simmerling C, Amaro RE, Freeman R. *GlycoGrip: Cell Surface-Inspired Universal Sensor for Betacoronaviruses.* ACS Cent Sci. 2022 Jan 26;8(1):22-42. Epub 2021 Dec 15. PMID: 35106370; PMCID: PMC8796303.Jan 26;8(1):22-42. doi: 10.1021/acscentsci.1c01080.

Fiona L. Kearns†, Daniel R. Sandoval†, Lorenzo Casalino†, Thomas M. Clausen†, **Mia A. Rosenfeld**†, Charlotte B. Spliid†, Rommie E. Amaro, Jeffrey D. Esko. Spike-heparan sulfate interactions in SARS-CoV-2 infection, Current Opinion in Structural Biology, Volume 76, 2022, 102439, ISSN 0959-440X, https://doi.org/10.1016/j.sbi.2022.102439.

Dommer, A., † Casalino, L.†, and Kearns, F. †; **Rosenfeld, M**.; Wauer, N.; Ahn, S-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A.; Trifan, A.; Brace, A.; Sztain, T.; Clyde, A.; Ma, H.; Chennubhotla, C.; Lee, H.; Turilli, M.; Khalid, S.; Tamayo-Mendoza, T.; Welborn, M.; Christensen, A.; Smith, D.; Qiao, Z.; Sirumalla, S.; O'Connor, M.; Manby, F.; Anandkumar, A.; Hardy, D.; Phillips, J.; Stern, A.; Romero, J.; Clark, D.; Dorrell, M.; Maiden, T.; Huang, L.; McCalpin, J.; Woods, C.; Grey, A.; Williams, M.; Barker, B.; Rajapaksha, H.; Pitts, R.; Gibbs, T.; Stone, J.; Zuckerman, D.; Mulholland, A.; Miller, T.; Jha, S.; Ramanathan, A.; Chong, L.; Amaro, R. #COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol. *Int J High Perform Comput Appl.* 2023 Jan;37(1):28-44. doi: 10.1177/10943420221128233. Epub 2022 Oct 2. PMID: 36647365; PMCID: PMC9527558.

Chang-ki Oh , Tomohiro Nakamura, Nathan Beutler, Xu Zhang, Juan Piña-Crespo, Maria Talantova, Swagata Ghatak, Dorit Trudler, Lauren N. Carnevale , Scott R. McKercher, Malina A. Bakowski, Jolene K. Diedrich, Amanda J. Roberts, Ashley K. Woods, Victor Chi, Anil K. Gupta, **Mia A. Rosenfeld**, Fiona L. Kearns, Lorenzo Casalino , Namir Shaabani, Hejun Liu , Ian A. Wilson , Rommie E. Amaro, Dennis R. Burton , John R. Yates III, Cyrus Becker , Thomas F. Rogers, Arnab K. Chatterjee , Stuart A. Lipton. Targeted protein S-nitrosylation of ACE2 inhibits SARS-CoV-2 infection. *Nature Chemical Biology* 19, 275–283 (2023). https://doi.org/10.1038/s41589-022-01149-6.

Kim SH†, Kearns FL†, **Rosenfeld MA**†, Votapka L, Casalino L, Papanikolas MJ, Amaro RE, Freeman R. Positively bound: SARS-CoV-2 Evolved Variants Optimize Binding to Cellular Glycocalyx. *Cell Reports Physical Science.* 2023. (accepted)

Joann Chongsaritsinsuk‡, Alexandra D. Steigmeyer‡, Keira E. Mahoney‡, **Mia A. Rosenfeld‡**, Taryn M. Lucas, Deniz Ince, Fiona L. Kearns, Alexandria S. Battison, Marie A. Hollenhorst, D. Judy Shon, Katherine H. Tiemeyer, Victor Attah, Catherine Kwon, Carolyn R. Bertozzi, Michael J. Ferracane, Rommie E. Amaro, Stacy A. Malaker. Glycoproteomic landscape and structural dynamics of TIM family immune checkpoints enabled by mucinase SmE. 2023. *bioRxiv* 2023.02.01.526488; doi: https://doi.org/10.1101/2023.02.01.526488

† Shared first-authorship.

ABSTRACT OF THE DISSERTATION

Computational Microscopy of Glycoproteins: from Viruses to T Cells

by

Mia Aurora Rosenfeld

Doctor of Philosophy in Biochemistry and Molecular Biophysics

University of California San Diego, 2023

Professor Rommie E. Amaro, Chair

Professor Elizabeth A. Komives, Co-Chair

Glycosylation is a post-translational modification of proteins in which carbohydrate chains, known as glycans, are linked to lipids or proteins. Many biological proteins are glycosylated, and these glycans are important in the necessary biological functionality of proteins. In this dissertation, the structure/function relationship of highly glycosylated proteins are explored using molecular dynamics simulations and experiment in concert, specifically: 1) the structural & mechanistic explanation of preventing COVID-19 infection with a novel inhaled antiviral compound, 2) the dynamics of the SARS-CoV-2 spike glycoprotein interacting with its cofactors on the host-cell and exploitation of said interaction to develop a rapid antigen test, 3) the mechanism of airborne transmission of the SARS-CoV-2 virion in a respiratory aerosol, and 4) the relationship between glycosylation and functional dynamics of T-cell immunoglobulin and mucin domain containing proteins.

Chapter 1: Introduction

1.1 Overview

This dissertation is organized into two parts. Part I is entitled “Treating, Diagnosing, and Transmission of COVID-19,” and includes four publications[1,2,3,4] with a unifying thread of the SARS-CoV-2 spike glycoprotein as the main protein of interest. In Part I of this dissertation, the following will be discussed: 1) the structural & mechanistic explanation of preventing COVID-19 infection with a novel inhaled antiviral compound, 2) the dynamics of the SARS-CoV-2 spike glycoprotein interacting with its cofactors on the host-cell and exploitation of said interaction to develop a rapid antigen test, 3) the mechanism of airborne transmission of the SARS-CoV-2 virion in a respiratory aerosol. In Part II, entitled “Mucin domain-containing Immunoprotein Modeling,” includes one publication[5] deposited into the biorXiv as a preprint. In this part of the dissertation, initial findings regarding mucin/spike interactions are built upon to streamline a mucin-domain modeling workflow, subsequently used to explore the relationship between glycosylation and functional dynamics of T-cell immunoglobulin and mucin domain containing proteins.

While Part I of the dissertation focuses on the dissertation author’s contribution to studying glycosylated viral proteins, a well established area of expertise of the Amaro Lab, the last chapter in Part I and all of Part II focuses on modeling and simulating a class of macromolecules which have never been simulated before: mucins. Part I sets the stage for highly glycosylated proteins, and Part II brings it all together by focusing on arguably the most highly glycosylated physiologically relevant class of molecules to exist.

All throughout the dissertation, both experimental and computational results are displayed. Although the dissertation author is only responsible for directly producing the computational results that are presented herein, it is necessary to include all results to 1) tell the entire scientific story and 2) to emphasize the importance of collaboration between computationalists and experimentalists to push the frontiers of scientific discovery. The dissertation author, although not involved in direct production of experimental results shared in this dissertation, was directly involved in the analysis, integration of results, and hypothesis generation, all of which takes into account both computational and experimental data. For example, in Chapter 2, a novel

inhaled compound is discovered by the Lipton lab and the dissertation author uses computational methods to posit a mechanism of action of the drug in vivo.

In Chapter 3 and 4, a collaboration between the dissertation author and the Freeman lab at UNC-Chapel Hill is summarized, in which the dissertation author used computational techniques to propose a novel binding ridge for the host cell glycocalyx on the SARS-CoV-2 spike protein. This information was subsequently exploited to produce rapid antigen test that is half of the price to make than a traditional antigen test, in addition to taking 80% less time to fine-tune to emerging SARS-CoV-2 variants of concern. Computational techniques were also applied to fine-tuning the rapid test for production, such as antibody selection and sample buffer ionic strength selection.

Chapter 4 represents a first-in-class simulation of a 1.1 billion atom virus-laden respiratory aerosol, including direct contributions from the dissertation author, specifically the SARS-CoV-2 spike glycoprotein variant models designed in Chapter 4, albumin, and respiratory mucins. This work including respiratory mucins represents the first of its kind, due to their complexity, the O-glycosylated regions of mucins have never before been constructed for molecular simulations.

Chapter 6 details novel discoveries about the structural properties and dynamics of TIMs, an extremely understudied class of immune checkpoint proteins. These molecules are understudied because they are extremely difficult to quantitatively characterize, as they contain highly heterogeneous mucin domains. The dissertation author builds off of their previous experiences and works hand-in-hand with mass spectrometrist, Prof, Stacy Malaker at Yale, who details the entire glycoprofile of TIM-3 and TIM-4. From there, the dissertation author constructs the first-ever molecular model of TIM-3 and TIM-4. Ultimately, the dissertation author proposes a mucin-domain dynamics related constant to further define the mucinome and mucinome membership, something that currently only includes static/non-dynamic data such as amino acid sequence and absolute density of glycosylation.

This dissertation represents firsts of many kinds in terms of computational achievements, as the systems described herein are between 800,000 and 1 billion atoms, but it also represents the advent of a new field

and frontier in science: the study of computational mucinology. Mucins are incredibly large and highly heterogeneous, and as such, there are very few quantitative studies on this class of macromolecules. The all-atom molecular dynamics simulations of respiratory mucins and TIMs in this dissertation represent first-in-class investigations, as these computational tools had never been applied to this field before. This work also highlights the importance of collaboration and co-first authorship. The computational models and findings explored in this dissertation are supplemented by experimental validation, and vice versa, enabling more rigorous characterization of physiologically relevant biomolecules that span scales and techniques. Both experimental and computational results are shared in the publications within this dissertation to both tell the entire scientific story and to emphasize the importance of cross-discipline collaboration. The exploration of scales and never-before-characterized macromolecules in this dissertation serves as a testament to the importance and efficiency of cross-discipline collaboration.

1.2 Glycoproteins

Proteins are molecular machines that have a diverse array of functions in vivo, such as enzymes that catalyze biochemical reactions, viral proteins that mediate host-cell entry, and histones that regulate gene expression and DNA organization. Proteins are made of amino acids that are bound together, where the amine and carboxylic acid of the residues are joined together via amide bond. Protein structure is thought of in four levels of canonical organization: 1) primary structure, the sequence of amino acids bound together, 2) secondary structure, local interactions of the peptide chain to form alpha helices and beta sheets, 3) tertiary structure, the three dimensional structure associated with a single protein chain, usually including beta sheets, alpha helices, and disordered regions, and 4) quaternary structure, the arrangement of multiple subunits with tertiary structure into a multi-protomer functional protein. The relationship between protein structure and protein function underlie the field of protein biophysics as a whole. Understanding a protein's function not only requires information about the three dimensional structure, but also the rigorous characterization of the conformational and dynamical landscape of a protein.

Often, proteins are post-translationally modified with covalent processing events, such as acetylation, methylation, or phosphorylation of amino acids, that result in altered function of the protein. One notable class of post-translational modifications is glycosylation, wherein an oligosaccharide is linked to the amino acid side chain of a protein in the endoplasmic reticulum. There are two distinct types of glycosylation: N-glycosylation and O-glycosylation. N-glycosylation is critical for proper protein folding and is highly conserved in all eukaryotes. O-glycosylation is much more diverse, and is responsible for a vast array of protein structure, such as modulating immune response, providing cartilage flexibility, and contributing to cell-surface or secreted mucins. Aberrant glycosylation is a hallmark of many diseases, such as cancer and congenital disorders. In this dissertation, discussion of the influence of glycans on protein conformation and function will be explored using both experimental and computational methods. Contribution of both N and O linked glycans will be explored in the context of viral glycan shields and mucins, respectively.

1.3 Molecular Dynamics

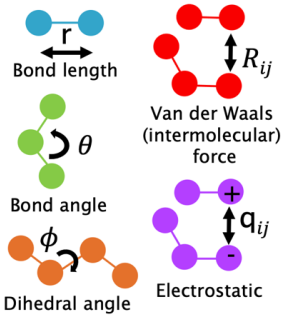
Molecular dynamics simulation is a computational tool in which the classical equations of motion are solved to generate trajectories that describe an atom's position over the course of time. We define N as atoms described in our system and calculate how these atoms move over a defined timestep. We use Newton's equation of motion (Eq. 1.1, below) to calculate the series of positions of a given atom over time with mass m and acceleration a. Specifically, $r_i(t)$ and $F_i(t)$ represent the position and force, respectively, of atom *i* at time *t*.

$$m_i a_i = m_i \frac{d^2 r_i}{dt^2} = F_i \quad i = 1 \ldots N, \tag{1.1}$$

The force of the atom is defined by the gradient of the potential energy function (Eq. 1.2, below), calculated based on the initial position of the atoms within the system. The potential is negative in this equation, as all atoms move in such a way to minimize their total energy.

$$F_i = -\nabla_{r_i} V\left(r_i, \ldots, r_N\right), \tag{1.2}$$

To calculate the potential, we use a set of equations that approximate the energetic contributions to the potential energy function (Fig 1.1), evaluating energetic contributions from bonded and non-bonded interactions with neighboring atoms. Bonded interactions are interactions that result from atoms being connected to one another through a chemical bond. Non-bonded interactions, on the other hand, are interactions that result from non-connected atoms that are physically nearby one another and interact via attractive, repulsive, and electrostatic interactions.



$$U(\vec{R}) = \sum_{bonds} k_i^{bond}(r_i - r_0)^2 + \sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2 + \sum_{dihed} k_i^{dihed}[1 + \cos(n_i \emptyset_i + \delta_i)] + \sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}$$

Figure 1.1: Depiction of bonded and non-bonded energetic contributions to the potential energy function, both describing the subsets of interactions and the equations used to solve them.

The constants associated with the bond, angle, rotation, and atom pairs for all types of interactions make up a force field, which is what allows us to solve the above equations and perform a molecular dynamics simulation. These constants are acquired through experimental observation. The forcefield used in this dissertation is CHARMM36m[6], as it is the only force field on the market that has successfully paramaterized the atoms and bonds associated with N and O glycans.

Through performing molecular dynamics simulations for satisfactorily long timescales, statistically significant data can be extracted from the resulting trajectories and used to surmise mechanistic hypotheses based on rigorous protein dynamic data at an atomistic level of detail. Oftentimes, these atomistic insights are simply impossible in experiment, and on the other hand, physiological timescale data is often impossible in computational simulations. The limitations of computational techniques discussed in this dissertation are complemented by the experimental technique accompanying it, and vice versa.

1.4 **References**

(1) Chang-ki Oh , Tomohiro Nakamura, Nathan Beutler, Xu Zhang, Juan Piña-Crespo, Maria Talantova, Swagata Ghatak, Dorit Trudler, Lauren N. Carnevale , Scott R. McKercher, Malina A. Bakowski, Jolene K. Diedrich, Amanda J. Roberts, Ashley K. Woods, Victor Chi, Anil K. Gupta, Mia A. Rosenfeld, Fiona L. Kearns, Lorenzo Casalino , Namir Shaabani, Hejun Liu , Ian A. Wilson , Rommie E. Amaro, Dennis R. Burton , John R. Yates III, Cyrus Becker , Thomas F. Rogers, Arnab K. Chatterjee , Stuart A. Lipton. Targeted protein S-nitrosylation of ACE2 inhibits SARS-CoV-2 infection. Nature Chemical Biology 19, 275–283 (2023). doi.org/10.1038/s41589-022-01149-6.

(2) Kim SH†, Kearns FL†, Rosenfeld MA†, Casalino L, Papanikolas MJ, Simmerling C, Amaro RE, Freeman R. GlycoGrip: Cell Surface-Inspired Universal Sensor for Betacoronaviruses. ACS Cent Sci. 2022 Jan 26;8(1):22-42. Epub 2021 Dec 15. PMID: 35106370; PMCID: PMC8796303.Jan 26;8(1):22-42. doi: 10.1021/acscentsci.1c01080.

(3) Kim SH†, Kearns FL†, Rosenfeld MA†, Votapka L, Casalino L, Papanikolas MJ, Amaro RE, Freeman R. Positively bound: SARS-CoV-2 Evolved Variants Optimize Binding to Cellular Glycocalyx. Cell Reports Physical Science. 2023. (accepted)

(4) Dommer, A., † Casalino, L.†, and Kearns, F. †; Rosenfeld, M.; Wauer, N.; Ahn, S-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A.; Trifan, A.; Brace, A.; Sztain, T.; Clyde, A.; Ma, H.; Chennubhotla, C.; Lee, H.; Turilli, M.; Khalid, S.; Tamayo-Mendoza, T.; Welborn, M.; Christensen, A.; Smith, D.; Qiao, Z.; Sirumalla, S.; O'Connor, M.; Manby, F.; Anandkumar, A.; Hardy, D.; Phillips, J.; Stern, A.; Romero, J.; Clark, D.; Dorrell, M.; Maiden, T.; Huang, L.; McCalpin, J.; Woods, C.; Grey, A.; Williams, M.; Barker, B.; Rajapaksha, H.; Pitts, R.; Gibbs, T.; Stone, J.; Zuckerman, D.; Mulholland, A.; Miller, T.; Jha, S.; Ramanathan, A.; Chong, L.; Amaro, R. #COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol. Int J High

Perform Comput Appl. 2023 Jan;37(1):28-44. doi: 10.1177/10943420221128233. Epub 2022 Oct 2. PMID: 36647365; PMCID: PMC9527558.

(5) Joann Chongsaritsinsuk‡, Alexandra D. Steigmeyer‡, Keira E. Mahoney‡, Mia A. Rosenfeld‡, Taryn M. Lucas, Deniz Ince, Fiona L. Kearns, Alexandria S. Battison, Marie A. Hollenhorst, D. Judy Shon, Katherine H. Tiemeyer, Victor Attah, Catherine Kwon, Carolyn R. Bertozzi, Michael J. Ferracane, Rommie E. Amaro, Stacy A. Malaker. Glycoproteomic landscape and structural dynamics of TIM family immune checkpoints enabled by mucinase SmE. 2023. bioRxiv 2023.02.01.526488; doi.org/10.1101/2023.02.01.526488

(6) Huang, J., Rauscher, S., Nawrocki, G. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14, 71–73 (2017). https://doi.org/10.1038/nmeth.4067

Part I: Treating, Diagnosing, and Transmission of COVID-19

Chapter 2: Targeted protein S-nitrosylation of ACE2 as a potential therapeutic for SARS-CoV-2

2.1 **Abstract**

Prevention of infection and propagation of SARS-CoV-2 is of high priority in the COVID-19 pandemic. Inhaled nitric oxide (NO) has shown promise in human clinical trials, but its mechanism of action remains unknown[1,2]. Here, we describe S-nitrosylation of multiple proteins involved in SARS-CoV-2 infection, including angiotensin converting enzyme 2 (ACE2), the receptor for viral entry[3]. This reaction prevents binding of ACE2 to the SARS-CoV-2 Spike protein, thereby inhibiting viral entry, infectivity, and cytotoxicity. Aminoadamantane compounds also inhibit coronavirus ion channels formed by envelope (E) protein[4,5]. Accordingly, we developed dual-mechanism aminoadamantane nitrate compounds that inhibit viral entry by S-nitrosylating ACE2, facilitated via E-protein channel blockade. These non-toxic compounds are active in vitro and in vivo in the Syrian hamster COVID-19 model, and thus provide a novel potential avenue for therapy.

2.2 **Introduction**

The process of SARS-CoV-2 infection first involves binding of the viral spike protein to a cell surface receptor, which has been shown to be ACE2 (ref. [1]). Viral entry into cells can be accomplished by fusion of the viral E protein, located on SARS-CoV-2 near the spike protein, to facilitate fusion with the cell surface, or endocytosis with subsequent envelope fusion to endosomal membranes[2–4]. We reasoned that using the E protein viroporin channel to target a molecular warhead to the ACE2 receptor to inhibit interaction with the spike protein could yield a novel mechanism for drug action to treat COVID-19.

Along these lines, we built upon our experience in developing the aminoadamantane drug, memantine, as a US Food & Drug Administration-approved treatment for Alzheimer's disease and synthesized several additional compounds with improved efficacy[5–9]. These new chemical probes, termed aminoadamantane nitrates, offer dual-allosteric inhibition of the ion channel associated with the N-methyl-d-aspartate (NMDA) type of glutamate receptor in the brain, with the aminoadamantane moiety providing channel block as well as targeted delivery of a nitric oxide (NO)-related group to S-nitrosylate, and, thus, further inhibit the receptor.

Moreover, these aminoadamantane nitrates have displayed no untoward side effects, such as hypotension or other NO-associated actions, in two-species toxicity studies[5–9]. Intriguingly, aminoadamantane drugs such as amantadine and memantine were originally developed as antiviral agents because they also block the ion channel found in the envelope of multiple viruses, including influenza and the β-coronaviruses, and anecdotal reports in humans suggest that they may possibly offer some efficacy for SARS-CoV-2, but definitive data are lacking[2,3,10–14]. Coupled with the facts that SARS-CoV and SARS-CoV-2 have also been shown to be susceptible to NO, in part by inhibiting their protease and replication cycle[15], and NO-based therapies have shown promise in human clinical trials for COVID-19 treatment[16], we postulated that the new aminoadamantane nitrate drugs might provide mechanistic information on the mode of action of NO against SARS-CoV-2 and offer improved antiviral activity. Indeed, we show here that the aminoadamantane moiety can block the ion channel in the envelope of SARS-CoV-2 to provide a guided missile to target a therapeutic warhead to ACE2 and, thus, prevent interaction with the nearby spike protein. Accordingly, we demonstrate that the viral receptor, ACE2, can be S-nitrosylated by NO-related species generated by the nitro adduct of the aminoadamantane nitrate compounds to inhibit viral entry.

2.3 **Results**

S-Nitrosylated ACE2 inhibits binding to SARS-CoV-2 spike protein

Initially, we investigated the molecular mechanism whereby NO-related species might inhibit SARS-CoV-2 activity. As assessed by biotin switch assay, we found that the host cell membrane protein receptor for SARS-CoV-2, ACE2, and a protease that cleaves the viral spike protein, transmembrane serine protease 2 (TMPRSS2), both of which are necessary for viral entry and infectivity[1,17], could be S-nitrosylated by the physiological small molecule NO donor and transnitrosylating agent S-nitrosocysteine (SNOC) (Fig. 1a–c). Interestingly, multiple cysteine residues have been shown to be of importance in ACE2 and TMPRSS2 activity, so S-nitrosylation might be expected to disrupt their activity[18,19].

We focused on S-nitrosylation of ACE2 (forming SNO-ACE2), reasoning that this nitrosylation reaction might prevent binding of SARS-CoV-2 S protein to ACE2, thus inhibiting viral infection. To test this premise,

we exposed HeLa cells stably expressing human ACE2 (HeLa-ACE2) to SNOC and assessed SNO-ACE2 formation by biotin switch assay. To evaluate binding of the spike protein to these HeLa-ACE2 cells, we then incubated the cells with purified recombinant SARS-CoV-2 spike protein (S1 + S2). Because NO dissipates very quickly from SNOC (<5 minutes at neutral pH), and spike protein was added sequentially after this period, we could rule out the possibility of direct S-nitrosylation of spike protein by SNOC under these conditions. We found that the formation of SNO-ACE2 was stable for at least 12 hours (Extended Data Fig. 1). The receptor-binding domain (RBD) in the S1 subunit of the SARS-CoV-2 spike glycoprotein binds to ACE2 expressed on the surface of host cells, whereas the C-terminal S2 membrane anchoring subunit functions to translocate virus into host cells[19,20]. After pre-incubation of HeLa-ACE2 cells with SNOC, we found significantly decreased binding of purified spike protein to HeLa-ACE2 cells (Fig. 1d,e), consistent with the notion that the cysteine residue(s) susceptible to S-nitrosylation in ACE2 affected spike protein binding.

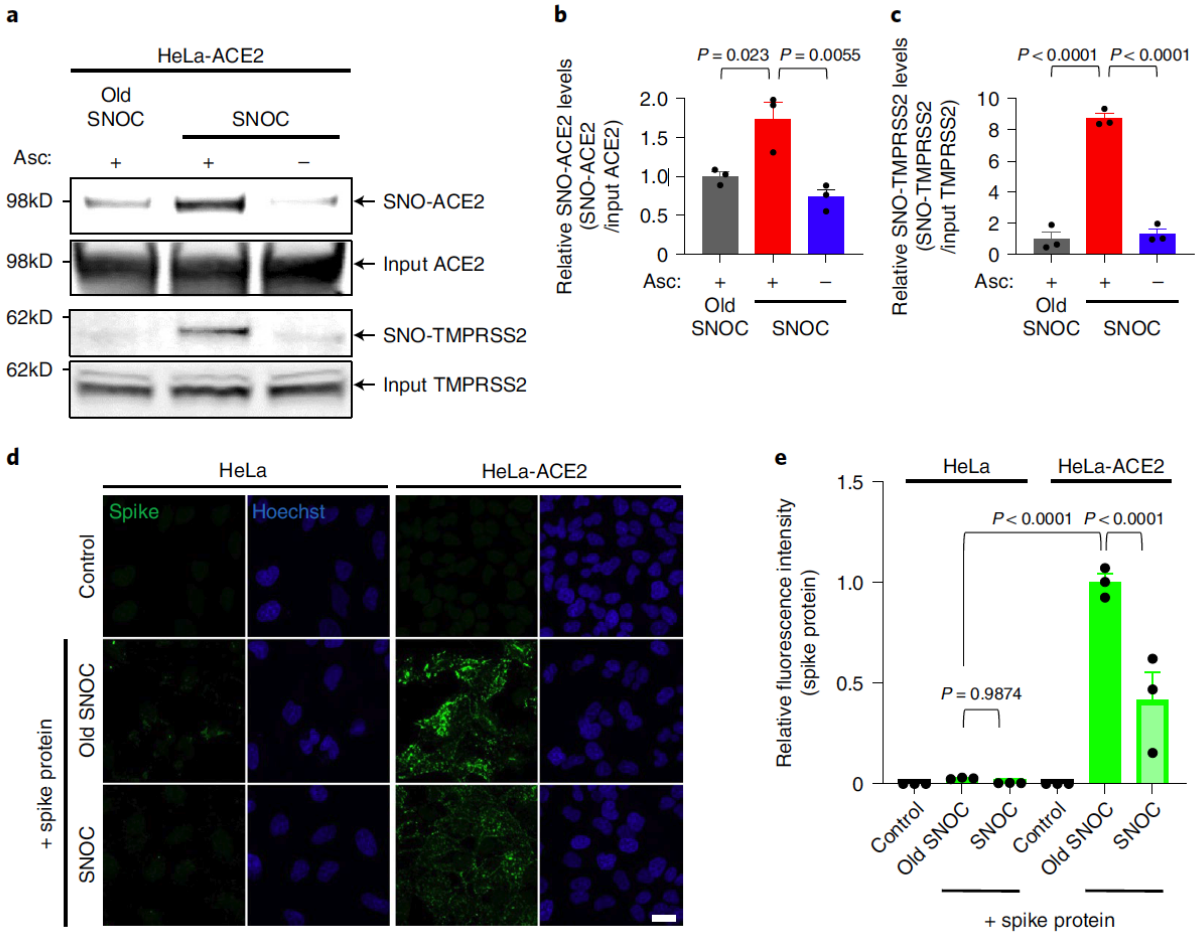


Figure 2.1: SNOC increases S-nitrosylation of ACE2 and inhibits binding of SARS-CoV-2 spike protein

a, Assay for SNO-ACE2 and SNO-TMPRSS2 in HeLa-ACE2 cells. Cells were exposed to 100 μM SNOC or, as a control, 'old' SNOC (from which NO had been dissipated). After 20 minutes, cell lysates were subjected to biotin switch assay to assess SNO and input (total) proteins detected by immunoblotting with cognate antibody. The ascorbate minus (Asc–) sample served as a negative control. b,c, Ratio of SNO-ACE2/input ACE2 protein and SNO-TMPRSS2/input TMPRSS2 protein. Data are mean + s.e.m. by one-way ANOVA with Tukey's multiple comparisons. n = 3 biological replicates. d, HeLa and HeLa-ACE2 cells were pre-exposed to 100 μM SNOC or old SNOC. After 30 minutes, 10 μg ml−1 of purified recombinant SARS-CoV-2 spike (S1 + S2) protein was incubated with the cells. After 1 hour, cells were fixed with 4% PFA for 15 minutes, and bound spike protein was detected by anti-spike protein antibody; nuclei were stained with Hoechst. Scale bar, 20 μm. e, Quantification of relative fluorescence intensity. Data are mean + s.e.m. by two-way ANOVA with Tukey's multiple comparisons. n = 3 biological replicates.

Human ACE2 protein contains eight cysteine residues, six of which participate in formation of three pairs of disulfide bonds, and the remaining two (Cys^{261} and Cys^{498}) are present as free thiols (or thiolates)

(Extended Data Fig. 2a)[19] and, thus, potentially available for S-nitrosylation via reversible nucleophilic attack on a nitroso nitrogen to form an SNO-protein adduct[21]. Accordingly, we performed site-directed mutagenesis of these cysteine residues in ACE2 and found that C261A, C498A or C261A/C498A mutation significantly inhibited SNOC-mediated S-nitrosylation on biotin switch assays, consistent with the notion that these two cysteine residues are targets of S-nitrosylation (Extended Data Fig. 2b,c). Moreover, mass spectrometry confirmed the presence of S-nitrosylated ACE2 at Cys^{261} and Cys^{498} after exposure to SNOC (Extended Data Fig. 2d).

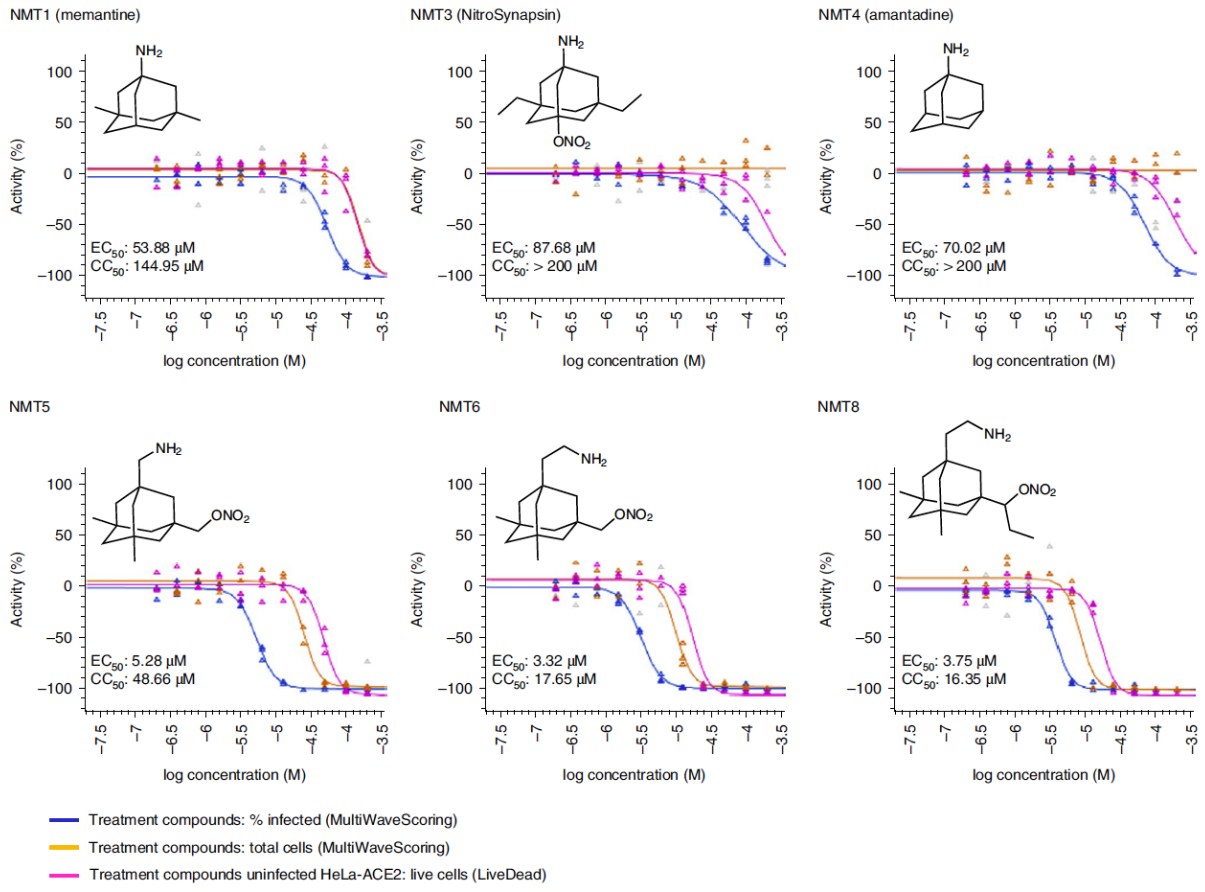


Figure 2.2 Dose response of drugs screened against SARS-CoV-2

Dose–response curves showing the EC_{50} of each compound against SARS-CoV-2 (% infected cells, blue), total cell counts (orange) in the infection experiment and the CC_{50} for uninfected host cell toxicity (magenta), as assessed in HeLa-ACE2 cells. See also Supplementary Dataset 1 for the full dataset.

SNO-ACE2 destabilizes dimerization and spike protein binding Notably, these S-nitrosylation sites on ACE2 are located near the collectrin-like domain (CLD) region rather than the spike protein-binding domain region (Extended Data Fig. 3a). This suggests that S-nitrosylation may affect the conformation of ACE2 protein at some distance from the S-nitrosylated cysteine residue(s) to diminish binding of ACE2 to trimeric spike protein[19,20,22]. Accordingly, explicitly solvated, all-atom molecular dynamics simulations of the SNO-ACE2/ RBD complex in plasma membrane show that the distance between each SNO-ACE2 protomer's center of mass (COM) is overall much longer and more broadly distributed than in simulations of wild-type (WT) ACE2 dimer (Extended Data Fig. 3a,b)[22]. This behavior indicates a certain extent of destabilization of the dimer interface imparted by S-nitrosylation, particularly of C^{498}. Specifically, at the beginning of the simulations, the SNO-ACE2/RBD model displays a hydrogen bond between $Q175_A$ and $Q139_B$, which is then interchanged with $D136_B$ (Extended Data Fig. 3c). This is the only interaction between the peptidase domains (PDs) of the two protomers, as also reported for the initial cryogenic electron microscopy (cryo-EM) structure[23]. Notably, over the course of our simulations, this interaction was progressively lost (Extended Data Fig. 3d), leading to partial disruption of the PD dimeric interface and transient detachment of the two protomers. Therefore, we hypothesize that the addition of S-nitrosylation at the sidechain of C^{498}, which is located in the vicinity of Q^{175}, could be sufficient to induce rearrangement in the packing of secondary structural elements of this region, leading, in turn, to the disruption of the only point of contact between the two PDs of ACE2. The loss of this contact may potentially trigger a further destabilization at the level of the dimeric interface between the neck domains. Alteration of ACE2 dimer stability has the potential to interfere with the SARS-CoV-2 spike binding[24], thus abrogating infection.

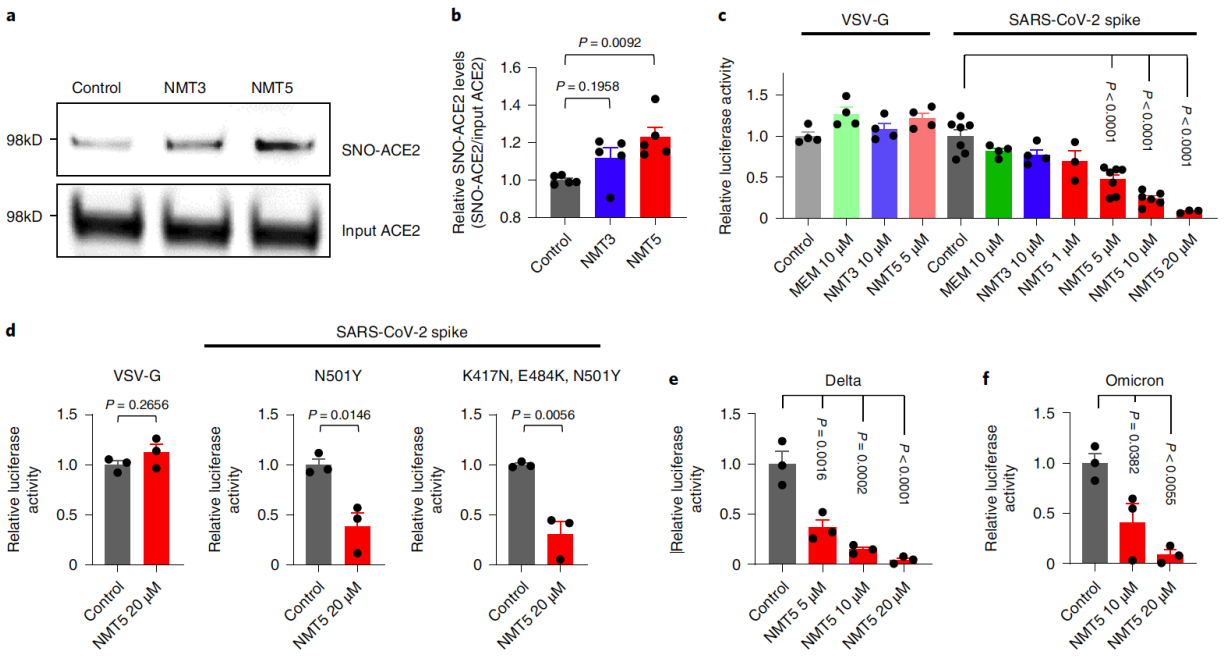


Figure 2.3: NMT5 inhibits SARS-CoV-2 pseudoviral entry

a, HeLa-ACE2 cells were treated with 10 μM NMT3 or 5 μM NMT5. After 1 hour, cell lysates were subjected to biotin switch assay for protein S-nitrosylation, detected by immunoblotting with anti-ACE2 antibody. b, Ratio of SNO-ACE2/input ACE2 protein. Data are mean + s.e.m. by one-way ANOVA with Tukey's multiple comparisons. n = 5 biological replicates. c, HeLa-ACE2 cells were incubated with SARS-CoV-2 spike (D614) or VSV-G (control) pseudovirus particles in the presence and absence of MEM (memantine), NMT3 or NMT5. After 48 hours, viral transduction efficiency was monitored by luciferase activity. Data are mean + s.e.m. by one-way ANOVA with Tukey's multiple comparisons. n = 3–7 biological replicates. d–f, HeLa-ACE2 cells were incubated in the presence and absence of NMT5 with SARS-CoV-2 N501Y spike, SARS-CoV-2 K417N/E484K/N501Y spike or VSV-G (control) pseudovirus particles (d) or with SARS-CoV-2 Delta variant (e) or Omicron variant pseudovirus particles (f). After 48 hours, viral transduction efficiency was monitored by luciferase activity. Data are mean + s.e.m. by two-tailed Student's t-test (d) or one-way ANOVA with Tukey's for multiple comparisons (e,f). n = 3 biological replicates.

Screening aminoadamantane nitrate compounds against SARS-CoV-2 infection

Next, we examined the effect of aminoadamantane nitrate compounds on SARS-CoV-2 entry into cells, causing infection. Aminoadamantaneshave been reported to directly bind to the viroporin ion channel formed by the SARS-CoV-2 E protein[10,25,26]. Therefore, we screened our series of aminoadamantane and nitro-aminoadamantane compounds[6–8] as potential therapeutic drugs against SARS-CoV-2. These latter chemical probes might be expected to bind to the viral channel, thus targeting S-nitrosylation to ACE2 to inhibit its interaction with spike protein and, thus, viral entry. Specifically, we tested, in a masked fashion, the

efficacy against live SARS-CoV-2 in HeLa-ACE2 cells of aminoadamantanes (memantine, blindly coded as NMT1, and amantadine/NMT4) and aminoadamantane nitrate compounds (NMT2, NMT3 and NMT5-NMT9) (Fig. 2 and Extended Data Fig. 4; full dataset shown in Supplementary Dataset 1). As positive controls, we used remdesivir, apilimod and puromycin (Supplementary Dataset 1)[27,28]. In determining the therapeutic potential of these compounds, we considered the selectivity index (SI) that compares a compound's half-maximal non-specific cytotoxicity (CC_{50}) in the absence of infection to its half-maximal effective antiviral concentration (EC_{50}) (CC_{50}/EC_{50}) (Supplementary Dataset 1). The SI can be considered an in vitro indicator of therapeutic index and, ideally, would approach 10. The aminoadamantane compounds alone (amantadine and memantine) offered no efficacy and, thus, were not studied further. In contrast, several of the aminoadamantane nitrate compounds offered some degree of protection from infection. However, NMT6 and NMT8 may have done this simply by killing the host cells irrespective of infection, as evidenced by its off-target killing of uninfected cells in the live/dead assay (Fig. 2 and Supplementary Dataset 1). Among the seven aminoadamantane nitrate compounds tested, NMT5 displayed the best combination of EC_{50} and CC_{50} (SI = 9.2) with an EC_{50} for protection against SARS-CoV-2 of 5.28 μM (Fig. 2 and SupplementaryDataset 1); this concentration of compound is well within the micromolar amounts attainable in human tissues at well-tolerated doses, as tested in two animal species[6–9,29]. Additionally, NMT3 (also known as NitroSynapsin), which was already being developed for central nervous system indications[6–9], displayed some degree of protection against SARS-CoV-2 with an EC_{50} of 87.7 μM, although this value may be artificially high due to the short half-life of NMT3 in aqueous solution under in vitro conditions[6,9,30]. Hence, these two compounds were advanced for further study. We next asked if NMT3 and NMT5 could S-nitrosylate ACE2. We found that NMT5 > NMT3 effectively S-nitrosylated ACE2 both in vitro in HeLa-ACE2 cells and in vivo in Syrian hamsters, as assessed by the biotin switch assay (Fig. 3a,b and Extended Data Fig. 5). Notably, a statistically significant increase in the level of S-nitrosylated ACE2 was observed in the SARS-CoV-2 target tissues of lung and kidney at 48 hours after oral administration of a single dose of drug at 10 mg kg^{-1} (Extended Data Fig. 5d–i). Consistent with the structure–activity relationship indicating that SNO-ACE2 was associated with

the antiviral effect of NMT5 and NMT3, the other aminoadamantane nitrates (including NMT6 and NMT8) did not S-nitrosylate ACE2 at low micromolar concentrations (Extended Data Fig. 6).

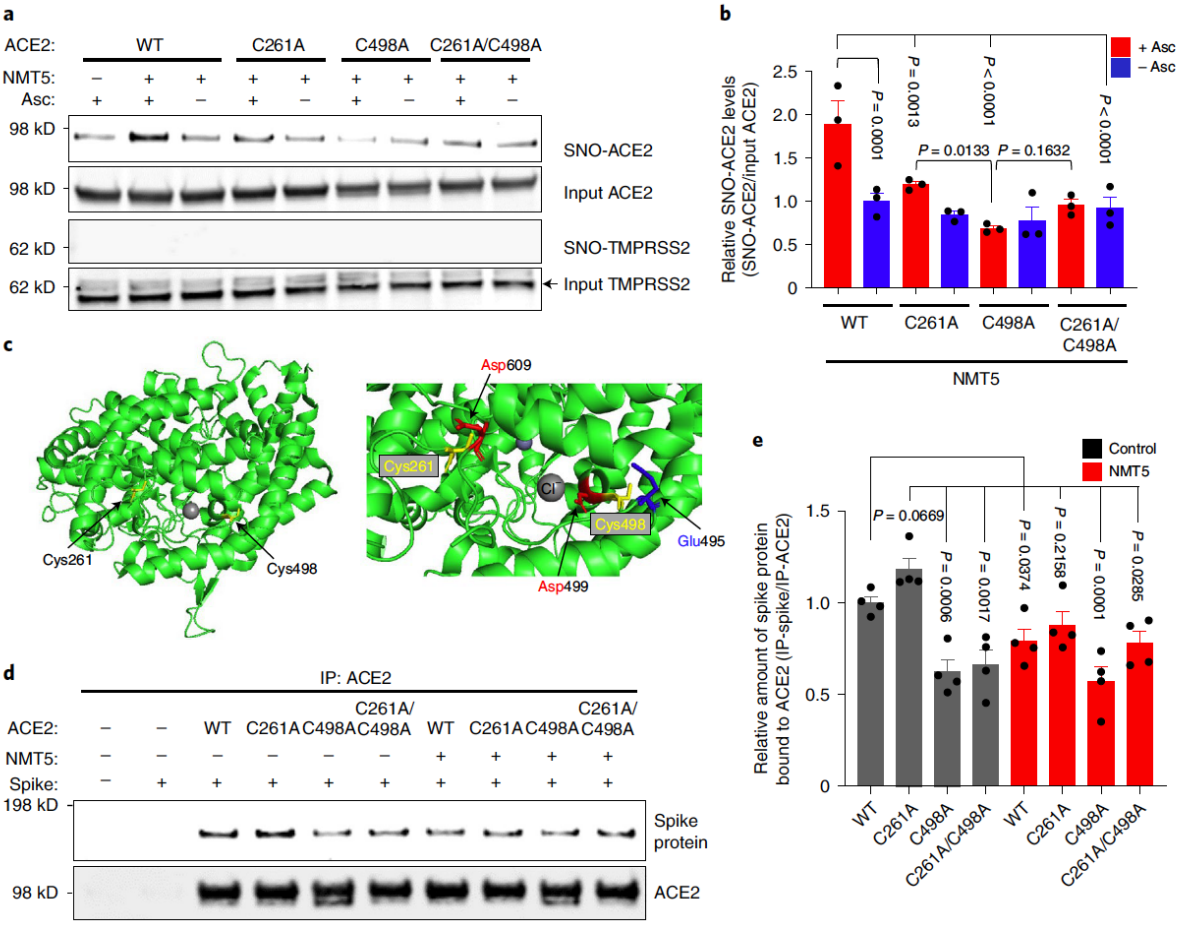


Figure 2.4: S-Nitrosylation of ACE2 by NMT5 inhibits binding to spike protein

a, HEK293T cells were transfected with plasmids encoding human WT ACE2 or non-nitrosylatable mutant ACE2 (C262A, C498A or C261A/C498A). Cells were treated with 10 μM NMT5 and subjected to biotin switch assay for detection of S-nitrosylated proteins by immunoblotting with anti-ACE2 and anti-TMPRSS2 antibodies. The absence of ascorbate (Asc–) served as a negative control. b, Ratio of SNO-ACE2/input ACE2. Data are mean + s.e.m. by two-way ANOVA with Fisher's LSD multiple comparisons. n = 3 biological replicates. c, Crystal structure of ACE2 (left panel; PDB ID: 6M0J) with enlarged view of S-nitrosylation sites of ACE2. Glu^{495} and Asp^{499}, acidic amino acid residues, surround Cys^{498} (right panel). d, HEK293T cells were transfected with plasmids encoding human WT ACE2 or non-nitrosylatable mutant ACE2 (C262A, C498A or C261A/C498A). Cells were exposed to 1 μg ml^{-1} of purified recombinant SARS-CoV-2 spike protein in the presence or absence of 5 μM NMT5; after 1 hour, cells were lysed and subjected to co-IP with anti-ACE2 antibody. Immunoprecipitated ACE2 and spike protein were detected by immunoblotting with anti-ACE2 and anti-spike protein antibodies. e, Ratio of IP-ACE2/IP-spike protein. Data are mean + s.e.m. by one-way ANOVA with Fisher's LSD multiple comparisons. n = 4 biological replicates.

SNO-ACE2 inhibits SARS-CoV-2 variant infectivity

Because we had found that S-nitrosylation of ACE2 inhibited the binding of SARS-CoV-2 spike protein, we next asked if NMT3-mediated or NMT5-mediated SNO-ACE2 formation could prevent viral entry into host cells. To test this premise, we employed a replication-deficient Maloney murine leukemia virus (MLV)-based SARS-CoV-2 spike protein pseudotyped virus, initially using the most prevalent strain of spike protein (D614) as of early 2020 (ref. [31]). We examined whether NMT3 and NMT5 could suppress infection with this SARS-CoV-2 pseudovirus. We found that NMT5 inhibited SARS-CoV-2 pseudoviral entry in a dose-dependent manner, with 5 μM inhibiting 53%, 10 μM inhibiting 76% and 20 μM inhibiting 92% (Fig. 2.3c). NMT3 showed more limited ability to suppress pseudovirus entry, ~24% at 10 μM. The fact that S-nitrosylation of ACE2 manifested inhibition in the pseudovirus assay (as shown in Fig. 2.3c) at approximately the same EC_{50} of 5 μM as found in the live virus infection assay (Fig. 2.2) strongly implies that SNO-ACE2 formation is indeed the predominant mechanism by which NMT5 prevents viral infection. As a control, the NMT5 metabolite lacking the nitro group did not suppress SARS-CoV-2 infection in the pseudovirus entry assay (Fig. 2.3c and Extended Data Fig. ED2.7). As a further control, the pseudovirus entry assay performed with vesicular stomatitis virus G protein (VSV-G) was unaffected by NMT3 or NMT5 (Fig. 2.3c).

We next examined if NMT5 could suppress viral infection from more recently identified SARS-CoV-2 variants, including N501Y spike protein, a common mutation in B.1.1.7 (or Alpha variant, United Kingdom) and B.1.351 (or Beta variant, South Africa), and P.1 (or Gamma variant, Brazil), encountered in the winter of 2020/2021. Additionally, we tested K417N, E484K and N501Y spike protein, as found in the B.1.351 and P.1 variants, the B.1.617.2 variant (or Delta, Indian; T19R, G142D, 156/157 DELETION, R158G, L452R, T478K, D614G, P681R and D950N) and the BA.1 variant (or Omicron, South Africa), found in 2021/2022, which are associated with higher transmissibility or severity as well as altered antigenicity[32]. We found that NMT5 was also effective in reducing infectivity of these SARS-CoV-2 variants, including the Delta and Omicron variants, by up to 95% (Fig. 2.3d–f). These results are consistent with the notion that NMT5 » NMT3-mediated S-nitrosylation of ACE2 can inhibit SARS-CoV-2 entry into host cells.

NMT5-induced SNO-ACE2 inhibits spike protein binding

Next, we sought to determine if NMT5 could modify ACE2 at both of the cysteine residues (Cys^{261} and Cys^{498}) that we demonstrated to be susceptible to S-nitrosylation by SNOC. Analysis by cysteine mutation revealed that NMT5 preferentially S-nitrosylated Cys^{498} over Cys^{261} (Fig. 2.4a,b). Interestingly, the crystal structure of ACE2 shows that an acid/ base motif (comprised of Glu^{495} and Asp^{499}), which, under some conditions, may facilitate S-nitrosylation, is present near Cys^{498}, whereas only a partial motif (represented by Asp^{609}) is found near Cys^{261} (Fig. 2.4c)[33–35]. This observation is consistent with previous findings that potent or supraphysiological amounts of NO, donors such as SNOC, can S-nitrosylate cysteine residues surrounded by no motif or only a partial SNO motif[30], whereas a full SNO motif can facilitate S-nitrosylation by less potent donors, presumably like NMT5. Mechanistically, cysteine thiol groups (in fact, thiolates) are nucleophiles that can perform reversible nucleophilic attack on an electrophilic nitroso nitrogen[36,37]. The local environment of the thiolate anion can kinetically favor one electrophile over another; moreover, the bulky R-group of NMT5, as an RNOx donor (x = 1 or 2) compared to the small molecule SNOC, could sterically hinder reactivity[30]. Notably, concentrations of NMT5 that significantly inhibited viral entry (~10 μM) failed to S-nitrosylate other proteins, including TMPRSS2, spike protein or E protein (Fig. 2.4a and Extended Data Fig. ED2.8), demonstrating relative selectivity of NMT5 for ACE2 at the cell surface. Although we cannot rule out the possibility that proteins associated with virus intracellular trafficking are S-nitrosylated, this would be less likely as aminoadamantane nitrate compounds are known to act on extracellular rather than intracellular targets[5–9].

To further investigate the effect of NMT5 on SARS-CoV-2 spike protein binding to ACE2, we performed co-immunoprecipitation (co-IP) experiments of these two proteins in the presence and absence of NMT5 using anti-ACE2 antibody for IP. As expected, the two proteins co-immunoprecipitated, as evidenced on immunoblots. NMT5 (5 μM) diminished this co-IP, consistent with the notion that the compound could inhibit binding of spike protein to ACE2 to a significant degree (Fig. 2.4d,e). As controls, the spike protein was not co-immunoprecipitated with cysteine mutant ACE2(C498A) or with double mutant ACE2(C261A/C498A) to a significant level, although mutant ACE2(C261A) was still co-immunoprecipitated. These data suggest that S-nitrosylation predominantly of C^{498} of ACE2 is important for spike protein binding to ACE2. Moreover,

NMT5 inhibited co-IP of the spike protein and ACE2(C261A) while having less effect on mutant ACE2(C498A) or ACE2(C261A/C498A) binding (Fig. 2.4d,e). Thus, the co-IP experiment suggests that the intact Cys residue is needed for full interaction/binding of spike protein with ACE2. Either Cys mutation or S-nitrosylation can significantly decrease the interaction of spike protein with ACE2. Taken together, these results are consistent with the notion that NMT5 inhibits SARS-CoV-2 spike protein from binding to ACE2 and, thus, virus entry into the cell via S-nitrosylation of ACE2.

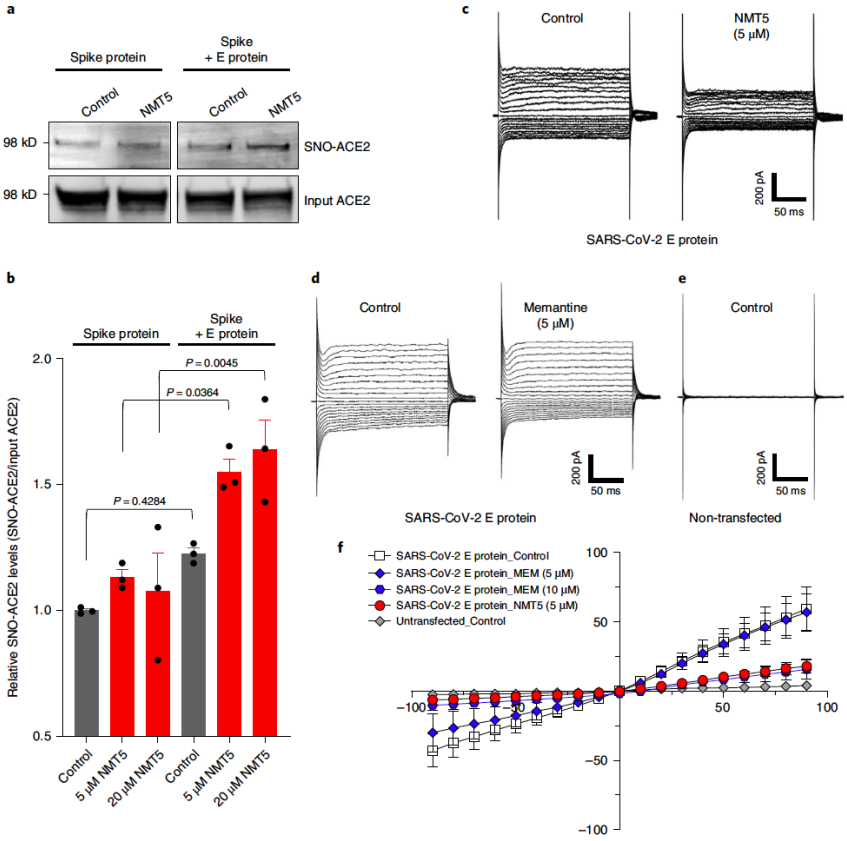


Figure 2.5: Targeted S-nitrosylation of ACE2 and inhibition of E viroporin protein channel by NMT5
a, E protein plasmid was transiently transfected into HEK293-spike protein cells. After 1 day, cells were harvested and plated onto HeLa-ACE2 cells in the presence or absence of 5 μM NMT5. After 30 minutes, cell lysates were subjected to biotin switch assay to monitor protein S-nitrosylation of ACE2, detected by immunoblotting. b, Ratio of SNO-ACE2/total input ACE2 protein. Data are mean + s.e.m. by two-way ANOVA with Tukey's multiple comparisons. n = 3 biological replicates. c–e, Representative traces of whole-cell currents from untransfected (n = 4) and transiently transfected (n = 15) HEK293T cells before and after application of memantine or NMT5 during patch-clamp recording. Whole-cell currents were generated by holding cells at 0 mV and applying voltage steps between –90 mV and +90 mV in increments of 10 mV. f, Current-voltage (I–V) curves from steady-state current density (pA/pF) versus holding potential (mV) for memantine (MEM, 5 μM and 10 μM) and NMT5 (5 μM). Data are mean ± s.e.m.; n = 16 cells recorded.

NMT5 targets SNO to ACE2 via blockade of the E protein channels

Intriguingly, we found that the presence of the E protein of SARS-CoV-2 appeared to target S-nitrosylation by NMT5 to nearby ACE2 receptor proteins (Fig. 2.5a,b and Extended Data Fig. ED2.9). To investigate this action further, we assessed the ability of the aminoadamantane compound, memantine, and the lead aminoadamantane nitrate candidate, NMT5, to block ion channel activity of the E protein[11] using the patch-clamp technique. To test direct interaction with the viroporin channel, we transiently transfected HEK293T cells with a construct encoding the E protein and assessed voltage-dependent currents (versus uninfected cells) in the presence and absence of drug (Fig. 2.5c–f). Under our conditions, we found that the presence of the E protein resulted in a robust voltage-dependent current carried by K^+ that was inhibited by memantine and with greater potency by NMT5. Notably, the low micromolar concentrations needed to see these effects are within attainable levels in mammalian plasma and tissues, as shown in pharmacokinetic (PK) studies, and have proven to be safe in animal toxicity studies[6–9,29].

NMT5 protects from SARS-CoV-2 infection in vivo

In preparation for in vivo efficacy testing of our chemical probes in a COVID-19 small animal model, we next performed 48-hour PK studies after a single oral dose of NMT3 or NMT5 at 10 mg kg^{-1} in ~150-g Syrian hamsters. We found a half-life in plasma for NMT3 of 7.9 hours and for NMT5 of 10.6 hours (Fig. 2.6a,b and Supplementary Table 1). The mean C_{max} for NMT5 was 0.2 μM and ~0.4 μM for NMT3; NMT3 also displayed a hydroxylated metabolite (full detailed PK dataset shown in Supplementary Datasets 2 and 3). The fact that

NMT5 was found to be more stable than NMT3 by mass spectrometry analysis (Fig. 2.6a,b and Supplementary Table 1) was also consistent with previous findings[6–9,29]. Moreover, these drugs are concentrated in tissues up to ~30-fold over plasma levels. Using a Bayesian-like adaptive clinical trial design, we determined the maximal tolerated dose (MTD) of NMT3 and NMT5 in vivo based on dose-ranging toxicity and efficacy studies in 52 Syrian hamsters. To assess treatment efficacy in the Syrian hamster model of COVID-19 at the MTD for NMT5, we administered by oral gavage 200 mg kg^{-1} in two equally divided doses separated by 12 hours, with the initial dose timed right after challenge with the virus and the second dose 12 hours later[31,38]. Based on the PK results, at this dose, drug levels in tissue should approach or exceed the EC_{50} found in our in vitro screens to significantly decrease viral infectivity. We found in the Syrian hamster model that this regimen of NMT5, but not NMT3, knocked down live viral titers of SARS-CoV-2 after infection by ~100-fold, as measured by plaque assay (Fig. 2.6c). In the absence of depletion by antibodies[31,39], virus can persist in lung tissue for several days even in the absence of infection and, thus, contribute to plaque assay titers; any significant decrement is encouraging as a potential treatment. More importantly in this model is the histological examination of the lungs for large hemorrhages related to actual SARS-CoV-2 infection, reflecting direct blood vessel damage as also seen in human lungs with fatal COVID-19 (ref. [40]). In this regard, on histological examination, NMT5 virtually eliminated large COVID-19-related hemorrhages in the lungs of infected hamsters compared to vehicle when examined up to 5 days after infection (Fig. 2.6d). This translational model revealed a striking absence of large SARS-CoV-2-induced hemorrhages in the lungs of NMT5-treated hamsters versus controls, with all controls displaying such hemorrhages, whereas no NMT-5 treated animals did so ($n = 12$; $P < 0.01$ by Fisher's exact test)[41]. The lethal effect of large hemorrhagic conversion was completely prevented by NMT5. Although some inflammatory changes were noted in the NMT5-treated lung tissue compared to uninfected controls, it was far less than in the infected/untreated tissue (Fig. 2.6d). In fact, pro-inflammatory cytokine and chemokine activation downstream of infection were abated with NMT5 treatment compared to vehicle control, as demonstrated by immunohistochemical staining of lung sections (Fig. 2.6e). Quantification of these results in an overall pathology score showed

significant protection by NMT5 (Fig. 2.6f). These findings benchmark favorably against our group's published studies on the same model using antibodies directed against spike protein[31,39].

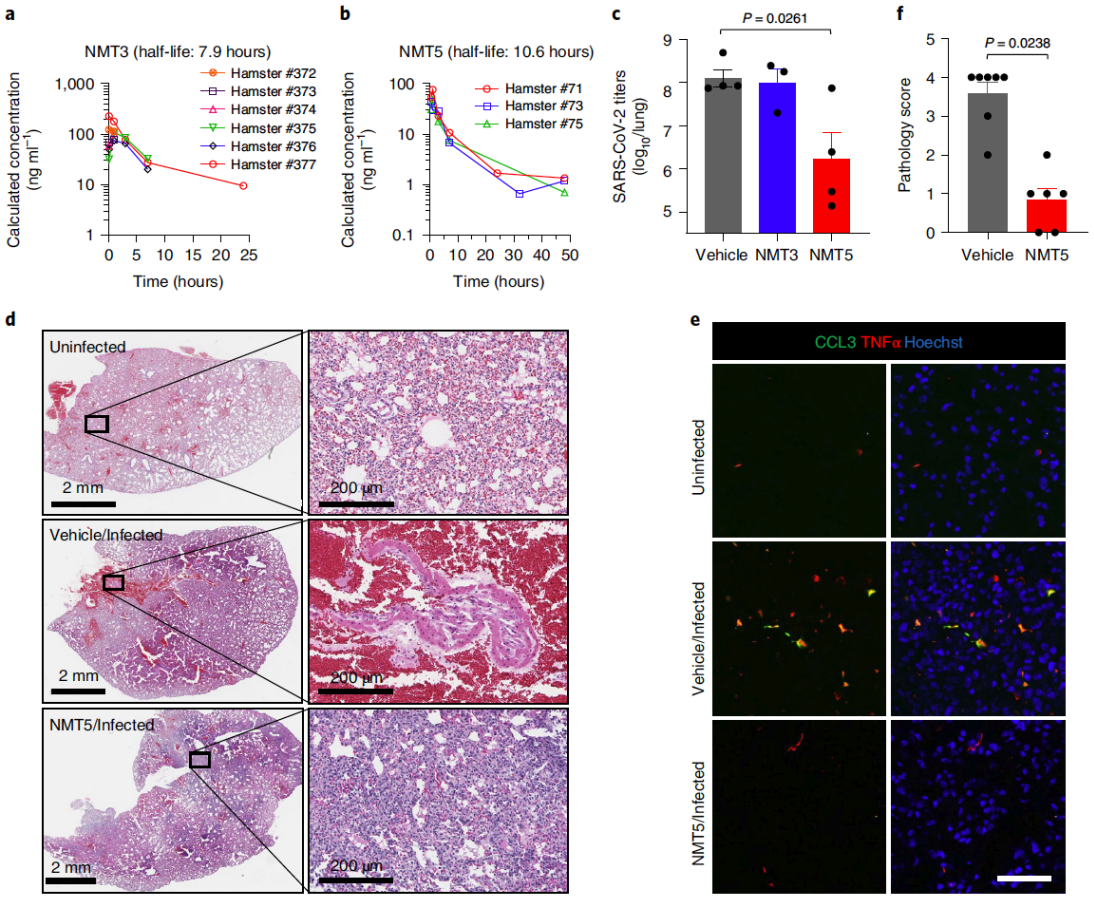


Figure 2.6: NMT5 inhibits SARS-CoV-2 infection in vivo in Syrian hamsters

a,b, PK data in plasma for NMT3 (n = 6) and NMT5 (n = 3) in Syrian hamsters after an oral dose of 10 mg kg^{-1}. c, Live viral load in Syrian hamsters monitored by plaque assay from lung tissue 2 days after infection after treatment with NMT3, NMT5 or vehicle (Control). Data are mean + s.e.m. by two-tailed Student's t-test (n = 8 Syrian hamsters tested). d,e, Syrian hamsters were sacrificed 5 days after infection with SARS-CoV-2 that were either untreated (labeled Vehicle/ Infected) or treated with oral NMT5 (NMT5/Infected) and compared to control (Uninfected). Representative H&E-stained sections showed virtually no areas of large hemorrhage (confluent bright red regions) in NMT5-treated hamster lungs compared to vehicle-treated. Scale bars, 2 mm; high magnification 200 μm (d). Representative immunohistochemistry of lung sections stained with anti-TNFα cytokine antibody and anti-CCL3 (MIP-1α) chemokine antibody from untreated and NMT5-treated hamsters 5 days after SARS-CoV-2 infection; uninfected sections are shown as a control. Merged image with Hoechst stain for DNA shown in right-hand panels. Scale bar, 40 μm (e). f, Lung pathology scores (Methods) for infected hamsters, either untreated or treated with NMT5. Treated hamsters showed significant improvement. Data are mean + s.e.m. by two-tailed Mann–Whitney U-test (n = 9 Syrian hamsters tested).

2.4 **Discussion**

In summary, development of an oral drug to combat acute SARS-Cov-2 infection remains a high priority to treat the COVID-19 pandemic, particularly for the unvaccinated segment of the world population. Our findings provide proof of concept that the cellular receptor of SARS-CoV-2, ACE2, can be S-nitrosylated. This nitrosylation reaction appears to inhibit binding of SARS-CoV-2 spike protein, thus inhibiting viral entry, infectivity and cytotoxicity. As emerging evidence for this mechanism of action, our binding, co-IP and Cys mutation experiments show a significant degree of disruption of binding of spike protein to ACE2 after S-nitrosylation of ACE2 or after mutation of the S-nitrosylation sites on ACE2. In fact, this mechanism was predicted in our molecular dynamics (MD) simulation experiments. We acknowledge, however, that, although this is a rational explanation, it remains a hypothesis for the observed mechanism of viral inhibition by chemical probes that S-nitrosylate ACE2.

Taking advantage of these findings, we developed a novel aminoadamantane nitrate compound, NMT5, as a chemical probe that provides inhibition of SARS-CoV-2 activity by protein S-nitrosylation. NMT5 contains a nitro group, and we provide evidence that this nitro group is targeted to ACE2 by aminoadamantane-mediated viroporin channel blockade of the E protein[2,3,10–12]. The discovery that ACE2 could be S-nitrosylated was quite unexpected, as most authorities had postulated that the beneficial effects of NO on patients with COVID-19 was due to a direct effect on the virus itself. These mechanistic insights provided by our chemical probes should facilitate development of aminoadamantane nitrate drugs for acute antiviral therapy for human COVID-19.

A key concept of this novel approach to ameliorating infection by SARS-CoV-2 is that these nitro-aminoadamantane compounds may prevent the viral spike protein from binding to the ACE2 receptor by S-nitrosylating the receptor in targeted fashion, apparently facilitated by blockade of the vicinal viroporin E protein (Fig. 2.5 and Extended Data Figs. ED2.9 and ED2.10). Hence, drugs like NMT5 should also prevent new variants of the spike protein from binding to ACE2 because ACE2 itself is blocked. In this manner, the development of aminoadamantane nitrates for COVID-19 drug therapy complements other drug, vaccine and antibody therapies, which are dependent on spike protein antigenic sites and, thus, may eventually be susceptible to evasion by further spike protein mutation. Critically, the binding of NMT5 to the viroporin channel may also confer the ability to block spread of SARS-CoV-2 from one host to another. To explain this mechanistically, we posit that NMT5 binds to the E protein viroporin channel on SARS-CoV-2 and then transfers NO to ACE2 on the host cell to prevent infection (Fig. 2.5a,b and Extended Data Fig ED2.9). However, if a patient is already infected and takes an NMT5-like drug, the newly produced viral particles will bind to the aminoadamantane moiety of NMT5 via their E protein viroporin channels; hence, viral infectivity should be limited when a new potential host is exposed to this virus because the new host's ACE2 target protein will be S-nitrosylated by the drug attached to the viral particles as the virus approaches ACE2 on the new host. Compounds like NMT5 thus provide tools for the development of a novel strategy to combat spread of COVID-19 from one host to another.

2.5 **Methods**

Cell lines

HEK293T (System Biosciences, LV900A-1) and HEK293-spike cells (SARS-CoV-2 spike (D614)-expressing 293 cells (293-SARS2-S cells, InvivoGen)) were maintained in DMEM with GlutaMAX (DMEM, high glucose, GlutaMAX Supplement, Life Technologies, 10566016) supplemented with 10% FBS (Sigma-Aldrich, F7524), 100 IU ml^{-1}, and 100 μg ml^{-1} of penicillin–streptomycin (Thermo Fisher Scientific, 10378016) at 37 °C in a 5% CO2 incubator. Transfections were carried out with Lipofectamine 2000 (Life

Technologies, 11668019) using the protocol recommended by the manufacturer. HeLa-ACE2 cells were a gift from David Nemazee (Scripps Research Institute)31. Monkey Vero E6 cells (American Type Culture Collection, CRL-1586) were maintained in complete DMEM (Corning, 15-013-CV) containing 10% FBS, 1× penicillin–streptomycin (Corning, 20-002-CL) and 2 mM l-glutamine (Corning, 25-005-CL) at 37 °C in a 5% CO_2 incubator.

Plasmids

hACE2 was a gift from Hyeryun Choe (Addgene plasmid, 1786; http:// n2t.net/addgene:1786; RRID: Addgene_1786)[42]. The C262A, C498A and C261/498A mutant ACE2 constructs were generated using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies, 210514), according to the manufacturer's protocol. pGBW-m4252984 (SARS-CoV-2 E) was a gift from Ginkgo Bioworks (Addgene plasmid, 153898; http://n2t.net/addgene:153898; RRID: Addgene_153898). MLV-gag/pol, MLV-CMV-luciferase, SARS-CoV-2 and VSV-G plasmids were a gift from David Nemazee (Scripps Research Institute)[31].

Aminoadamantane and aminoadamantane nitrate drugs

Aminoadamantane nitrate compounds (blindly coded NMT2, NMT3, NMT5-NMT9 and NMT5-met (metabolite, sans nitro group)) were synthesized by and obtained from EuMentis Therapeutics and have been described previously[6–9,29]. Additionally, note that these compounds were prepared as hydrochloride salts to increase stability and aqueous solubility. The aminoadamantane compounds, memantine and amantadine (blindly coded NMT1 and NMT4), were also obtained from EuMentis Therapeutics. Detailed protocols for chemical synthesis are provided in the Supplementary Note in the Supplementary Information. All compounds were sent to the Scripps Research Institute for testing in a masked fashion, and compound identities were not revealed until after experiments were completed and analyzed.

Biotin switch assays and immunoblotting

For analysis of S-nitrosylated proteins, we performed the biotin switch assay as previously described[43]. In brief, cells or tissue samples were lysed with HENTS buffer (100 mM HEPES, pH 7.4, 1 mM EDTA, 0.1 mM neocuproine, 0.1% SDS and 1% Triton X-100) containing 10 mM methyl methanethiosulfonate (MMTS). SDS solution (25% w/v) was added to lysed samples to a final concentration of 1% and incubated for 20 minutes at 45 °C with frequent vortexing to facilitate blockade of free thiol groups. After removing excess MMTS by acetone precipitation, S-nitrosothiols were reduced to thiols with 20 mM ascorbate. Newly formed thiols were then linked with the sulfhydryl-specific biotinylating reagent N-[6-(biotinamido)-hexyl]-l′-(2′-pyridyldithio) propionamide (biotin-HPDP; Dojindo, SB17-10). Unreacted biotin-HPDP was removed by acetone precipitation, and the pellet was resuspended in HENS buffer (100 mM HEPES, pH 7.4, 1 mM EDTA, 0.1 mM neocuproine and 1% SDS), neutralized and centrifuged to clear any undissolved debris. Five percent of the supernatant was used as the input for the loading control. Biotinylated proteins were pulled down with High Capacity NeutrAvidin Agarose beads (Thermo Fisher Scientific, 29202) and analyzed by immunoblotting for S-nitrosylated ACE2, TMPRSS2 or spike protein. Protein samples were subjected to Bolt Bis-Tris Plus (Thermo Fisher Scientific, NW04122BOX) gel electrophoresis and transferred to PVDF membranes (Millipore, IPFL00010). Membranes were blocked with Odyssey blocking buffer (Li-Cor, 927-40000) for 30 minutes at room temperature and then probed with primary antibodies against ACE2 (1:3,000, Abcam, ab15348; 1:3,000, Proteintech, 21115-1-AP), TMPRSS2 (1:1,000, Santa Cruz Biotechnology, sc-515727) or SARS-CoV-2 spike protein (1:2,000, Abcam, ab275759). After incubation with secondary antibodies (IR-dye 680LT-conjugated goat anti-mouse (1:20,000; Li-Cor, 926-68020) or IR-dye 800CW-conjugated goat anti-rabbit (1:15,000; Li-Cor, 926-32211)), membranes were scanned with an Odyssey infrared imaging system (Li-Cor). Image Studio (Li-Cor) software was used for densitometric analysis of immunoblots.

Immunocytochemistry for SARS-CoV-2 spike protein

Purified recombinant SARS-CoV-2 spike (S1 + S2) protein (10 μg ml−1, Sino Biological, 40589-V08B1) exposed cells were fixed with 4% paraformaldehyde (PFA) for 15 minutes at room temperature, washed three times with PBS and blocked (3% BSA and 0.3% Triton X-100 in PBS) for 30 minutes at room temperature. Cells were incubated with anti-SARS-CoV-2 spike protein antibodies (1:200, Sino Biological, 40150-R007)

overnight at 4 °C, followed by incubation with Alexa Fluor 488-conjugated secondary antibody (Thermo Fisher Scientific, A21206). Cells were counterstained with 1 μg ml−1 of Hoechst dye 33342 (Invitrogen). Cell images were acquired with a Nikon A1 confocal microscope using a ×20/0.75 air objective (1-μm Z-stack). Maximum intensity projection of images was generated with NIS-Elements AR, and fluorescence intensity was analyzed with ImageJ software (https://imagej.nih.gov/ ij/download.html) as previously described[44].

Expression and purification of human ACE2 protein

The N-terminal peptidase domain of human ACE2 (residues 19–615, GenBank: BAB40370.1) was cloned into phCMV3 vector and fused with C-terminal His-tag. The plasmid was transiently transfected into Expi293F cells using ExpiFectamin 293 Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. The supernatant was collected at day 7 after transfection. The His-tagged ACE2 protein was then purified by Ni-NTA (Qiagen) affinity purification, followed by size-exclusion chromatography. The ACE2 preparation then was buffer-exchanged to 1× PBS for the S-nitrosylation assay.

Co-IP experiments

Cultured cells were harvested and lysed in 1% Triton X-100 in PBS. Equivalent protein quantities were immunoprecipitated with anti-ACE2 antibody (Abcam, ab15348)-conjugated magnetic beads (Dynabeads Protein A; Thermo Fisher Scientific, 10002D) for 90 minutes at room temperature. Immunoprecipitants were eluted and subjected to immunoblotting with anti-ACE2 antibody (1:1,000, Cell Signaling Technology, 15983) and anti-SARS-CoV-2 spike protein antibody (1:2,000, Abcam, ab275759).

Mass spectrometry analysis of S-nitrosylated ACE2 protein

Biotin switch was performed as described above. Biotinylated proteins were then precipitated with iced acetone, pelleted and solubilized in HENS buffer (100 mM HEPES, pH 7.4, 1 mM EDTA, 0.1 mM neocuproine and 1% SDS). The samples were desalted using a ZebaSpin desalting column (Thermo Fisher Scientific) pre-equilibrated with PBS, and biotinylated ACE2 protein was immunoprecipitated as described above. Immunoprecipitated ACE2 was eluted in 1% SDS solution and precipitated using methanol-chloroform. Dried pellets were dissolved in 8 M urea/100 mM triethylammonium bicarbonate

(TEAB, pH 8.5). Samples were diluted to 2 M urea/100 mM TEAB, and proteins were trypsin digested overnight at 37 °C. The digested ACE2 peptides were enriched a second time by biotin-avidin affinity to enrich biotinylated peptides representing the initial SNO sites. After avidin enrichment, peptides were eluted by reduction using tris(2-carboxyethyl)phosphine (TCEP).

Samples were analyzed on an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific). Samples were injected directly onto a 25-cm, 100-μm ID column packed with ethylene bridged hybrid (BEH) 1.7-μm C18 resin (Waters). Samples were separated at a flow rate of 300 nl min−1 on an nLC 1200 (Thermo Fisher Scientific) using a gradient of solution A (0.1% formic acid in water and 5% acetonitrile) and solution B (80% acetonitrile and 0.1% formic acid). Specifically, a gradient of 1–25% B over 100 minutes, an increase to 40% B over 20 minutes, an increase to 100% B over another 10 minutes and held at 90% B for 10 minutes was used for a 140-minute total run time. Peptides were eluted directly from the tip of the column and nanosprayed directly into the mass spectrometer by application of 2.5 kV voltage at the back of the column. The Orbitrap Eclipse mass spectrometer was operated in data-dependent mode. Full MS1 scans were collected in the Orbitrap at 120,000 resolution. The cycle time was set to 3 seconds, and, within these 3 seconds, the most abundant ions per scan were selected for high-energy collisional dissociation (HCD) with detection in the Orbitrap. Monoisotopic precursor selection was enabled, and dynamic exclusion was used with exclusion duration of 5 seconds.

Protein and peptide identification were done with Integrated Proteomics Pipeline–IP2 (Integrated Proteomics Applications). Tandem mass spectra were extracted from raw files using RawConverter[45] and searched with ProLuCID46 against the UniProt human database. The search space included all fully-tryptic and half-tryptic peptide candidates. Data were searched with 50-p.p.m. precursor ion tolerance and 600-p.p.m. fragment ion tolerance. Identified proteins were filtered to 10-p.p.m. precursor ion tolerance using DTASelect47 using a target-decoy database search strategy to control the false discovery rate to 1% at the protein level48.

MD simulations

The fully glycosylated, Cys^{261}/Cys^{498}-S-nitrosylated model of human ACE2 dimer bound to two SARS-CoV-2 spike RBDs (with a 1:1 stoichiometry) is based on the cryo-EM structure of the ACE2/RBD/B0AT1 complex (Protein Data Bank (PDB) ID: 6M17)[23]. B0AT1 dimer chaperone coordinates were manually removed, and N-glycans were added on both ACE2 and RBD in the same fashion as in Barros et al.[22]. The sidechain of Cys^{261} and Cys^{498} was S-nitrosylated in both ACE2 protomers using Schrödinger Maestro (Schrödinger Release 2020-4). ParamChem web interface was used to generate CHARMM36 suitable parameters for the S-N = O moiety[49–51]. The glycosylated and S-nitrosylated ACE2/RBD construct was inserted into a lipid bilayer patch of 350 Å × 350 Å with a composition similar to that of mammalian cell membranes (56% POPC, 20% CHL, 11% POPI, 9% POPE and 4% PSM)[52,53]. Finally, the resulting system was embedded into an orthorhombic box of explicit waters and Na^+/Cl^- ions at a concentration of 150 mM. MD simulations were performed on the Frontera supercomputer at the Texas Advanced Supercomputing Center using NAMD 2.14 (ref. [54]) and CHARMM36m all-atom additive force fields[55,56]. Excluding initial minimization and equilibration, a total of 310 ns were collected for analysis. We note that, except for Cys^{261} and Cys^{498} S-nitrosylation, the model described here is the same as the ACE2/ RBD complex presented in Barros et al.[22]. Therefore, we refer to that work for a complete description of system setup procedures and simulation protocol.

Analysis of COM distance was performed with the compute_ center_of_mass function within MDTraj[57]. COM for each ACE2 protomer was calculated taking into account the amino acid backbones of residues 18–742. The distance between COMs was evaluated at each frame along a 310-ns trajectory for both the WT ACE2/RBD complex by Barros et al.[22] and the model of the SNO-ACE2/RBD complex presented here. As a reference, the distance between COMs from the cryo-EM structure (PDB ID: 6M17) was also calculated. The simulations were visually inspected using VMD[58].

SARS-CoV-2 virus generation

Monkey Vero E6 cells were plated in a T225 flask with complete DMEM containing 10% FBS, 1× penicillin–streptomycin, 2 mM l-glutamine and incubated for overnight at 37 °C in a humified atmosphere of

5% CO2. The medium in the flask was removed, and 2 ml of complete DMEM containing the WA1 strain of SARS-CoV-2 (USA-WA1/2020 (BEI Resources, NR-52281)) was added to the flask at a multiplicity of infection (MOI) of 0.5. After incubation for 30 minutes at 34 °C in a 5% CO2 incubator, 30 ml of complete DMEM was added to the flask. The flask was then placed in a 34 °C incubator with 5% CO2 for 5 days. On day 5 after infection, the supernatant was harvested and centrifuged at 1,000g for 5 minutes. The supernatant was filtered through a 0.22-μM filter and stored at −80 °C.

SARS-CoV-2/HeLa-ACE2 high-content imaging assay for infection

Control compounds solvated in DMSO were transferred into 384-well μClear-bottom plates (Greiner, 781090-2B) using the ADS Labcyte Echo liquid handler. Aminoadamantane and aminoadamantane nitrate compounds to be screened were solvated in saline solution on ice immediately before use and transferred into the assay plates in 5 μl of DMEM with 2% FBS (assay medium). HeLa-ACE2 cells were added to plates in assay medium at a density of 1.0×10^3 cells per well to a 13-μl volume. Plated cells were transported to the Biosafety Level 3 facility at Scripps Research Institute, and, within 1 hour, 13 μl of SARS-CoV-2 diluted in assay media was added at MOI = 0.65. Cells were incubated for 24 hours at 34 °C in a 5% CO_2 incubator and then fixed with 8% formaldehyde for 1 hour. Human polyclonal sera diluted at 1:500 in Perm/Wash buffer (BD Biosciences, 554723) were added to the plate and incubated at room temperature for 2 hours. Then, 6 μg ml^{-1} of goat anti-human H+L conjugated Alexa Fluor 488 (Thermo Fisher Scientific, A11013) together with 8 μM of Antifade DAPI (Thermo Fisher Scientific, D1306) in SuperBlock T20 (PBS) buffer (Thermo Fisher Scientific, 37515) were added to the plate and incubated at room temperature for 1 hour in the dark. Four fields were imaged per well using the ImageXpress Micro Confocal High-Content Imaging System (Molecular Devices) with a ×10 objective. Images were analyzed using the Multi-Wavelength Cell Scoring Application Module (MetaXpress) where DAPI staining was used to identify host cell nuclei (the total number of cells in the images) and the SARS-CoV-2 immunofluorescence signal for identification of infected cells.

Uninfected host cell cytotoxicity counterscreen

Compounds were prepared and plated in 384-well plates as for the infection assay. HeLa-ACE2 cells were seeded in the assay-ready plates at 1.6×10^3 cells per well in assay medium, and plates were incubated for 24 hours at 37 °C with 5% CO_2. To assess cell viability, the Image-iT DEAD Green reagent (Thermo Fisher Scientific, I10291) was used according to the manufacturer's instructions. Cells were fixed with 4% PFA and counterstained with DAPI. Fixed cells were imaged using the ImageXpress Micro Confocal High-Content Imaging System (Molecular Devices) with a ×10 objective, and total live cells per well quantified in the acquired images using the Live Dead Application Module (MetaXpress).

Data analysis of the compound screening results

The in vitro infection assay and the host cell cytotoxicity counterscreen data were uploaded to Genedata Screener, version 16.0. Data were normalized to neutral (DMSO) minus inhibitor controls (2.5 μM remdesivir for antiviral effect and 10 μM puromycin for infected host cell toxicity). For the uninfected host cell cytotoxicity counterscreen, 10 μM puromycin (Sigma-Aldrich) was used as a positive control. For dose–response experiments, compounds were tested in technical triplicates, and dose curves were fitted with the four-parameter Hill equation. Replicate data were analyzed using median condensing. The full dataset is provided in Supplementary Dataset 1.

Pseudoviral entry assay

To measure SARS-CoV-2 viral infectivity, we performed pseudo viral entry assay as previously described[31]. In brief, HEK293 Tcells were transiently co-transfected with MLV-gag/pol, MLV-CMV-luciferase plasmid and SARS-CoV-2 spike (D614) or VSV-G plasmid. Two days later, supernatants containing pseudotyped virus particles were collected. To assay for pseudoviral entry, HeLa-ACE2 cells were seeded in 96-well plates at 10,000 cells per well (PerkinElmer, 6005680). One day later, cells were incubated with diluted pseudovirus. After 48 hours to allow for viral transduction, cells were lysed and assayed for luciferase activity by Steady-Glo (Promega, E2510) according to the manufacturer's instructions. Luminescence was quantified using a Luminoskan Ascent plate leader (Thermo Fisher Scientific). To assess the ability of NMT5 to reduce viral infection in human-ACE2 cells with SARS-CoV-2 variants, lenti-SARS-CoV-2 N501Y spike, SARS-CoV-2

K417N, E484K and N501Y spike and VSV-G pseudoviral particles were obtained from the Rhode Island Hospital lentivirus construct core[59], or lenti-SARS-CoV-2 Delta and Omicron variant pseudoviral particles were obtained from PBS Bioscience (78215 and 78348, respectively). SARS-CoV-2 variant preparations were diluted in assay media and added at 4.0×10^6 TU/1×10^4 or 1.5×10^2 TU/1×10^4 cells (for Delta variant), or 9×10^2 TU/1×10^4 cells (for Omicron variant), in the presence or absence of 20 μM NMT5. Plates were centrifuged at 600g for 1 hour at room temperature, and then cells were incubated at 37 °C in a 5% CO_2 incubator. After 48 hours to allow for viral transduction, cells were lysed and assayed for luciferase activity by Steady-Glo. Luminescence was quantified using a Luminoskan Ascent plate reader.

PK testing and analysis

Golden Syrian hamsters (110–150 g, n = 3 for each compound) were dosed by oral gavage at 10 mg kg^{-1}. Blood samples were obtained at 30 minutes and at 1, 3, 7, 24, 32 and 48 hours. The blood samples were collected in BD Vacutainers (BD, 366664) containing sodium heparin and centrifuged. The processed plasma samples were stored at −20 °C until high-performance liquid chromatography–tandem mass spectrometry (HPLC–MS/MS) analysis. Animals were then sacrificed, and lung and kidney tissues were harvested for biotin switch analysis for SNO-ACE2. To quantify the test compound in the collected plasma samples, a plasma calibration curve was generated by spiking aliquots of drug-free plasma with the test compound at the specified concentration levels. Spiked and collected plasma samples were treated with an aliquot of acetonitrile containing a known concentration of an internal standard (IS). The extraction solvent was analyzed on an Agilent 1100 LC mated to an AB Sciex 4000 Q TRAP MS. Separation of the analytes was achieved by using a Phenomenex Kinetex EVO C18 50 ×2.1-mm column (Phenomenex, 00B-4633-AN) and a mobile phase consisting of (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. The LC gradient consisted of a 0.5 ml min^{-1} flow rate starting at 100% of (A). The LC gradient then ramped to 90% (B) over 0.1 minutes and held for 2.5 minutes. The gradient reverted to 100% (A) over 0.1 minute and allowed for 2 minutes of re-equilibration time. Ionization spray voltage was set to 4,000, with the source temperature set at 400 °C. Analytes were monitored by multiple reaction monitoring (MRM) in positive ionization mode. Peak areas were recorded, and the concentrations of the test compound in the unknown

plasma samples were quantified by the calibration curve using Sciex Analyst software (PE Sciex). Phoenix WinNonLin 8.1 software (Certara) was used in NCA mode to determine the PK parameters from the LC–MS/MS measured values. NMT5 PK analysis was performed by 'serial sampling' where we followed the PK of the drug in each individual animal (animals 71, 73 and 75) and then averaged those PK values. We took this approach because all the animals had values for the full time course available, as detailed in Supplementary Dataset 2. For NMT3 PK, the 'sparse sampling' procedure was used to account for plasma concentrations from animals that did not have a full time course profile available (animals 372–374), in part due to the short half-life and rapid metabolism to the hydroxylated form with release of the nitrate group, as detailed in Supplementary Dataset 3. The differences in methodology results in 'serial sampling' averaging the PK values and 'sparse sampling' averaging the concentrations were used to determine the PK values. Sparse sampling was also necessary because tissue samples were obtained at two of the timepoints (1 hour and 24 hours), so the study parameters had to be modified. Additionally, both parent drug (NMT3) and metabolite were detected, making the samples sparser. Summary PK data are shown in Supplementary Table 1, and the full datasets are provided in Supplementary Dataset 2 (for NMT5) and Supplementary Dataset 3 (for NMT3).

Syrian hamster COVID-19 model

Eight-week-old golden Syrian hamsters (110–150 g, Charles River Laboratories) were challenged with a dose of 1×10^5 or 1×10^6 plaque-forming units (PFU) of SARS-CoV-2 (USA-WA1/2020) by intranasal administration in a volume of 100 µl of DMEM as previously described[31]. For the experiments shown, the hamsters were subsequently administered via oral gavage either vehicle or the aminoadamantane nitrate candidate drugs NMT3 or NMT5 in a volume of 500 µl, with the initial dose timed for delivery right after challenge with SARS-CoV-2 and a second dose administered 12 hours later. At 2 days and 5 days after infection, lung tissue was collected for assessing viral titers and histology as described[31]. All animal experiments were approved by the Scripps Research Institute Animal Care and Use Committee.

Plaque assay for SARS-CoV-2 viral titers

SARS-CoV2 titers were measured 2 days and 5 days after infection by homogenizing hamster lungs in DMEM 2% FBS using 100-μm cell strainers (Myriad, 2825-8367). Homogenized lungs were titrated 1:10 over six steps and layered over Vero cells. After 1 hour of incubation at 37 °C, a 1% methylcellulose in DMEM overlay was added, and the cells were incubated for 3 days at 37 °C. Cells were then fixed with 4% PFA, and plaques were counted by crystal violet staining and expressed as PFU. PFU after treatment with NMT3 was statistically equal to that of vehicle control.

Lung histology and immunohistochemistry

Lung tissue from golden Syrian hamsters was stored in zinc formalin for ~72 hours. The tissue was processed for paraffin embedding, and sections cut at a thickness of 5 μm. The tissue was then stained with hematoxylin and eosin (H&E). The slides were scanned at ×20 using an Aperio AT2 whole slide scanner. For immunohistochemistry, lung sections were fixed with 4% PFA for 15 minutes and washed three times with PBS. Sections were blocked with 3% BSA and 0.3% Triton X-100 in PBS for 30 minutes. Sections were incubated with primary antibody in blocking solution overnight at 4 °C and then washed with PBS. The appropriate Alexa Fluor 488-conjugated (Thermo Fisher Scientific, A21206) and Alexa Fluor 555-conjugated (Thermo Fisher Scientific, A31570) secondary antibodies were used at 1:500, plus Hoechst 33342, trihydrochloride and trihydrate dye (1:1,000, Thermo Fisher Scientific, H3570) to visualize nuclei, for 1 hour at room temperature. Primary antibodies and dilutions were as follows: mouse anti-TNFα (5 μg ml^{-1}, Abcam, ab1793) and rabbit anti-macrophage inflammatory protein 1α (MIP-1α)/CCL3 + CCL3L1 (1:250, Abcam, ab259372). Slides were imaged using a C2 confocal microscope (Nikon) in a masked fashion. For quantification of lung injury, pathology scores were applied based on the percentage of interstitial hemorrhage and cytokines present using the following criteria: (0) normal; (1) less than 10% of tissue affected; (2) 10–25% affected; (3) 26–50% affected; and (4) more than 50% affected, similar to the scoring system used previously for COVID-19 models[60].

Patch-clamp analysis of SARS-CoV-2 E protein viroporin ion channel activity

Using the patch-clamp technique, we recorded whole-cell currents from untransfected HEK293T cells and cells transfected with SARS-CoV-2 E protein. Cells were recorded at room temperature in HEPES-supplemented HBSS with the following composition (in mM): NaCl 138, KCl 5.3, KH_2PO_4 0.4, Na_2HPO_4 0.3, $NaHCO_3$ 4.2, $MgCl_2$ 0.5, $MgSO_4$ 0.4, $CaCl_2$ 2.0, d-glucose 5.6 and HEPES 10; pH 7.4, osmolality 300 mOsm. Patch pipettes were filled with an intracellular pipette solution composed of (in mM): 113 K-gluconate, 6 KCl, 4.6 $MgCl_2$, 1.1 $CaCl_2$, 10 HEPES, 10 EGTA, 4 Na_2-ATP, 0.4 Na_2-GTP, pH adjusted to 7.3 with KOH; osmolality 291 mOsm. Whole-cell currents were recorded using Clampex software (Molecular Devices) on a computer-controlled patch-clamp amplifier (Multiclamp 700B, Molecular Devices) and analyzed with Clampfit software (Molecular Devices). Traces were filtered at 2 kHz and sampled at 20 kHz. Drugs were dissolved in water and stored as 100 mM stocks at −20 °C. During experiments, drugs were dissolved in extracellular solution and tested at concentrations of 5 μM and 10 μM. HEK293T cells were plated on 12-mm-diameter glass coverslips coated with a mixture of rat tail type I collagen and poly-d-lysine. For transient expression in HEK293T cells, we used a transfection reagent (FuGENE HD, Promega) to co-transfect plasmids containing cDNAs for SARS-CoV-2 E protein (pGBW-m4133502, Addgene) and green fluorescent protein (GFP) at a ratio of 1:0.1 (0.5:0.05 μg per well, respectively).

SNO-ACE2 targeting assay via E protein viroporin channel

SARS-CoV-2 E plasmid was transiently transfected into HEK293-spike cells with Lipofectamine 2000 (controls received Lipofectamine 2000 vehicle alone). One day later, cells were harvested by tapping the plate in pre-warmed PBS. These cells were then added onto pre-plated HeLa-ACE2 cells in the presence or absence of NMT5. After 30 minutes, all cells were collected and subjected to biotin switch assay and immunoblotting with anti-ACE2 antibody to assess the levels of SNO-ACE2 and total input ACE2.

Quantification and statistical analysis

A power analysis of our prior data was used to determine the number of replicates needed for statistical purposes. The number of replicates or experiments is indicated in the individual figure legends. Data are expressed as mean ± s.e.m. In general, for multiple comparisons on data with a normal distribution,

differences between experimental groups were evaluated using an ANOVA followed by a post hoc Tukey's or Fisher's least significant difference (LSD) test. For comparison between two groups, a Student's t-test (for normally distributed data) or a non-parametric Mann–Whitney U-test (for data not known to be normally distributed) was employed. Additionally, for the statistical assessment of lung pathology in the Syrian hamster model of COVID-19, a Fisher's exact test with a 2 × 2 contingency table was used. P values less than 0.05 were considered significant. Statistical analyses were performed using GraphPad Prism software.

2.6 **Extended Data**

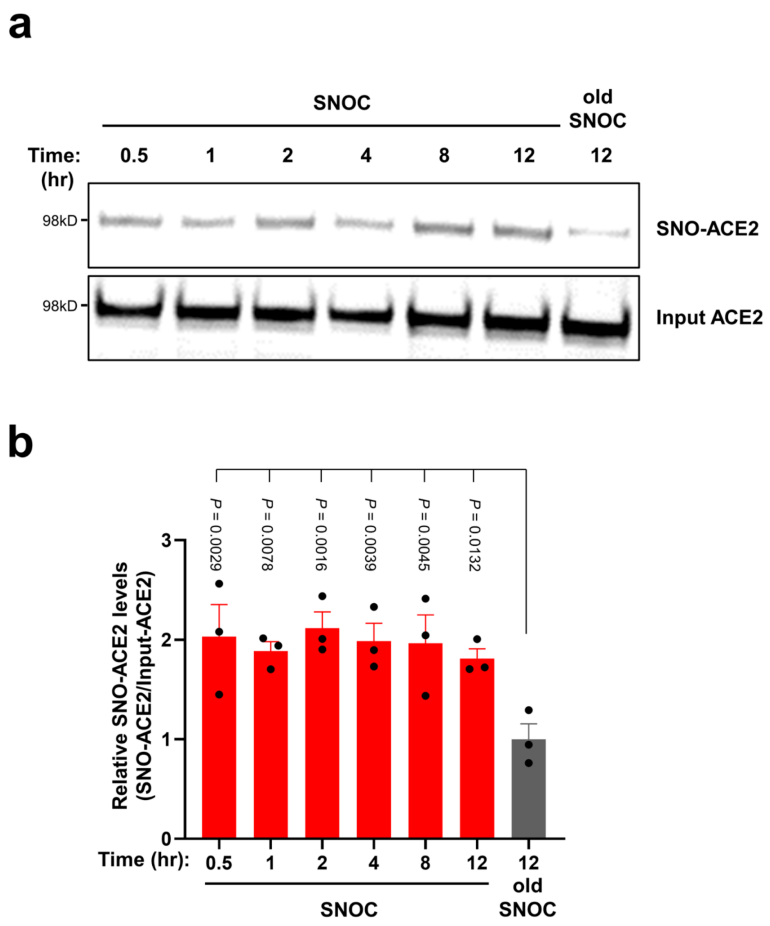


Figure ED2.1: S-Nitrosylation of ACE2 persists for at least 12 hours

a, HeLa-ACE2 cells were exposed to 100 μM SNOC; 30 min later the cells were incubated in serum free medium for the time periods indicated. Cell lysates were then subjected to biotin-switch assay to assess protein S-nitrosylation, which was detected by immunoblotting with anti-ACE2 antibody. b, Ratio of SNO-ACE2/ input ACE2 protein. Data are mean + s.e.m. by one-way ANOVA with Fisher's LSD multiple comparisons. n = 3 biological replicates.

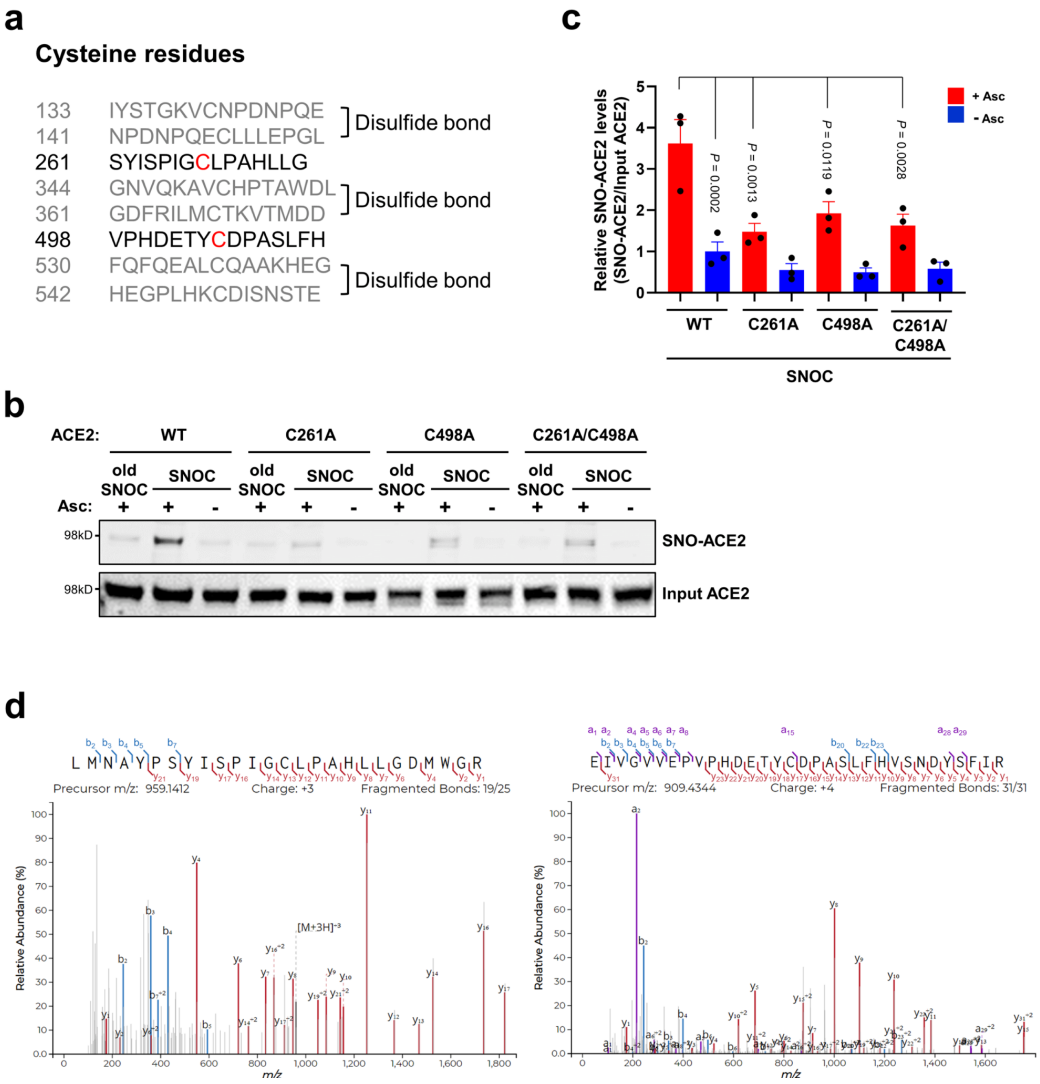


Figure ED2.2: Identification of cystene residues in ACE2 that are S-nitrosylated

a, List of human ACE2 peptides (± 7 amino acid residues flanking a cysteine residue); gray: peptides involved in disulfide bond formation; black: peptides containing a free cysteine thiol (red) that could potentially be S-nitrosylated. b, HEK293T cells were transiently transfected with plasmids containing human WT ACE2 or cysteine mutant ACE2 (C261A, C498A, or C261/498A). One day after transfection, cells were exposed to 100 μM SNOC. After 20 minutes, cells were subjected to biotin-switch assay. Absence of ascorbate (Asc -) served as a negative control. c, Ratio of SNO-ACE2/input ACE2. Data are mean + s.e.m. by two-way ANOVA with Tukey's multiple comparisons. n = 3 biological replicates. d, HEK293T cells expressing ACE2 were exposed to SNOC and subjected to biotin switch. The peptides were eluted by reduction for subsequent detection by LC-MS/MS. Representative MS/MS spectra of detected peptides from human ACE2 containing Cys^{261} (left) or Cys^{498} (right) are shown.

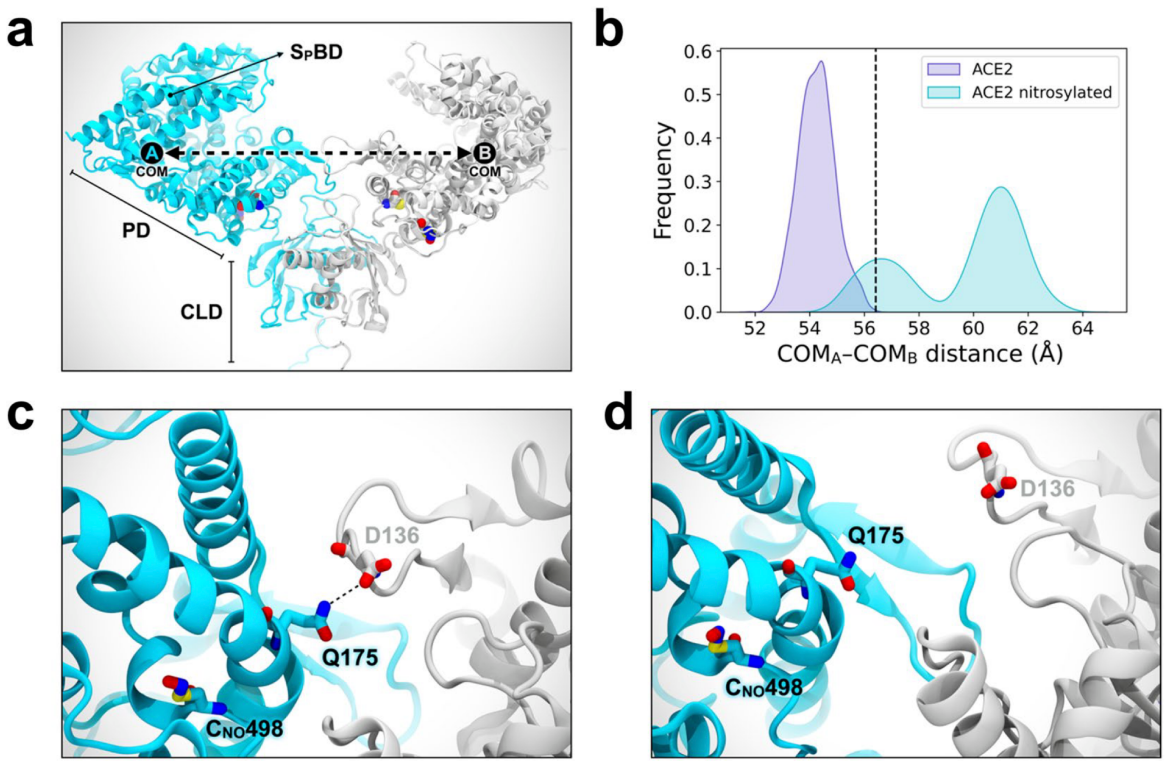


Figure ED2.3: Molecular dynamics simulation of S-nitrosylation of ACE2

a, Molecular representation of the S-nitrosylated-ACE2/RBD model upon transient detachment at the level of the peptidase domain dimeric interface. SNO-Cys261 and SNO-Cys498 are shown with Van der Waals spheres. The black dots indicate qualitative placement of centers of mass (COM) for each ACE2 protomer, and the dashed arrow represents the distance between COMs. Spike's RBDs and N-glycans, which were included in the simulation, are hidden for image clarity. SpBD, Spike binding domain; CLD, collectrin-like domain; PD, peptidase domain. b, Distribution of the distance between COMs from molecular dynamics simulations of WT ACE2/RBD (purple) vs. nitrosylated-ACE2/RBD (cyan). Dashed black line at approximately 56.5 Å indicates the reference distance between COMs calculated from the cryo-EM structure (PDB: 6M17). S-Nitrosylated-ACE2/ RBD shows an overall larger distance between COMs with a bimodal distribution. c, Close-up image illustrating Q175A to D136B interaction present in starting conformations of the S-nitrosylated-ACE2 system. d, Close-up image illustrating the disruption of the interaction between Q175A and D136B occurring along the dynamics of the S-nitrosylated-ACE2 system.

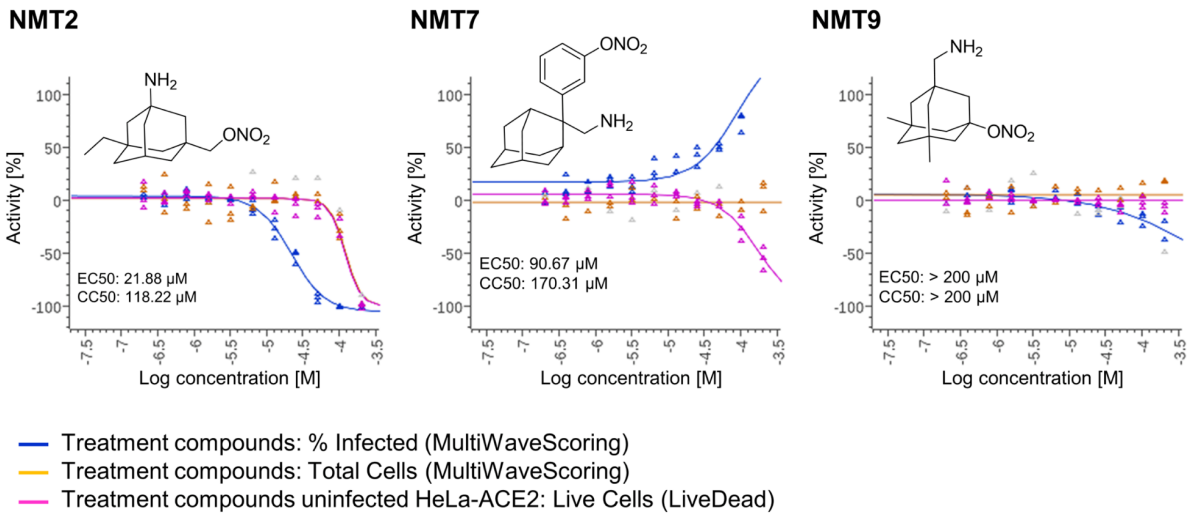


Figure ED2.4: Dose-response of drugs screened against SARS-CoV-2 infection

Dose-response curves showing the EC50 of each compound against SARS-CoV-2 (% infected cells, blue), total cell counts (orange) in the infection experiment and the CC50 for uninfected host cell toxicity (magenta), as assessed in HeLa-ACE2 cells. See also Supplementary Dataset 1 for full dataset. Continued from Fig. 2.

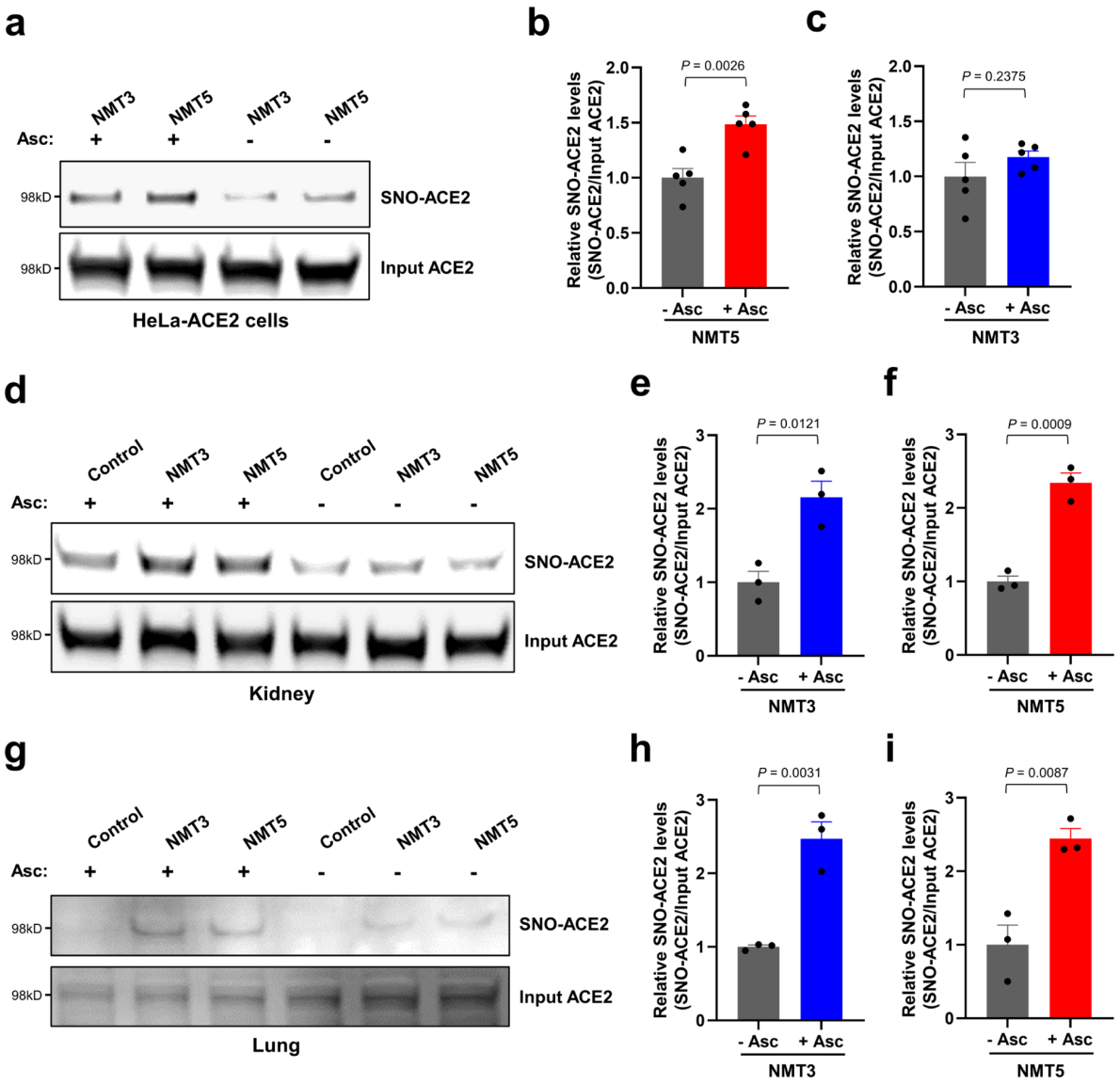


Figure ED2.5: NMT5 S-nitrosylates ACE2 in vitro and in vivo

a, Detection of SNO-ACE2 in vitro. HeLa-ACE2 cells were exposed to 10 μM NMT3 or 5 μM NMT5. After 1 h, cells were subjected to the biotin-switch assay in the presence or absence of ascorbate. SNO-ACE2 and input ACE2 were detected by immunoblotting with anti-ACE2 antibody. b, c, Ratio of SNO-ACE2/input ACE2. Data are mean + s.e.m. by two-tailed Student's t test. n = 5 biological replicates. d–i, Detection of SNO-ACE2 in vivo. Syrian hamsters received 10 mg/kg of NMT3 or of NMT5 by oral gavage and were sacrificed 48 h later. Kidney and lung tissues were subjected to biotin-switch assay in the presence or absence of ascorbate. Note that in some samples, low levels of SNO-ACE2 were observed in control tissue, suggesting endogenous S-nitrosylation of ACE2 may occur at low levels. Graphs show ratio of SNO-ACE2/input ACE2. Data are mean + s.e.m. by two-tailed Student's t test. n = 3 Syrian hamsters for each condition.

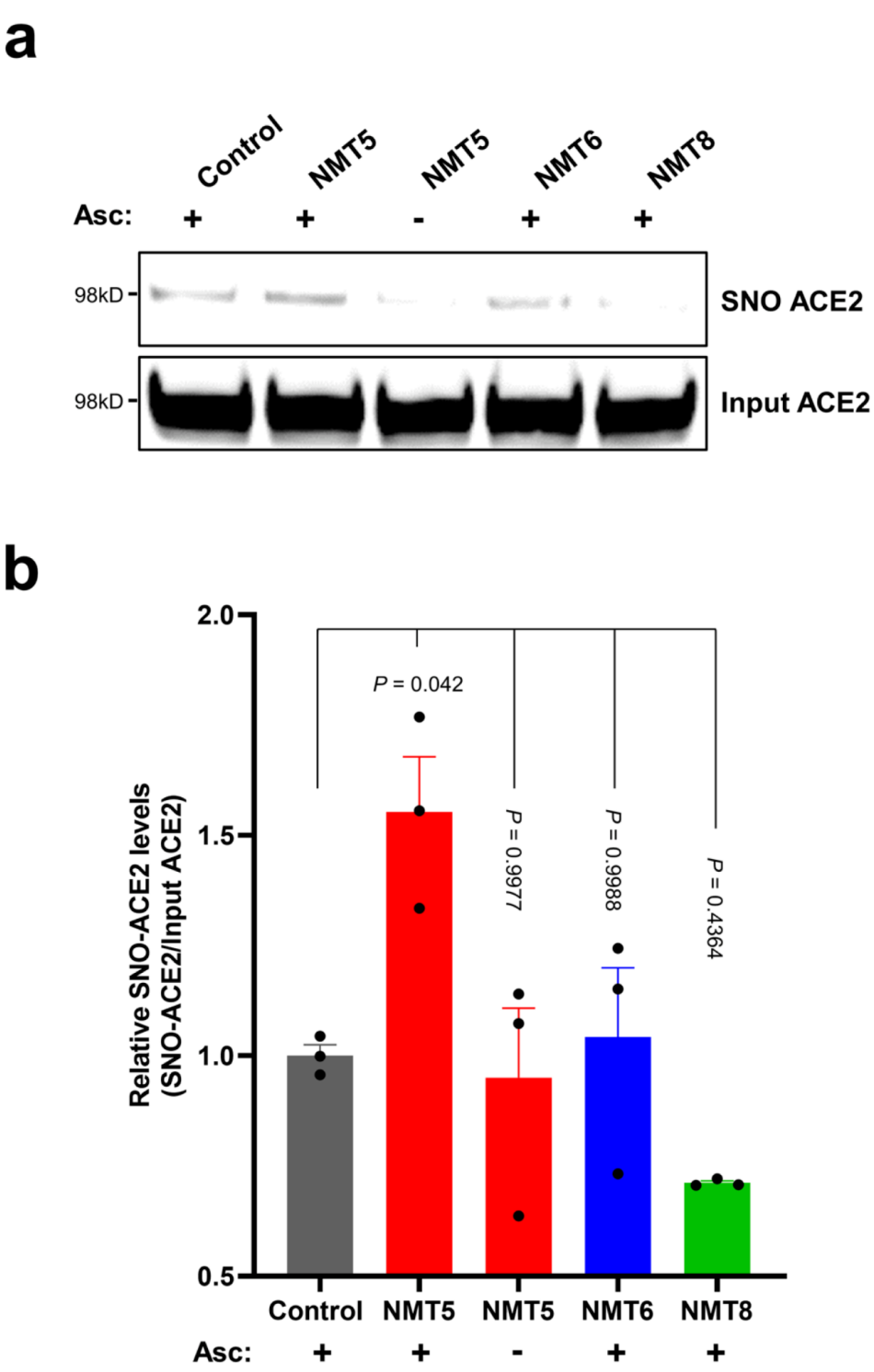


Figure ED2.6: Protein S-nitrosylation of ACE2 by NMT5

a, HeLa-ACE2 cells were treated with 10 μM NMT5, NMT6, or NMT8. After 1 h, cell lysates were subjected to biotin-switch assay for protein S-nitrosylation, detected by immunoblotting with anti-ACE2 antibody. The ascorbate minus (Asc-) sample served as a negative control. b, Ratio of SNO-ACE2/input ACE2 protein. Data are mean + s.e.m. by one-way ANOVA with Tukey's multiple comparisons. n = 3

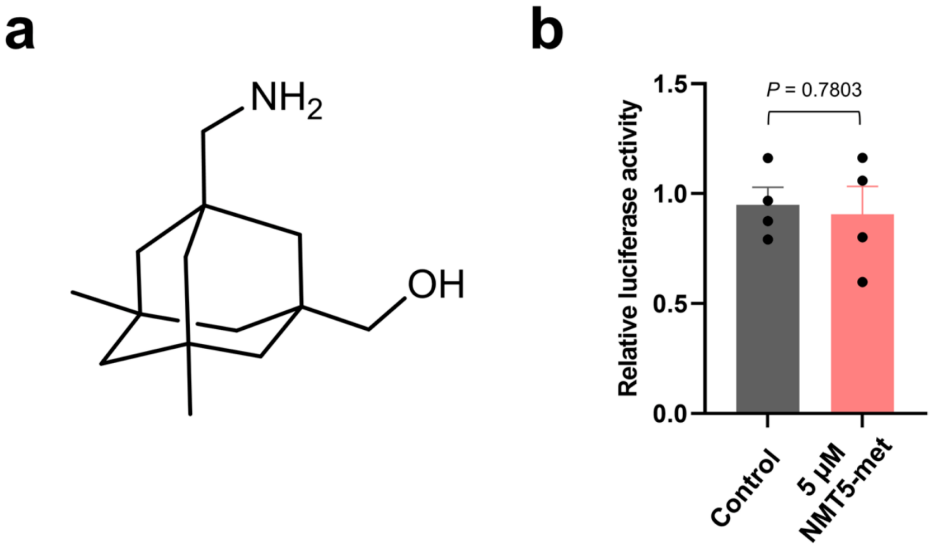


Figure ED2.7 Critical role of nitro group of NMT5 suppressing SARS-CoV-2 infection on pseudovirus entry assay.

a, Chemical structure of NMT5 metabolite (NMT5-met, lacking the nitro group). b, HeLa-ACE2 cells were incubated in the presence and absence of 5 μM NMT5-met with SARS-CoV-2 Spike (D614) pseudovirus particles. After 48 h, viral transduction efficiency was monitored by luciferase activity. Inhibitory activity was lost in the absence of the nitro group (compare with Fig. 3c,e). Data are mean + s.e.m. by two-tailed Student's t test. n = 4 biological replicates.

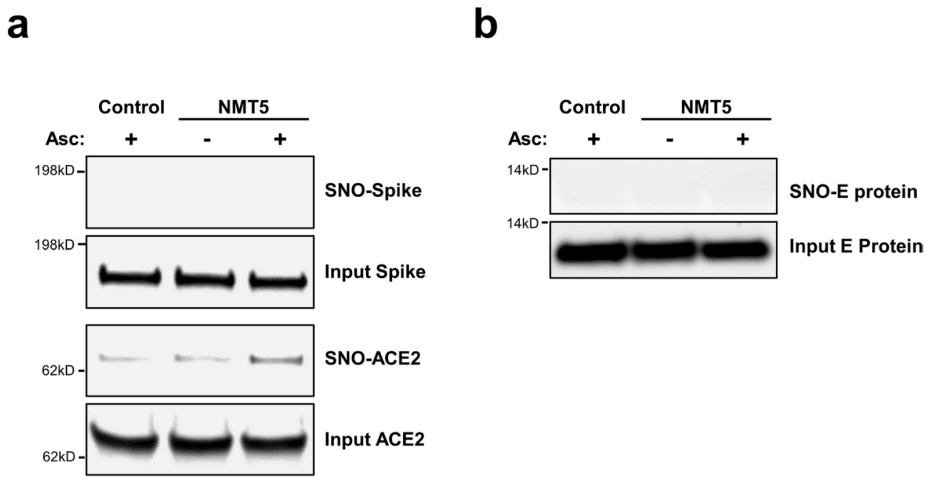


Figure ED2.8: Lack of S-nitrosylation of Spike protein and E protein

a, Purified recombinant SARS-CoV-2 Spike (S1 + S2) protein and ACE2 protein were exposed to 100 μM SNOC; 30 min later, samples were subjected to biotin switch assay in the presence or absence of ascorbate (Asc) to assess protein S-nitrosylation. n = 3 biological replicates. b, Lack of E protein S-nitrosylation by NMT5. HA-tagged E protein plasmid was transiently transfected into HEK293 cells. One day after transfection, cells were exposed to 10 μM NMT5. After 1 hour, the cells were harvested and subjected to biotin-switch assay in the presence or absence of ascorbate (Asc) to assess protein S-nitrosylation, which was detected by immunoblotting with anti-HA antibody. n = 3 biological replicates.

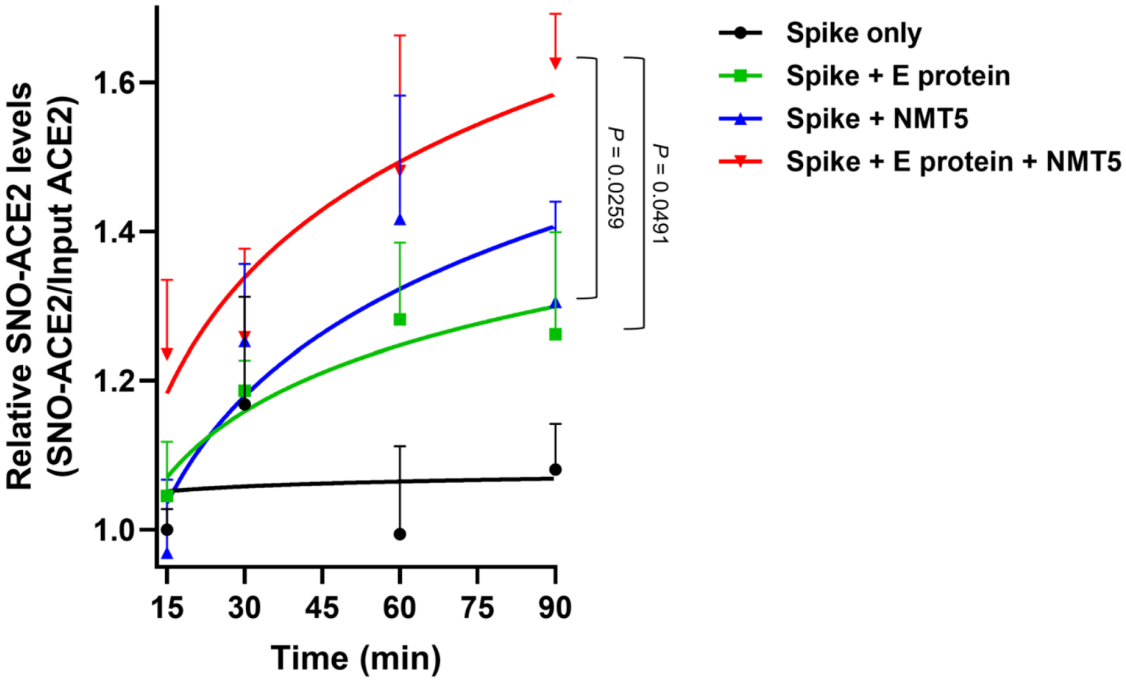
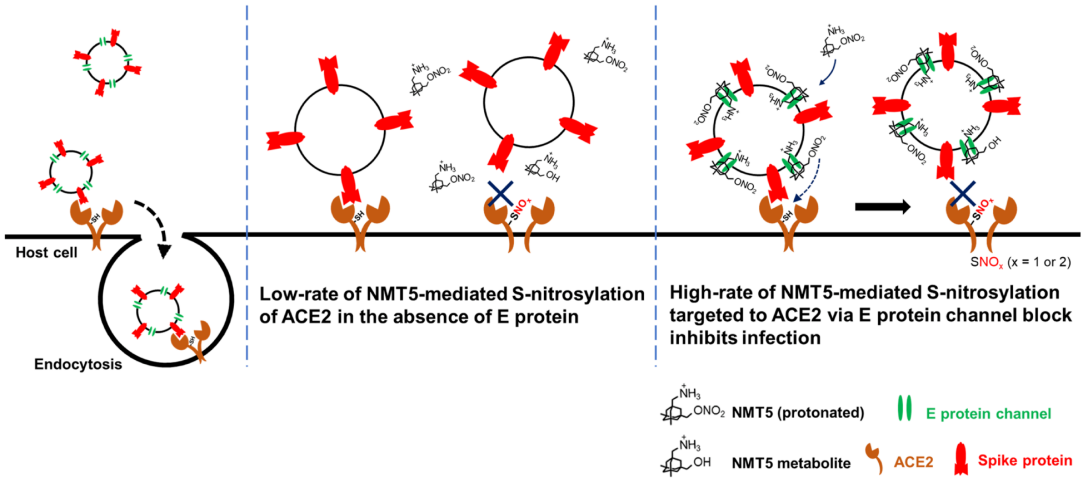


Figure ED2.9: Targeted S-nitrosylation of ACE2 by NMT5 in the presence of envelope (E) viroporin protein

Kinetic analysis of S-nitrosylation of ACE2. E protein plasmid was transiently transfected into HEK293 cells stably expressing Spike protein cells. Subsequently, the cells were harvested and plated onto ACE2-expressing HeLA cells in the presence or absence of 20 μM NMT5. At the indicated timepoints, cells were subjected to biotin-switch assay. Ordinate shows ratio of SNO-ACE2/total input ACE2 protein above the baseline value, defined as 1.0. Data are mean + s.e.m. by two-way ANOVA with Fisher's LSD test. n = 5 biological replicates.

Scheme ED2.1: Schematic of NMT5 targeting of SNO-ACE2 via SARS-CoV-2 E protein

S-Nitrosylation of ACE2 on the host cell inhibits SARS-CoV-2 entry and thus infection. Note that at physiological pH, the bridgehead amine of NMT5, like other aminoadamantanes, is generally protonated to NH_3^+.



2.7 **Acknowledgements**

We thank D. Nemazee (Scripps Research Institute) for providing HeLa-ACE2 cells and plasmids for pseudovirus. This work was supported, in part, by National Institutes of Health (NIH) grants RF1 AG057409, R01 AG056259, R01 DA048882, R35 AG071734 and DP1 DA041722 (to S.A.L.), R01 AG061845, R61 NS122098 and RF1NS123298 (to T.N.), UM1 AI144462 (to D.R.B.) and P41 GM103533 (to J.R.Y.), HOPE T32 training grant T32AI007384 (to L.N.C.); California Institute for Regenerative Medicine grant DISC2 COVID19-11811; COVID-19 awards from Fast Grants (to S.A.L.); and grants from the Bill & Melinda Gates Foundation, OPP1107194 (to Calibr) and INV-004923 (to I.A.W). The molecular dynamics simulations were supported by NIH R01 GM132826, National Science Foundation (NSF) RAPID (MCB-2032054), an award from RCSA Research Corporation, a UC San Diego Moores Cancer Center 2020 SARS-CoV-2 seed grant (to R.E.A.) and the Interfaces Graduate Training Program, NIH T32 EB009380 (to M.A.R). We thank the Texas Advanced Computing Center Frontera team and acknowledge computer time made available through a Director's Discretionary Allocation (made possible by NSF award OAC-1818253). We also thank O. Liang of the Lentivirus Core of the COBRE Center for Stem Cells and Aging at Brown University, supported by NIH P20 GM119943.

Chapter 2, in full, is a modified reprint of the material as it appears in "Chang-ki Oh , Tomohiro Nakamura, Nathan Beutler, Xu Zhang, Juan Piña-Crespo, Maria Talantova, Swagata Ghatak, Dorit Trudler, Lauren N. Carnevale , Scott R. McKercher, Malina A. Bakowski, Jolene K. Diedrich, Amanda J. Roberts, Ashley K. Woods, Victor Chi, Anil K. Gupta, Mia A. Rosenfeld , Fiona L. Kearns, Lorenzo Casalino , Namir Shaabani, Hejun Liu , Ian A. Wilson , Rommie E. Amaro, Dennis R. Burton , John R. Yates III, Cyrus Becker , Thomas F. Rogers, Arnab K. Chatterjee , Stuart A. Lipton. *Targeted protein S-nitrosylation of ACE2 inhibits SARS-CoV-2 infection.* Nature Chemical Biology 19, 275–283 (2023)." The dissertation author was a primary co-investigator of all computational findings in this work, and a co-author of this work.

2.8 **References**

1. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181, 271–280 (2020).

2. Kozakov, D., Chuang, G. Y., Beglov, D. & Vajda, S. Where does amantadine bind to the influenza virus M2 proton channel? Trends Biochem. Sci. 35, 471–475 (2010).

3. Torres, J. et al. Conductance and amantadine binding of a pore formed by a lysine-flanked transmembrane domain of SARS coronavirus envelope protein. Protein Sci. 16, 2065–2071 (2007).

4. Asarnow, D. et al. Structural insight into SARS-CoV-2 neutralizing antibodies and modulation of syncytia. Cell 184, 3192–3204 (2021).

5. Lipton, S. A. Paradigm shift in neuroprotection by NMDA receptor blockade: memantine and beyond. Nat. Rev. Drug Discov. 5, 160–170 (2006).

6. Wang, Y. et al. The pharmacology of aminoadamantane nitrates. Curr. Alzheimer Res. 3, 201–204 (2006).

7. Talantova, M. et al. Aβ induces astrocytic glutamate release, extrasynaptic NMDA receptor activation, and synaptic loss. Proc. Natl Acad. Sci. USA 110, E2518–E2527 (2013).

8. Tu, S. et al. NitroSynapsin therapy for a mouse MEF2C haploinsufficiency model of human autism. Nat. Commun. 8, 1488 (2017).

9. Takahashi, H. et al. Pharmacologically targeted NMDA receptor antagonism by NitroMemantine for cerebrovascular disease. Sci. Rep. 5, 14781 (2015).

10. Pervushin, K. et al. Structure and inhibition of the SARS coronavirus envelope protein ion channel. PLoS Pathog. 5, e1000511 (2009).

11. Schoeman, D. & Fielding, B. C. Coronavirus envelope protein: current knowledge. Virol. J. 16, 69 (2019).

12. Nieva, J. L., Madan, V. & Carrasco, L. Viroporins: structure and biological functions. Nat. Rev. Microbiol. 10, 563–574 (2012).

13. Mandala, V. S. et al. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. Nat. Struct. Mol. Biol. 27, 1202–1208 (2020).

14. Butterworth, R. F. Potential for the repurposing of adamantane antivirals for COVID-19. Drugs R D 21, 267–272 (2021).

15. Akaberi, D. et al. Mitigation of the replication of SARS-CoV-2 by nitric oxide in vitro. Redox Biol. 101734 (2020).

16. Adusumilli, N. C., Zhang, D., Friedman, J. M. & Friedman, A. J. Harnessing nitric oxide for preventing, limiting and treating the severe pulmonary consequences of COVID-19. Nitric Oxide 103, 4–8 (2020).

17. Li, Y. et al. SARS-CoV-2 and three related coronaviruses utilize multiple ACE2 orthologs and are potently blocked by an improved ACE2-Ig. J. Virol. 94, e01283–20 (2020).

18. Thunders, M. & Delahunt, B. Gene of the month: TMPRSS2 (transmembrane serine protease 2). J. Clin. Pathol. 73, 773–776 (2020).

19. Cai, Y. et al. Distinct conformational states of SARS-CoV-2 spike protein. Science 369, 1586–1592 (2020).

20. Huang, Y., Yang, C., Xu, X. F., Xu, W. & Liu, S. W. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol. Sin. 41, 1141–1149 (2020).

21. Nakamura, T. et al. Noncanonical transnitrosylation network contributes to synapse loss in Alzheimer's disease. Science 371, eaaw0843 (2021).

22. Barros, E. P. et al. The flexibility of ACE2 in the context of SARS-CoV-2 infection. Biophys. J. 120, 1072–1084 (2020).

23. Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367, 1444–1448 (2020).

24. Casalino, L. et al. AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. Int. J. High Perform. Comput. Appl. https://journals.sagepub.com/doi/10.1177/10943420211006452 (2021).

25.Singh Tomar, P. P. & Arkin, I. T. SARS-CoV-2 E protein is a potential ion channel that can be inhibited by gliclazide and memantine. Biochem. Biophys. Res. Commun. 530, 10–14 (2020).

26. Wilson, L., Gage, P. & Ewart, G. Hexamethylene amiloride blocks E protein ion channels and inhibits coronavirus replication. Virology 353, 294–306 (2006).

27. Wang, M. et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 30, 269–271 (2020).

28. Ou, X. et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat. Commun. 11, 1620 (2020).

29. Okamoto, S. et al. Balance between synaptic versus extrasynaptic NMDA receptor activity influences inclusions and neurotoxicity of mutant huntingtin. Nat. Med. 15, 1407–1413 (2009).

30. Stomberski, C. T., Hess, D. T. & Stamler, J. S. Protein S-nitrosylation: determinants of specificity and enzymatic regulation of S-nitrosothiol-based signaling. Antioxid. Redox Signal. 30, 1331–1351 (2019).

31. Rogers, T. F. et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. Science 369, 956–963 (2020).

32. Funk, T. et al. Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. Eur. Surveill. 26, 2100348 (2021).

33. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220 (2020).

34. Stamler, J. S., Toone, E. J., Lipton, S. A. & Sucher, N. J. (S)NO signals: translocation, regulation, and a consensus motif. Neuron 18, 691–696 (1997).

35. Greco, T. M. et al. Identification of S-nitrosylation motifs by site-specific mapping of the S-nitrosocysteine proteome in human vascular smooth muscle cells. Proc. Natl Acad. Sci. USA 103, 7420–7425 (2006).

36. Smith, B. C. & Marletta, M. A. Mechanisms of S-nitrosothiol formation and selectivity in nitric oxide signaling. Curr. Opin. Chem. Biol. 16, 498–506 (2012).

37. Lancaster, J. R. Jr. How are nitrosothiols formed de novo in vivo? Arch. Biochem. Biophys. 617, 137–144 (2017).

38. Chan, J. F. et al. Simulation of the clinical and pathological manifestations of Coronavirus Disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility. Clin. Infect. Dis. 71, 2428–2446 (2020).

39. Cho, H. et al. Bispecific antibodies targeting distinct regions of the spike protein potently neutralize SARS-CoV-2 variants of concern. Sci. Transl. Med. 13, eabj5413 (2021).

40. D'Agnillo, F. et al. Lung epithelial and endothelial damage, loss of tissue repair, inhibition of fibrinolysis, and cellular senescence in fatal COVID-19. Sci. Transl. Med. 13, eabj7790 (2021).

41. Doi, M., Takahashi, F. & Kawasaki, Y. Bayesian noninferiority test for 2 binomial probabilities as the extension of Fisher exact test. Stat. Med. 36, 4789–4803 (2017).

42. Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426, 450–454 (2003).

43. Uehara, T. et al. S-nitrosylated protein-disulphide isomerase links protein misfolding to neurodegeneration. Nature 441, 513–517 (2006).

44. Ghatak, S. et al. Mechanisms of hyperexcitability in Alzheimer's disease hiPSC-derived neurons and cerebral organoids vs isogenic controls. eLife 8, e50333 (2019).

45. He, L., Diedrich, J., Chu, Y. Y. & Yates, J. R. III Extracting accurate precursor information for tandem mass spectra by RawConverter. Anal. Chem. 87, 11361–11367 (2015).

46. Xu, T. et al. ProLuCID: an improved SEQUEST-like algorithm with enhanced sensitivity and specificity. J. Proteom. 129, 16–24 (2015).

47. Tabb, D. L., McDonald, W. H. & Yates, J. R. III DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J. Proteome Res. 1, 21–26 (2002).

48. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. & Gygi, S. P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC–MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. 2, 43–50 (2003).

49. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Jr. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. J. Chem. Inf. Model. 52, 3155–3168 (2012).

50. Vanommeslaeghe, K. & MacKerell, A. D. Jr. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. J. Chem. Inf. Model. 52, 3144–3154 (2012).

51. Yu, W., He, X., Vanommeslaeghe, K. & MacKerell, A. D. Jr. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. J. Comput. Chem. 33, 2451–2468 (2012).

52. Casares, D., Escriba, P. V. & Rossello, C. A. Membrane lipid composition: effect on membrane and organelle structure, function and compartmentalization and therapeutic avenues. Int. J. Mol. Sci. 20, 2167 (2019).

53. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. Nat. Rev. Mol. Cell Biol. 9, 112–124 (2008).

54. Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. J. Chem. Phys. 153, 044130 (2020).

55. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods 14, 71–73 (2017).

56. Guvench, O., Hatcher, E. R., Venable, R. M., Pastor, R. W. & Mackerell, A. D. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. J. Chem. Theory Comput. 5, 2353–2370 (2009).

57. McGibbon, R. T. et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. Biophys. J. 109, 1528–1532 (2015).

58. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. J. Mol. Graph. 14, 27–38 (1996). 33-38.

59. Zhou, L. et al. MEK inhibitors reduce cellular expression of ACE2, pERK, pRb while stimulating NK-mediated cytotoxicity and attenuating inflammatory cytokines relevant to SARS-CoV-2 infection. Oncotarget 11, 4201–4223 (2020).

60. Shapira, T. et al. A TMPRSS2 inhibitor acts as a pan-SARS-CoV-2 prophylactic and therapeutic. Nature 605, 340–348 (2022).

Chapter 3: GlycoGrip: Cell Surface-Inspired Universal Sensor for Betacoronaviruses

3.1 **Abstract**

Inspired by the role of cell-surface glycoproteins as co-receptors for pathogens, we report the development of GlycoGrip: a glycopolymer-based lateral flow assay for detecting SARS-CoV-2 and its variants. GlycoGrip utilizes glycopolymers for primary capture, and anti-spike antibodies labeled with gold nanoparticles for signal-generating detection. A lock-step integration between experiments and computation has enabled efficient optimization of GlycoGrip test strips which can selectively, sensitively, and rapidly detect SARS-CoV-2 and its variants in biofluids. Employing the power of the glycocalyx in a diagnostic assay has distinct advantages over conventional immunoassays as glycopolymers can bind to antigens in a multivalent capacity and are highly adaptable for mutated strains. As new variants of SARS-CoV-2 are identified, GlycoGrip will serve as a highly reconfigurable biosensor for their detection. Additionally, via extensive ensemble-based docking simulations to incorporate protein and glycan motions, we have elucidated important clues as to how heparan sulfate and other glycocalyx components may bind the spike glycoprotein during the SARS-CoV-2 host-cell invasion. GlycoGrip is a promising and generalizable alternative to costly, labor-intensive RT-PCR, and we envision it will be broadly useful, including for rural or low-income populations that are historically undertested and underreported in infection statistics.

3.2 **Introduction**

The glycocalyx, a dense “sugary” matrix that coats epithelial tissue cells, is responsible for cell-cell adhesion, extracellular communication, growth factor monitoring, defense against exogenous pathogens, and much more.[1,2] The major components of the glycocalyx include enzymatic glycoproteins, glycolipids, and proteoglycans. Proteoglycans are heavily glycosylated membrane proteins whose glycan components are mainly glycosaminoglycans (GAGs), or long, linear polysaccharides with repeating units of a uronic acid sugar and an amino sugar.[3]

Heparan sulfate (HS) proteoglycans are the most abundant component of epithelial glycocalyx, making up 50-90% of the total sugar composition, followed by chondroitin sulfate (CS).[4] HS is made up of repeating dimeric units of a uronic acid (either glucuronic or iduronic acid) and N-acetylglucosamine, and CS consists of repeating dimeric units of glucuronic acid and N-acetylgalactosamine. Stereochemical composition (i.e., proportions of glucuronic acid versus iduronic acid) and sulfation rates in HS and CS vary greatly depending on tissue types, as well as other physiological conditions such as healthy or diseased tissue status.[5–9] Viral pathogens often hijack glycocalyx receptor trafficking and signal transduction mechanisms to facilitate entry into host cells (Figure 3.1).[10–12] An important example of this is SARS-CoV-2 (severe acute respiratory coronavirus 2) viral replication cycle and its resultant disease known as COVID-19 (coronavirus disease 2019).

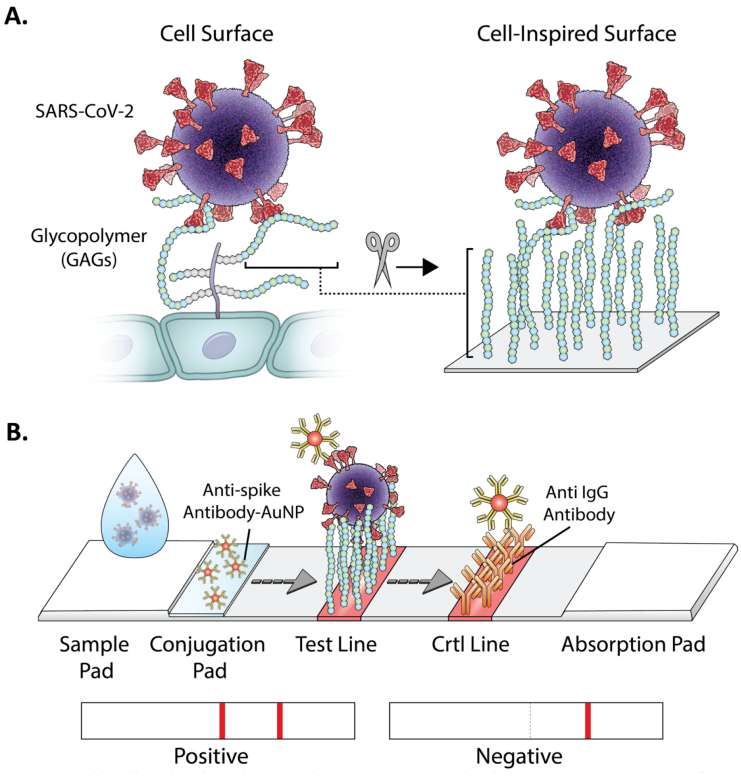


Figure 3.1: Graphical illustration of (A) virus interaction with GAGs on the cell surface, and (B) on GlycoGrip lateral flow (LF) biosensor for detecting SARS-CoV-2

The sample is deposited on the sample pad and migrates towards the conjugate. The conjugated antibodies bind the virus and migrate to the test line, where the bound target analyte is captured by the glycopolymers. Possible results and interpretation of the test are shown below.

SARS-CoV-2 is a member of the betacoronavirus genus within the Coronaviridae family of viruses; it is a lipid-enveloped, positive-sense, single-stranded RNA virus studded with approximately 30 structural S or "spike" glycoproteins (Figure 3.2). The spike is a homotrimeric protein composed of many interworking domains. Two of these domains are particularly important to this work: the Receptor Binding Domain (RBD, residues 330 to 530) and the N-terminal Domain (NTD, residues 13 to 296) (Figure 3.2A). Furthermore, the spike surface is heavily shielded with 66 N-linked glycans and a varying number of O-glycans. [13,14] SARS-CoV-2 spike's main function is to incite the membrane fusion process by binding to angiotensin converting enzyme 2 (ACE2), situated on the surface of ciliated lung epithelial cells. [15–19] To bind ACE2, the spike must be in an "open" conformation, with at least one RBD in the "up" state (Figure 3.2A, Figure S3.1). [20,21] In particular, the RBD moves into the "up" state to reveal the Receptor Binding Motif (the RBM), the spike region that makes direct contact with ACE2. Recent works have described the role of the spike's glycans in facilitating and stabilizing the conformational transition from down RBDs to up RBDs[15,19,22] thereby facilitating host-cell invasion. Intriguingly, glycocalyx glycopolymers may also help facilitate SARS-CoV-2 invasion: Skidmore and coworkers identified a GAG binding site on the spike RBD,[23] and Esko and coworkers have illustrated that spike binding to HS in the glycocalyx facilitates interaction with ACE2, and incubation with heparin (HEP) induces an increase in spike populations with up versus down RBDs.[15,19,22,24–32] Furthermore, other work by Linhardt and coworkers,[33] Fadda and coworkers,[34] Wade and coworkers,[35] and Gandhi and coworkers,[36] have posited HEP binding sites on the spike surface (Figure 3.2B).

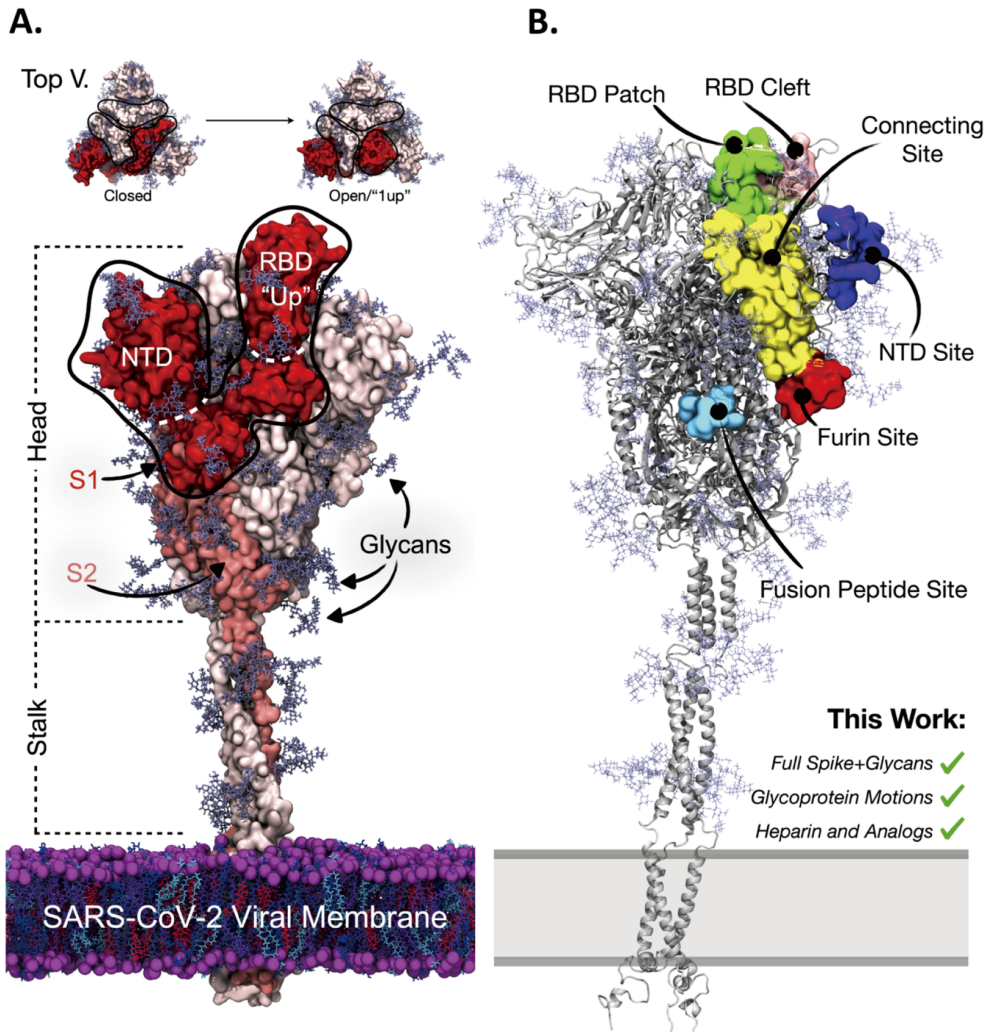


Figure 3.2: Molecular representation of SARS-CoV-2 spike proteins

(A) Molecular representation of SARS-CoV-2 in the "1-up" conformational state. The spike protein is represented with red, salmon, and light pink surfaces. Spike glycan atoms are shown with light blue licorice representation. Top panel illustrates, from a top-down view the "closed" to "open"/ "1-up" RBD conformational change required for host-cell invasion. The spike's S1 domain is highlighted in red surface representation, while the spike's S2 domain is in salmon surface representation. The spike's RBD and NTD are outlined for reference. (B) Molecular representation of SARS-CoV-2 spike in closed conformation depicting literature proposed HEP binding sites. Green surface: the "RBD patch", a site proposed by Skidmore and coworkers[23] and supported by Esko and coworkers to have high affinity for heparin.[27] Pink surface: the "RBD cleft" a site proposed by Fadda and coworkers to have high affinity for polysaccharides.[34] Red surface: the furin cleavage site.[37–39] Light blue surface: the fusion peptide site proposed by Linhardt et al. [33] Yellow surface: the connecting site proposed by Wade et al. [35] Purple surface: the NTD site identified by Schuurs et al.[36]

GAGs overall have been underappreciated as potential bioreceptors in biosensors due to their complexity, heterogeneity in sulfation patterns, and variable binding specificity (i.e., one GAG can bind many analytes)

compared to highly targeted antibodies.[11,12,40,41] Biologically, HS and CS serve as cellular staging grounds: they bind many analytes while the targeted co-receptors find optimal orientation on the cell surface. [42] We argue that GAG's abilities to bind multiple analytes [KF1] -- inspired by glycobiology – can be leveraged to design highly generalizable sensors for the sensitive detection of viruses and viral antigens, and we applied this approach for sensing SARS-CoV-2 and its variants in a Lateral Flow Strip-based Assay (LFSA) (Figure 3.1B).

LFSA is an attractive platform for detecting viruses, especially in limited resource settings, due to its simplicity, low cost, and rapid signal generation. Typically, "sandwich-type" detection using lateral flow (LF) strips utilize two bioreceptors that bind to the target molecule simultaneously.[43–46] One of the bioreceptors is usually immobilized on the nitrocellulose membrane surface to capture the target analyte, while the other bioreceptor is typically labeled with reporter molecules (e.g., gold nanoparticles, fluorescent dyes, enzymes, etc.) to signal the formation of a sandwich complex in the presence of the analyte. Using traditional bioreceptors, such as antibodies and aptamers as immobilizing agents in LFSA design, requires these receptors to be screened and optimized, as well as necessitating that LF strips be reconfigured, for every desired viral analyte and potential mutant strains. Furthermore, in the case of LFSA sandwich-type detection, pairs of antibodies or aptamers must be screened for both capturing and reporting, which delays sensor design and development.[47,48] In contrast, GAGs could serve as universal capture agents for various viral analytes including mutant strains, resulting in an easily generalizable LF platform. This could greatly reduce the assay's antibody screening and optimization time, which typically takes anywhere from 12 to 16 weeks.[49,50] This could enable fast adaptation of GAG-based LF strips for current and emerging viruses and provide cheap and simple to administer viral antigen tests for critically under-tested communities during times of global health crises. In the case of COVID-19, nucleic acid detection methods, such as RT-PCR, are the gold standard in viral testing. However, there is a significant time gap between testing and obtaining results due to the PCR testing capacity limitations. Additionally, areas without accessible RT-PCR testing capabilities are predominantly lower income and/or rural, making this a public health priority as well. Thus,

self-administered antigen-based rapid testing would be able to fill the time and resource gaps of the nucleic acid-based testing method for monitoring and containment.[51–53]

Using GAGs as capture probes in LF biosensors introduces two major design challenges: (1) Due to their molecular flexibility, specific binding mechanisms between GAGs and the spike protein are largely unknown and experimentally-determined bound structures remain elusive, and (2) due to their highly heterogeneous composition -- i.e., varying stereochemical ratios, sulfation rates, and chain lengths -- it is challenging to optimize sensitive and selective LF strips in a robust and reproducible fashion. To address these issues herein, we have integrated rigorous computational methods and extensive experimental system development to create GlycoGrip: a highly sensitive and selective LF strip biosensor for a rapid detection of SARS-CoV-2 spike protein (Figure 3.1B). Our GlycoGrip LF biosensor is inspired by eons old, biological signaling technology: the glycocalyx. It utilizes GAGs as primary bioreceptors anchored to a test trip to capture the spike protein, while anti-spike monoclonal antibodies labeled with gold nanoparticles (AuAb) are used as reporters. In the presence of the virus, GAGs and AuAb co-bind to the target virus forming the “sandwich” ternary complex and generating color on both the test and control lines in under 30 minutes (Figure 3.1B, ‘positive’). In the absence of target virus, only color on the control line emerges as AuAb are captured by the control anti-IgG antibodies (Figure 3.1B, ‘negative’).

Through a recursive feedback loop of experiments and simulations, we have fine-tuned our GlycoGrip sensor and elucidated key mechanisms of how polysaccharides, specifically HS, bind spike during SARS-CoV-2 host cell invasion. Our computational methods uniquely integrated fully flexible, ensemble-based docking procedures considering the entirety of the spike head, with a complete spike glycoprofile, and modeled several GAGs including HS, HEP, CS and dextran sulfate (DEX, a synthetic HEP analog). Our GlycoGrip technology is uniquely tailored to capture and detect SARS-CoV-2 and its rapidly emerging variants and could be applied to detect other pathogenic proteins.

3.3 **Results and Discussion**

Establishing glycocalyx polymers as capture agents for spike binding

The concept of taking advantage of GAGs' ability to bind spike in a multivalent manner and employ them as surface-anchored capture agents for detecting SARS-CoV-2 serves as the basis for our work. Skidmore and coworkers were the first to show, via circular dichroism, surface plasmon resonance, and molecular modeling, a clear GAG binding site on the spike RBD.[23] Furthermore, they illustrate that SARS-CoV-2 cellular invasion could be inhibited by introduction of exogenous heparin. Esko and coworkers elaborated on this by showing HS is a necessary co-receptor for SARS-CoV-2 viral infection, positing that HEP inhibition is caused by "distraction" from cellular HS.[27] Esko and coworkers showed that binding of SARS-CoV-2 spike to ACE2 in mammalian cell lines is drastically reduced upon introduction of heparan lyase, and the presence of ACE2 alone on a mammalian cell surface is not sufficient for SARS-CoV-2 host-cell invasion, suggesting HS is required.[27,54] Interestingly, Fadda and coworkers,[34] through investigation of evolutionary loss of a glycan at position N370, posit that a cleft at the tip of the RBD is uniquely suited for binding oligosaccharides. Skidmore's and Esko's computational modeling focused attention on interactions between HS/HEP and a patch along the spike's RBD,[23,27] whereas Wade and coworkers,[35] Linhardt and coworkers,[33] and Gandhi and coworkers,[36] extended their search ranges to include the entire spike head. Wade and coworkers identified a cleft connecting the RBD and furin cleavage site that can be stably occupied by long chain GAGs.[35] Linhardt and coworkers identified several binding sites of interest including a site located near the fusion peptide.[26] Gandhi and coworkers identified an otherwise unidentified site on the NTD.[36] Additionally, Boons and coworkers and Desai and coworkers have conducted extensive mappings of HS stereochemistry and sulfation patterns to identify key "heparin-like" motifs optimal for targeting the SARS-CoV-2 spike glycoprotein. Taken together, these works indicate that the structure of SARS-CoV-2 spike might have evolved to be uniquely tailored for binding glycocalyx GAGs.

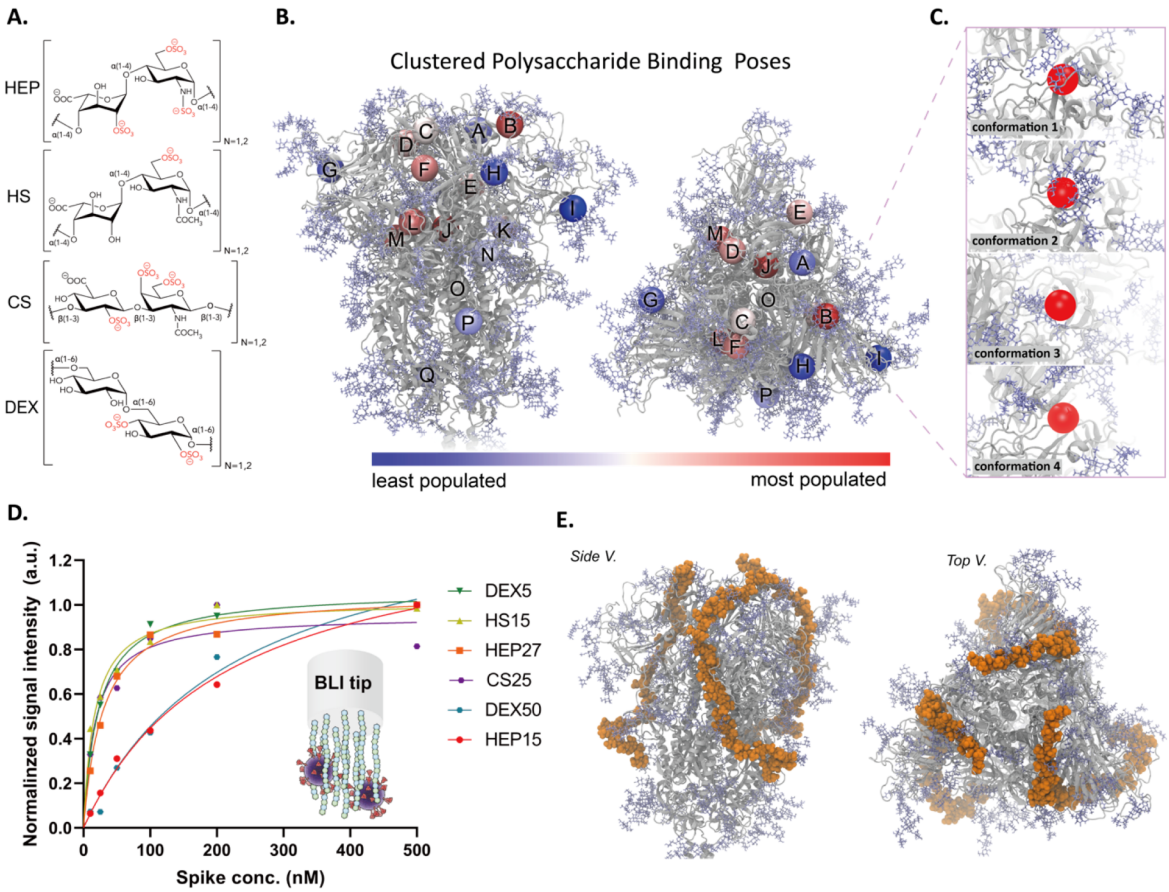


Figure 3.3: Depiction of clustered saccharide binding sites and trivalent HS binding mode

(A) Molecular representation of each GAG candidate considered in this work: HEP, H6S (heparan sulfate sulfated at the 6-O position), CS, and DEX. All four GAGs were modeled in both dimeric (n = 1) and tetrameric (n = 2) forms. (B) Molecular representations (side and top views) of spike head (grey ribbons) with spike glycans (light blue, licorice atoms). Colored spheres shown on the spike head illustrate the centers of mass of each GAG binding site as predicted by our ensemble-based docking studies. We predicted over 12,800 spike-GAG binding modes and clustered these binding modes into 17 distinct binding sites. These sites were each ranked by our binding site importance score. Spheres are colored to indicate the relative importance of each site according to our binding site importance score: red spheres indicating relatively important sites, blue spheres indicating less important sites. (C) Molecular snapshots of binding site B, corresponding to a supersite formed between the RBD patch and the RBD cleft, for each protein/glycan conformation used in ensemble-based docking. Four spike (protein and glycan) conformations were used in docking studies to incorporate a degree of spike flexibility, the degree of conformational diversity can be observed in these zoom-in images of site B. (D) Experimentally calculated binding affinities between spike and tested GAGs (steady-state analysis of BLI data to determine K_D values). (E) Molecular representation of our constructed spike head-trivalent hep40mer model.

Modeling GAG chemical diversity

We first sought to identify spike-GAG binding sites to exploit for our sensor. We used an ensemble-based docking protocol that includes docking of multiple GAG identities into various well-sampled/varied protein and glycan conformations, and unbiased searching of the entire spike head. In addition to the most abundant GAGs on the cell surface, HS and CS,[4] low molecular weight HEP has been shown to bind effectively to the SARS-CoV-2 spike, as well as induce the RBD down (where the RBM is shielded, thus cannot bind ACE2) to RBD up (RBM exposed, thus ready to bind ACE2) conformation.[26,27,29,35] To capture sufficient GAG diversity, we modeled HEP, HS, CS, and DEX to identify which GAG would best capture the SARS-CoV-2 spike on an LF test strip (Figure 3.3A). Cellular HS is incredibly heterogeneous in both uronic acid identity and degree of sulfation. [23,26,27,29,55–58] To better match HS considered in our LFSA design and testing (where HS was purchased from Sigma and reported to be only 5-7% sulfation by mass), we constructed 6-O-sulfated heparan sulfate, from here on referred to as our H6S model. Although this model does not capture the full heterogeneity of cellular HS, it does better reflect the low sulfation rate of HS considered experimentally in this work. As docking long polysaccharide chains is intractable -- due to combinatorial enumeration of rotational degrees of freedom -- we chose to model dimeric (n = 1) and tetrameric (n = 2) forms of each of our four candidates. Dimeric forms were included to capture highly localized interactions, while tetrameric forms were included to elucidate slightly longer-range effects, i.e., inaccessibility for longer polysaccharide chains. We used MatrixDB[59–62] to build dimeric and tetrameric HEP and H6S, and CHARMM-GUI[63–66] to build CS and low sulfated (~6% sulfated) DEX. (See Methods for complete chemical description of each molecule, and Figure 3.3A for ChemDraw images.)

Accounting for spike conformations and surface accessibility for binding

To incorporate protein and glycan flexibility in our docking protocol, we docked all GAG models (8 total molecule models) into four different spike conformations extracted from Casalino et al.'s trajectories (Figure 3.3B-C).[15] We used accessibility of the furin-cleavage site as a metric to identify four conformationally unique

frames to serve as receptor structures for docking. The polybasic furin-cleavage site (spike residues 674 to 685) is one of the most flexible regions of the spike protein (Figures S3.2-4) and postulated to bind a myriad of physiological cofactors.[32,35,37] Therefore, we selected four spike conformations (conformations 1 through 4 in Figure 3.3C, S3.3-S3.4) based on their degree of accessibility to the furin cleavage site (Table S1, see Supporting Information Methods for calculation details). Furthermore, given each of these GAGs is highly flexible, we thoroughly sampled polysaccharide rotational degrees of freedom by predicting 400 poses per ligand and protein conformation pair (see Supporting Information methods for complete details). At 8 GAG models, four protein conformations, and 400 predicted poses per pair, this resulted in a total of 12,800 resultant binding modes.

Selecting favorable spike-gag binding sites

To organize our 12,800 predicted binding poses into discernable binding "sites," we clustered them by their centers of mass (Figure S3.5, Table S3.2, see SI Methods). From clustering, we determined 17 distinct spike-GAG binding sites (Figure 3.3B). For easy reference, we have indexed the sites by letter, A through Q (see Scheme S3.1, Figure S3.6, and SI Methods for complete detail on our docking and analytical clustering methods). We then derived a "binding site importance score" (Equation 1) to describe the "importance" of each predicted binding site.

$$\textit{Binding Site Importance Score} = |\textit{(avg binding score of cluster)}| * \textit{fraction populated} \quad \text{(Equation 1)}$$

Out of the 17 unique binding sites, three sites (J, O, and Q) were omitted from further analysis as they are highly buried within the spike and would not be accessible to long-chain GAGs (Scheme S1, Figure S3.7). From the remaining 14 surface binding sites, we identified 7 novel binding sites (F, G, I, K, L, M, N) and validated 7 previously identified binding sites (A, B, C, D, E, H, P) (Figure S3.8). Sites C, B, and D correspond to a "supersite" formed between the RBD patch[23,27] and RBD cleft[34] sites (C, B, and D are analogous sites

centered on one of each spike protomers) and sites E and H correspond to the connecting ridge posited by Wade and coworkers[35] which connects the RBD supersite down to the furin-cleavage site. Our results support the importance of these sites for GAG binding and reaffirm the need to focus on these regions when studying the role of HS in SARS-CoV-2 host-cell invasion mechanism. Novel sites F, L, and M have yet to be described in the literature, but they represent an alternative binding mode for long-chain GAGs on the spike surface. As shown in Figure 3.2B, the connecting ridge connects the RBD sites (green and pink) down to the furin cleavage site along a ridge formed between the left protomer's RBD and the right protomer's NTD. Sites M and L represent sites that could, instead, be used to connect the RBD supersite down to the furin cleavage site between a ridge formed by the right protomer's RBD and the left protomer's NTD (Figure S3.9), a path similar to that predicted by Schuurs et al.[36] Thus, there are potentially two distinct surface paths along which a long GAG could bind to interact with both the RBD supersite and the furin cleavage site. This finding highlights the importance of GAGs' multivalent binding mode in attaching to spike surface. For all 14 surface binding sites, we estimated their relative "importance" through our binding site importance score (Equation 1). Sites B, D, E, F, L, and M, all have relatively high importance with scores 0.9, 0.6, 0.6, 0.7, 0.8, 0.9, respectively, (see Table S3-S7 for all docking results), indicating a high likelihood for GAG binding at these sites.

Experimental Characterization of spike-GAG binding affinities by biolayer interferometry

Our ensemble-based docking indicated that all dimer and tetrameric GAGs bound with relatively similar predicted binding energies in each binding site (Tables S2-S6). To better discriminate their affinity to spike, we turned to biolayer interferometry. As the glycocalyx composition is heterogeneous with respect to GAG identity and GAG length, we were interested to evaluate the binding affinity of various GAGs at different chain lengths to the spike: 15 and 27 kDa HEP (HEP15, HEP27), 15 kDa HS (HS15), 25 kDa CS (CS25), and 5 and 50 kDa DEX (DEX5, DEX50). As shown in Figure 3.3D, all GAGs, at all chain lengths, bound to spike with relatively high affinity. HS15 and CS25 exhibited the highest binding affinities out of all GAGs tested, with K_D

of 16.7 nM [95% CI; 8 nM-29 nM] , 17.9 nM [95% CI; 5 nM-43 nM] respectively. These are followed in order of K_D by HEP27, DEX5, HEP15, and DEX50 (Figure S3.10, Table S3.8 for summary of all BLI results).[67–69] As HS15 shows the highest affinity for spike and considering the ubiquitous presence of HS in the glycocalyx, coexisting around ACE2, it is possible the SARS-CoV-2 spike sequence has undergone selective pressure with respect to binding ACE2 as well as HS. In the context of our sensor construction, although all BLI-tested GAGs showed high affinity for SARS-CoV-2 and could serve as capture agents, we chose to focus on HS and HEP in our current device.

Constructing a long-chain HEP-spike model

Although using short (e.g., dimeric and tetrameric) GAG models was necessary to conduct extensive ensemble-based docking protocols, these short GAGs do not fully capture the extent of steric hindrance and torsional constraints that would arise as SARS-CoV-2 spike approaches a long-chain GAG either in an LF test strip or during host cell invasion. Thus, we used our list of highly populated binding sites and literature sites to guide the construction of a more relevant computational model of spike-GAG binding (Figure 3.3E). We used octameric HEP units as a molecular “thread” to “sew” continuously through binding sites B, H, and down to the furin cleavage site, ultimately connecting all these sites with a 40-meric HEP (hep40mer) molecule (SI Methods). Given the symmetrical nature of spike, we repeated this process for all three spike protomers to illustrate the multivalent GAG-binding potential of spike under biological and in vitro assay conditions (Figure 3.3E). In this current work, our model serves to connect our docking results, at the atomic (molecular-) scale, to our experimental results and the biological context, at the macro-scale.

Taken together, our computational models and biolayer interferometry measurements reveal that glycocalyx-derived GAGs are a promising class of molecules to act as capture agents for SARS-CoV-2, as they have high affinity for its spike protein and can bind the spike in multiple binding sites, with various spike conformations, at high valency.

Mechanistic insights: GAGs adapt to spike conformation.

Interestingly, the iterative process of computation and experiments for identifying spike-GAG binding has enabled us to unravel mechanistic insights into SARS-CoV-2 host-cell invasion. Accounting for protein and glycan flexibility via ensemble-based docking elucidated two key mechanistic hypotheses: (1) GAGs are likely to bind in multiple compensatory modes to accommodate changes in spike conformation, and (2) the spike's glycans compete with GAGs for certain binding sites on the spike surface yet stabilize other GAG-spike interactions.

GAGs can adapt multiple binding modes, adjusting to spike conformations

As mentioned, we predict site B to be a hotspot for GAG binding; according to our binding site importance score it is our #1 ranked site overall, and it is relatively accessible over the course of 1.8 μs (Figure S3.7). Furthermore, B sits at crucial spot for the spike: "behind" the RBM, and at the interface between the RBD and the neighboring protomer's NTD (Figure S3.11). Esko posits interactions between low molecular weight HEP and this site could induce transition from a down to up RBD.[27] Fadda elaborates this site has high affinity for negatively charged oligosaccharides.[34] Additionally, Casalino et al.'s simulations illustrate the RBM is one of the most flexible regions in the spike head, second only to the furin cleavage site (Figure S3.3).[15] The RBM's/site B's conformational diversity is exemplified clearly in Figure 3.3C. Despite drastic differences in protein and glycan topography around the RBM, site B is highly populated in all four protein conformations for all GAG models (Table S3-7). Thus, we hypothesize site B may accommodate multiple interconverting GAG binding modes. The differing binding modes at site B may reveal how HEP facilitates the RBD's transition to the up state: by alleviating tight interfacial interactions between the RBD and NTD, and by remaining bound despite drastic RBD conformational changes (Figure S3.11).

Spike glycans can compete with GAGs or stabilize GAGs at surface binding sites

Some GAG-spike binding sites, most notably site E, are entirely unoccupied by GAGs in specific spike conformations but highly populated in others. Our analysis reveals that spike glycans may compete with, or stabilize GAG binding at the spike surface. In spike conformations 1 and 3, site E is highly populated for all GAGs. When considering data from only spike conformations 1 and 3, site E is ranked #1, above site B, according to our binding site importance score. However, in spike conformations 2 and 4, site E was shown to be completely inaccessible to GAG binding: 0 binding modes for all modeled molecules (Table S4, S6). Careful inspection of this site reveals that in spike conformations 2 and 4 the glycan at $N331_B$ directly occupies site E, whereas in conformations 1 and 3 $N331_B$ glycan moves away from site E (Figure S3.12). Notably, while $N331_B$ moves away from site E in conformations 1 and 3, it is still neighboring site E and thus can provide stabilizing hydrogen bonding interactions to GAGs at this site (Figure S3.12E). Thus, glycan $N331_B$ serves to shield GAGs from binding at site E, but when site E becomes available, $N331_B$, $N122_C$, and $N165_C$ serve to stabilize GAGs bound at this site. This analysis suggests yet another role for glycans in the spike infection process. Casalino et al. have shown that glycans can shield key spike antigenic regions from recognition by potent antibodies, but glycans N165 and N234 can also prop up and stabilize the RBD in the up state.[15] Furthermore, Sztain et al. have shown that N343 facilitates movement to the up state by "pushing" the RBD up through hydrophobic interactions with RBD residues.[19] Here, we show that glycans serve a dual role: they can both shield the spike surface and stabilize GAGs that make it to the spike surface.

NTD is ideal for co-binding spike using GAG-bound test strips

Next, to generate a robust sandwich-based lateral flow assay, we tested a few spike-binding proteins and antibodies for their ability to co-bind spike with surface-bound GAGs. We experimentally screened the following candidates: (1) REGN10933 (RBD Ab), one half of the powerful REGN-COV2 antibody cocktail which binds to a sub-region of the spike RBM,[70] (2) ACE2, the enzyme responsible for binding to spike RBM, which initiates host cell invasion,[71,72] and (3) an NTD binding antibody (NTD Ab) (Figure 3.4A). As the NTD binding antibody we have used here does not have a specific known epitope, for our computational

modeling we have chosen to use the 4A8 NTD binding antibody as a structural stand-in.[73] Initial experimental screening revealed that all candidates form the sandwich-style complex with GAGs when exposed to spike protein for 30 min (Figure 3.4B). To investigate which candidate would generate a positive signal in a shorter detection time, we reduced the incubation time to 5 minutes. Interestingly, using NTD Ab resulted in, by far, the most intense LFSA signal compared to RBD Ab and ACE2 after 5 min, suggesting that the RBD site might be the most ideal for co-binding spike with GAGs (Figure 3.4B). To explore this result at the atomic scale, we constructed simple spatial models of these complexes: (1) 3xhep40mer with one RBD open spike protomers and bound RBD antibody REGN10933 (Figure S3.13A),[70] (2) 3xhep40mer with one RBD open spike protomers and bound ACE2 monomer (Figure S3.13B),[71] and finally (3) 3xhep40mer with all down/closed spike protomers and bound NTD antibody 4A8 (Figure 3.4C).[73] From these spatial models, even in the 3xhep40mer-spike complexes, binding of REGN10933, ACE2, or 4A8 could all be theoretically accommodated. Thus, static models alone could not explain why NTD antibodies would provide a much higher LFSA signal relative to RBD-binding biomolecules (REGN10933 and ACE2). None of the predicted binding sites, nor our long-range HEP model, nor the literature proposed sites, overlap with the REGN10933 epitope, the RBM/ACE2 binding domain, or the 4A8 epitope (as illustrated by Figures S3.14A-B, and 3.2C, respectively). Therefore, we did not suspect that choice of antibody would have such dramatic impact on the intensity of LFSA signal.

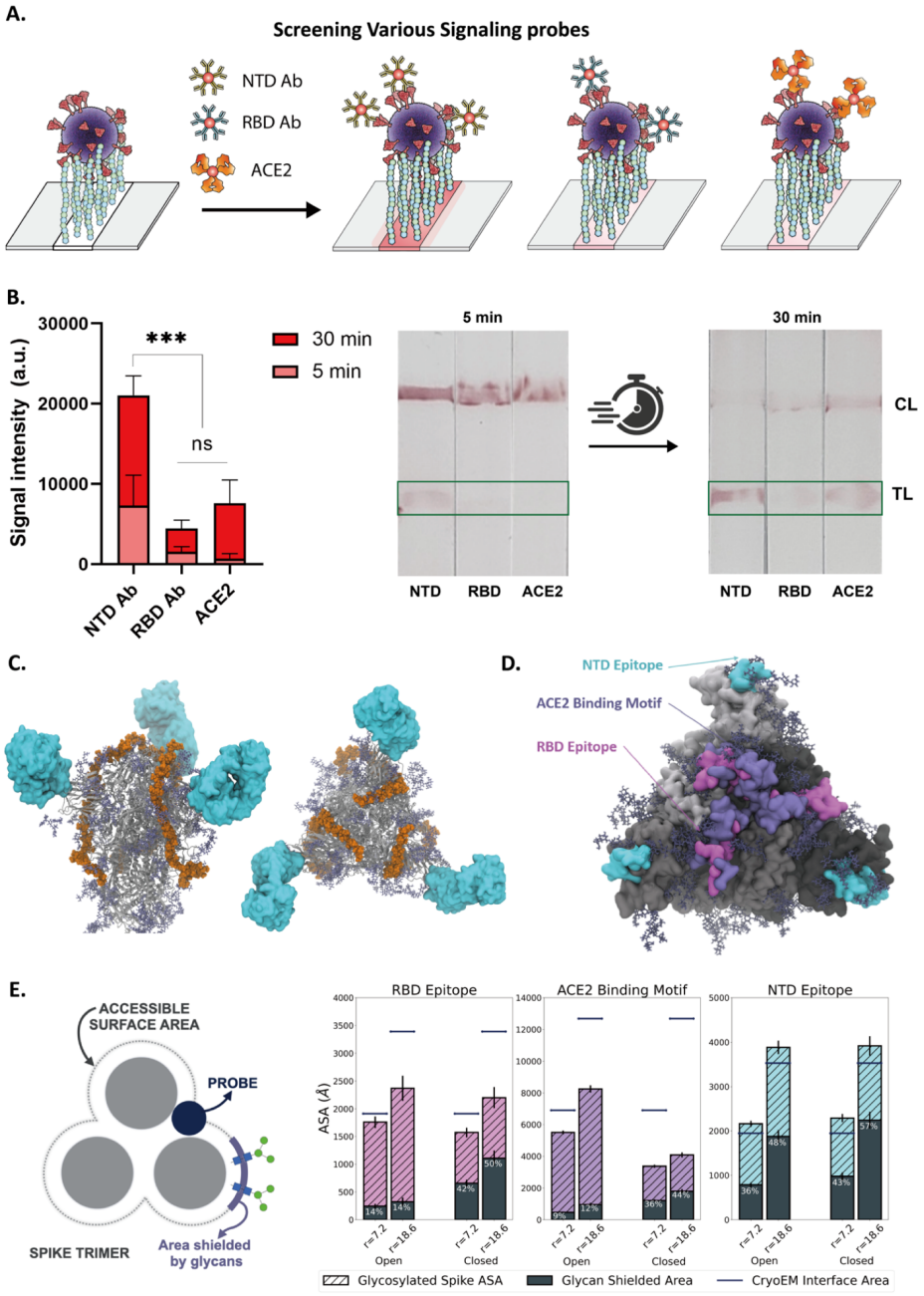


Figure 3.4: Antibody selection analysis for LFSA design

(A) Schematic illustration of our screening of various signaling probe candidates on HEP-based LF strip and (B) corresponding screening result with 30 min and 5 min incubation. (C) Alignment of hep40mer and NTD Ab to the Spike. (D) Computational model of NTD and RBD epitopes (4A8 and REGN10933 respectively) along with the ACE2 binding motif. (E) Accessible surface area calculated from the RBD epitope (REGN10933), ACE2 binding motif, and NTD epitope (4A8). Dark blue bars indicate the size of the interface area as seen in Cryo-EM structures for the RBD, ACE2, and NTD binding footprints (6XDG, 6M17, and 7C2L, respectively). p values < 0.05 (*), 0.01 (**) and 0.001 (***) determined using a two-way ANOVA with Tukey's post hoc test.

To investigate why an NTD binding antibody might be more suitable than an RBD binding antibody for generating strong LFSA signals in our device, we utilized Casalino et al.'s simulations[15] to interrogate the relative accessibilities of epitopes for REGN10933, ACE2, and 4A8 via accessible surface area (ASA) calculations (Figure 3.4D-E). We used the Shrake-Rupley algorithm[74] to calculate ASA for each of these antibody's epitopes (Figure 3.4E) in both the closed and open spike conformations. We also calculated the size of a "reference" interface from cryo-EM structures, to estimate the required exposed surface area for a successful binding event. To calculate the Cryo-EM reference interface sizes, we used the Protein Data Bank structures for each antibody bound to the spike (PDBID's as follows: 6XDG for REGN10933, 6M17 for ACE2, and 7C2L for 4A8). We then removed the antibody binding partner and calculated surface area of the same interface with the Shrake-Rupley algorithm, as done for Casalino et al.'s simulations.

The resulting ASA plots reveal a high degree of protein self-shielding and glycan shielding exhibited by the two RBD based epitopes (REGN10933 and RBM/ACE2), whereas the NTD provides a consistently exposed epitope for antibody binding (Figure 3.4E). The reference values for REGN10933 and ACE2 epitopes are out of range of the simulation-visited surface areas for both closed and open spike conformations, with a more pronounced effect in the closed states. In contrast, the NTD epitope reference value is well within the range of simulation-visited surface areas for all probe radii, regardless of closed/open spike conformations. This analysis indicates that the NTD epitope is easily and consistently accessed by AuNP-NTD antibodies, making it a superior choice for LFSA, both for specificity and accessibility. In contrast, many antigenic regions of the RBD are sequestered within the spike trefoil, and only become accessible after the down to up transition is triggered.[75–77] This is likely due to selective pressure on the spike which has driven sequence and structural changes to hide the spike RBD antigenic regions to potentially limit exposure until close-range interactions can occur between the RBD and ACE2.[27] Conversely, NTD structure and accessibility is not affected by spike conformation and thus there are always 3 NTD epitopes available. Even in the best case for an RBD binding antibody, where a spike has begun moving into an open conformation, there may be only 1 or 2 RBDs in the up conformation. Our analysis of Casalino et al.'s simulations provide a reasonable explanation

for our observed LFSA results. Taken together, these results indicate that an NTD Ab, of the Abs tested, is an ideal partner for spike recognition with GlycoGrip test strips.

Sensor optimization: spike and GAG binding is driven by electrostatics and tuned by hydrogen bonding

Recent work has posited that binding of HEP and other GAGs to the spike is driven by electrostatic interactions between negatively charged GAGs and positively charged patches on the spike surface.[26,27,34,35] Further exploring the nature of spike-GAG binding will allow us to better optimize conditions for sensitive and specific sensing.

Spike glycans tune its surface electrostatics, shielding electrostatically and sterically

Past works that have commented on the electrostatic potential of the spike surface have ignored electrostatic contributions of spike glycans.[27,33] To better elucidate spike's surface electrostatic profile, we have calculated electrostatic potential (ESP) maps for wild type (WT) spike with Adaptive Poisson-Boltzmann Solver (APBS) including and excluding glycan contributions. When considering closed spike surface alone (i.e., without including glycans, Figure S3.14A), we indeed see large positively charged regions. These regions tend to contour interfacial regions between protomers. For example, in the closed spike conformation (Figure S3.14A) there is a large positive region extending from the top of the spike head, between the RBD and neighboring protomer's NTD, down along this interface between protomers, and then toward the furin cleavage site below the NTD. As expected, these positive regions are postulated by Skidmore, Esko, Linhardt, and Wade as primary sites for long-chain GAG interaction with spike, as shown in Figure 3.3B.[23,26,27,35] Interestingly, when glycans are included in ESP maps (Figure S3.14B), the positive regions remain but they are not as pronounced. Glycans are decorated primarily with electron-rich hydroxyl groups and surround positively charged spike surface regions with significant electron density. Our electrostatic maps underscore the need to include glycans at every step of glycoprotein investigation, as glycans may

shield a protein surface sterically or electrostatically. If GAG binding to the spike is dominated by electrostatic interactions, accounting for the electrostatic nature of glycans is important as electron dense glycans may compete with negatively charged GAGs. This pattern of an electron-poor protein surface crowded by electron-dense glycans is also observed in the open spike conformation (Figure 3.5A, S3.15). Without considering glycans, this putative long-range binding motif – starting at the RBD, running between the RBD and neighboring NTD, down along the protomer interface, and finally to the furin-cleavage site – is evident. In fact, with the exposure of the RBD in the open spike conformation, the positively charged surface on the RBD becomes more evident and extends to the top of the up RBD. But as in the closed structure, upon including glycans, one can see these positive protein surfaces are crowded by electron-rich glycans (Figure S3.15B).

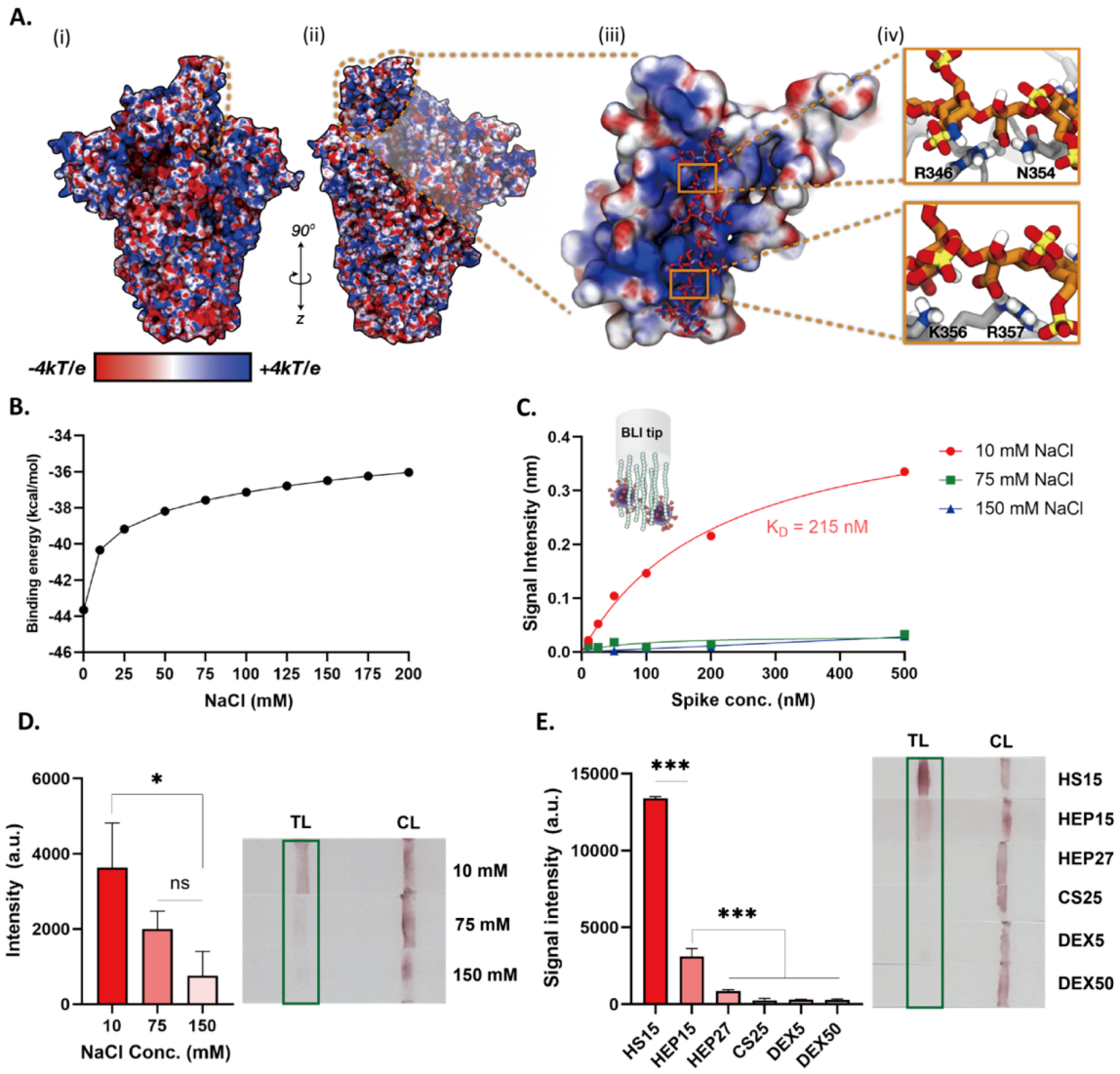


Figure 3.5: Buffer concentration selection for LFSA design

(A(i)) Electrostatic potential map (ESP) of 1up spike without glycans included for simple illustration. ESP is illustrated on a range from -4eV to +4eV. The RBD supersite is highlighted with an orange dashed line. (A(ii)) Rotated viewpoint of electrostatic potential map of 1up spike, again ESP is illustrated on a range from -4eV to +4eV and the RBD HS supersite is highlighted with an orange dashed line. (A(iii)) Close up view of spike RBD (surface representation) and bound HEP octamer (hep8mer, licorice representation). Both spike RBD and hep8mer are colored according to their corresponding electrostatic potentials (ranging -4eV to +4eV). (A(iv)) Close-ups of key interactions between hep8mer (licorice representation, orange carbon atoms) and spike RBD (gray cartoon representation) facilitated by R346, N354, K356, and R357. (B) Computational calculation of binding energy of hep8mer to spike RBD over a range of implicit solvent ionic strengths. (C) BLI results of HEP to spike in three different concentrations of NaCl (10 mM, 75 mM, 150 mM). (D) Response of the lateral flow test in different ionic concentrations (10 mM, 75 mM, 150 mM). (E) Screening results of HS15, HEP15, and HEP27 using LFSA. p values < 0.05 (*), 0.01 (**) and 0.001 (***) determined using a one-way ANOVA with Tukey's post hoc test.

Solution ionic strength modulates GAGs and spike binding

As encounter complex formation and electrostatically driven binding affinities can be modulated by solution ionic strength, we computationally predicted the binding affinity of HEP to spike RBD over a range of ionic concentrations using APBS (see SI Methods).[78–81] These results show a clear trend: binding affinity between HEP and spike RBD decreases with increasing solution ionic strength (Figure 3.5B). BLI results confirm our computational predictions: HEP binding affinity to spike dramatically decreased that K_D values could not be determined under the concentration range of spike tested as ionic strength increased (Figure 3.5C, S3.16). We have also tested the effect of ionic strength on binding affinity between spike and the NTD antibody. The measured binding affinity between NTD Ab and spike slightly decreased (changed from 52 nM to 83 nM) as NaCl concentration decreased (Figure S3.17), suggesting that the Ab binding affinity is much less dominated by electrostatics. Next, we monitored the intensity of LFSA signals under three NaCl concentrations (10 mM, 75 mM, 150 mM). As shown in Figure 3.5D, 10 mM NaCl solution conditions resulted in a 4.7 times more intense signal compared to 150 mM solution conditions ($p < 0.05$). Thus, lower ionic strength yields a higher signal intensity in our LFSA device.

Electrostatics isn't everything: HS binds with higher affinity to spike than HEP.

As mentioned, the prevailing hypothesis in HEP/HS interaction with SARS-CoV-2 spike is that binding is electrostatically driven.[23,25–29,33,35,82] One intriguing result of ours complicates this, otherwise, straightforward spike-GAG electrostatic binding model. Our BLI measurements show that HS15 binds to spike with a greater affinity (16.7 nM K_D) than the highly sulfated/highly charged HEP15 of the same molecular weight: (215 nM K_D). Due to the varied nature of the spike electrostatic surface -- i.e., large, positively charged patches obscured by electron dense or electroneutral glycans -- this likely indicates that, while electrostatics is a major initial driving force of spike-GAG binding, it is not the only driving force. In fact, at close range, hydrogen bonding interactions and appropriate moderation of electrostatics likely allows HS to bind with

higher affinity to spike than highly charged HEP. Additionally, due to the highly charged nature of HEP, the spike-HEP interaction may incur a higher desolvation penalty than the less negatively charged HS.[83–85]

HS and HEP are optimal glycocalyx-inspired LFSA capture agents

Applying our optimized conditions elucidated by computational and experimental investigations thus far -- i.e., use of NTD conjugated AuNP and low ionic strength conditions -- we compared test strips employing HS15, HEP15, and HEP27 as capture probes (Figure 3.5E). HS15 showed the highest signal intensity, in agreement with our BLI and results illustrating HS binds spike with highest affinity (lowest K_D). However, BLI affinity alone could not explain the comparison results of GAGs on the LFSA - i.e., the reason of higher intensity of HS15 compared to HEP15 and HEP27, and low signal intensity from CS25, DEX50, and DEX5. The signal intensity of LFSA can be affected by several factors such as binding affinity of GAGs and antibodies, formation of sandwich-type binding of both bioreceptors, adsorption efficiency of GAGs onto the nitrocellulose membrane. To delve into deeper, firstly, we have conducted ELISA assay using GAGs as primary capture probes. HS15, HEP15, HEP27, and DEX5 showed strong signal intensity than CS25 and DEX50 which indicates that CS25 and DEX50 are not matching pair with NTD Ab for sandwich-type binding (Figure S3.18). This is likely due to their long chain or stereochemical differences such as sulfation pattern or branching of polysaccharides affected the binding of the secondary antibody. Moreover, short chain length DEX5 exhibited lower signal intensity on LFSA than could be expected from BLI and ELISA results. We hypothesize that this may be due to the negatively charged nature of the nitrocellulose membrane repelling negatively charged GAGs during strip preparation.[86] To check this hypothesis, we have immobilized the biotinylated HS15, HEP15, and DEX5 onto the nitrocellulose membrane and flowed the streptavidin coated AuNP. HS15 exhibited strong band on the nitrocellulose membrane while HS15 and DEX5 showed very weak or no bands were observed, respectively (Figure S3.19). Considering that HS15 (approximately 0.8 sulfate per disaccharide) is less negatively charged than HEP15 (approximately 2.3 sulfate per disaccharide) and DEX5 (average 1.9 sulfate per glucosyl residue), this result indicates that adsorption of negatively charged

GAGs on nitrocellulose membrane could be hindered due to the repulsion. This is also might explains the reason why HS15 showed the higher intensity on LFSA than HEP15 since there is difference in the number of GAGs on the surface. Thus, further studies are needed to optimize the adsorption conditions of GAGs onto the nitrocellulose membrane to improve the sensing performance.

Since both HS and HEP showed robust and rigorous LFSA bands, we tested the analytical performance of both. Presence of spike protein was detectable as low as 78 ng/reaction (3.13 µg/ml, 25 µl), and the detectable range was 78 ng/reaction - 1250 ng/reaction (3.13 µg/ml - 50 µg/ml) with the naked eye for both HEP (Figure 3.6A) and HS (Figure S3.20) based LFSA. The limit of the detection (LOD) for HS and HEP were similar (Figure 3.6A, Figure S3.20). Despite our BLI indications that HS may be a better spike binder, SPR measurement indicates that binding affinity of HS and HEP were similar (Figure S3.21), suggesting that the LOD is most likely dictated by the NTD binding affinity. To determine the efficiency of NTD Ab binding compared between HS and HEP, we used immunofluorescence to estimate the proportion of NTD Ab bound to captured spike proteins on HEP and HS surfaces (Figure S3.22, S3.23). There are more spike protein bindings to the heparan sulfate compare to the heparin, however, the NTD binding was not significantly different between HS and HEP which indicates that the LOD was set by the NTD binding. Since HEP LFSA signals provide a larger range of observably intense signals to the naked eye, and it is more cost effective than HS, which is certainly important when looking to mass produce testing kits for viral outbreaks. Therefore, we continued to optimize our GlycoGrip biosensor with HEP as the surface-anchored GAG.

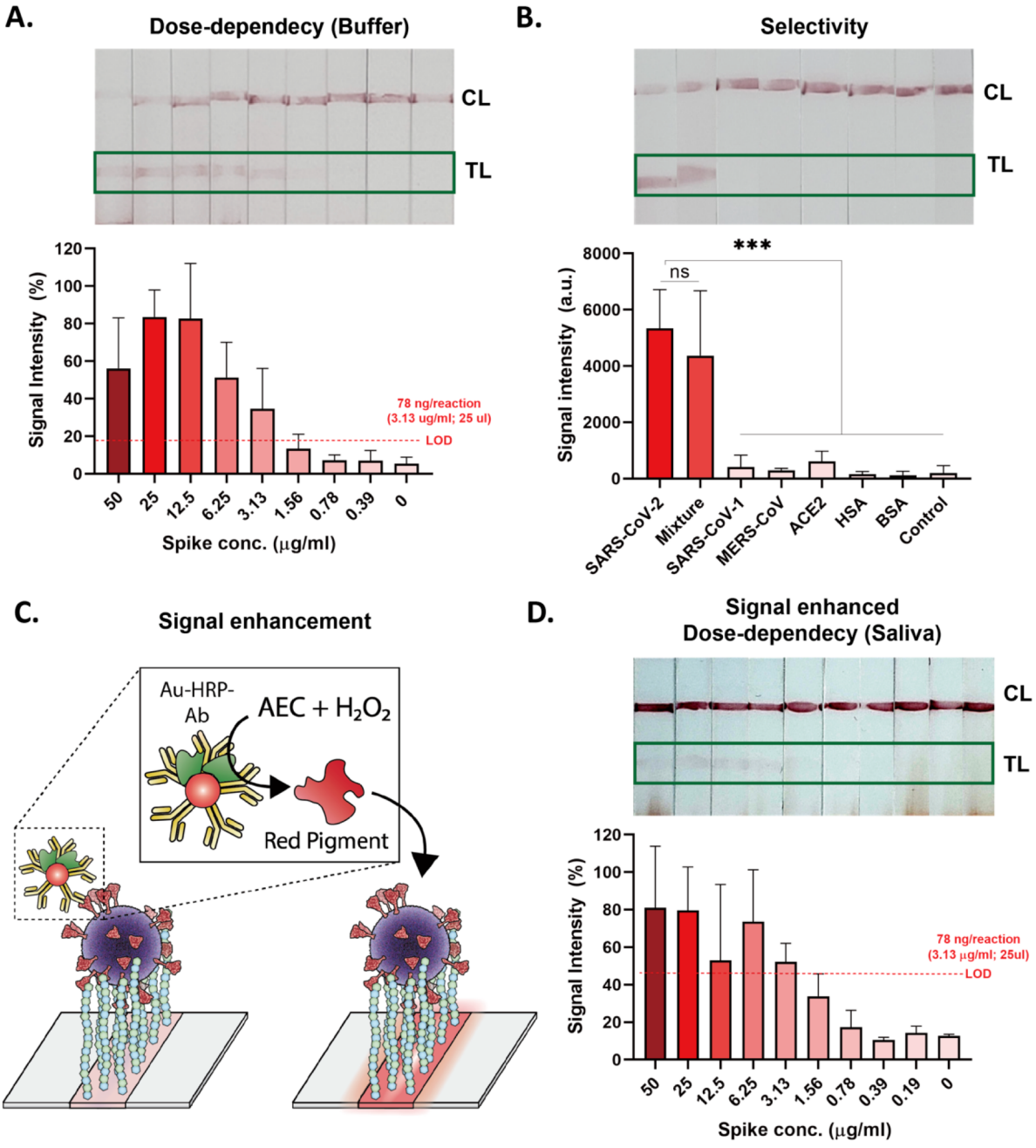


Figure 3.6: LFSA performance summary

(A) Analytical performance of HEP15 based GlycoGrip LF biosensor in buffer conditions. (B) Selectivity of HEP15 based GlycoGrip LF biosensor was tested with different counter targets: SARS-CoV, MERS-CoV, ACE2, human serum albumin (HSA), and bovine serum albumin (BSA). (C) Schematic illustration of the signal enhancement using HRP and AEC. (D) Analytical performance of the signal enhanced GlycoGrip LF biosensor in human saliva condition. p values < 0.05 (*), 0.01 (**) and 0.001 (***) determined using a one-way ANOVA with Tukey's post hoc test.

GlycoGrip is a rapid, sensitive, stable, and selective assay for the detection of SARS-CoV-2

We tested the selectivity of our GlycoGrip biosensor against related beta-coronavirus spike glycoproteins (SARS-CoV spike and MERS-CoV spike S1 domain), as well as biologically relevant proteins likely to be found in patient samples (ACE2, Bovine Serum Albumin, BSA, and Human Serum Albumin, HSA). As shown in Figure 3.6B, positive bands were observed only when GlycoGrip LF strips were treated with SARS-CoV-2 spike, whereas treatment with other beta-coronavirus spike proteins and biologically relevant "distractors" did not indicate positive test results (Video S1). Moreover, when tested with a mixture of SARS-CoV-2, SARS-CoV, MERS-CoV spike glycoproteins, band intensity was similar to the pure SARS-CoV-2 spike band. These results suggest that our GlycoGrip LF biosensor can selectively detect the SARS-CoV-2 in more complex solutions, potentially minimizing the possibility of undesirable false positive test results.

To further enhance the sensitivity of our GlycoGrip LF biosensor, we incorporated a reporter system (horseradish peroxidase (HRP) and 3-amino-9-ethylcarbazole (AEC)) (Figure 3.6C). NTD Ab tethered with multiple HRP enzymes were used as a signaling probe to enhance the sensitivity by catalytic reaction. Through the enzymatic reaction, water-insoluble red colored chromogenic products were released on the test line which enhances the signal intensity. The LOD with this enzymatic signal enhancement mechanism was estimated to be 19.5 ng/reaction (0.78 μg/ml, 25 μl), enhanced 4-fold compared with unamplified results (Figure S3.24).

GlycoGrip detects spike in human saliva samples

GlycoGrip can serve as a rapid test whereby samples can be self-collected from one's own saliva in a simple, non-invasive fashion without the need for specialized equipment or personnel. The benefits of this are two-fold: (1) LFSA tests have the potential to reach a wider testing population, and (2) removing the specialized personnel requirement reduces extra cost and eliminates direct contact between infected and non-infected persons. Moreover, collecting saliva samples as opposed to nasal samples has a higher likelihood of indicating positively for both symptomatic and asymptomatic SARS-CoV-2 carriers, as nasal

collection has shown high variability in sample integrity due to sampling procedure differences on an individual basis.[87,88] To test GlycoGrip performance in human saliva samples, we introduced a range of spike concentrations into human saliva samples. Sensing proteins in complex fluid such as saliva can be challenging as it contains many other biomolecules that could limit or compete with spike-GAG binding interactions. In the context of sensing SARS-CoV-2, it has been recently reported that glycoproteins in saliva, such as mucin proteins (MUC7, MUC5B) and neutrophil defensins, may bind to and interact with the spike.[89] Thus, these complex glycoproteins may then compete with HEP or Ab for binding sites on the spike surface.[89] Remarkably, the LOD of our GlycoGrip in saliva samples was 78 ng/reaction (3.13 μg/ml, 25 μl) comparable to buffer conditions (LOD: 19.5 ng/reaction) (Figure 3.6D, see Table S9, Table S10 for summary of GlycoGrip LF results and comparison with reported spike protein detection LF). These results indicate the strong feasibility of applying GlycoGrip technology to clinical samples.

In addition to sensitivity and selectivity, sensor stability is a vital factor for sensor distribution and storage. To test GlycoGrip's stability, we stored same-day-fabricated GlycoGrip sensor strips in a plastic bag with desiccant at room temperature and tested signal intensity after varying lengths of storage (0, 11, and 47 days). We saw no significant decrease in the signal intensity over 47 days, which indicates that our GlycoGrip LF biosensor is stable for at least 47 days, and most likely much longer (Figure S3.25).

GlycoGrip detects the presence of SARS-CoV-2 spike variants

New strains of SARS-CoV-2 began emerging as early as summer of 2020 and, at the time of this publication, the World Health Organization has highlighted four variants of concern (Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), and Delta (B.1.617.2)) and four variants of interest (Eta (B.1.525), Iota (B.1.526), Kappa (B.1.617.1), and Lambda (C.37)).[90,91] These variants exhibit key mutations in the spike protein (as well as mutations in other SARS-CoV-2 structural and non-structural proteins) that are postulated to directly translate to increased infectivity and/or increased immune-system evasion ability. To ensure our GlycoGrip

technology would signal positively for patients infected with these emerging SARS-CoV-2 strains, we have tested variant full spikes with our LFSA technology.

Characteristic mutations of Alpha, Beta, Delta spike variants do not interfere with GAG binding sites

We hoped to assess at the molecular scale whether variations in spike sequence and structure, as seen in emerging SARS-CoV-2 variants, could impact binding to HEP on the LFSA test strip. While, at the time of this publication, there are several cryo-EM structures of Alpha, Beta, and Delta SARS-CoV-2 spikes in the Protein Data Bank,[92,93] unfortunately, each of these structures contain unresolved regions corresponding to highly flexible loops. Thus, we constructed computational models of the Alpha, Beta, and Delta spike variants from our refined closed WT spike structure. We then aligned our trivalent spike-hep40mer complex to these variant structures and visually inspected for overlap between mutation points with proposed hep40mer binding regions (Figures 3.7A, S3.26). From these structures, we observed no overlap or clash between HEP and mutations characterizing Alpha and Beta spike variants. This suggests that GAGs will still likely capture Alpha and Beta spikes under LFSA conditions, as we have proposed in the WT case. In the case of the Delta spike, interestingly, we observed there are several mutations (L452R, T478K, and P681R) that increase the number of positively charged residues along the posited HEP binding cleft. Furthermore, WT spike protein and its glycans constitute an overall absolute charge of –32, while the overall charge of the Delta spike protein, with the same glycoprofile, is –17. Thus, the Delta spike exhibits a drastic increase in total charge as a result of changes in sequence.

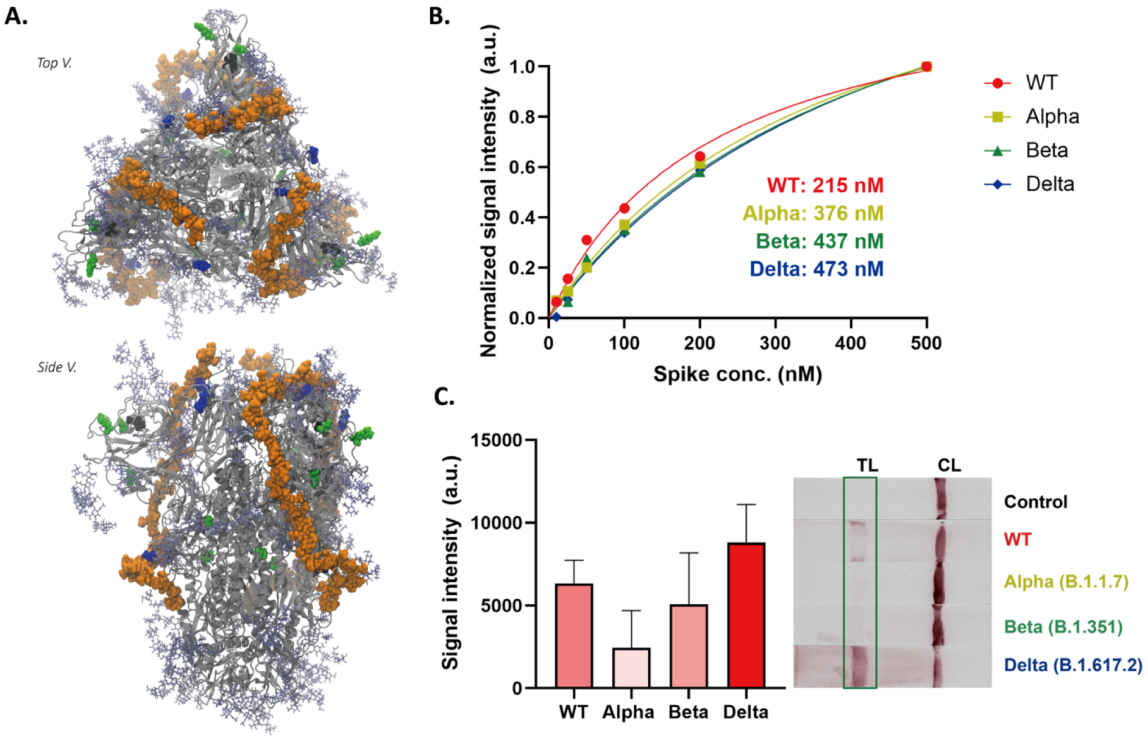


Figure 3.7: GlycoGrip LFSA performance with SARS-CoV-2 emerging variants

(A) Trivalent hep40mer aligned to Delta spike variant. Light green van der Waals spheres represent single point mutation sites, black van der Waals spheres represent locations of NTD deletion sites, blue van der Waals spheres represent mutations along the hep40mer binding site L452K, T478R, P681R (B) BLI results of the HEP to spike proteins (Wild type, Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2). (C) Response of the Alpha, Beta, and Delta variants in GlycoGrip LF biosensor. Statistical analysis was performed using a one-way ANOVA with Tukey's post hoc test.

As discussed, past work has highlighted the importance of large positively charged regions on the spike surface.[26,27,35] To ensure mutation points in spike variants do not disrupt these positive patches, we have also calculated electrostatic potential maps for Alpha, Beta, and Delta spikes. While there are some differences in surface electrostatic potential compared to WT (Figures S3.14, S3.15 versus Figures S3.27, S3.28, S3.29), noticeably in the trefoil region between neighboring RBDs, our postulated hep40mer binding region remains largely positively charged at the protein surface for Alpha, Beta, and Delta spike variants. This again supports that binding of HEP to spike could be electrostatically driven, and these electrostatic interactions are not affected by the key mutations in new spike variants.[94] One could further hypothesize that an evolutionary advantage exists to maintain the spike's ability to bind glycocalyx polymers. This would be supported by the fact that in Alpha and Beta spike variants there exist no single point mutations along the putative HEP

binding cleft. Additionally, the Delta spike variant exhibits three mutations to positive residues along the putative HEP binding cleft (L452R, T478K, P681R) which would further potentiate the ability of spike to bind to negatively charged GAGs. Finally, although we have predicted a long-chain model for HEP binding on the spike surface based on docking studies with the WT, we note that our results indicate that there are at least 14 sites on the WT spike surface where GAGs can bind. Thus, if new spike variants emerge with mutations that interfere with our proposed long-chain HEP binding site, there is still potential for GAGs binding to spike via other long-chain modes according to our proposed multi-site binding model.

Biolayer interferometry confirms Alpha, Beta, Delta spike variants bind to HEP

To confirm this in silico prediction, i.e., that GlycoGrip can bind and signal positively for emerging SARS-CoV-2 spike variants, we measured binding affinity of HEP15 to full-length trimeric Alpha, Beta, and Delta spikes with BLI (Figure 3.7B, S3.30). Alpha, Beta, and Delta spikes all bound to HEP15 with comparable binding affinity to WT spike, despite their characteristic point mutations and deletions. This result supports both our posited hep40mer binding mode and our electrostatic potential maps: HEP binding is not perturbed by mutations exhibited in the SARS-CoV-2 Alpha, Beta, and Delta strains. This result underscores the power of using GAGs as the capture probe for SARS-CoV-2 spike sensing. To complete the profile of our sandwich-style LFSA detector, we measured the binding affinity between NTD Ab and each spike variant. Binding affinities of variant spikes to the NTD Ab were decreased compared to the WT spike (Figure S3.31), however, this is to be expected as each variant exhibits characteristic mutations in the NTD. From all variants, the alpha variant showed the lowest binding affinity, most likely due to two key deletions in the NTD which are characteristic of the Alpha variant, one of which is within the N2 loop and one within the N3 loop, with both loops being key for antibody recognition. However, these results reflect another key feature of GlycoGrip: modularization. Since the choice of capture probe and signaling antibody are decoupled in GlycoGrip's design, selecting a new signaling antibody largely does not impact the performance of GAGs to capture analytes.

Alpha, Beta, Delta spike variant detection on GlycoGrip strips

To confirm the sandwich-type binding of HEP15 and NTD Ab to variant strains of SARS-CoV-2 spike via an orthogonal methodology, we performed ELISA using full-length trimeric Alpha, Beta, and Delta spikes (Figure S3.32). Alpha, Beta, and Delta spikes all bound to HEP15 and NTD Abs in a sandwich-type complex with the highest signal intensity shown for the Delta variant spikes. Encouraged by our BLI and ELISA results, we then tested Alpha, Beta, and Delta spikes with our GlycoGrip LF test strip. As shown in Figure 3.7C, all variants exhibited positive bands on the LF test line corresponding to the ELISA, and the currently circulating highly infectious Delta variant, can be detected, demonstrating the universality of using glycopolymers as viral capture agents (Video S2). Thus, our GlycoGrip LF could be rapidly adaptable to newly emerging SARS-CoV-2 variants which is an important aspect of point-of-care sensing platforms for viral pathogen detection.

3.4 **Conclusions**

We have harnessed the power of the glycocalyx and its glycosaminoglycans to serve as capturing agents within an LFSA “sandwich” binding assay to develop our GlycoGrip sensor. Our rigorously applied lock-step integration of experiments and computational modeling allowed us to achieve optimized conditions for GlycoGrip and provide mechanistic insights into GAG and spike binding interactions. We have demonstrated the first use of GlycoGrip for detecting wild type SARS-CoV-2 spike as well as the newly emerged SARS-CoV-2 spike variants, Alpha, Beta, and Delta. We have shown that SARS-CoV-2 spikes can be detected on GlycoGrip LF strips and, due to specificity of the chosen NTD-based signaling antibody, we saw no cross-reactivity to SARS-CoV, MERS-CoV, HSA, or ACE2 in buffer or in human saliva. We have also seen that optimizing solution ionic strength and GAG length can enhance LFSA signals, along with traditional signal enhancement systems.

In addition to sensor design, we used our extensive ensemble-based docking results to provide biologically relevant mechanistic insights into SARS-CoV-2 host cell invasion mediated by spike-GAG binding. We have confirmed literature proposed sites as well as provided identified seven novel GAG binding sites on the spike surface. Collectively, a clear picture emerges: GAGs in the glycocalyx bind tightly to spike, at multiple sites, and with potential for mutli-valent long-chain GAG binding. Additionally, our work highlights, once more, the importance of modeling glycans when studying spike dynamics and interactions. We predict spike glycans may play a role in shielding the spike surface from incoming GAGs, but once GAGs reach the surface, glycans are likely to support GAGs via hydrogen bonding interactions.

The power of GlycoGrip lies in its modularity and generalizability. Many pathogens -- including viruses, bacteria, and parasites -- exploit the glycocalyx for cell adhesion. Thus, these pathogens, and/or their characteristic antigens, have the potential to be captured by GAGs on a GlycoGrip strip. To achieve a selective detection, one simply needs to optimize an appropriate signaling antibody that will pair with GAGs. The need for one antibody instead of two (as required for constructing traditional sandwich-type LFSA sensors) will significantly shorten the screening time[47,48] when applied towards a new pathogen or variant, as well as drastically lowering the cost, potentially 10-fold, as compared to current LFSA technologies.[95,96]

GlycoGrip has remarkable promise as a widespread tool for capturing and detecting current and emerging viruses. While SARS-CoV-2 was of preeminent concern at the time of writing this manuscript, GlycoGrip can be easily extended for rapid screening and detection of future pathogenic infections. Loss of life prevention in public health crises requires quick detection and disease containment. We specifically note that traditionally medically underserved, and therefore undertested, populations have the hardest time identifying communal outbreaks because they lack access to RT-PCR, a technique that requires highly skilled laboratory staff and expensive instrumentation. In the case of SARS-CoV-2 spike, we have shown that GlycoGrip can detect rapidly emerging variants. This not only speaks to the generalizability of GlycoGrip, but also to its robust longevity over the course of a real-time sustained global health crisis. One could envision GlycoGrip as a synthetic glycocalyx able to trap pathogenic antigens, and coupled with antibodies, used to test patients within minutes.

In summary, we have retooled the glycocalyx, an essential component of the host cell surface, into a rapid and sensitive biosensor for viral antigens. GlycoGrip is a novel, biologically inspired, generalizable assay that has the potential to be inexpensive to manufacture, easy to distribute, and simple to operate.

3.5 **Materials and Methods**

3.5.1 Computational Methods

3.5.1.1 SARS-CoV-2 spike and sulfated polysaccharide docking.

In this work we modeled heparin (HEP, specifically dimeric LIdoA2S-α(1-4)-DGlcNS6S-αOH and tetrameric LIdoA2S-α(1-4)-DGlcNS6S- α(1-4)-LIdoA2S-α(1-4)-DGlcNS6S-αOH), 6-O sulfated heparan sulfate (H6S, specifically dimeric LIdoA-α(1-4)-DGlcNAc6S-αOH and tetrameric LIdoA-α(1-4)-DGlcNAc6S-α(1-4)-LIdoA-α(1-4)-DGlcNAc6S-αOH), chondroitin sulfate (CS, specifically dimeric DGlcA2S-β(1-3)-DGalNAc4S6S-βOH and tetrameric DGlcA2S-β(1-3)-DGalNAc4S6S-β(1-3)-DGlcA2S-β(1-3)-DGalNAc4S6S-βOH), dextran sulfate, (DEX, specifically dimeric DGlc-α(1-6)-DGlc2S4S-αOH and tetrameric DGlc-α(1-6)-DGlc2S4S-α(1-6)-DGlc-α(1-6)-DGlc2S4S-αOH) as potential binding partners for SARS-CoV-2 spike. H6S was chosen as a defined sequence for modeling and docking as it captures the 5-7% sulfation rate reported by Sigma Aldrich while also capturing potentially important interactions facilitated by the 6-O sulfation position reported by several groups.[23,26,27,29,55,58] As docking of long chain polysaccharides to large protein structures is combinatorically intractable, we modeled small dimer and tetramer structures of each sulfated polysaccharide. Dimeric sulfated polysaccharides were modeled with the intention of capturing highly localized interactions, while tetrameric sulfated polysaccharides were modeled with the intention of capturing steric hindrance effects encountered with larger substrates. HEP and H6S dimeric and tetrameric structures were constructed with MatrixDB,[59–62] CON and DEX dimeric and tetrameric structures were built with CHARMM-GUI Glycan Builder.[63–66]

To predict potential locations of sulfated polysaccharide binding to spike, we conducted extensive unbiased docking with AutoDock Vina.[97,98] Using exposure of the furin cleavage site as a metric to detect conformational changes, we selected four spike coordinates from Casalino et al. closed spike trajectories (https://amarolab.ucsd.edu/covid19.php). These spike structures were prepared as described by Casalino et al.[15]

Two structures must be well defined in any docking protocol: the receptor/macromolecule to be docked into and the ligands/small molecule to be docked. To avoid biasing docking results to only certain regions of the spike, for each of the four spike structures, we defined the "receptor" to be any location on/within the trimeric spike head (residues 13 to 1140 of chains A, B, and C). To define these receptors in AutoDock Vina, we generated grids centered on the trimeric spike head, with large enough dimensions to encompass the entire head, and with default grid spacings. To characterize the structural diversity of each of these protein structures we have calculated the root mean square deviation (RMSD) between chains for each spike sub-domain of interest (Supporting Information Results, Table S1). We prepared all molecules (dimers and tetramers of HEP, H6S, CS, and DEX) with AutoDockTools, and all polysaccharide torsions were treated flexibly with AutoDock Vina conformational sampling and scoring function. For complete details of receptor grid coordinates, grid values, and all other docking settings, see input scripts included with shared docking output files in the Supporting Information.

To identify as many binding sites as possible, we conducted 20 replicates of each docking procedure and requested AutoDock return 20 predicted binding modes per docking replicate -- where one docking procedure would entail docking one dimer or tetramer sulfated polysaccharide to each spike conformation. As a result, we predicted 400 binding modes for each dimeric or tetrameric sulfated polysaccharide in each spike conformation, resulting in 12,800 predicted binding modes total: 400 binding modes per molecule, by 8 molecules (dimeric and tetrameric versions of each sulfated polysaccharide), by 4 protein conformations, is 12,800 total. To parse all 12,800 predicted binding modes into discernable binding "sites", we clustered all these resultant poses by their centers of mass using k-means clustering through python's scikitlearn. A knee/elbow locator algorithm was used to identify the optimal number of clusters.[99] We then derived a

binding site “importance” metric to rank binding sites according to average binding score and relative population in that site. The top “important” binding sites were then inspected by eye through Visual Molecular Dynamics (VMD)[100] to determine important binding factors governing each of these sites.

3.5.1.2 SARS-CoV-2 spike and hep40mer model system.

The fully glycosylated SARS-CoV-2 spike model used in our hep40mer modeling is based on an experimental cryo-EM structure of the spike in the closed state where all RBDs are in the down conformation (PDB: 6VXX).[101] To improve the accuracy of our model, fully resolved RBD and NTD loops were incorporated from another closed spike structure (PDB: 7JJI).[102] We note that utilization of the 7JJI structure in its entirety was not ideal as this structure is known to be more compact than 6VXX due to the presence of a fatty acid ligand resolved in the RBD.[102] The complete glycoprofile was replicated from Casalino et al.[15] Protonation state assignment was performed for spike glycoprotein with complete glycans modeled with stand-alone PropKa (so that glycan atoms could be considered during calculation), [103,104] histidine protonation states were assigned via PropKa through Schrodinger’s Protein Preparation Wizard.[103–105] Protonation states of all titratable residues were then compared to those assigned in Casalino et al.[15] for consistency. To propose a long range, HEP binding site along the spike surface, we considered both literature-proposed binding sites, as well as proposed binding sites from our own docking simulations. We considered only surface sites, and the most highly ranked sites were prioritized for inclusion in long range binding mode construction. To generate relaxed HEP conformations for building hep40mer, we conducted 6 replicates of 50 ns of NVT equilibrium molecular dynamics simulations of hep8mer in a water box with NAMD.[106,107] From the resultant 300 ns of HEP simulation, we clustered those frames according to conformation. 95% of all hep8mer coordinates from these simulations could be clustered according to six conformational clusters. Hep8mer coordinates representing the frame closest to each cluster center were used as hep8mer units to fill necessary coordinates between docked poses predicted from AutoDock Vina. Molefacture, a VMD based modeling tool, was used to ensure there were no clashes between protein atoms and ligand atoms.

3.5.1.3 Accessible surface area (ASA) analysis.

ASA analysis was done using the measure sasa command built-in to VMD[100] along with extra protocol established by Casalino et al.[15] The ASA analyses were performed by considering the antigenic regions in the NTD (residues 143-153, 245-259) and the RBD (residues 403-406, 416-422, 453-456, 473-478, 484-498). Additionally, ASA analysis was performed on the canonical RBM/ACE2 binding site (residues 437-508). Calculated ASAs are shown for two probe radii: 7.2 Å and 18.6 Å.[108] The reference interface areas were calculated from cryo-EM structures as follows: REGN10933 antibody bound to spike-RBD (PDB: 6XDG[70]), 4A8 NTD antibody bound to spike-RBD (PDB: 7C2L[73]), ACE2 bound to spike-RBD (PDB: 6M17[71]).

3.5.1.4 System construction for Ionic Concentration Effect Monitoring:

To investigate the effect of ionic concentration on HEP binding affinity, a HEP octamer (hep8mer) was docked to the RBD of an RBD/ACE2 complex (PDB: 6M17) using Schrodinger's Induced Fit Docking protocol.[109–111] This cryo-EM structure was prepared by removing the B^0AT1 dimer chaperone coordinates manually with VMD,[100] and N-glycans were added on the ACE2 and RBD as done in Barros et al. [20] The ACE2/RBD/hep8mer construct was inserted into a lipid bilayer patch of 225 Å x 225 Å with a composition similar to that of mammalian cell membranes (56% POPC, 20% CHL, 11% POPI, 9% POPE, and 4% PSM).[112] The resulting system was then embedded into an orthorhombic box of explicit TIP3P waters.[113] The system was ionized with Na+/Cl- ions at 150mM for all simulations, unless otherwise specified. All-atom MD simulations were performed on the Frontera supercomputer at the Texas Advanced Supercomputing Center (TACC) using NAMD 2.14[106,107] and CHARMM36m all-atom additive force fields.[65,114,115] Minimization and equilibration were performed in four steps. In the first step, while keeping all the coordinates fixed but the lipid tails, the system was subjected to an initial minimization of 10,000 steps using the conjugate gradient energy approach, followed by an NVT equilibration of 0.5 ns at 1 fs/step, where the temperature was gradually increased from 10 to 310 K. In the second step, positional constraints on lipids head, water and ions were lifted, and the system was NPT-equilibrated for 0.5 ns at 1.01325 bar and 310K with the protein, glycans,

and hep8mer harmonically restrained using a spring constant of 1 kcal/mol/Å^2. Then, the restraints on protein and glycan atoms were removed and the equilibration was extended by 10 ns. Next, restraints on hep8mer atoms were removed to allow the entire system to equilibrate for an additional 10 ns. Finally, MD simulation production runs were performed and 3 replicas of ~500 ns each were collected (Figure S3.33).

APBS was used to estimate binding affinity between hep8mer and spike RBD at varying ionic concentrations. Binding affinity was calculated according to an appropriate thermodynamic cycle by calculating binding energy in a homogeneous reference medium (dielectric constant = 4), and then by calculating the solvation free energy difference between the homogenous reference state and non-homogenous target state (dielectric constant = 78). (Ref https://apbs.readthedocs.io/en/latest/using/examples/binding-energies.html and the Supporting Information for a complete listing of all APBS options used in this work). Binding energies were calculated for RBD-hep8mer complexes in the following NaCl concentrations: 0.0M, 0.01M, 0.025M, 0.05 M, 0.075M, 0.10M, 0.125M, 0.150M, 0.175M, 0.200M.

3.5.2 Experimental Methods

3.5.2.1 Materials

The same source for each of the following GAGs is used throughout our set of experiments including BLI, ELISA, and LF assays. HEP15 (B9806, porcine mucosa), HS15(H7640, bovine kidney) are purchased from sigma-aldrich. CS25 (CS-Biotin-25k, porcine cartilage), DEX5 (DES-Biotin-5k), and DEX50 (DES-Biotin-50k) were purchased from HAWORKS. HEP27 (HP-207, porcine mucosa) was purchased from Creative PEGWorks. Human Serum Albumin (A3782), Sucrose (S0389), AEC staining kit (AEC101), streptavidin coated gold nanoparticles (53134) were purchased from Sigma-Aldrich. Biotin-PEG3-amine (BG-17) was purchased from G-Biosciences. Tween 20 (J20605-AP) was purchased from thermo scientific. Sodium chloride (BDH9286) was purchased from VWR. Bovine serum albumin (105033) was purchased from MP biomedicals. Gold nanoparticles (15703-20) were purchased from Ted Pella Inc. N-terminal domain binding

antibody (LT-2000) and HRP modified N-terminal domain binding antibody (LT2010) were purchased from Leinco Technologies. Receptor domain binding (RBD) antibody (Clone REGEN10933; CPC511B) was purchased from cell sciences. Rabbit Anti-Human IgG (ab6715) and Goat Anti-Mouse IgG (ab6708) were purchased from abcam. Mouse Anti-6x-His Tag Monoclonal Antibody Alexa Fluor 488 (MA1-21315A488) and Goat Anti-Human Alexa 594 Antibody (A-11014) were purchased from ThermoFisher Scientific. Human saliva pooled from human donors (991-05-P) were purchased from LEE Biosolutions. Nitrocellulose membrane (FF120HP), sample pad (Whatman CF4 dipstick pad), and absorbent pad (Whatman standard 17) were purchased from Cytiva. SARS-CoV-2 Spike protein (40589-V08H4), SARS-CoV S1 (40150-V05H1), MERS-CoV S1 (40069-V08H), HRP modified anti-human antibody (10702-T16-H) were purchased from Sino biological. SARS-CoV-2 Alpha (B.1.1.7) spike (10796-CV-100), SARS-CoV-2 Beta (B.1.351) spike (10786-CV-100), and SARS-CoV-2 Delta (B.1.617.2) spike (10878-CV-100) were purchased from R&D systems. Fc tagged human ACE2 (AC2-H5257) were purchased from Acro Biosystems. Streptavidin modified BLI biosensor tips (18-5019), and anti-human IgG Fc Capture (AHC) BLI biosensor tips (18-5060) were purchased from Sartorious. Immu-Mount (9990402) was purchased through Fisher Scientific.

3.5.2.2 Biolayer interferometry

To measure the binding affinities of polysaccharides, biolayer interferometry (BLI) was used. Polysaccharide modified tips were prepared by the streptavidin-biotin methods. Streptavidin tips were functionalized with 1 mg/ml of biotin-polysaccharides (40 µl) in a kinetic buffer (10 mM HEPES, 10 mM NaCl, 0.05% Tween 20, pH 7.4) for 180 sec. Polysaccharides modified tips were incubated with various concentrations of spike proteins from 0 to 500 nM in a kinetic buffer for 400 sec. Then, dissociation was measured for 500 sec. Dissociation constants (K_D) were analyzed with steady-state analysis using the HT 11.1 software provided with instruments. In case of NTD antibody (NTD Ab), anti-human IgG Fc capture (AHC) tips were functionalized with 5 µg/ml of NTD Ab in a kinetic buffer, and the same measurement procedure was applied. For comparison study of salt effect, kinetic buffers containing different NaCl concentrations (75 mM, 150 mM) were used.

3.5.2.3 Preparation of streptavidin modified polysaccharides

To immobilize the polysaccharides into nitrocellulose membrane, polysaccharides were conjugated to streptavidin by biotin-streptavidin interaction. Biotin modified polysaccharides were conjugated to streptavidin (1 mg/ml) with molar ratio of 4:1 (polysaccharides: streptavidin). After incubation for 1hr at room temperature, the mixture solutions were purified to remove excess polysaccharides by using the amicon filter (30k, 50k, 100k) depending on the size of the polysaccharides.

3.5.2.4 Preparation of antibody modified gold nanoparticles

For naked-eye detection, antibodies were conjugated to gold nanoparticles (AuNP) as a signaling probe. To prepare antibody-AuNP conjugates, NTD antibody (5 µl of 1 mg/ml), RBD antibody (5 µl of 1 mg/ml), ACE2 (10 µl of 0.63 mg/ml) were each added to 1 ml of AuNP (10 nm) with 0.1 ml of borate buffer (0.1 M, pH 8.5). After 1 hr incubation at room temperature, BSA (100 µl of 10 mg/ml) was introduced and incubated for 30 min to reduce the non-specific adsorption by blocking the surface of the gold nanoparticles. Then, the mixture solution was centrifuged at 22,000g and 4ºC for 45 min. Supernatant was removed and AuNP solution was resuspended in 1 ml of BSA (1 mg/ml). Centrifugation and suspension process was repeated twice. Finally, antibody-AuNP conjugate was stored in the storage buffer (10 mM HEPES, 10 mM NaCl, 1 mg/ml BSA, pH 7.4) at 4ºC.

For signal enhancement testing, gold nanoparticles were modified with horseradish peroxidase (HRP) conjugated NTD ab (NTD Ab-HRP). To prepare the gold nanoparticle modified with NTD-HRP (NTD-HRP-AuNP), 10 µl of NTD Ab-HRP (0.5 mg/ml) was added to the 1 ml of AuNP (10 nm) with 0.1 ml of borate buffer (0.1 M, pH 8.5). Then, the same procedure was utilized to prepare the NTD-HRP-AuNP.

4.2.5 Preparation of polysaccharides based lateral flow strip biosensor

Figure 3.1 shows the general design of the polysaccharides based lateral flow strip biosensor. Polysaccharides conjugated with streptavidin (1 mg/ml) and rabbit anti-human IgG (1 mg/ml) were dispensed on the nitrocellulose membrane (FF120HP.). Dispensed nitrocellulose membrane was dried at 65ºC for 3 min. After drying, nitrocellulose membrane was blocked with a blocking buffer (1% BSA, 0.05% Tween-20 in 10

mM HEPES, 10 mM NaCl, pH 7.4). Finally, the sample pad (Whatman CF4 dipstick pad) and the absorbent pad (Whatman standard 17) were assembled onto the nitrocellulose membrane. Assembled strips were stored at room temperature with desiccant before use.

3.5.2.5 Screening optimal epitope and buffer using lateral flow assay

To screen the optimal antibody for the sandwich-type detection of spike protein that will work along with polysaccharides, two antibodies which bind to different epitopes of spike protein (i.e., N-terminal domain binding antibody (NTD Ab) and receptor binding domain binding antibody (RBD Ab)), and ACE2 receptor were tested. The lateral flow strip for screening optimal epitope was prepared as described in the previous section. Dipstick method was used for testing all the lateral flow strips using 96 well plates. For the comparison study, 625 ng of SARS-CoV- 2 spike was incubated with each signaling probe (20 nM) in the kinetic buffer (10 mM HEPES, 10 mM NaCl, 0.05% Tween 20, pH 7.4) for 5 min or 30 min at room temperature. Mixture solutions were loaded to the 96 well plate and prepared lateral strips were dipped for 20 min. After 20 min, red signals were observed by the naked eye and smartphone camera. Signals were quantitatively analyzed by imageJ software.

To test the effect of the NaCl on lateral flow assay, lateral flow strips, signaling probes, and SARS-CoV-2 spike were prepared by using HEPES buffers containing different concentrations of NaCl (10 mM, 75 mM, 150 mM). 25 µl of SARS-CoV-2 spike (25 µg/ml) and 25 ul of signaling probes (20 nM) prepared in HEPES buffers containing different concentrations of NaCl (10 mM, 75 mM, 150 mM) were incubated for 5 min. Then, the previously described dipstick method was used for the lateral flow assay.

3.5.2.6 Selectivity and sensitivity analysis

For the selectivity test, 25 ul of each proteins SARS-CoV spike (50 µg/ml), MERS-CoV spike (50 µg/ml), ACE2 (50 µg/ml), human serum albumin (50 mg/ml), and bovine serum albumin (50 mg/ml), and the mixture of SARS-CoV spike (25 µg/ml), MERS-CoV spike (25 µg/ml), and SARS-CoV-2 spike (25 µg/ml) were incubated with 25 µl of signaling probe (NTD Ab-AuNP; 20 nM) for 5 min. Then, resulting solutions were loaded to the 96 well plate and lateral flow strip were dipped for 20 min.

For sensitivity test, 25 µl of various concentrations of spike (0, 0.39, 0.78, 1.56, 3.13, 6.25, 12.5. 25, 50 µg/ml) were incubated with 25 µl of signaling probe (NTD Ab-AuNP; 20 nM) for 5 min. Then, the same procedure of dipstick method was used for the lateral flow assay. The test line signals were quantitatively analyzed by imageJ software. The limit of the detection (LOD) was calculated by using blank + 3 standard deviations.

3.5.2.7 Signal enhancement analysis

For signal enhancement tests, a mixture of NTD Ab-AuNP and NTD Ab-HRP-AuNP were used as a signaling probe. Molar ratio of the mixture and reaction time was optimized (Figure S3.34), and 1:1 molar ratio of NTD Ab-AuNP and NTD Ab-HRP-AuNP with 15 min reaction time were chosen for the signal enhancement testing. The assay was conducted in the buffer and spiked-in-human saliva condition. In the case of the buffer, 25 µl of various concentrations of spike (0, 0.05, 0.10, 0.20, 0.39, 0.78, 1.56, 3.13, 6.25, 12.5. 25, 50 µg/ml) were incubated with 25 µl of signaling probe mixture (20 nM of NTD Ab-AuNP and NTD Ab-HRP-AuNP) for 5 min. Resulting solutions were loaded to the 96 well plate and prepared lateral strips were dipped for 20 min. Subsequently, 100 µl of AEC solution was introduced to enhance the signal for 15 min. For human saliva conditions, various concentrations of SARS-CoV-2 spike spiked in 1/50 diluted human saliva were used as a testing sample following the same test procedure, which was used in buffer condition. The test line signals were quantitatively analyzed by imageJ software. The limit of the detection (LOD) was calculated by using blank + 3 standard deviations.

3.5.2.8 Detection of mutant strain (SARS-CoV-2 Alpha, Beta, and Delta)

For the mutant strain testing, 25 µl of each proteins SARS-CoV-2 spike (50 µg/ml), Alpha strain spike- (50 µg/ml), Beta strain spike (50 µg/ml), and Delta strain spike (50 µg/ml) were incubated with 25 µl of signaling probe (NTD Ab-AuNP; 20 nM) for 5 min. Then, resulting solutions were loaded to the 96 well plate and lateral flow strip were dipped for 20 min. The test line signals were quantitatively analyzed by imageJ software.

3.5.2.9 Immunofluorescence staining of heparin and heparin-sulfate surfaces

For immunofluorescence, APTES coated coverslips were incubated with streptavidin (200 nM) for 10 minutes prior to a fixation step (2.5% Glutaraldehyde, 2.0% Paraformaldehyde in PBS) for another 15 minutes. Samples were washed three times for 2 minutes with PBS (3x2:00). Surfaces were then incubated with biotin-heparin and biotin-heparin sulfate (800 nM) in PBS for 30 minutes. Samples were washed again with PBS (3x2:00). Samples were then blocked with bovine serum albumin (4% in PBS) for another 30 minutes. After blocking, samples were washed three times using the kinetic buffer (10 mM HEPES, 10 mM NaCl, 0.05% Tween 20 pH 7.4). Spike protein (50 nM) was incubated with Mouse Anti 6x-His Tag Antibody conjugated with Alexa Fluor 488 (1 µg/mL) for 1 hour. Surfaces were washed with kinetic buffer (3x2:00) and were then incubated with mixture of Spike-Ab for 1 hour. Samples were then incubated with NTD Ab (1 µg/mL) in lateral flow assay buffer for 1 hour. Finally, surfaces were incubated with Anti-human Alexa 594 in PBS for 1 hour. Samples were washed with PBS and mounted onto glass slides using Immu-Mount. Samples were then imaged on a Ziess 710 confocal microscope.

Data analysis was done using FIJI. Prior to processing, immunofluorescence images were first blurred using a Gaussian threshold (diameter: 2 pixels) and a rolling pin filter for background subtraction (50 pixels). Protein locations were then identified through an automatic threshold using either a max entropy or triangle algorithm. Single randomly bright pixels were then removed using the “analyze particles” function to remove particles smaller than 0.5 μm^2. Proteins with both NTD Ab and His-Tag Ab binding were then found using the “AND” function in the “image calculator” function of FIJI. Particles were then analyzed on all three channels (NTD, His-Tag, and Combined) to determine the percentage of particles displaying both NTD and His-Tag signals.

3.5.2.10 Immobilization and binding of GAGs to spike proteins

Nunc maxisorp flat bottom 96 well plates were coated with streptavidin (200 nM; 50 µl) overnight at 4°C. The plates were blocked with 2% BSA for 1hr and biotinylated GAGs (800 nM; 50 µl) were added to the plates for 1hr. Unbound GAGs were thoroughly washed with 200 µl of 1xPBST (0.05% Tween 20) for three times. Spike proteins (100nM; 50 µl) diluted in the kinetic buffer (10 mM HEPES, 10 mM NaCl, 0.05% Tween

20, pH 7.4) were added to the plate and incubated for 1hr. Unbound spike protein were washed three times with 1xPBST and NTD Abs (2 µg/ml; 50 µl) diluted in the kinetic buffer were added for sandwich-type binding. Unbound NTD Abs were washed three times with 1xPBST and incubated with 50 µl each of 0.1 µg/ml anti-human-HRP (Sinobiological, 10702-T16-H) for 30 min at room temperature. The wells were washed thoroughly 5 times with 200 µl of 1xPBST. Finally, 100 µl of TMB substrate (Thermo scientific, 34028) was added to each well to develop color. The reaction was stopped by adding 50 ul of stop solution (Thermo scientific, N600), and absorbance was measured at 450 nm.

3.5.2.11 Immobilization and binding of heparin to variant spike proteins

Nunc maxisorp flat bottom 96 well plates were coated with streptavidin (200 nM; 50 µl) overnight at 4°C. The plates were blocked with 2% BSA for 1hr and biotinylated heparin (800 nM; 50 µl) was added to the plates for 1hr. Unbound GAGs were thoroughly washed three times with 200 µl of 1xPBST (0.05% Tween 20). Variant spike proteins along with wild-type spike protein (100nM; 50 µl) diluted in the kinetic buffer (10 mM HEPES, 10 mM NaCl, 0.05% Tween 20, pH 7.4) were added to the plate and incubated for 1hr. Unbound spike protein were washed three times with 1xPBST and NTD Abs (2 µg/ml; 50 µl) diluted in the kinetic buffer were added for sandwich-type binding. Unbound NTD Abs were washed three times with 1xPBST and incubated with 50 µl each of 0.1 µg/ml anti-human-HRP (Sino biological, 10702-T16-H) for 30 min at room temperature. The wells were washed thoroughly 5 times with 200 µl of 1xPBST. Finally, 100 µl of TMB substrate (Thermo scientific, 34028 Thermo scientific, 34028) was added to each well to develop color. The reaction was stopped by adding 50 µl of stop solution (Thermo scientific, N600), and absorbance was measured at 450 nm.

3.5.2.12 Surface Plasmon Resonance (SPR)

To measure the binding affinity of HEP15 and HS15, a single cycle kinetic analysis method was used. Briefly, biotinylated HEP15 and HS15 were immobilized to a streptavidin (SA) chip by flowing 0.2 mg/ml of HEP15 and HS15 in kinetic buffer (10 mM HEPES, 10 mM NaCl, 0.05% Tween 20, pH 7.4) to the flow cell of the SA chip at a flow rate of 10 µl/min for 60 sec. Various concentration (10, 25, 50, 100, 200, 500 nM) of

wild type spike protein samples were prepared by serial dilution in kinetic buffer. Spike protein samples were sequentially injected at a flow rate of 30 µl/min for 120 sec without regeneration and dissociation was measured at the end using the same buffer for 6 min. Binding affinity was calculated using the steady-state analysis method.

3.5.2.13 Comparing adsorption of GAGs onto the nitrocellulose membrane

To test the adsorption of HS15, HEP15, and DEX5 onto the nitrocellulose membrane, biotinylated HS15, HEP15, and DEX5 (140 µM; 1 µl) in lateral flow assay buffer (10 mM HEPES, 10 mM NaCl, pH 7.4) was dispensed to the nitrocellulose membrane (FF120HP). Dispensed nitrocellulose membrane was dried at 65ºC for 3 min. After drying, sample pad (Whatman CF4 dipstick pad) and the absorbent pad (Whatman standard 17) were assembled onto the nitrocellulose membrane. HS15, HEP15, and DEX5 adsorbed strip was dipped into the streptavidin coated AuNP (1 nM; 50 µl) for 5 min. Resulting image was analyzed with Image J.

3.6 **Supporting Information**

3.6.1 Supplementary Computational Methods

3.6.1.1. Selection of protein conformations from “Beyond Shielding” simulations.

As described in the main text, protein and glycan motions could impact GAG binding at specific sites on the spike surface. To ensure we incorporated a degree of protein and glycan flexibility during docking simulations, we selected four protein conformations from Casalino et al.’s freely available trajectories (https://amarolab.ucsd.edu/covid19.php).[1] The first conformation selected was Casalino et al.’s closed state psf/pdb pair (named conformation “2” in the main text). Given the furin cleavage site’s high degree of flexibility, we used the accessible surface area of this site as an unbiased, numerical, metric to select three more receptor coordinates for docking simulations. We calculated accessible surface area of residues 674 to 685 per chain A, B, C, according to the Shrake-Rupley algorithm,[2] and plotted the accessibility of these residues at the r=1.4 probe radii over the course of the aggregate 1.7 us closed simulations. From these results we identified three additional conformations via selecting one frame in which chain A’s furin site was

relatively exposed (named “conformation 4” in the main text), one conformation in which chain B’s furin site was relatively exposed (named “conformation 1” in the main text) and one conformation in which chain C’s furin site was relatively exposed (named “conformation 3” in the main text). See Figures S3.3-4 for visual summaries of these conformations. Additionally, all receptor and ligand structures, as well as docking input files, have been provided in a compressed file included with this supporting information.

3.6.1.2. RMSD calculations between protein conformations.

While the furin cleavage site’s ASA was used as the simple metric to select protein conformations for docking, other regions of the spike head also differed in conformation as a result of molecular dynamics simulations conducted by Casalino et al.[1] We thus calculated RMSD for key domains on the spike head between each selected conformation, to describe the conformational diversity between each receptor structure. To calculate RMSD between two protein conformations (as listed in Table S1) an in-house TCL script was prepared. We first controlled for global translational and rotational degrees of freedom by aligning the central helices of both spike conformations (Cα atoms of residues 747 to 784, 946 to 967, 986 to 1034) onto one another. “Spike Head” RMSD (Table S1, row 1) was calculated by calculating the heavy atom RMSD for every residue within the spike head (residues 16 to 1140) between each protein conformation. All other RMSD’s (Table S1, rows 2-13) were calculated on a per-chain basis. For per-chain RMSDs, we calculated the heavy atom RMSD for each denoted region (Table S1) per chain between each given protein conformation and then averaged all per-chain RMSDs to represent the results as averages with standard deviations. We used per-chain RMSDs to investigate if there were large differences between RMSD at the chain-by-chain level. Each region of interest was selected as follows:

· Spike Head: protein and resid 16 to 1140 and not hydrogens

· Chain A, B, C: chain X and resid 16 to 1140 and not hydrogens

· RBD: chain X and resid 330 to 530 and not hydrogens

· RBM: chain X and resid 427 to 508 and not hydrogens

· RBD beta-sheets: chain X and (resid 375 to 380 or resid 394 to 404 or resid 431 to 438 or resid 508 to 517) and not hydrogens

· NTD: chain X and resid 16 to 296 and not hydrogens

· Central Helix: chain X and (resid 747 to 784 or resid 946 to 967 or resid 986 to 1034) and not hydrogens

· Furin site: chain X and resid 674 to 685 and not hydrogens

· Fusion peptide: chain X and resid 817 to 834 and not hydrogens

· Connecting Domain: chain X and resid 1080 to 1135 and not hydrogens

· Esko site: chain X and resid 346 354 to 357 444 466 509 and not hydrogens

· Fadda site: chain X and resid 348 to 352 449 to 452 470 484 490 to 494 and not hydrogen

· Head glycans: glycans located at positions N17, N61, N74, N122, N149, N165, N234, N282, N331, N343, N603, N616, N657, N709, N717, N801, N1074, N1098, N1134, O323, O325 and not hydrogens (chains B and C do not have a glycan O325)

3.6.1.3. Unbiased docking of sulfated polysaccharides into Spike conformations with AutoDock Vina:

Protein conformations were prepared for docking according to default protocols in AutoDockTools.[3,4] A receptor grid was generated (centered on x=-1.417, y=0.947, z=200.541, x-width=126Å, y-width=126 Ang, z-width=126Å) for each spike conformation. These large grids were consistently generated to allow for unbiased docking into any site within the spike head structure, as done by Linhardt et al.[5,6]

To predict as many binding sites and binding modes on the spike as possible, we conducted thorough sampling during each docking procedure. A “docking procedure” being defined as one GAG model docked into one spike conformation, e.g., dimeric HEP docked into spike conformation 1. Autodock Vina settings were specified as follows: energy_range = 30, exhaustiveness = 80, num_modes = 100.[3–6] In addition to setting explicitly exhaustive settings in each docking procedure, we repeated each docking procedure 20 times per spike conformation/GAG model pair. Thus, per GAG model and per spike conformation we returned 400 output binding modes (20 per procedure x repeated 20 times = 400). Upon aggregating results

from all GAG models (in both dimeric and tetrameric states) and all four spike conformations, we obtained 12,800 spike-GAG binding modes.

3.6.1.4. K-means clustering of predicted GAG binding modes in Spike conformations:

We clustered all 12,800 predicted spike-GAG binding modes to identify unique binding "sites". We used scikitlearn's k-means clustering module to perform clustering.[7] To search for the optimal number of clusters, we first clustered by initializing random cluster centers and specifying k=1 to k=100, where k is the total number of cluster centers. For each k, i.e, for each total number of cluster centers, we performed clustering according to 10 distinct randomized cluster center coordinates, and per k the cluster centers with lowest sum of square error (SSE, which is the distance between each cluster member and the optimized cluster center) were reported. We then used an elbow/knee location module called "kneed"[7] to determine the optimal number of clusters, i.e., the smallest number of clusters that still achieves minimized SSE. From K-means clustering we identified 17 unique clusters from our 12,800 predicted spike-GAG binding modes.

3.6.1.5. Calculation of important contacting residues per binding site.

All binding modes were clustered into 17 binding sites according to K-means clustering as described above (A through Q). To identify key residues in each binding site, we parsed all binding modes by binding site, then tabulated all residues that made contact with the ligand bound in each site and in each binding mode. A "contact" was defined as any protein or glycan residue within 3.5 Å of the ligand structure. We then tabulated the frequency of contact per residue in each binding site to determine those residues with most importance for binding. The total number of binding modes (for all ligands and all four spike receptors) was used to normalize the frequency of per residue contacts as a percent. Only those residues with greater than or equal to 25% contact frequency were selected as "important contacting residues" per binding site. Those residues are tabulated in Table S2. For a complete list of all residues per binding site, please see the data directory in the shared supporting information .tar.gz file.

3.6.1.6. Setup trivalent hep40mer model spike system (spike-3xhep40mer):

Heparan sulfate (HS) proteoglycans in the glycocalyx have anywhere from 40 to 500 saccharide monomeric units.[8] Docking of such long GAGs directly to SARS-CoV-2 spike would be intractable due to combinatorial enumeration of torsional degrees of freedom. However, to interpret experimental GlycoGrip results at the atomic scale, we needed a proposed binding long-chain GAG binding model to spike. To construct such a model, we used information about the relative importance of each of our identified binding sites, A through Q, as calculated from our binding site importance score (Equation 1), as well as past literature evidence for these sites.[5,6,9–12] From these results we were able to construct an outline of a HEP/HS binding site on the spike surface by connecting several sites that were both highly "important" and overlapped well with literature proposed sites: sites B, H, and the furin cleavage site.

According to our results, site B is the #1 ranked binding site on the spike surface and site B also corresponds to the Esko and Fadda supersite.[5,9,11] Additionally, site H is near site B and represents an "bridging" site predicted by Wade and coworkers.[12] While site H was not ranked highly according to our importance metric, it is analogous in location to the highly important site E: site H sits at the RBD/NTD interface of chain A and B while site E sits at the RBD/NTD interface between chains B and C. Site E's results were particularly interesting as these indicated positions of N331B could either compete with or stabilize GAG binding at this site. In our selected spike structures, N331A happens to populate this site more heavily and thus we did not predict site H to be as highly populated as site E. However, we can hypothesize that this site would become populated if N331A moves out of the way, thus we considered site H in our long-chain heparin (HEP) model building. Finally, while we did not elucidate a significant number of binding modes at the furin cleavage site, experimental and past computational research suggests HEP/HS should interact with and bind to this site. In all, we sought to construct a HEP/spike model that incorporated HEP binding at site B, site H, and the polybasic (RRAR) furin cleavage site.

To generate a long (40meric) HEP for placement along the spike surface, we conducted molecular dynamics simulations of a HEP octamer and clustered the resulting simulations to find relaxed conformations. We constructed the HEP octamer (hep8mer) with MatrixDB according to the following stereochemistry and sulfation pattern: [IdoA2S-(1-->4)-GlcNS6S-(1-->4)]4-OH .

Hep8mer was solvated in a water box of size x=41.8Å, y=41.8Å, x=68.3Å. To introduce physiologically relevant ionic conditions and to neutralize the -16 total formal charge of hep8mer, 0.15 M NaCl was added to the water box. Ten replicas of hep8mer molecular dynamics simulations were conducted according to the following protocol: All molecular dynamics simulations used an integration time step of 2fs and SHAKE was applied to constrain all bonds between heavy atoms and hydrogen atoms to their parameter values. Periodic boundary conditions were used to mimic bulk aqueous conditions, and Particle Mesh Ewald was used to efficiently calculate long range electrostatic interactions (PME options set as such: PME interpolation order = 4 (cubic interpolation order) and PME grid spacing = 1.0Å). A switching function was used to gradually switch pairwise non-bonded (LJ and Coulomb) interactions off between 10 and 12Å pair distances. Pairwise non-bonded lists were updated once per integration cycle and included all atoms within 13.5Å. Each replica was first minimized with 10,000 steps of conjugate gradient with line search algorithm minimization (NAMD default minimizer). After minimization, velocities were initialized at 310K and 25,000,000 steps of MD were conducted at a 2fs timestep, for a total of 50ns. A Langevin thermostat and pressure piston was used to maintain constant temperature at 310K and constant pressure at 1Atm. Coordinates were saved every 20ps (i.e., every 10,000th step) resulting in 2,500 hep8mer conformations collected per replica. Aggregating all 10 replicas of 50ns trajectories resulted in a total of 25,000 hep8mer conformations to be clustered.

To prepare simulation results for clustering, we concatenated all resultant 50ns hep8mer production trajectories into one trajectory and aligned all frames by hep8mer heavy atoms to control for global translational/rotational degrees of freedom. We used GROMACS tools[13] to cluster hep8mer coordinates according to RMSD. Clustering was performed for the RMSD cutoffs from 1.0Å to 3.0Å, incrementing every 0.1Å. An RMSD cutoff at r=1.2Å resulted in optimal clustering, with 95% of coordinates classified within the first 5 clusters, and the remaining 5% of structures populating 55 unique clusters. For each of the top 5 clusters, we saved the structure closest to the cluster center, and took these coordinates as a representative structure of each cluster. We then used these 5 structures as coordinates to be manipulated on the spike surface for hypothesizing a long-chain hep40mer binding mode on the spike surface.

To generate a long-chain HEP binding mode, we first aligned hep8mer conformations to sites B, H, and the furin cleavage site using alignment tools in VMD. These alignments served as "anchors" on the spike surface. We then used VMD alignment tools to connect the anchor points via "linking" hep8mers. We then used Molefacture within VMD if conformational changes were required to maneuver around glycans (particularly important in the case of moving around $N331_A$). We then used psfgen to "glue" the five individually aligned hep8mers together, creating one large hep40mer. We then triplicated the hep40mer onto the other putative HEP binding sites, i.e., the analogous sites at the interface between chains B and C and chains C and A). We again used Molefacture to adjust hep40mers where needed to avoid clashes with glycan or protein residues. While this trivalent spike-hep40mer (spike-3xhep40mer) structure is speculative, and we only used it as a spatial model in our current work, we hope to use it for extensive molecular dynamics simulations in future directions.

3.6.1.7. Setup of spike-3xhep40mer with antibody and ACE2 models:

To assess whether the spike surface could accommodate multivalent binding of long-chain GAGs as well as signaling antibodies, we constructed a proposed model of these systems via alignment in VMD. Vendors would not specify either protein structure, or sequence of the monoclonal NTD antibody (Clone 2146 from Leinco Technologies) that we purchased for experimental testing (see Experimental Methods in the main text). Thus, we considered 4A8 as a representative NTD binding antibody. We downloaded a structure of three 4A8's bound to spike head via the three NTDs from the Protein Data Bank (PDB: 7C2L).[14] We aligned each light and heavy chain 4A8 to our chain A NTD of spike-3xhep40mer by aligning our chain A spike NTD to 7C2L's chain A NTD through RMSD alignment in VMD: aligning via Cα atoms of chain A's residues 16 to 296 between our spike structure and 7C2L's spike structure and saving coordinates of 4A8 in this position. We then repeated this procedure for chain B's NTD and chain C's NTD to result in 3 positionings of 4A8, one at each NTD. Given this was merely intended to be a spatial model, to ensure the spike surface could accommodate both long-chain GAG as well as NTD-binding antibodies, we did not minimize or geometrically optimize the structure further. The same procedure was then repeated to construct models for RBD-binding

antibody REGN10933 (PDB: 6XDG, Figure S3.14A)[15] and ACE2 bound to the spike RBM (PDB: 6M17, Figure S3.14B).[16]

3.6.1.8. Accessible surface area analyses:

To investigate the discrepancies between signal intensity observed between RBD-based antibodies and NTD-based antibodies, we calculated the accessible surface area (ASA) of their corresponding antigenic regions on the spike surface from Casalino et al.'s available trajectories.[1] ASA is the surface area of a molecule that is accessible to either solvent molecules (in which case ASA would specifically correspond to solvent accessible surface area which is SASA) or other larger binding partners. We used the Shrake-Rupley algorithm[2] which calculates ASA by effectively rolling a probe sphere of given radius size over a site of interest. Two parameters must be well defined for Shrake-Rupley ASA calculations: the site of interest, i.e., either the whole molecular surface or a specific region, and the probe radius size. Different probe radii can be used to approximate binding partners of different sizes: a probe radius of 1.4Å roughly approximates water molecules, a probe radius of 7.2Å approximates the size of a small molecule or an antibody's hypervariable loop, and a probe radius of 18.6Å approximates the size of a small protein binding partner or an antibody's variable fragment domain. In our work we probed the accessibility of spike antigenic regions to antibody hypervariable loops (r=7.2Å) and variable fragment regions (r=18.6Å). We used the closed spike trajectories from Casalino et al.[1] to calculate ASAs for three regions: the 4A8/NTD epitope (chains A, B, C residues 143-153 and 245-259), the REGN10933/RBD epitope (chain A, B, C residues 403-406, 416-422, 453-456, 473-478, 484-498), and the RBM or ACE2 binding motif (chains A, B, C, residues 437 to 508). We calculated ASA for each region at every 200th step of Casalino et al.'s simulations. Additionally, we calculated a reference interface size of each spike-antibody pair by calculating the size of these same antigenic regions (at each of the probe radii) from Cryo-EM structures (REGN10933 from PDB 6XDG,[15] 4A8 from 7C2L,[14] RBM/ACE2 binding motif from 6M17.[16] To calculate the Cryo-EM reference interface sizes, we used the Protein Data Bank structures for each antibody bound to the spike (PDBID's as follows: 6XDG for REGN10933, 6M17 for ACE2, and 7C2L for 4A8). We then removed the antibody binding partner and

calculated surface area of the same interface with the Shrake-Rupley algorithm, as done for Casalino et al.'s simulations.

The Shrake-Rupley algorithm was also applied to calculate the ASA of each clustered binding site, as shown in Figure S3.7. Each site, A through Q, was defined as all residues within 10Å of the cluster's center. ASA was plotted, in this case, only for r=7.2, but ASA for all sites over all radii can be made available upon request.

3.6.1.9. Electrostatic potential map calculations with APBS:

To confirm the presence of large, positively charged regions on the spike surface we used Adaptive Poisson Boltzmann Solver (APBS) to calculate the electrostatic potential at equally spaced grid points along the spike surface. We calculated electrostatic potential maps for the WT spike in closed and open states as well as for Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2) variant full-length spike structures in the closed state. For all structures we calculated electrostatic potential maps according to the following options, and example input scripts can be found in the downloadable tar.gz file associated with this supporting information:

```
elec name complex
         mg-auto
         dime 321 321 321
         cglen 246.4983 240.6809 557.8890
         fglen 164.9990 161.5770 348.1700
         cgcent mol 1
         fgcent mol 1
         mol 1
         lpbe
         bcfl sdh
         ion charge 1 conc 0.150 radius 1.36375
         ion charge -1 conc 0.150 radius 2.27
         pdie 4.0
         sdie 78.00
         chgm spl2
         srfm smol
         srad 1.4
         swin 0.3
         sdens 10.0
         temp 298.15
         gamma 0.105
         calcenergy total
         write pot dx test
end
```

3.6.1.10. Setup of system for APBS ionic strength calculations:

To model the experimentally observed relationship between solution ionic strength and binding affinity between HEP and spike, we utilized simulation trajectories from a related, unpublished project. In this tangential (unpublished) work, we investigated the potential ternary complex formed between HEP, the spike RBD, and ACE2, as proposed by Esko and coworkers.[9] Our goal in this investigation was to: (1) confirm Esko's docking results via our own flexible docking studies, (2) confirm that HEP would remain tightly bound to the spike RBD even under the conditions of molecular dynamics, (3) identify key points of contact between the spike RBD and HEP that drive binding. Esko and coworkers used rigid receptor/flexible ligand docking studies focused only on the RBD to predict that HEP would bind tightly to a positively charged site on the RBD, shown in Figure 3.1a of the main text. We used Schrödinger's Induced Fit Docking protocol -- which balances Glide Rigid Docking with Prime Loop Refinement to incorporate protein induced fit motions during docking -- to flexibly dock a hep8mer onto the spike RBD surface.

Glide and Schrödinger's Induced Fit Docking does not treat glycans, thus, before docking of HEP into the RBD, all glycan atoms were deleted. This was deemed an acceptable treatment as there are only two glycans on the RBD and these glycans, as well as ACE2 glycans, are far from the Esko-predicted HEP binding site considered for this flexible docking with Schrödinger's IFD. Schrödinger's IFD options were set as follows: The OPLS3e force field was used at all steps of Schrödinger's Induced Fit Docking and, unless described here default options were chosen during Schrödinger's IFD procedure. The inner Glide grid box was set at 10 Å x 10 Å x 10 Å, and the outer Glide grid box was set to 56 Å x 56 Å x 56 Å, with the grid centered on the center of mass of the following RBD residues: 346, 354-357, 444, 466, and 509. During Glide receptor grid generation, receptor Coulombic forces were scaled by 0.25, and receptor van der Waals forces were scaled by 0.75. Initial Glide docking was conducted with Standard Precision (SP) docking, ligand Coulombic forces and van der Waals forces were scaled by 0.15, and 0.50, respectively. The following RBD residues were selected for mutation to Alanine (a.k.a. "trimming") and backmutation after ligand replacement

(the procedure that models protein flexibility in Schrödinger's IFD): 341, 346, 348, 352, 354-357, 396-399, 466. All residues within 5Å of initial ligand placement were selected for Prime refinement, i.e., energetic minimization after ligand placement, and the following residues were specified for Prime refinement regardless of distance from initial ligand placement: 341, 346, 348, 352, 354-357, 396-399, 466. After Prime refinement, poses were redocked with Glide Extra Precision (XP). During Glide XP redocking, receptor Coulombic and vdW energy terms were scaled by 0.25, and 1.00, respectively, while ligand Coulombic and vdW energy terms were scaled by 0.15, and 0.80, respectively. Glide XP descriptors were also written alongside successful poses. Resultant poses were ranked by Glide XP Score and inspected visually.

As a result of these docking studies with Schrödinger's IFD, we predicted 30 poses of hep8mer on the spike RBD surface that all replicate the predicted Esko binding mode, i.e., that all align along the positively charged cleft on the spike RBD. Use of flexible docking in this case allowed us to optimize key contacts between hep8mer's sulfate groups and key charged residues on the spike RBD surface. We then selected the top binding mode (lowest predicted Glide score) to continue with molecular dynamics simulations.

As Barros et al. has shown,[17] ACE2 exhibits a significant degree of large-scale flexibility: the ACE2 dimer waves back and forth on the cell surface and binding of the spike RBD to ACE2 does not impede these large-scale motions. Pecora de Barros et al. posit that this large-scale motion facilitates binding to the spike by increasing angular degrees of freedom and potentially facilitates the S1 shedding via applied forces.[17] Thus, if a ternary complex were to form between the spike RBD, ACE2, and HS, as Esko posits, one would expect that binding between the spike RBD and HEP must be strong enough to withstand these potential large degrees of "waving" by ACE2. Thus, in MD simulations of the spike RBD with hep8mer, not only did we want to include the native flexibility of the spike RBD, but we also wanted to incorporate the high degree of movement imbued by binding of spike RBD to ACE2. We conducted MD simulations of the whole ternary complex using our top predicted binding mode from Schrödinger IFD.[17]

We followed the system setup procedure for the spike RBD+ACE2 system as published by Pecora de Barros et al.,[17] however instead of starting from the cryo-EM structure (PDB 6M17) we started directly with

Pecora de Barros et al.'s freely available PSF/PDB pair (https://amarolab.ucsd.edu/covid19.php). We then used VMD tools to align our top RBD+hep8mer binding mode from Schrödinger IFD to the spike RBDs on Pecora de Barros's structure. In system designe: (1) we replaced the RBD coordinates because Schrödinger IFD protocol treats specified regions of the protein structure flexibly, thus the resultant RBD structure was modified relative to the unbound state, and (2) as we were modeling ACE2 in the dimeric state, we have two RBDs bound to ACE2, and we have two hep8mers one each bound to each RBD, and (3) we incorporated the same glycosylation profile on ACE2 and the spike RBD that Pecora de Barros et al. used on their ACE2+RBD model. We used PropKa to predict protonation states, which remained consistent to those predicted protonation states from Pecora de Barros et al. and Casalino et al.[1,17] We used psfgen to generate topology files and save coordinates suitable for starting molecular dynamics simulations. The ACE2/RBD/hep8mer ternary complex was added to a lipid bilayer patch of x_length = 225 Å by y_length = 225 Å, with a composition similar to that of mammalian cell membranes (56% POPC, 20% CHL, 11% POPI, 9% POPE, and 4% PSM).[1,17] The resulting system was then embedded into an orthorhombic box of explicit TIP3P water molecules for use with periodic boundary conditions (PBC): x_length = 187.46Å, y_length = 189.29Å, z_length = 250.84Å. The complete system was then neutralized at a concentration of 150 mM NaCl.

Three replicas of MD simulations were conducted according to the following protocols: All MD simulations were performed on the Frontera supercomputer at the Texas Advanced Supercomputing Center (TACC) using NAMD 2.14[18,19] and CHARMM36m all-atom additive force fields.[20–26] All molecular dynamics simulations used an integration time step of 2fs and SHAKE was applied to constrain all bonds between heavy atoms and hydrogen atoms to their parameter values. Periodic boundary conditions were used to mimic bulk aqueous conditions, and Particle Mesh Ewald was used to efficiently calculate long range electrostatic interactions (PME options set as such: PME interpolation order = 4 (cubic interpolation order) and PME grid spacing = 1.0Å). A switching function was used to gradually switch pairwise non-bonded (LJ and Coulomb) interactions off between 10 and 12Å pair distances. Pairwise non-bonded lists were updated once per integration cycle and included all atoms within 13.5Å. Minimization and equilibration were

performed in four steps. In the first step, while keeping all atoms fixed except for the lipid tails, the system was subjected to an initial minimization of 10,000 steps using the conjugate gradient energy approach (NAMD default minimizer). Following minimization, lipid tails were gradually melted from 10K to 310K over the course of 200ps incrementing by 25K every 20ps; during lipid tail melting, constraints remained on for protein, glycan, and water atoms. After lipid-tail melting, each replica was subjected to 0.5ns NpT equilibration (1Atm, 310K) where protein, glycan, and hep8mer atoms were restrained according to a force constant of 1kcal/mol/Å2. During this equilibration, an in-house NAMD script was used to keep water molecules from penetrating the lipid membrane. Next, 0.5ns of additional NpT (1Atm, 310K) equilibration was conducted where restraints on protein and glycan atoms were released, but restraints were maintained on hep8mer atoms (according to force constant of 1kcal/mol/Å2) to ensure hep8mer and RBD atoms were thoroughly optimized. Finally, a 20ns free, unrestrained, equilibration was conducted under NpT conditions (310K) to ensure unbiased production sampling. After equilibration steps were complete, each replica was simulated under NpT conditions for a total of 300ns, resulting in an aggregate of 900ns of simulation time; coordinates were saved to trajectory once every 50,000th step (every 100 ps) resulting in a total of 2,000 frames saved per replica, or 6,000 frames saved in total. These three replicate trajectories were aggregated, and the final frame was saved as a PDB for ionic concentration effect monitoring with APBS.

3.6.1.11. APBS Ionic Strength Effect Computational Modeling:

Adaptive Poisson Boltzmann Solver (APBS) was used to estimate the binding affinity between hep8mer and the spike RBD at varying solvent ionic strengths. Binding affinity was calculated according to an appropriate thermodynamic cycle by calculating binding energy in a homogeneous reference medium (dielectric constant = 4), and then by calculating the difference in solvation free energy between the homogenous reference state and non-homogenous target state (dielectric constant = 78).[27] For a complete detailed description on calculating binding energies with APBS, including the appropriate thermodynamic cycle, please see the APBS 3.1 manual (https://apbs.readthedocs.io/_/downloads/en/latest/pdf/).

The final frame from ACE2+RBD+hep8mer simulations was used as the structure for these ionic concentration effect monitoring. As the effect was observed only in the context of spike binding to HEP, for these calculations, we took the structure from simulations and deleted ACE2 and kept only one of the RBD+hep8mer pairs. Thus, this minimal spike RBD-hep8mer structure was used as our model for APBS calculations. Binding energies were calculated for the RBD-hep8mer complex at the following NaCl concentrations: 0.0M, 0.01M, 0.025M, 0.05M, 0.075M, 0.10M, 0.125M, 0.150M, 0.175M, 0.200M. For complete options used in all calculations required to complete the above thermodynamic cycle, please see the example input scripts included in the attached tar.gz file associated with this Supporting Information.

3.6.2 Supplemental Computational Results

3.6.2.1. Structural differences between selected spike conformations:

While we have used the furin site's accessibility as a metric for selecting spike conformations, other sites throughout the protein also differ between these selected conformations. We calculated RMSD between each conformation for all heavy atoms in specific regions of interest in chain A, and did the same for chain B, and chain C, then we averaged these RMSDs and are reporting RMSD alongside standard deviations (see SI Methods section 2) for complete description on RMSD calculations). When considering all spike head heavy atoms (residues 13 to 1140), Conformations 2, 3, and 4 differ from Conformation 1 by 4.6 ± 0.4 Å, 5.2 ± 0.4 Å, and 6.2 ± 1.0 Å, respectively (Table S1, row labeled "Chains A, B, C). The receptor binding motif (RBM), furin site, and glycans were the most varied between protein conformations: RBMs differed on average by 8.2 ± 2.3 Å, furin sites 11.5 ± 0.5 Å, and glycans 14.7 ± 7.2 Å. The RBD β-sheets and spike central helices were, as could be expected, the most similar between protein conformations: RBD β-sheets differed on average by 2.7 ± 0.8 Å, central helices 2.2 ± 0.3 Å. These regions are known from past work to be the most stable, thus small degrees of variation in these sites indicate reasonable spike structures were selected. (See Table S1 for complete RMSD comparison between each protein conformation.)

3.6.2.2. ASA per predicted binding site:

To determine the accessibility of each of the 17 identified binding sites, we calculated the accessible surface area (ASA) of each of these sites from the trajectories of Casalino et al.[1] (Figure S3.7). While sites M and L are inaccessible at larger probe radii, sites B and F are accessible over all probe radii, however there is significant glycan shielding at these sites. Thus, GAGs may have to compete with the glycan shield for accessible binding sites on the spike surface.

3.6.3 Supplemental Tables

Table S3.1: Description of conformational differences between each spike conformation selected to serve as a receptor in docking studies

All values are RMSDs listed in Å. See Supporting Information Methods section 2 for details on how each value is calculated.

	Confs 1 v. 2	Confs 1 v. 3	Confs 1 v. 4	Confs 2 v. 3	Conf 2 v. 4	Conf 3 v. 4	Avg
Spike Head	4.7	5.0	6.3	5.5	6.2	4.9	5.4 ± 0.7
Chains A, B, C	4.6 ± 0.4	5.2 ± 0.4	6.2 ± 1.0	5.4 ± 0.6	6.2 ± 0.9	4.8 ± 1.0	5.0 ± 0.9
RBD	5.3 ± 1.4	5.7 ± 1.9	6.8 ± 1.0	6.7 ± 2.3	7.1 ± 0.8	5.5 ± 1.2	5.8 ± 1.4
RBM	7.4 ± 2.4	7.9 ± 3.1	9.4 ± 1.8	9.4 ± 3.6	10.2 ± 1.0	7.9 ± 2.0	8.2 ± 2.3
RBD β-sheets	2.4 ± 0.2	2.9 ± 1.7	3.3 ± 0.3	3.5 ± 1.4	3.2 ± 0.8	2.5 ± 0.4	2.7 ± 0.8
NTD	5.4 ± 0.6	7.1 ± 0.6	8.3 ± 1.4	7.1 ± 0.5	8.7 ± 1.8	5.7 ± 1.8	6.4 ± 1.7
Central Helix	2.3 ± 0.1	2.4 ± 0.2	2.5 ± 0.4	2.4 ± 0.3	2.1 ± 0.4	2.5 ± 0.3	2.2 ± 0.3
Furin Site	10.3 ± 5.1	10.3 ± 3.1	13.0 ± 8.4	10.4 ± 3.0	14.1 ± 4.2	14.8 ± 6.5	11.5 ± 5.1
Fusion Peptide	5.2 ± 1.1	4.6 ± 0.9	5.9 ± 1.4	4.7 ± 0.3	5.6 ± 2.3	6.0 ± 2.3	5.2 ± 1.4
Connecting Domain	3.6 ± 0.9	2.7 ± 0.1	3.4 ± 0.2	2.6 ± 0.3	2.3 ± 0.3	2.5 ± 0.4	2.8 ± 0.7
Esko Site	3.4 ± 0.3	3.5 ± 1.4	4.3 ± 0.3	4.3 ± 2.1	4.6 ± 0.7	3.4 ± 0.8	3.7 ± 1.0
Fadda Site	6.9 ± 2.8	7.3 ± 3.3	7.5 ± 2.1	9.4 ± 3.6	9.3 ± 1.2	7.2 ± 1.9	7.5 ± 2.4
Head Glycans	14.2 ± 5.8	14.7 ± 7.6	16.1 ± 6.7	14.6 ± 7.3	14.4 ± 7.1	13.4 ± 7.6	14.7 ± 7.2

Table S3.2: Complete list of residues found in each binding site

Residues are named as $AXXX_B$ where A is the one letter residue name, XXX is the residue number, and B is chain name.

	Residues
A	$F456_A$, $S477_A$, $Q493_A$, $S494_A$, $G496_A$, $Q498_A$, $T500_A$, $N501_A$, $N370_B$, $S371_B$, $A372_B$, $S373_B$, $W436_B$, $N437_B$, $S438_B$, $N440_B$, $L441_B$, Glycan $N343_B$
B	$R346_A$, $A348_A$, $S349_A$, $Y351_A$, $A352_A$, $R355_A$, $N354_A$, $R454_A$, $N450_A$, $I472_A$, $R466_A$, $D467_A$, $S469_A$, $Y489_A$, $F490_A$, $K113_B$, $T114_B$, $Q115_B$, $E132_B$, $N165_B$, $T167_B$, Glycan $N149_B$, Glycan $N165_B$
C	$Y369_A$, $N370_A$, $S371_A$, $A372_A$, $S373_A$, $S375_A$, $T376_A$, $F377_A$, $R403_C$, $D405_C$, $E406_C$, $Q409_C$, $T415_C$, $G416_C$, $K417_C$, $F456_C$, $Y505_C$, $S494_C$, $G496_C$, $Y453_C$, Glycan $N343_A$
D	$R403_B$, $T415_B$, $G416_B$, $D420_B$, $K424_B$, $L425_B$, $D427_B$, $Y449_B$, $R457_B$, $K458_B$, $S494_B$, $Y495_B$, $G496_B$, $F497_B$, $Q498_B$, $N501_B$, $A372_C$, $S373_C$, $S375_C$, $F377_C$, $Y369_C$, $W436_C$, $N437_C$, $N439_C$, $N440_C$, Glycan $N343_C$
E	$P337_B$, $K356_B$, $R357_B$, $S359_B$, $N360_B$, $T393_B$, $N394_B$, $E516_B$, $A520_B$, $P521_B$, $T523_B$, $P561_B$, $F562_B$, $R577_B$, $P579_B$, $Q580_B$, $L582_B$, $K41_C$, $Y170_C$, $S172_C$, $Q173_C$, $L226_C$, $V227_C$, $D228_C$, Glycan $N331_B$, Glycan $N122_C$, Glycan $N165_C$
F	$S366_A$, $Y369_A$, $N370_A$, $P384_A$, $T385_A$, $K386_A$, $N388_A$, $D389_A$, $K528_A$, $D985_B$, $Y421_C$, $R457_C$, $K458_C$, $S459_C$, $N460_C$, $K462_C$, $Y473_C$, Glycan $N234_A$, Glycan $S325_A$
G	$K129_A$, $S155_A$, $E156_A$, $V159_A$, $Y160_A$, $S162_A$, $A163_A$, $E169_A$, Glycan $N122_A$, Glycan $N149_A$
H	$P330_A$, $N331_A$, $I332_A$, $R357_A$, $S359_A$, $N360_A$, $T393_A$, $N394_A$, $A520_A$, $P521_A$, $A522_A$, $T523_A$, $P561_A$, $F562_A$, $R577_A$, $P579_A$, $Q580_A$, $L582_A$, $K41_B$, $Y170_B$, $Q173_B$, $L226_B$, $V227_B$, Glycan $N331_A$
I	$A27_B$, $Y28_B$, $T29_B$, $N30_B$, $W64_B$, $H66_B$, $A67_B$, $I68_B$, $H69_B$, $F186_B$, $A262_B$, $A263_B$, Glycan $N74_B$, Glycan $N122_B$
J	$L754_A$, $Q755_A$, $G757_A$, $S758_A$, $L518_B$, $R567_B$, $D571_B$, $D40_C$, $V42_C$,$R44_C$, $H49_C$, $S50_C$, $T51_C$, $Q52_C$, $K304_C$, $K964_C$, $S967_C$, $S968_C$, $N969_C$, $S974_C$, $S975_C$, $V976_C$,
K	$D571_A$, $S50_B$, $H49_B$, $Q52_B$, $T274_B$, $E298_B$, $C301_B$, $T302_B$, $K304_B$, $T315_B$, $S316_B$, $R319_B$, $Q628_B$, $T632_B$, $W633_B$, $K964_B$, $S967_B$, $S968_B$, $N969_B$, $D737_C$, $C738_C$, $T739_C$, $L753_C$, $L754_C$, $G757_C$, $S758_C$, $C760_C$, $T761_C$, $N764_C$, $R765_C$, Glycan $N61_B$
L	$H49_A$, $S50_A$, $Q52_A$, $T274_A$, $C291_A$, $E298_A$ $C301_A$, $T302_A$, $K304_A$, $T315_A$, $S316_A$, $N317_A$, $R319_A$, $Q321_A$, $P322_A$, $T549_A$, $D745_B$, $K964_A$, $S967_A$, $S968_A$, $T739_B$, $T747_B$, $S750_B$, $L754_B$, $Q755_B$, $S758_B$, $T761_B$, $N764_B$, Glycan $N234_A$, Glycan $S325_A$
M	$G744_A$, $F855_A$, $L966_A$, $V976_A$, $L977_A$, $R1000_A$, $P57_C$, $S60_C$, $N61_C$, $Y269_C$, $Q271_C$, $R273_C$, $L293_C$, $Q321_C$, $G548_C$, $T549_C$, $T572_C$, $H625_C$, $Q628_C$, Glycan $N17_C$, Glycan $N61_C$, Glycan $N234_C$
N	$I569_A$, $S46_B$, $V47_B$, $L48_B$, $K278_B$, $E281_B$, $L303_B$, $K304_B$, $S305_B$, $F306_B$, $T307_B$, $E309_B$, $T827_B$, $I834_B$, $G838_B$, $D839_B$, $C840_B$, $L841_B$, $G842_B$, $D843_B$, $V952_B$, $N953_B$, $A956_B$, $Q957_B$, $N960_B$, Glycan $N616_A$, Glycan $N603_B$
O	$Q762_A$, $R765_A$, $A766_A$, $G769_A$, $I770_A$, $E773_A$, $L1012_A$, $R1019_A$, $Q954_C$, $A958_C$, $Q1010_C$, $I1013_C$, $R1014_C$, $E1017_C$, $R1019_C$
P	$T791_B$, $P793_B$, $I794_B$, $K795_B$, $F797_B$, $N801_B$, $F802_B$, $S803_B$, $Q804_B$, $L806_B$, $D808_B$, $P809_B$, Glycan $N801_B$
Q	$D796_A$, $F797_A$, $G798_A$, $G799_A$, $W886_A$, $P897_A$, $F898_A$, $A899_A$, $Q920_A$, $K921_A$, $G1035_A$, $Q1036_A$, $K1038_A$, $Y707_C$, $N709_C$, $G908_C$, $K1038_C$, $V1040_C$, $Y1047_C$, $P1079_C$, $R1107_C$, $I1130_C$, $G1131_C$, Glycan $N709_C$

Table S3.3: Complete heparin docking results

AutoDock Vina docking scores (Binding energy, BE) are reported in kcal/mol. Pop, or population is number of poses in each site. Score is binding site importance score.

		Conf 1			Conf 2			Conf 3			Conf 4			Avg		
	Site	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Total Pop	Score
	A	--	--	0.0	--	--	0.0	--	--	0.0	-8.2 ± 0.2	41	0.8	-8.2 ± 0.2	41	0.2
	B	-7.5 ± 0.1	9	0.2	-8.3 ± 0.2	78	1.6	-7.8 ± 0.1	5	0.1	-7.9 ± 0.2	10	0.2	-8.1 ± 0.3	102	0.5
	C	-7.8 ± 0.3	75	1.5	--	--	0.0	-8.2 ± 0.4	28	0.6	-8.3 ± 0.4	49	1.0	-8.0 ± 0.4	152	0.8
	D	--	--	0.0	-8.3 ± 0.2	78	1.6	-8.1 ± 0.2	45	0.9	-8.4 ± 0.0	2	0.0	-8.2 ± 0.2	125	0.6
	E	-8.0 ± 0.3	96	1.9	--	--	0.0	-7.8 ± 0.2	13	0.3	--	--	0.0	-8.0 ± 0.3	109	0.5
	F	-7.6 ± 0.2	5	0.1	--	--	0.0	-8.3 ± 0.5	19	0.4	-8.3 ± 0.3	12	0.2	-8.2 ± 0.5	36	0.2
	G	--	--	0.0	--	--	0.0	--	--	0.0	-8.4 ± 0.3	58	1.2	-8.4 ± 0.3	58	0.3
	H	--	--	0.0	--	--	0.0	-8.1 ± 0.3	47	1.0	-8.1 ± 0.2	27	0.5	-8.1 ± 0.3	74	0.4
2mer	I	--	--	0.0	-8.3 ± 0.1	38	0.8	-7.9 ± 0.3	8	0.2	-8.0 ± 0.0	1	0.0	-8.2 ± 0.2	47	0.2
	J	-7.8 ± 0.3	40	0.8	-8.6 ± 0.2	15	0.3	-7.9 ± 0.2	55	1.1	-8.2 ± 0.3	92	1.9	-8.0 ± 0.4	202	1.0
	K	-7.4 ± 0.1	6	0.1	--	--	0.0	-8.0 ± 0.3	79	1.6	-8.0 ± 0.1	14	0.3	-7.9 ± 0.3	99	0.5
	L	--	--	0.0	-8.5 ± 0.2	53	1.1	-8.0 ± 0.0	1	0.0	-8.2 ± 0.1	4	0.1	-8.4 ± 0.2	58	0.3
	M	-7.5 ± 0.2	3	0.1	-8.4 ± 0.2	100	2.1	-8.0 ± 0.1	5	0.1	-8.3 ± 0.4	75	1.6	-8.3 ± 0.3	183	1.0
	N	--	--	0.0	-8.3 ± 0.2	3	0.1	-7.8 ± 0.2	38	0.7	-7.8 ± 0.0	3	0.1	-7.9 ± 0.2	44	0.2
	O	-7.6 ± 0.2	54	1.0	-8.2 ± 0.1	30	0.6	-7.8 ± 0.2	20	0.4	-8.0 ± 0.1	6	0.1	-7.8 ± 0.3	110	0.5
	P	-7.8 ± 0.2	71	1.4	-7.9 ± 0.1	5	0.1	-7.8 ± 0.2	37	0.7	--	--	0.0	-7.8 ± 0.2	113	0.6
	Q	-7.7 ± 0.2	41	0.8	--	--	0.0	--	--	0.0	-8.1 ± 0.1	6	0.1	-7.8 ± 0.2	47	0.2
	A	-9.2 ± 0.2	2	0.0	-9.3 ± 0.3	12	0.3	-8.6 ± 0.2	10	0.2	-8.8 ± 0.0	1	0.0	-9.0 ± 0.4	25	0.1
	B	-8.9 ± 0.5	52	1.2	-9.5 ± 0.3	32	0.8	-8.9 ± 0.4	55	1.2	-9.0 ± 0.3	109	2.5	-9.0 ± 0.4	248	1.4
	C	-8.7 ± 0.4	3	0.1	-9.0 ± 0.0	1	0.0	-9.3 ± 0.2	3	0.1	--	--	0.0	-9.0 ± 0.4	7	0.0
	D	-8.7 ± 0.4	18	0.4	-9.4 ± 0.3	34	0.8	--	--	0.0	-8.6 ± 0.0	1	0.0	-9.1 ± 0.5	53	0.3
	E	-9.1 ± 0.5	69	1.6	--	--	0.0	-8.7 ± 0.3	106	2.3	--	--	0.0	-8.9 ± 0.4	175	1.0
	F	-8.7 ± 0.3	166	3.6	--	--	0.0	-8.7 ± 0.3	22	0.5	-9.1 ± 0.3	10	0.2	-8.8 ± 0.3	198	1.1
	G	-8.8 ± 0.3	20	0.4	--	--	0.0	-8.6 ± 0.3	9	0.2	-8.7 ± 0.2	7	0.2	-8.7 ± 0.3	36	0.2
	H	--	--	0.0	-9.0 ± 0.1	6	0.1	--	--	0.0	-8.7 ± 0.0	1	0.0	-8.9 ± 0.1	7	0.0
4mer	I	--	--	0.0	--	--	0.0	--	--	0.0	-8.9 ± 0.3	11	0.2	-8.9 ± 0.3	11	0.1
	J	-8.8 ± 0.5	19	0.4	--	--	0.0	-8.5 ± 0.1	5	0.1	-9.2 ± 0.4	102	2.4	-9.1 ± 0.4	126	0.7
	K	-9.3 ± 0.0	1	0.0	--	--	0.0	-8.7 ± 0.3	46	1.0	-9.1 ± 0.3	84	1.9	-9.0 ± 0.3	131	0.7
	L	--	--	0.0	-9.3 ± 0.4	197	4.6	-8.6 ± 0.3	31	0.7	--	--	0.0	-9.2 ± 0.4	228	1.3
	M	-8.6 ± 0.1	7	0.1	-9.2 ± 0.2	83	1.9	-8.5 ± 0.2	16	0.3	-9.1 ± 0.3	23	0.5	-9.0 ± 0.3	129	0.7
	N	-8.5 ± 0.1	5	0.1	--	--	0.0	-8.8 ± 0.4	88	1.9	-9.0 ± 0.3	34	0.8	-8.8 ± 0.4	127	0.7
	O	-8.8 ± 0.3	16	0.4	-9.0 ± 0.3	7	0.2	-8.5 ± 0.2	7	0.1	-9.0 ± 0.2	11	0.2	-8.8 ± 0.3	41	0.2
	P	--	--	0.0	-9.2 ± 0.2	28	0.6	--	--	0.0	-8.7 ± 0.2	6	0.1	-9.1 ± 0.3	34	0.2
	Q	-8.6 ± 0.3	22	0.5	--	--	0.0	-8.4 ± 0.1	2	0.0	--	--	0.0	-8.5 ± 0.3	24	0.1
	A	-9.2 ± 0.2	2	0.0	-9.3 ± 0.3	12	0.1	-8.6 ± 0.2	10	0.1	-8.2 ± 0.2	42	0.4	-8.5 ± 0.5	66	0.2
	B	-8.7 ± 0.7	61	0.7	-8.6 ± 0.6	110	1.2	-8.8 ± 0.5	60	0.7	-8.9 ± 0.4	119	1.3	-8.8 ± 0.5	350	1.0
	C	-7.8 ± 0.4	78	0.8	-9.0 ± 0.0	1	0.0	-8.3 ± 0.5	31	0.3	-8.3 ± 0.4	49	0.5	-8.1 ± 0.5	159	0.4
	D	-8.7 ± 0.4	18	0.2	-8.6 ± 0.6	112	1.2	-8.1 ± 0.2	45	0.5	-8.4 ± 0.1	3	0.0	-8.5 ± 0.5	178	0.5
	E	-8.4 ± 0.7	165	1.7	--	--	0.0	-8.6 ± 0.4	119	1.3	--	--	0.0	-8.5 ± 0.6	284	0.8
	F	-8.7 ± 0.4	171	1.9	--	--	0.0	-8.5 ± 0.4	41	0.4	-8.6 ± 0.5	22	0.2	-8.7 ± 0.4	234	0.6
	G	-8.8 ± 0.3	20	0.2	--	--	0.0	-8.6 ± 0.3	9	0.1	-8.5 ± 0.3	65	0.7	-8.5 ± 0.4	94	0.3
	H	--	--	0.0	-9.0 ± 0.1	6	0.1	-8.1 ± 0.3	47	0.5	-8.1 ± 0.2	28	0.3	-8.2 ± 0.3	81	0.2
Avg	I	--	--	0.0	-8.3 ± 0.1	38	0.4	-7.9 ± 0.3	8	0.1	-8.8 ± 0.3	12	0.1	-8.3 ± 0.4	58	0.2
	J	-8.1 ± 0.6	59	0.6	-8.6 ± 0.2	15	0.2	-7.9 ± 0.3	60	0.6	-8.7 ± 0.6	194	2.1	-8.5 ± 0.7	328	0.9
	K	-7.7 ± 0.7	7	0.1	--	--	0.0	-8.2 ± 0.5	125	1.3	-8.9 ± 0.5	98	1.1	-8.5 ± 0.6	230	0.6
	L	--	--	0.0	-9.2 ± 0.5	250	2.9	-8.6 ± 0.3	32	0.3	-8.2 ± 0.1	4	0.0	-9.1 ± 0.5	286	0.8
	M	-8.2 ± 0.5	10	0.1	-8.7 ± 0.4	183	2.0	-8.4 ± 0.3	21	0.2	-8.5 ± 0.5	98	1.0	-8.6 ± 0.5	312	0.8
	N	-8.5 ± 0.1	5	0.1	-8.3 ± 0.2	3	0.0	-8.5 ± 0.5	126	1.3	-8.9 ± 0.4	37	0.4	-8.6 ± 0.5	171	0.5
	O	-7.9 ± 0.6	70	0.7	-8.4 ± 0.3	37	0.4	-8.0 ± 0.4	27	0.3	-8.6 ± 0.5	17	0.2	-8.1 ± 0.6	151	0.4
	P	-7.8 ± 0.2	71	0.7	-9.0 ± 0.5	33	0.4	-7.8 ± 0.2	37	0.4	-8.7 ± 0.2	6	0.1	-8.1 ± 0.6	147	0.4
	Q	-8.0 ± 0.5	63	0.6	--	--	0.0	-8.4 ± 0.1	2	0.0	-8.1 ± 0.1	6	0.1	-8.0 ± 0.4	71	0.2

Table S3.4: Complete heparan sulfate docking results

AutoDock Vina docking scores (Binding energy, BE) are reported in kcal/mol. Pop, or population is number of poses in each site. Score is binding site importance score.

		Conf 1			Conf 2			Conf 3			Conf 4			Avg		
	Site	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Total Pop	Score
	A	--	--	0.0	--	--	0.0	--	--	0.0	-8.8 ± 0.3	20	0.4	-8.8 ± 0.3	20	0.1
	B	--	--	0.0	-8.1 ± 0.2	44	0.9	--	--	0.0	-8.6 ± 0.4	3	0.1	-8.1 ± 0.2	47	0.2
	C	-7.6 ± 0.3	70	1.3	--	--	0.0	-8.0 ± 0.3	51	1.0	-8.2 ± 0.5	38	0.8	-7.9 ± 0.4	159	0.8
	D	-7.5 ± 0.0	1	0.0	-8.2 ± 0.3	94	1.9	-8.1 ± 0.4	140	2.9	--	--	0.0	-8.2 ± 0.4	235	1.2
	E	-7.8 ± 0.3	43	0.8	--	--	0.0	-8.7 ± 0.3	7	0.2	--	--	0.0	-8.0 ± 0.4	50	0.2
	F	-7.9 ± 0.3	34	0.7	--	--	0.0	-8.4 ± 0.5	20	0.4	-8.3 ± 0.2	7	0.1	-8.1 ± 0.5	61	0.3
	G	--	--	0.0	--	--	0.0	--	--	0.0	-8.2 ± 0.4	41	0.8	-8.2 ± 0.4	41	0.2
	H	--	--	0.0	--	--	0.0	-8.1 ± 0.3	33	0.7	-8.2 ± 0.4	36	0.7	-8.2 ± 0.4	69	0.4
2mer	I	--	--	0.0	-7.8 ± 0.1	6	0.1	--	--	0.0	--	--	0.0	-7.8 ± 0.1	6	0.0
	J	-8.2 ± 0.6	99	2.0	-8.2 ± 0.1	63	1.3	-8.3 ± 0.4	50	1.0	-8.5 ± 0.4	117	2.5	-8.3 ± 0.5	329	1.7
	K	-7.9 ± 0.4	13	0.3	--	--	0.0	-8.0 ± 0.3	20	0.4	-8.3 ± 0.0	1	0.0	-8.0 ± 0.4	34	0.2
	L	--	--	0.0	-7.9 ± 0.2	19	0.4	--	--	0.0	-8.2 ± 0.4	20	0.4	-8.1 ± 0.3	39	0.2
	M	--	--	0.0	-8.1 ± 0.2	130	2.6	-8.7 ± 0.4	9	0.2	-8.5 ± 0.6	109	2.3	-8.3 ± 0.5	248	1.3
	N	--	--	0.0	-8.1 ± 0.2	26	0.5	-7.9 ± 0.3	13	0.3	-7.6 ± 0.0	1	0.0	-8.0 ± 0.3	40	0.2
	O	-7.6 ± 0.2	42	0.8	-7.9 ± 0.1	18	0.4	-8.3 ± 0.5	42	0.9	-8.4 ± 0.0	3	0.1	-8.0 ± 0.5	105	0.5
	P	-8.2 ± 0.4	29	0.6	--	--	0.0	-8.0 ± 0.3	15	0.3	--	--	0.0	-8.1 ± 0.4	44	0.2
	Q	-7.8 ± 0.3	69	1.3	--	--	0.0	--	--	0.0	-7.8 ± 0.4	4	0.1	-7.8 ± 0.3	73	0.4
	A	-8.8 ± 0.5	3	0.1	-9.5 ± 0.4	16	0.4	--	--	0.0	-9.4 ± 0.0	2	0.0	-9.4 ± 0.5	21	0.1
	B	-9.0 ± 0.3	8	0.2	-9.4 ± 0.4	87	2.0	-9.3 ± 0.5	69	1.6	-9.3 ± 0.4	70	1.6	-9.3 ± 0.4	234	1.4
	C	-9.0 ± 0.5	7	0.2	--	--	0.0	-9.7 ± 0.4	11	0.3	-9.1 ± 0.3	17	0.4	-9.3 ± 0.5	35	0.2
	D	-9.1 ± 0.3	27	0.6	-9.2 ± 0.2	47	1.1	--	--	0.0	--	--	0.0	-9.1 ± 0.3	74	0.4
	E	-9.1 ± 0.4	66	1.5	--	--	0.0	-9.2 ± 0.4	94	2.2	--	--	0.0	-9.2 ± 0.4	160	0.9
	F	-9.0 ± 0.4	142	3.2	--	--	0.0	-9.3 ± 0.6	15	0.4	-9.4 ± 0.2	9	0.2	-9.1 ± 0.5	166	0.9
	G	--	--	0.0	--	--	0.0	-9.3 ± 0.4	24	0.6	-9.1 ± 0.3	6	0.1	-9.3 ± 0.4	30	0.2
	H	--	--	0.0	--	--	0.0	-9.1 ± 0.5	4	0.1	-8.8 ± 0.2	2	0.0	-9.0 ± 0.4	6	0.0
4mer	I	--	--	0.0	--	--	0.0	--	--	0.0	-9.1 ± 0.4	4	0.1	-9.1 ± 0.4	4	0.0
	J	-9.4 ± 0.6	33	0.8	-8.8 ± 0.0	1	0.0	-9.5 ± 0.5	35	0.8	-9.2 ± 0.4	143	3.3	-9.3 ± 0.4	212	1.2
	K	-8.6 ± 0.3	3	0.1	-9.1 ± 0.1	3	0.1	-9.3 ± 0.6	39	0.9	-9.1 ± 0.3	45	1.0	-9.2 ± 0.5	90	0.5
	L	-8.8 ± 0.0	1	0.0	-9.2 ± 0.3	172	3.9	--	--	0.0	-9.5 ± 0.6	5	0.1	-9.2 ± 0.3	178	1.0
	M	-8.3 ± 0.0	2	0.0	-9.1 ± 0.3	70	1.6	-9.1 ± 0.5	54	1.2	-9.1 ± 0.3	19	0.4	-9.1 ± 0.4	145	0.8
	N	--	--	0.0	--	--	0.0	-9.3 ± 0.5	44	1.0	-9.1 ± 0.2	36	0.8	-9.2 ± 0.4	80	0.5
	O	-9.0 ± 0.4	13	0.3	-9.0 ± 0.2	4	0.1	-9.4 ± 0.7	7	0.2	-9.2 ± 0.4	26	0.6	-9.1 ± 0.5	50	0.3
	P	-9.0 ± 0.4	27	0.6	--	--	0.0	--	--	0.0	-9.0 ± 0.3	16	0.4	-9.0 ± 0.3	43	0.2
	Q	-9.1 ± 0.4	68	1.5	--	--	0.0	-8.9 ± 0.5	4	0.1	--	--	0.0	-9.1 ± 0.4	72	0.4
	A	-8.8 ± 0.5	3	0.0	-9.5 ± 0.4	16	0.2	--	--	0.0	-8.9 ± 0.3	22	0.2	-9.1 ± 0.5	41	0.1
	B	-9.0 ± 0.3	8	0.1	-8.9 ± 0.7	131	1.5	-9.3 ± 0.5	69	0.8	-9.2 ± 0.4	73	0.8	-9.1 ± 0.6	281	0.8
	C	-7.8 ± 0.5	77	0.7	--	--	0.0	-8.3 ± 0.7	62	0.6	-8.5 ± 0.6	55	0.6	-8.1 ± 0.7	194	0.5
	D	-9.0 ± 0.4	28	0.3	-8.5 ± 0.5	141	1.5	-8.1 ± 0.4	140	1.4	--	--	0.0	-8.4 ± 0.5	309	0.8
	E	-8.6 ± 0.7	109	1.2	--	--	0.0	-9.2 ± 0.5	101	1.2	--	--	0.0	-8.9 ± 0.7	210	0.6
	F	-8.8 ± 0.6	176	1.9	--	--	0.0	-8.8 ± 0.7	35	0.4	-8.9 ± 0.6	16	0.2	-8.8 ± 0.6	227	0.6
	G	--	--	0.0	--	--	0.0	-9.3 ± 0.4	24	0.3	-8.3 ± 0.5	47	0.5	-8.7 ± 0.6	71	0.2
	H	--	--	0.0	--	--	0.0	-8.2 ± 0.4	37	0.4	-8.2 ± 0.4	38	0.4	-8.2 ± 0.4	75	0.2
Avg	I	--	--	0.0	-7.8 ± 0.1	6	0.1	--	--	0.0	-9.1 ± 0.4	4	0.0	-8.3 ± 0.7	10	0.0
	J	-8.5 ± 0.8	132	1.4	-8.2 ± 0.2	64	0.7	-8.8 ± 0.7	85	0.9	-8.9 ± 0.5	260	2.9	-8.7 ± 0.7	541	1.5
	K	-8.0 ± 0.5	16	0.2	-9.1 ± 0.1	3	0.0	-8.9 ± 0.8	59	0.7	-9.1 ± 0.3	46	0.5	-8.9 ± 0.7	124	0.3
	L	-8.8 ± 0.0	1	0.0	-9.1 ± 0.5	191	2.2	--	--	0.0	-8.5 ± 0.7	25	0.3	-9.0 ± 0.5	217	0.6
	M	-8.3 ± 0.0	2	0.0	-8.4 ± 0.5	200	2.1	-9.0 ± 0.5	63	0.7	-8.6 ± 0.6	128	1.4	-8.6 ± 0.6	393	1.1
	N	--	--	0.0	-8.1 ± 0.2	26	0.3	-9.0 ± 0.7	57	0.6	-9.0 ± 0.3	37	0.4	-8.8 ± 0.7	120	0.3
	O	-7.9 ± 0.6	55	0.5	-8.1 ± 0.4	22	0.2	-8.5 ± 0.7	49	0.5	-9.1 ± 0.5	29	0.3	-8.3 ± 0.7	155	0.4
	P	-8.6 ± 0.6	56	0.6	--	--	0.0	-8.0 ± 0.3	15	0.1	-9.0 ± 0.3	16	0.2	-8.5 ± 0.6	87	0.2
	Q	-8.4 ± 0.8	137	1.4	--	--	0.0	-8.9 ± 0.5	4	0.0	-7.8 ± 0.4	4	0.0	-8.4 ± 0.8	145	0.4

Table S3.5: Complete chondroitin sulfate docking results

AutoDock Vina docking scores (Binding energy, BE) are reported in kcal/mol. Pop, or population is number of poses in each site. Score is binding site importance score.

		Conf 1			Conf 2			Conf 3			Conf 4			Avg		
	Site	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Total Pop	Score
	A	-7.8 ± 0.3	19	0.4	-7.8 ± 0.0	1	0.0	--	--	0.0	-7.6 ± 0.1	10	0.2	-7.8 ± 0.3	30	0.1
	B	--	--	0.0	-8.6 ± 0.5	70	1.5	-7.7 ± 0.2	38	0.7	-7.8 ± 0.2	24	0.5	-8.2 ± 0.6	132	0.7
	C	-7.8 ± 0.2	56	1.1	--	--	0.0	-7.9 ± 0.2	14	0.3	-8.1 ± 0.3	35	0.7	-7.9 ± 0.3	105	0.5
	D	-7.7 ± 0.3	40	0.8	-8.2 ± 0.2	101	2.1	-8.1 ± 0.2	25	0.5	--	--	0.0	-8.1 ± 0.3	166	0.8
	E	-8.2 ± 0.4	78	1.6	--	--	0.0	-7.8 ± 0.2	53	1.0	--	--	0.0	-8.0 ± 0.4	131	0.7
	F	-7.6 ± 0.3	17	0.3	--	--	0.0	-7.8 ± 0.2	31	0.6	-9.0 ± 0.3	3	0.1	-7.8 ± 0.4	51	0.2
	G	-7.3 ± 0.2	4	0.1	--	--	0.0	-7.4 ± 0.0	1	0.0	-8.1 ± 1.2	41	0.8	-8.1 ± 1.2	46	0.2
	H	--	--	0.0	--	--	0.0	-8.0 ± 0.2	15	0.3	-8.1 ± 0.3	12	0.2	-8.1 ± 0.2	27	0.1
2mer	I	--	--	0.0	-8.2 ± 0.1	22	0.4	--	--	0.0	--	--	0.0	-8.2 ± 0.1	22	0.1
	J	-7.9 ± 0.4	48	0.9	-8.5 ± 0.2	11	0.2	-7.8 ± 0.3	45	0.9	-8.1 ± 0.4	47	0.9	-7.9 ± 0.4	151	0.7
	K	-7.5 ± 0.1	20	0.4	--	--	0.0	-7.8 ± 0.2	81	1.6	-7.8 ± 0.2	62	1.2	-7.8 ± 0.2	163	0.8
	L	--	--	0.0	-8.0 ± 0.1	75	1.5	-7.6 ± 0.0	3	0.1	--	--	0.0	-8.0 ± 0.2	78	0.4
	M	--	--	0.0	-8.5 ± 0.4	84	1.8	-7.6 ± 0.1	15	0.3	-8.4 ± 0.4	54	1.1	-8.3 ± 0.5	153	0.8
	N	--	--	0.0	-8.1 ± 0.0	1	0.0	-7.6 ± 0.1	29	0.6	-7.8 ± 0.2	52	1.0	-7.7 ± 0.2	82	0.4
	O	-7.7 ± 0.2	39	0.7	-8.1 ± 0.1	32	0.6	-7.8 ± 0.2	35	0.7	-7.9 ± 0.2	24	0.5	-7.9 ± 0.2	130	0.6
	P	-7.7 ± 0.2	23	0.4	-7.9 ± 0.0	1	0.0	-7.6 ± 0.0	2	0.0	-7.7 ± 0.2	36	0.7	-7.7 ± 0.2	62	0.3
	Q	-7.7 ± 0.3	56	1.1	-8.0 ± 0.0	2	0.0	-7.8 ± 0.1	13	0.3	--	--	0.0	-7.7 ± 0.3	71	0.3
	A	-8.2 ± 0.3	4	0.1	--	--	0.0	--	--	0.0	-8.2 ± 0.2	4	0.1	-8.2 ± 0.2	8	0.0
	B	-8.4 ± 0.4	78	1.6	-9.3 ± 0.5	23	0.5	-8.6 ± 0.3	31	0.7	-8.6 ± 0.4	88	1.9	-8.6 ± 0.5	220	1.2
	C	-7.9 ± 0.4	13	0.3	-8.7 ± 0.3	2	0.0	--	--	0.0	-8.4 ± 0.3	9	0.2	-8.1 ± 0.5	24	0.1
	D	-8.3 ± 0.3	9	0.2	-9.0 ± 0.2	14	0.3	--	--	0.0	-8.1 ± 0.2	7	0.1	-8.6 ± 0.5	30	0.2
	E	-8.4 ± 0.4	45	0.9	--	--	0.0	-8.8 ± 0.3	118	2.6	--	--	0.0	-8.7 ± 0.4	163	0.9
	F	-8.5 ± 0.5	137	2.9	--	--	0.0	-8.2 ± 0.2	2	0.0	-8.6 ± 0.4	56	1.2	-8.5 ± 0.4	195	1.0
	G	-8.1 ± 0.3	18	0.4	--	--	0.0	-8.4 ± 0.1	3	0.1	-8.6 ± 0.3	14	0.3	-8.3 ± 0.4	35	0.2
	H	--	--	0.0	--	--	0.0	--	--	0.0	--	--	0.0	--	--	0.0
4mer	I	--	--	0.0	--	--	0.0	--	--	0.0	--	--	0.0	--	--	0.0
	J	-8.4 ± 0.4	20	0.4	--	--	0.0	-8.3 ± 0.0	1	0.0	-8.8 ± 0.3	75	1.7	-8.7 ± 0.3	96	0.5
	K	-8.5 ± 0.3	25	0.5	-8.7 ± 0.1	3	0.1	-8.7 ± 0.3	74	1.6	-8.6 ± 0.4	85	1.8	-8.6 ± 0.3	187	1.0
	L	-8.0 ± 0.2	5	0.1	-9.2 ± 0.4	163	3.7	-8.7 ± 0.3	47	1.0	-8.3 ± 0.3	9	0.2	-9.0 ± 0.5	224	1.3
	M	-8.5 ± 0.2	8	0.2	-9.0 ± 0.3	158	3.6	-8.7 ± 0.4	25	0.5	-9.1 ± 0.4	5	0.1	-9.0 ± 0.4	196	1.1
	N	-8.4 ± 0.1	8	0.2	-8.4 ± 0.0	2	0.0	-8.6 ± 0.2	66	1.4	-8.6 ± 0.3	38	0.8	-8.6 ± 0.3	114	0.6
	O	-8.2 ± 0.4	11	0.2	-8.8 ± 0.2	24	0.5	-8.6 ± 0.3	32	0.7	-8.5 ± 0.2	2	0.0	-8.6 ± 0.3	69	0.4
	P	-8.2 ± 0.2	5	0.1	-8.8 ± 0.2	11	0.2	-8.3 ± 0.0	1	0.0	-8.6 ± 0.4	8	0.2	-8.6 ± 0.4	25	0.1
	Q	-8.4 ± 0.1	14	0.3	--	--	0.0	--	--	0.0	--	--	0.0	-8.4 ± 0.1	14	0.1
	A	-7.9 ± 0.3	23	0.2	-7.8 ± 0.0	1	0.0	--	--	0.0	-7.8 ± 0.3	14	0.1	-7.8 ± 0.3	38	0.1
	B	-8.4 ± 0.4	78	0.8	-8.7 ± 0.6	93	1.0	-8.1 ± 0.5	69	0.7	-8.5 ± 0.5	112	1.2	-8.4 ± 0.5	352	0.9
	C	-7.8 ± 0.3	69	0.7	-8.7 ± 0.3	2	0.0	-7.9 ± 0.2	14	0.1	-8.1 ± 0.3	44	0.4	-7.9 ± 0.3	129	0.3
	D	-7.8 ± 0.4	49	0.5	-8.3 ± 0.3	115	1.2	-8.1 ± 0.2	25	0.3	-8.1 ± 0.2	7	0.1	-8.2 ± 0.4	196	0.5
	E	-8.2 ± 0.4	123	1.3	--	--	0.0	-8.5 ± 0.5	171	1.8	--	--	0.0	-8.4 ± 0.5	294	0.8
	F	-8.4 ± 0.5	154	1.6	--	--	0.0	-7.8 ± 0.2	33	0.3	-8.6 ± 0.4	59	0.6	-8.4 ± 0.5	246	0.6
	G	-7.9 ± 0.4	22	0.2	--	--	0.0	-8.2 ± 0.5	4	0.0	-8.3 ± 1.1	55	0.6	-8.2 ± 0.9	81	0.2
	H	--	--	0.0	--	--	0.0	-8.0 ± 0.2	15	0.2	-8.1 ± 0.3	12	0.1	-8.1 ± 0.2	27	0.1
Avg	I	--	--	0.0	-8.2 ± 0.1	22	0.2	--	--	0.0	--	--	0.0	-8.2 ± 0.1	22	0.1
	J	-8.0 ± 0.5	68	0.7	-8.5 ± 0.2	11	0.1	-7.8 ± 0.3	46	0.4	-8.5 ± 0.5	122	1.3	-8.2 ± 0.5	247	0.6
	K	-8.1 ± 0.5	45	0.5	-8.7 ± 0.1	3	0.0	-8.2 ± 0.5	155	1.6	-8.3 ± 0.5	147	1.5	-8.2 ± 0.5	350	0.9
	L	-8.0 ± 0.2	5	0.0	-8.8 ± 0.7	238	2.6	-8.6 ± 0.4	50	0.5	-8.3 ± 0.3	9	0.1	-8.7 ± 0.6	302	0.8
	M	-8.5 ± 0.2	8	0.1	-8.8 ± 0.4	242	2.7	-8.3 ± 0.6	40	0.4	-8.4 ± 0.5	59	0.6	-8.7 ± 0.5	349	0.9
	N	-8.4 ± 0.1	8	0.1	-8.3 ± 0.2	3	0.0	-8.3 ± 0.5	95	1.0	-8.1 ± 0.5	90	0.9	-8.2 ± 0.5	196	0.5
	O	-7.8 ± 0.3	50	0.5	-8.4 ± 0.4	56	0.6	-8.2 ± 0.4	67	0.7	-8.0 ± 0.2	26	0.3	-8.1 ± 0.5	199	0.5
	P	-7.8 ± 0.3	28	0.3	-8.7 ± 0.3	12	0.1	-7.8 ± 0.4	3	0.0	-7.9 ± 0.4	44	0.4	-8.0 ± 0.5	87	0.2
	Q	-7.9 ± 0.4	70	0.7	-8.0 ± 0.0	2	0.0	-7.8 ± 0.1	13	0.1	--	--	0.0	-7.9 ± 0.4	85	0.2

Table S3.6 Complete dextran sulfate docking results

AutoDock Vina docking scores (Binding energy, BE) are reported in kcal/mol. Pop, or population is number of poses in each site. Score is binding site importance score.

		Conf 1			Conf 2			Conf 3			Conf 4			Avg		
	Site	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Total Pop	Score
	A	--	--	0.0	--	--	0.0	--	--	0.0	-8.0 ± 0.4	26	0.5	-8.0 ± 0.4	26	0.1
	B	--	--	0.0	-8.0 ± 0.3	79	1.6	-7.8 ± 0.0	2	0.0	-8.4 ± 0.5	16	0.3	-8.1 ± 0.4	97	0.5
	C	-7.5 ± 0.2	42	0.8	-7.9 ± 0.0	1	0.0	-7.8 ± 0.2	26	0.5	-7.7 ± 0.2	43	0.8	-7.7 ± 0.2	112	0.5
	D	-7.9 ± 0.2	6	0.1	-8.1 ± 0.2	56	1.1	-8.1 ± 0.3	58	1.2	--	--	0.0	-8.1 ± 0.3	120	0.6
	E	-7.6 ± 0.3	73	1.4	--	--	0.0	-8.2 ± 0.3	14	0.3	--	--	0.0	-7.7 ± 0.4	87	0.4
	F	-7.7 ± 0.1	9	0.2	--	--	0.0	-7.9 ± 0.2	50	1.0	-7.8 ± 0.3	21	0.4	-7.9 ± 0.2	80	0.4
	G	--	--	0.0	--	--	0.0	--	--	0.0	-7.8 ± 0.3	43	0.8	-7.8 ± 0.3	43	0.2
	H	--	--	0.0	--	--	0.0	-7.9 ± 0.2	22	0.4	-7.8 ± 0.2	69	1.3	-7.8 ± 0.2	91	0.4
2mer	I	-7.6 ± 0.3	19	0.4	-8.2 ± 0.3	56	1.1	--	--	0.0	--	--	0.0	-8.0 ± 0.4	75	0.4
	J	-8.0 ± 0.5	79	1.6	-8.3 ± 0.3	34	0.7	-7.7 ± 0.1	23	0.4	-7.9 ± 0.3	73	1.4	-8.0 ± 0.4	209	1.0
	K	--	--	0.0	--	--	0.0	-7.8 ± 0.3	60	1.2	-7.6 ± 0.1	6	0.1	-7.8 ± 0.3	66	0.3
	L	-7.9 ± 0.0	3	0.1	-7.9 ± 0.2	99	2.0	-8.0 ± 0.1	10	0.2	-8.0 ± 0.2	11	0.2	-8.0 ± 0.2	123	0.6
	M	-7.5 ± 0.2	11	0.2	-8.0 ± 0.3	51	1.0	-8.3 ± 0.2	6	0.1	-8.0 ± 0.3	91	1.8	-8.0 ± 0.3	159	0.8
	N	--	--	0.0	-8.3 ± 0.2	13	0.3	-7.8 ± 0.1	21	0.4	--	--	0.0	-8.0 ± 0.3	34	0.2
	O	-7.6 ± 0.3	52	1.0	-7.7 ± 0.2	3	0.1	-7.9 ± 0.2	70	1.4	--	--	0.0	-7.8 ± 0.3	125	0.6
	P	-7.5 ± 0.2	54	1.0	-8.0 ± 0.2	8	0.2	-7.8 ± 0.1	22	0.4	-7.2 ± 0.0	1	0.0	-7.6 ± 0.2	85	0.4
	Q	-7.5 ± 0.3	52	1.0	--	--	0.0	-7.7 ± 0.1	16	0.3	--	--	0.0	-7.5 ± 0.3	68	0.3
	A	-7.5 ± 0.2	10	0.2	-8.4 ± 0.4	9	0.2	-9.1 ± 0.1	3	0.1	-9.2 ± 0.6	36	0.8	-8.8 ± 0.8	58	0.3
	B	-7.7 ± 0.4	86	1.7	-9.4 ± 0.4	121	2.8	-8.9 ± 0.4	56	1.2	-8.7 ± 0.5	19	0.4	-8.7 ± 0.8	282	1.5
	C	-7.5 ± 0.3	6	0.1	-9.0 ± 0.0	1	0.0	-8.5 ± 0.5	10	0.2	-8.3 ± 0.5	3	0.1	-8.2 ± 0.7	20	0.1
	D	-7.5 ± 0.2	9	0.2	-9.1 ± 0.4	38	0.9	-8.9 ± 0.3	5	0.1	--	--	0.0	-8.8 ± 0.7	52	0.3
	E	-7.6 ± 0.4	31	0.6	--	--	0.0	-8.6 ± 0.5	19	0.4	--	--	0.0	-7.9 ± 0.7	50	0.2
	F	-7.7 ± 0.4	215	4.1	--	--	0.0	-8.7 ± 0.4	16	0.3	-8.9 ± 0.5	32	0.7	-7.9 ± 0.6	263	1.3
	G	-7.4 ± 0.2	4	0.1	--	--	0.0	-8.9 ± 0.2	5	0.1	-9.0 ± 0.4	4	0.1	-8.5 ± 0.8	13	0.1
	H	--	--	0.0	--	--	0.0	--	--	0.0	--	--	0.0	--	--	0.0
4mer	I	--	--	0.0	--	--	0.0	--	--	0.0	--	--	0.0	--	--	0.0
	J	-7.7 ± 0.3	7	0.1	--	--	0.0	-9.0 ± 0.4	63	1.4	-9.0 ± 0.3	94	2.1	-8.9 ± 0.4	164	0.9
	K	-7.4 ± 0.0	1	0.0	-8.4 ± 0.6	4	0.1	-8.9 ± 0.6	90	2.0	-8.8 ± 0.6	74	1.6	-8.8 ± 0.6	169	0.9
	L	--	--	0.0	-9.0 ± 0.4	185	4.1	-8.7 ± 0.3	4	0.1	--	--	0.0	-8.9 ± 0.4	189	1.1
	M	--	--	0.0	-9.1 ± 0.3	20	0.5	-9.2 ± 0.6	34	0.8	-8.8 ± 0.5	21	0.5	-9.1 ± 0.5	75	0.4
	N	--	--	0.0	--	--	0.0	-8.7 ± 0.5	31	0.7	-8.8 ± 0.4	77	1.7	-8.8 ± 0.4	108	0.6
	O	-7.6 ± 0.4	22	0.4	-8.5 ± 0.4	17	0.4	-9.1 ± 0.8	8	0.2	-8.9 ± 0.3	26	0.6	-8.4 ± 0.7	73	0.4
	P	-7.3 ± 0.1	3	0.1	-8.5 ± 0.3	5	0.1	-8.7 ± 0.6	34	0.7	-8.6 ± 0.3	14	0.3	-8.6 ± 0.6	56	0.3
	Q	-7.7 ± 0.4	6	0.1	--	--	0.0	-8.4 ± 0.5	22	0.5	--	--	0.0	-8.3 ± 0.6	28	0.1
	A	-7.5 ± 0.2	10	0.1	-8.4 ± 0.4	9	0.1	-9.1 ± 0.1	3	0.0	-8.7 ± 0.8	62	0.7	-8.6 ± 0.8	84	0.2
	B	-7.7 ± 0.4	86	0.8	-8.9 ± 0.8	200	2.2	-8.8 ± 0.5	58	0.6	-8.5 ± 0.5	35	0.4	-8.6 ± 0.8	379	1.0
	C	-7.5 ± 0.2	48	0.5	-8.4 ± 0.5	2	0.0	-8.0 ± 0.4	36	0.4	-7.8 ± 0.3	46	0.4	-7.8 ± 0.4	132	0.3
	D	-7.7 ± 0.3	15	0.1	-8.5 ± 0.6	94	1.0	-8.2 ± 0.4	63	0.6	--	--	0.0	-8.3 ± 0.6	172	0.4
	E	-7.6 ± 0.3	104	1.0	--	--	0.0	-8.4 ± 0.5	33	0.3	--	--	0.0	-7.8 ± 0.5	137	0.3
	F	-7.7 ± 0.4	224	2.1	--	--	0.0	-8.1 ± 0.4	66	0.7	-8.5 ± 0.7	53	0.6	-7.9 ± 0.5	343	0.8
	G	-7.4 ± 0.2	4	0.0	--	--	0.0	-8.9 ± 0.2	5	0.1	-7.9 ± 0.5	47	0.5	-7.9 ± 0.5	56	0.1
	H	--	--	0.0	--	--	0.0	-7.9 ± 0.2	22	0.2	-7.8 ± 0.2	69	0.7	-7.8 ± 0.2	91	0.2
Avg	I	-7.6 ± 0.3	19	0.2	-8.2 ± 0.3	56	0.6	--	--	0.0	--	--	0.0	-8.0 ± 0.4	75	0.2
	J	-8.0 ± 0.5	86	0.9	-8.3 ± 0.3	34	0.4	-8.7 ± 0.7	86	0.9	-8.5 ± 0.6	167	1.8	-8.4 ± 0.6	373	1.0
	K	-7.4 ± 0.0	1	0.0	-8.4 ± 0.6	4	0.0	-8.5 ± 0.7	150	1.6	-8.7 ± 0.6	80	0.9	-8.5 ± 0.7	235	0.6
	L	-7.9 ± 0.0	3	0.0	-8.6 ± 0.6	284	3.1	-8.2 ± 0.4	14	0.1	-8.0 ± 0.2	11	0.1	-8.6 ± 0.6	312	0.8
	M	-7.5 ± 0.2	11	0.1	-8.3 ± 0.6	71	0.7	-9.1 ± 0.6	40	0.5	-8.1 ± 0.5	112	1.1	-8.3 ± 0.6	234	0.6
	N	--	--	0.0	-8.3 ± 0.2	13	0.1	-8.3 ± 0.6	52	0.5	-8.8 ± 0.4	77	0.8	-8.6 ± 0.5	142	0.4
	O	-7.6 ± 0.3	74	0.7	-8.4 ± 0.5	20	0.2	-8.0 ± 0.5	78	0.8	-8.9 ± 0.3	26	0.3	-8.0 ± 0.6	198	0.5
	P	-7.5 ± 0.2	57	0.5	-8.2 ± 0.4	13	0.1	-8.3 ± 0.6	56	0.6	-8.5 ± 0.5	15	0.2	-8.0 ± 0.6	141	0.4
	Q	-7.5 ± 0.3	58	0.5	--	--	0.0	-8.1 ± 0.5	38	0.4	--	--	0.0	-7.7 ± 0.5	96	0.2

Table S3.7: Summary of all docking results

AutoDock Vina docking scores (Binding energy, BE) are reported in kcal/mol. Pop, or population is number of poses in each site. Score is binding site importance score.

		Conf 1			Conf 2			Conf 3			Conf 4			Avg		
	site	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Pop	Score	BE	Total Pop	Score
	A	-7.8 ± 0.3	19	0.1	-7.8 ± 0.0	1	0.0	--	--	0.0	-8.2 ± 0.4	97	0.5	-8.1 ± 0.4	117	0.1
	B	-7.5 ± 0.1	9	0.0	-8.3 ± 0.4	271	1.4	-7.7 ± 0.2	45	0.2	-8.0 ± 0.4	53	0.3	-8.1 ± 0.4	378	0.5
	C	-7.7 ± 0.3	243	1.2	-7.9 ± 0.0	1	0.0	-8.0 ± 0.3	119	0.6	-8.1 ± 0.4	165	0.8	-7.9 ± 0.4	528	0.7
	D	-7.7 ± 0.3	47	0.2	-8.2 ± 0.2	329	1.7	-8.1 ± 0.3	268	1.4	-8.4 ± 0.0	2	0.0	-8.1 ± 0.3	646	0.8
	E	-7.9 ± 0.4	290	1.4	--	--	0.0	-8.0 ± 0.3	87	0.4	--	--	0.0	-7.9 ± 0.4	377	0.5
	F	-7.8 ± 0.3	65	0.3	--	--	0.0	-8.0 ± 0.4	120	0.6	-8.1 ± 0.4	43	0.2	-8.0 ± 0.4	228	0.3
	G	-7.3 ± 0.2	4	0.0	--	--	0.0	-7.4 ± 0.0	1	0.0	-8.2 ± 0.7	183	0.9	-8.1 ± 0.7	188	0.2
	H	--	--	0.0	--	--	0.0	-8.1 ± 0.3	117	0.6	-8.0 ± 0.3	144	0.7	-8.0 ± 0.3	261	0.3
2mer	I	-7.6 ± 0.3	19	0.1	-8.2 ± 0.2	122	0.6	-7.9 ± 0.3	8	0.0	-8.0 ± 0.0	1	0.0	-8.1 ± 0.3	150	0.2
	J	-8.0 ± 0.5	266	1.3	-8.3 ± 0.3	123	0.6	-7.9 ± 0.4	173	0.9	-8.2 ± 0.4	329	1.7	-8.1 ± 0.4	891	1.1
	K	-7.6 ± 0.3	39	0.2	--	--	0.0	-7.9 ± 0.3	240	1.2	-7.8 ± 0.2	83	0.4	-7.8 ± 0.3	362	0.4
	L	-7.9 ± 0.0	3	0.0	-8.1 ± 0.3	246	1.2	-7.9 ± 0.2	14	0.1	-8.1 ± 0.3	35	0.2	-8.1 ± 0.3	298	0.4
	M	-7.5 ± 0.2	14	0.1	-8.2 ± 0.3	365	1.9	-8.1 ± 0.5	35	0.2	-8.3 ± 0.5	329	1.7	-8.2 ± 0.4	743	1.0
	N	--	--	0.0	-8.2 ± 0.2	43	0.2	-7.8 ± 0.2	101	0.5	-7.8 ± 0.2	56	0.3	-7.9 ± 0.3	200	0.2
	1O	-7.6 ± 0.3	187	0.9	-8.1 ± 0.2	83	0.4	-8.0 ± 0.4	167	0.8	-8.0 ± 0.2	33	0.2	-7.8 ± 0.4	470	0.6
	P	-7.8 ± 0.3	177	0.9	-7.9 ± 0.1	14	0.1	-7.8 ± 0.2	76	0.4	-7.7 ± 0.2	37	0.2	-7.8 ± 0.3	304	0.4
	Q	-7.7 ± 0.3	218	1.0	-8.0 ± 0.0	2	0.0	-7.7 ± 0.1	29	0.1	-7.9 ± 0.3	10	0.0	-7.7 ± 0.3	259	0.3
	A	-8.0 ± 0.7	19	0.1	-9.2 ± 0.6	37	0.2	-8.7 ± 0.3	13	0.1	-9.1 ± 0.6	43	0.2	-8.9 ± 0.7	112	0.2
	B	-8.3 ± 0.6	224	1.2	-9.4 ± 0.4	263	1.5	-9.0 ± 0.5	211	1.2	-8.9 ± 0.4	286	1.6	-8.9 ± 0.6	984	1.4
	C	-8.2 ± 0.7	29	0.1	-8.8 ± 0.3	4	0.0	-9.1 ± 0.7	24	0.1	-8.8 ± 0.5	29	0.2	-8.7 ± 0.7	86	0.1
	D	-8.7 ± 0.6	63	0.3	-9.2 ± 0.3	133	0.8	-8.9 ± 0.3	5	0.0	-8.2 ± 0.2	8	0.0	-9.0 ± 0.5	209	0.3
	E	-8.7 ± 0.7	211	1.2	--	--	0.0	-8.9 ± 0.4	337	1.9	--	--	0.0	-8.8 ± 0.5	548	0.8
	F	-8.4 ± 0.7	660	3.5	--	--	0.0	-8.8 ± 0.5	55	0.3	-8.8 ± 0.5	107	0.6	-8.5 ± 0.7	822	1.1
	G	-8.4 ± 0.5	42	0.2	--	--	0.0	-9.0 ± 0.5	41	0.2	-8.8 ± 0.3	31	0.2	-8.7 ± 0.6	114	0.2
	H	--	--	0.0	-9.0 ± 0.1	6	0.0	-9.1 ± 0.5	4	0.0	-8.8 ± 0.2	3	0.0	-9.0 ± 0.3	13	0.0
4mer	I	--	--	0.0	--	--	0.0	--	--	0.0	-8.9 ± 0.3	15	0.1	-8.9 ± 0.3	15	0.0
	J	-8.9 ± 0.7	79	0.4	-8.8 ± 0.0	1	0.0	-9.1 ± 0.5	104	0.6	-9.1 ± 0.4	414	2.4	-9.1 ± 0.5	598	0.8
	K	-8.5 ± 0.4	30	0.2	-8.7 ± 0.5	10	0.1	-8.9 ± 0.5	249	1.4	-8.9 ± 0.4	288	1.6	-8.8 ± 0.5	577	0.8
	L	-8.1 ± 0.3	6	0.0	-9.2 ± 0.4	717	4.1	-8.6 ± 0.3	82	0.4	-8.7 ± 0.7	14	0.1	-9.1 ± 0.4	819	1.2
	M	-8.5 ± 0.2	17	0.1	-9.1 ± 0.3	331	1.9	-9.0 ± 0.5	129	0.7	-9.0 ± 0.4	68	0.4	-9.0 ± 0.4	545	0.8
	N	-8.4 ± 0.1	13	0.1	-8.4 ± 0.0	2	0.0	-8.8 ± 0.5	229	1.3	-8.9 ± 0.4	185	1.0	-8.8 ± 0.4	429	0.6
	O	-8.3 ± 0.7	62	0.3	-8.8 ± 0.3	52	0.3	-8.8 ± 0.6	54	0.3	-9.0 ± 0.4	65	0.4	-8.7± 0.6	233	0.3
	P	-8.7 ± 0.6	35	0.2	-9.0 ± 0.4	44	0.2	-8.6 ± 0.6	35	0.2	-8.8 ± 0.4	44	0.2	-8.8 ± 0.5	158	0.2
	Q	-8.8 ± 0.5	110	0.6	--	--	0.0	-8.5 ± 0.5	28	0.1	--	--	0.0	-8.8 ± 0.5	138	0.2
	A	-7.9 ± 0.5	38	0.1	-9.1 ± 0.6	38	0.1	-8.7 ± 0.3	13	0.0	-8.5 ± 0.7	140	0.4	-8.5 ± 0.7	229	0.2
	B	-8.2 ± 0.6	233	0.6	-8.8 ± 0.7	534	1.5	-8.7 ± 0.7	256	0.7	-8.8 ± 0.5	339	0.9	-8.7 ± 0.7	1362	0.9
	C	-7.7 ± 0.4	272	0.7	-8.7 ± 0.4	5	0.0	-8.2 ± 0.6	143	0.4	-8.2 ± 0.5	194	0.5	-8.0 ± 0.5	614	0.4
	D	-8.3 ± 0.7	110	0.3	-8.5 ± 0.5	462	1.2	-8.1 ± 0.4	273	0.7	-8.2 ± 0.2	10	0.0	-8.3 ± 0.5	855	0.6
	E	-8.3 ± 0.7	501	1.3	--	--	0.0	-8.7 ± 0.5	424	1.1	--	--	0.0	-8.4 ± 0.7	925	0.6
	F	-8.3 ± 0.7	725	1.9	--	--	0.0	-8.3 ± 0.6	175	0.5	-8.6 ± 0.5	150	0.4	-8.4 ± 0.6	1050	0.7
	G	-8.3 ± 0.6	46	0.1	--	--	0.0	-9.0 ± 0.5	42	0.1	-8.3 ± 0.7	214	0.6	-8.4 ± 0.7	302	0.2
	H	--	--	0.0	-9.0 ± 0.1	6	0.0	-8.1 ± 0.3	121	0.3	-8.0 ± 0.3	147	0.4	-8.1 ± 0.4	274	0.2
Avg	I	-7.6 ± 0.3	19	0.0	-8.2 ± 0.2	122	0.3	-7.9 ± 0.3	8	0.0	-8.9 ± 0.4	16	0.0	-8.2 ± 0.4	165	0.1
	J	-8.2 ± 0.7	345	0.9	-8.3 ± 0.3	124	0.3	-8.4 ± 0.7	277	0.7	-8.7 ± 0.6	743	2.0	-8.5 ± 0.7	1489	1.0
	K	-8.0 ± 0.6	69	0.2	-8.7 ± 0.5	10	0.0	-8.4 ± 0.7	489	1.3	-8.6 ± 0.6	371	1.0	-8.5 ± 0.6	939	0.6
	L	-8.0 ± 0.3	9	0.0	-8.9 ± 0.6	963	2.7	-8.5 ± 0.4	96	0.3	-8.3 ± 0.6	49	0.1	-8.8 ± 0.6	1117	0.8
	M	-8.0 ± 0.5	31	0.1	-8.6 ± 0.5	696	1.9	-8.8 ± 0.6	164	0.5	-8.4 ± 0.5	397	1.0	-8.6 ± 0.6	1288	0.9
	N	-8.4 ± 0.1	13	0.0	-8.2 ± 0.2	45	0.1	-8.5 ± 0.6	330	0.9	-8.6 ± 0.6	241	0.6	-8.5 ± 0.6	629	0.4
	O	-7.8 ± 0.5	249	0.6	-8.3 ± 0.4	135	0.4	-8.2 ± 0.5	221	0.6	-8.7 ± 0.6	98	0.3	-8.1 ± 0.6	703	0.4
	P	-7.9 ± 0.5	212	0.5	-8.8 ± 0.6	58	0.2	-8.1 ± 0.5	111	0.3	-8.3 ± 0.6	81	0.2	-8.1 ± 0.6	462	0.3
	Q	-8.1 ± 0.7	328	0.8	-8.0 ± 0.0	2	0.0	-8.1 ± 0.5	57	0.1	-7.9 ± 0.3	10	0.0	-8.1 ± 0.6	397	0.3

Table S3.8: Summary of all BLI assay results

Bioreceptor	Spike	NaCl Conc. (mM)	K_D (nM)
HS15	WT	10	16.7
CS25	WT	10	17.8
HEP27	WT	10	29.5
DEX5	WT	10	22.9
DEX50	WT	10	229
HEP15	WT	10	215
HEP15	WT	75	ND
HEP15	WT	150	ND
HEP15	Alpha	10	376
HEP15	Beta	10	437
HEP15	Delta	10	473
NTD Ab	WT	150	52
NTD Ab	WT	10	83
NTD Ab	Alpha	10	1219
NTD Ab	Beta	10	436
NTD Ab	Delta	10	492

Table S3.9: Summary of all GlycoGrip LF assay results

Bioreceptor (Test line)	Spike	Signal enhancement	Samples	LOD (ng/reaction)
HS15	WT	x	Buffer	78
HEP15	WT	x	Buffer	78
HEP15	WT	O	Buffer	19.5
HEP15	WT	O	Saliva	39

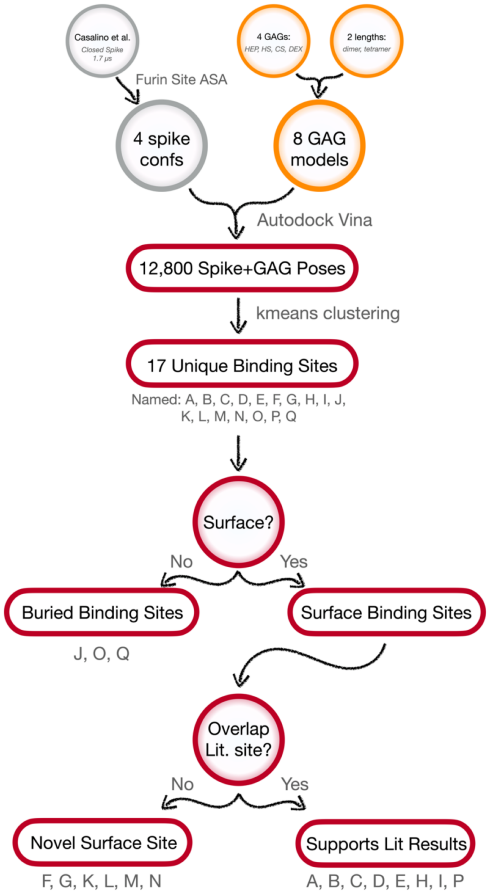
Table S3.10: Comparison of the analytical performance of LFSA for spike detection

Target	Bioreceptor	Detection type	Limit of Detection (ng/reaction)	Ref.
Spike	Primary: ACE2 Secondary: CR3022 Ab	sandwich-type	5 ng/reaction	Lee2020[28]
Spike	Primary: NeuNAc[a] Secondary: NueNAc-PHEA$_{50}$[b]	sandwich-type	200 ng/reaction	Baker2020[29]
Spike	Primary: Heparin Secondary: NTD Ab	sandwich-type	19.5 ng/reaction (Buffer) 78 ng/reaction (Saliva)	This study

[a]NeuNAc: Neuraminic acid, [b]PHEA: Poly(N-hydroxyethyl acrylamide)

3.6.4 Supplemental Schemes

Scheme S3.1: Flowchart demonstrating the complete ensemble-based docking workflow and analysis of resultant binding poses



3.6.5 Supplemental Figures

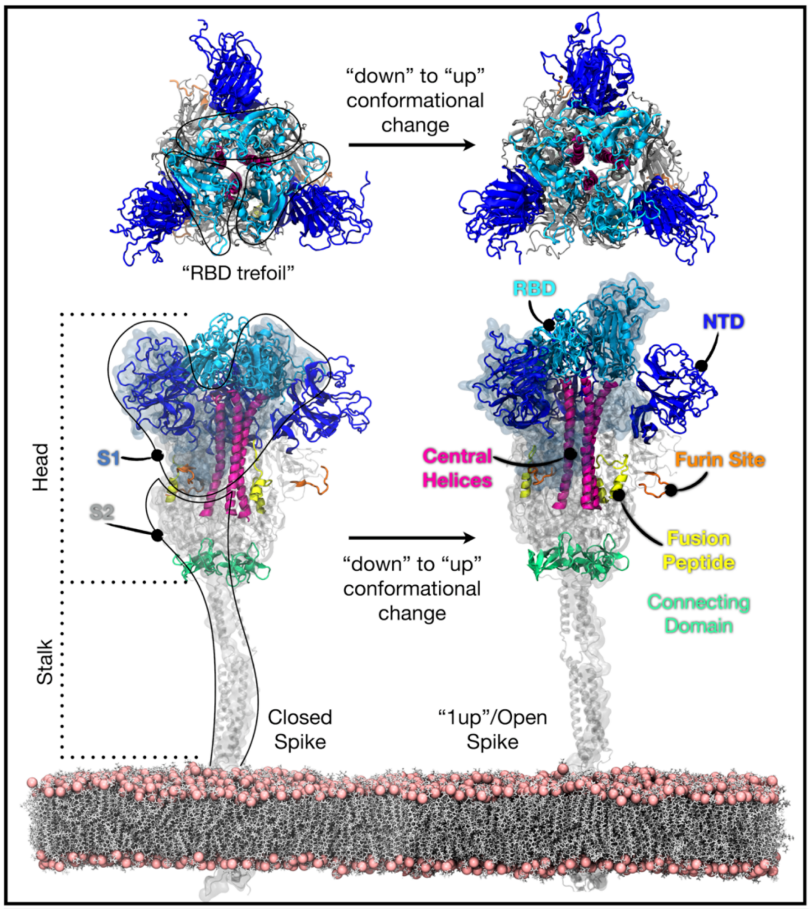


Figure S3.1: Anatomy of the SARS-CoV-2 spike protein with key protein regions highlighted

Spike Receptor Binding Domain (RBD) "down"/closed to "up"/open conformational change shown in top-down view and side-view. See the list presented in Supporting Information Methods section 2 for residue ranges corresponding to each protein region.

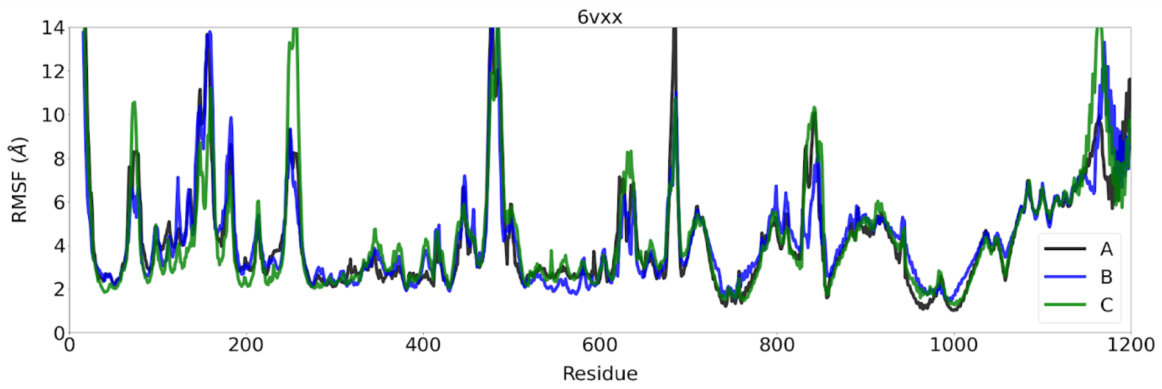


Figure S3.2: RMSF of all residues in spike over the course of 1.8 us simulations from Casalino et al.[1]

The furin cleavage site is located from residues 674 to 685 on each chain.

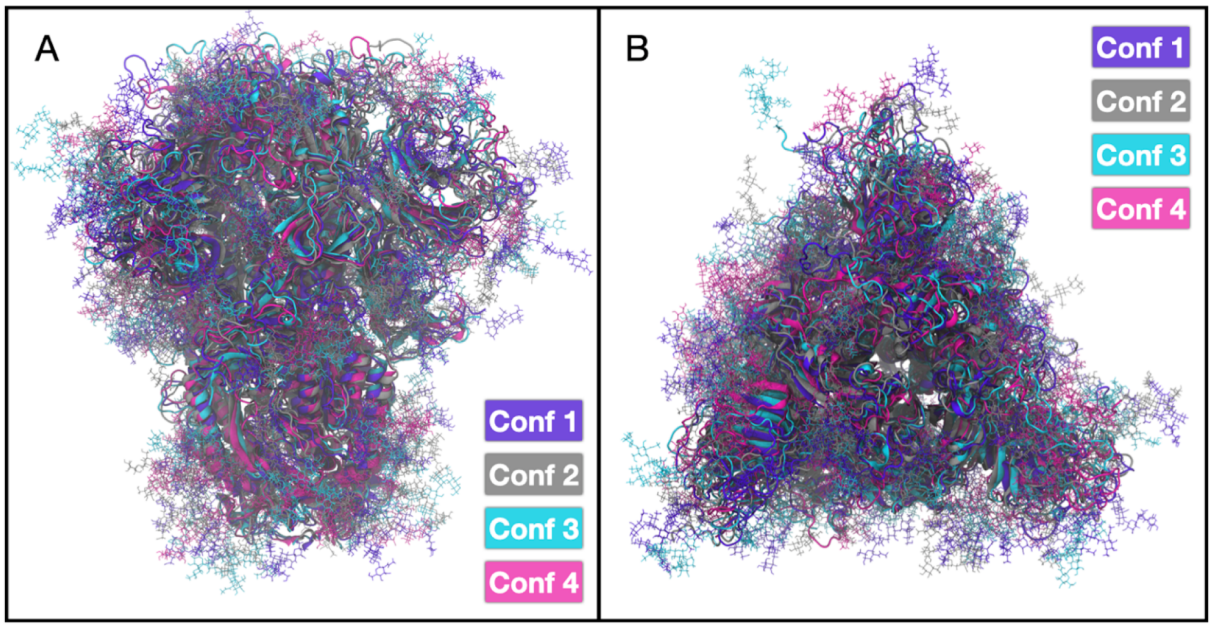


Figure S3.3: Comparison of spike conformations 1 through 4 used in ensemble-based docking studies

Figure was captured after aligning Cα atoms of all structures onto conformation 1. Spike protein structures are colored according to spike conformation number. As one can see, each conformation differs slightly in many areas.

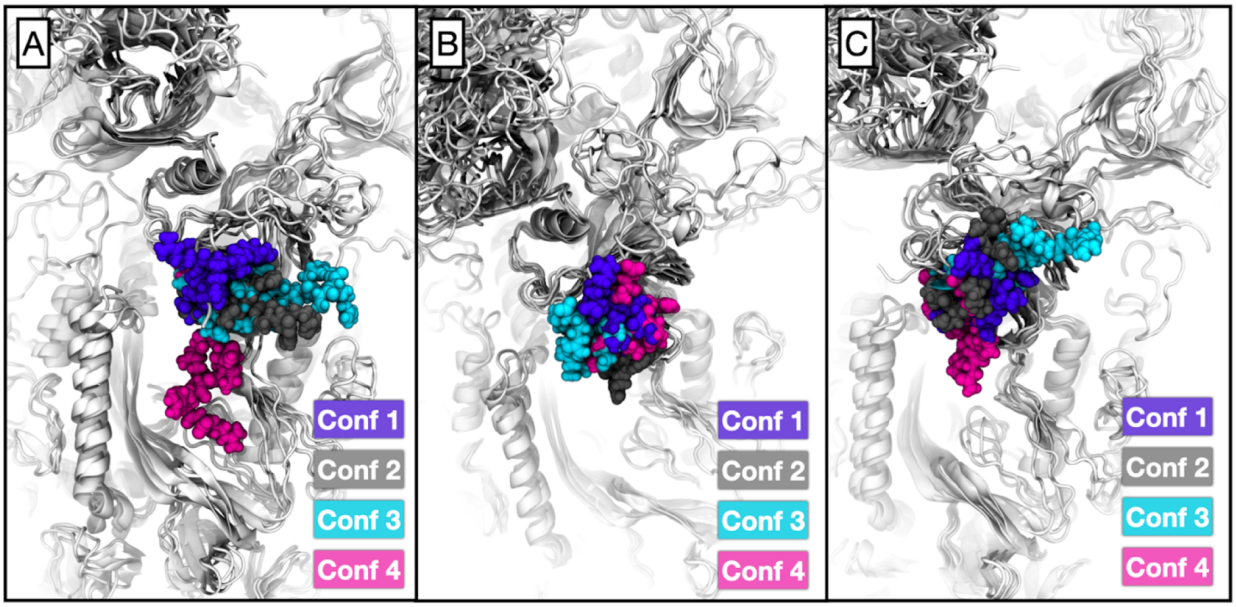


Figure S3.4: Comparison of furin site structures observed in spike conformations 1 through 4 used in ensemble-based docking studies

Figure was captured after aligning all structures onto confrontation 1 by Cα atoms. Furin site residues for each conformation are highlighted in van der Waals representations and colored according to spike conformation number.

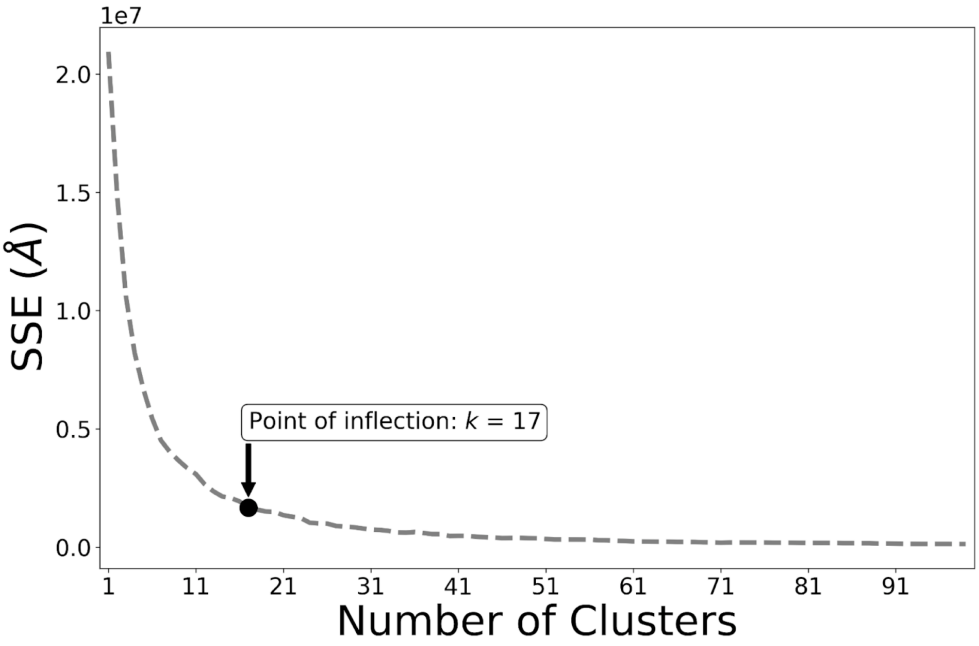


Figure S3.5: Sum of squared error (SSE) for each iteration of K-means clustering over different cluster sizes

"kneed", a scikitlearn module, was used to locate the optimal number of clusters at the SSE inflection point.

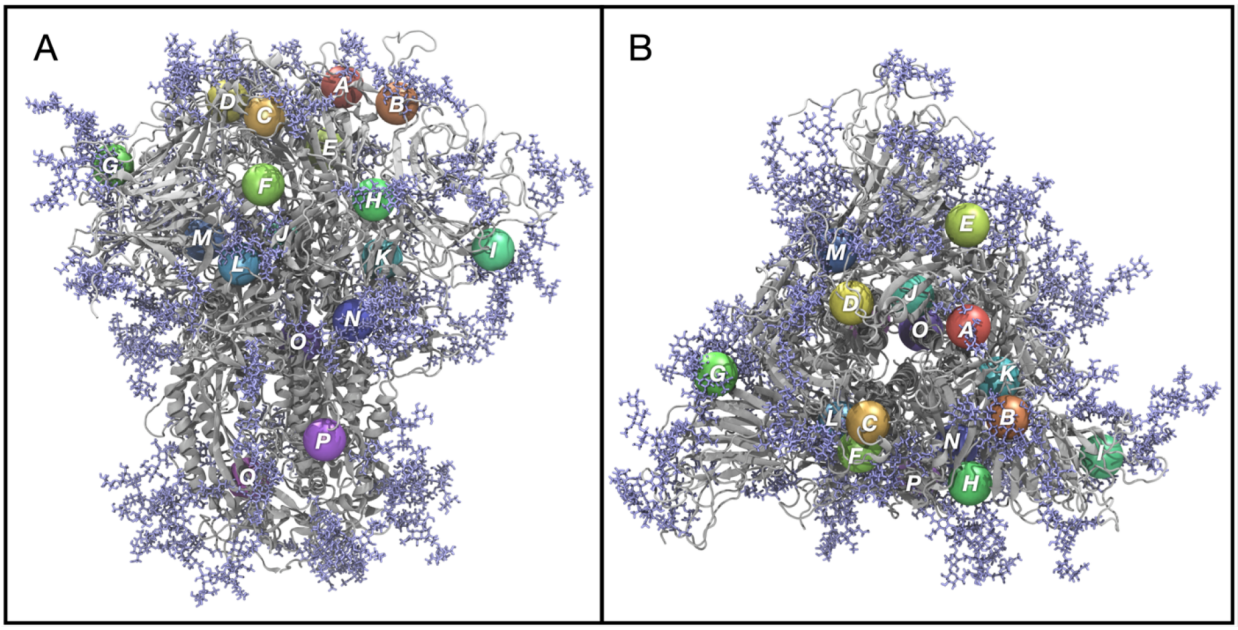


Figure S3.6: Cartoon image of spike with spheres rendered to indicate centers of mass of each predicted GAG binding site, identified by clustering docking results

Spheres are colored by height (z-directional coordinate) along the spike, sites are indexed A through Q and correspond to site indexes as listed in Tables S2-S6.

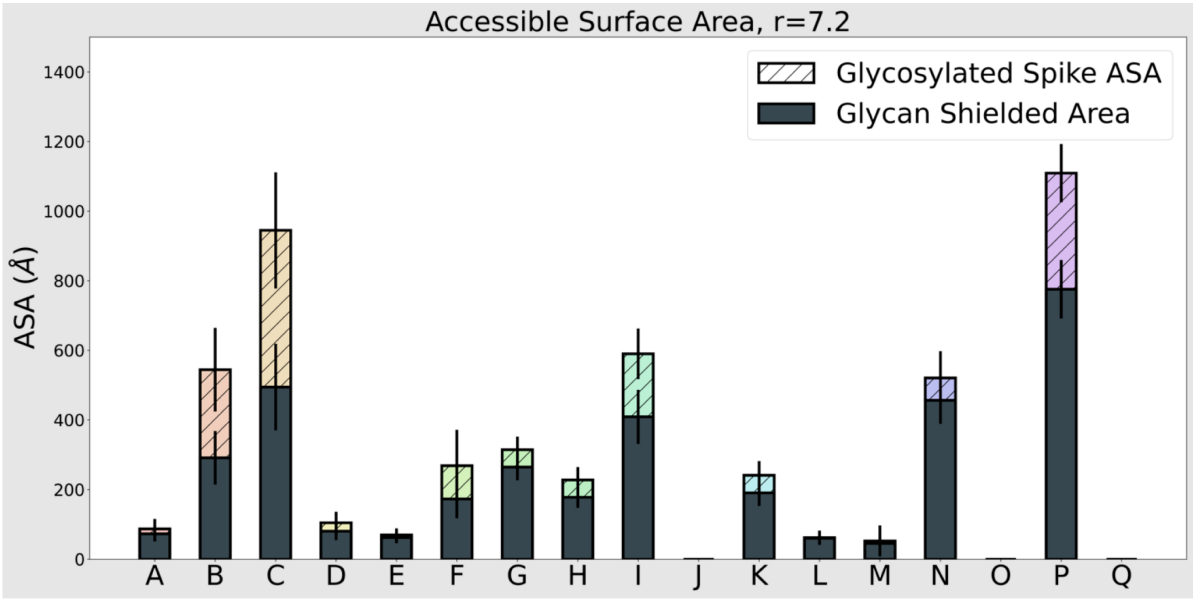


Figure S3.7: ASA calculated at r=7.2 for all identified GAG binding sites, A through Q

As one can see, for sites J, O, Q, this site is completely inaccessible at probe radii consistent with polysaccharide chains (r=7.2 Å). Thus, these sites are excluded from our discussions as they are not surface sites. See SI Methods section 7 for complete details in calculating ASA for each binding site, and SI Results section 2 for discussion of these results.

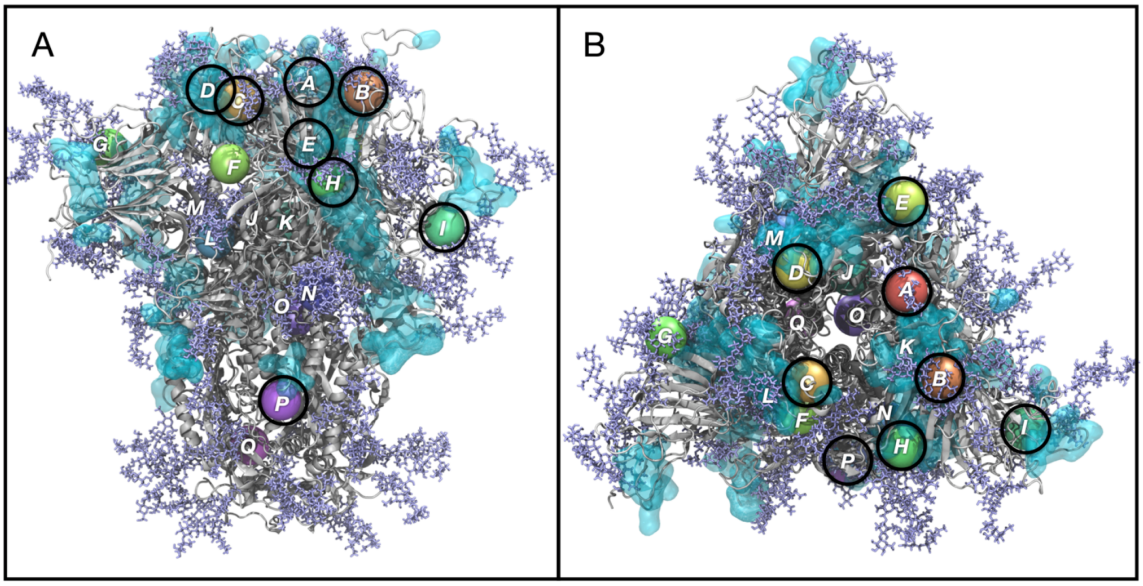


Figure S3.8: Image highlighting which of our 17 predicted binding sites overlaps with literature proposed sites

Rainbow balls indicate cluster centers as found by K-means clustering. Balls are colored on a rainbow spectrum according to height on the spike. Transparent cyan surface representation indicates literature binding site regions. Sites A, B, C, D, E, H, I, and P all overlap with literature proposed binding sites and are highlighted with black circles.

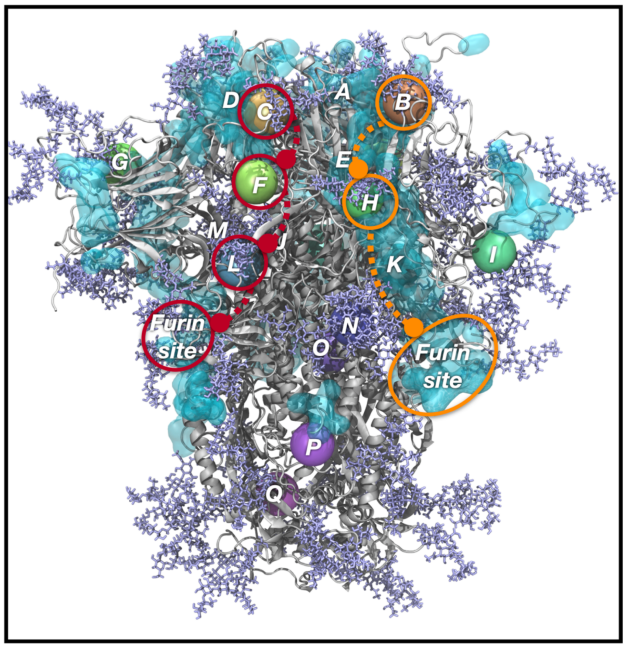


Figure S3.9: Cartoon representation illustrating the two proposed long-chain GAG binding modes which could connect the Esko-Fadda supersite down to the furin cleavage site

The path connecting sites B to H and then down to the furin cleavage site (orange circles and lines) is proposed by Wade and coworkers.[12]

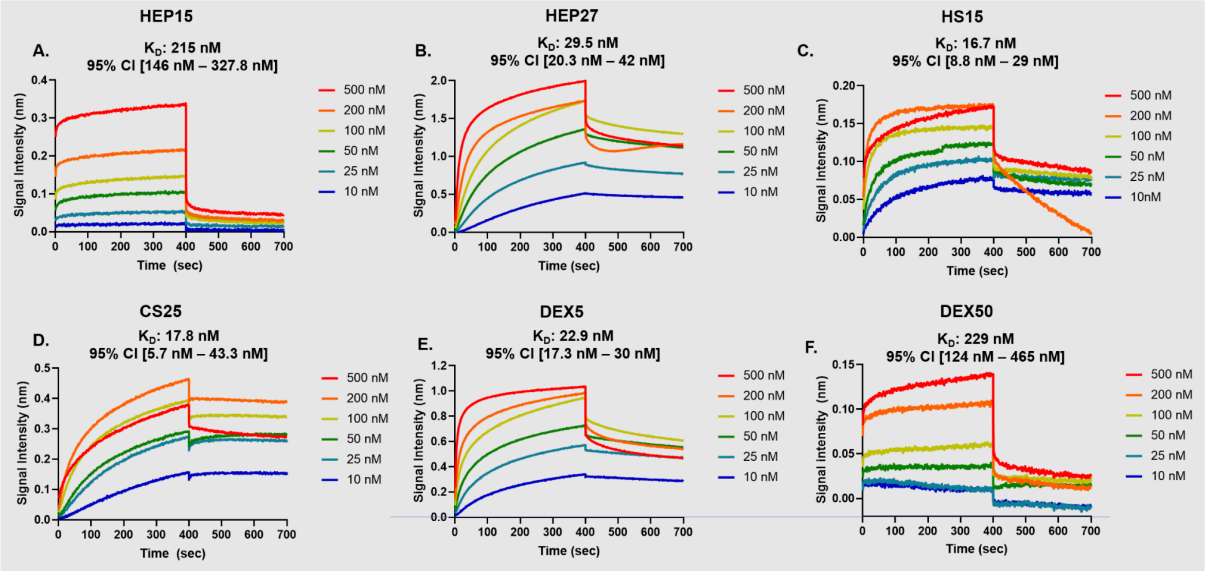


Figure S3.10: BLI curves of various polysaccharides to SARS-CoV-2 spike

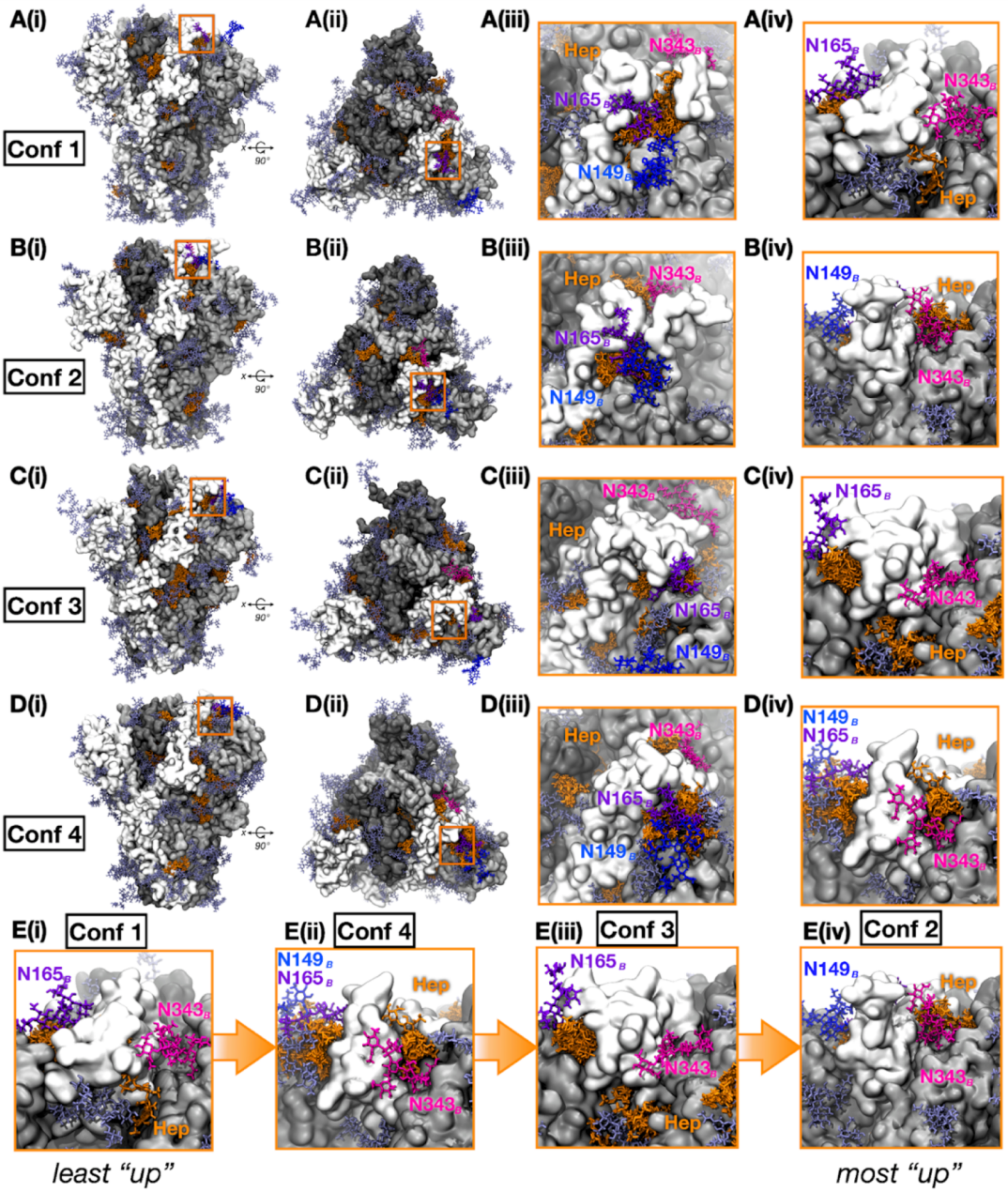


Figure S3.11: Comparison of docking results from site B

For all images: white, grey, and darker grey surface representations indicate spike chains A, B, and C, respectively. Orange licorice atoms indicate docked poses of HEP. Purple, blue, and magenta licorice atoms indicate N165B, N149B, and N343B glycans respectively. Light blue atoms indicate neighboring glycans which are not specifically discussed in the text. Images are organized primarily by spike conformation: (A(i-iv)) conformation 1, (B(i-iv)) conformation 2, (C(i-iv)) conformation 3, (D(i-iv)) conformation 4, (E(i-iv)) a proposed ordering of conformations along a potential `least up" to `most up" pathway. Images A through D are then organized as follows: (X(i)) side view of spike conformation highlighting site B location. (X(ii)) top view of spike conformation highlighting site B location. (X(iii)) top-down zoom-in image of site B. (X(iv)) rotated zoom-in view of site B.

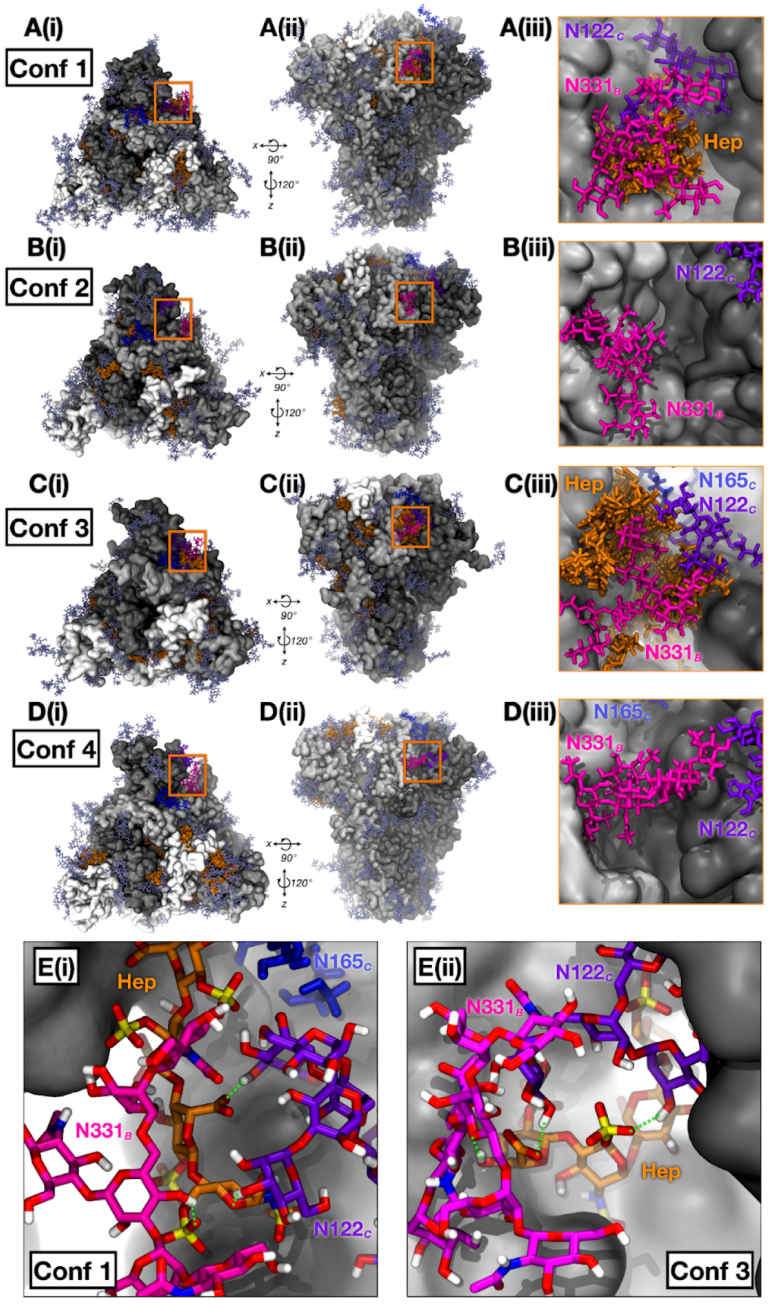


Figure S3.12: Comparison of docking results in site E

For all images: white, grey, and darker grey surface representations indicate spike chains A, B, and C, respectively. Orange licorice atoms indicate docked poses of HEP. Purple, blue, and magenta licorice atoms indicate N122C, N165B, and N331B glycans respectively. Light blue atoms indicate neighboring glycans which are not specifically discussed in the text. Images A through D are organized primarily by spike conformation: (A(i-iii)) conformation 1, (B(i-iii)) conformation 2, (C(i-iii)) conformation 3, (D(i-iii)) conformation 4. Images A through D are then sub-organized as follows: (X(i)) side view of spike conformation highlighting site B location. (X(ii)) top view of spike conformation highlighting site B location. (X(iii)) top-down zoom-in image of site B. (X(iv)) rotated zoom-in view of site B. (E(i-ii)) Highlights of a HEP tetramer pose at site E in (i) spike conformation 1 and (ii) spike conformation 3. (E(i-ii)) illustrate the hydrogen bonding network formed between glycans and GAGs once GAGs do reach this site.

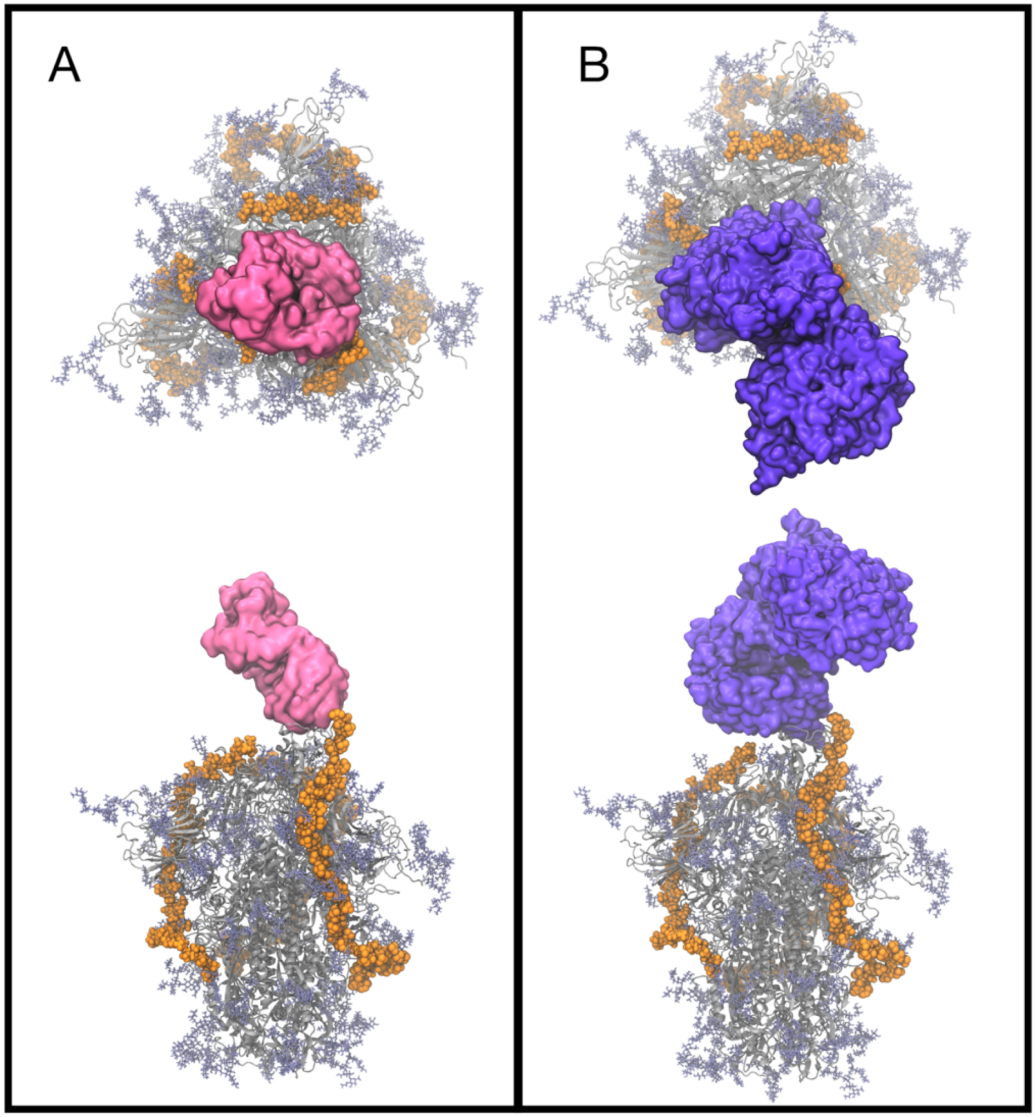


Figure S3.13: Illustration spike-3xhep40mer structure in the 1-up spike conformational state bound to antibodies

Antibodies are (A) REGN10933 Antibody (pink surface) and (B) ACE2 (purple surface). For both A and B: grey ribbons indicate spike protein structure, orange van der Waals atoms indicate 3xhep40mer, and light blue licorice atoms indicate spike glycans.

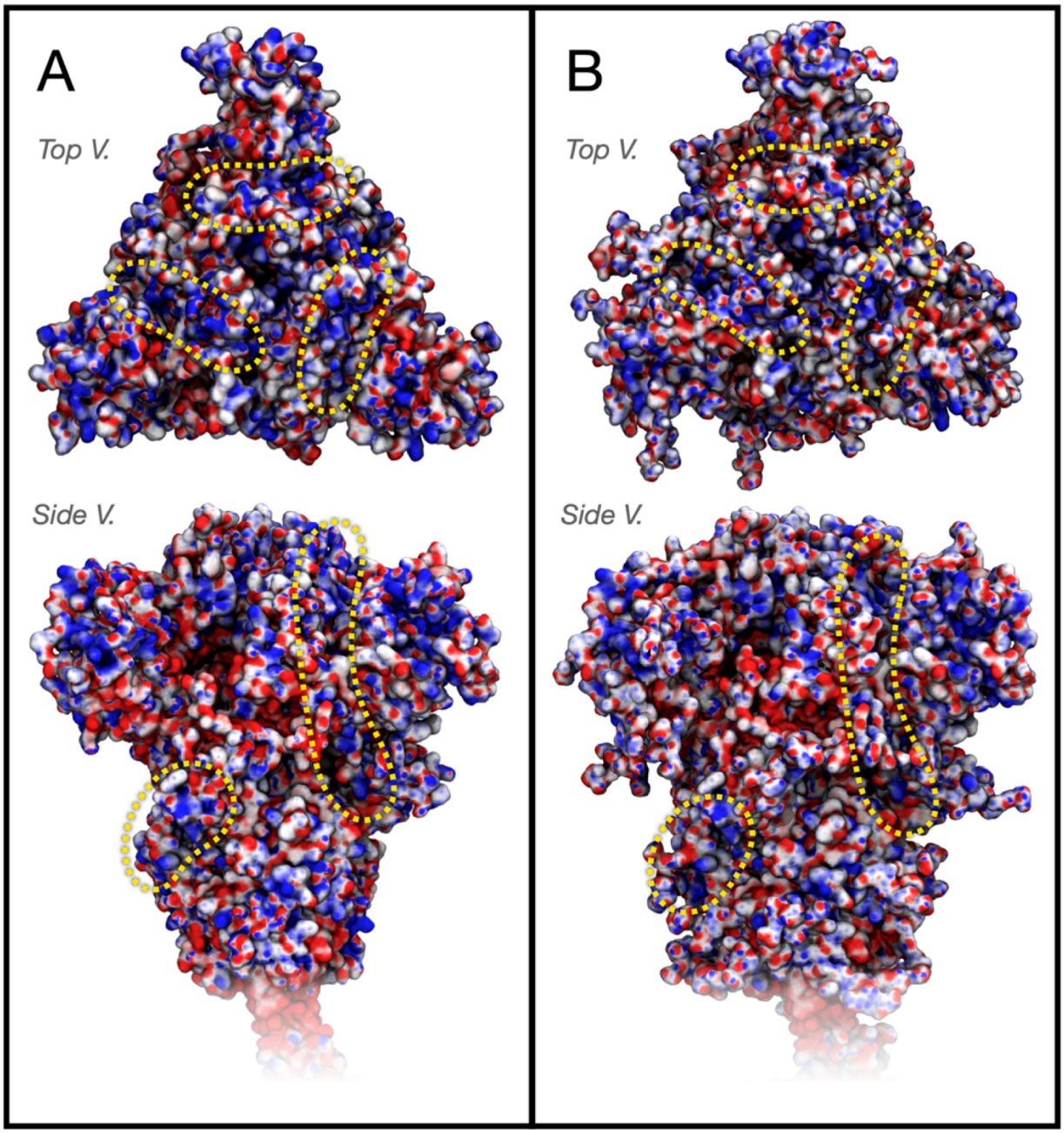


Figure S3.14: ESP maps of WT spike in the “all down” or “closed” state

(A) is without glycans and (B) is with glycans. Yellow indicators outline key partial positive surface patches posited to have high affinity for GAGs.

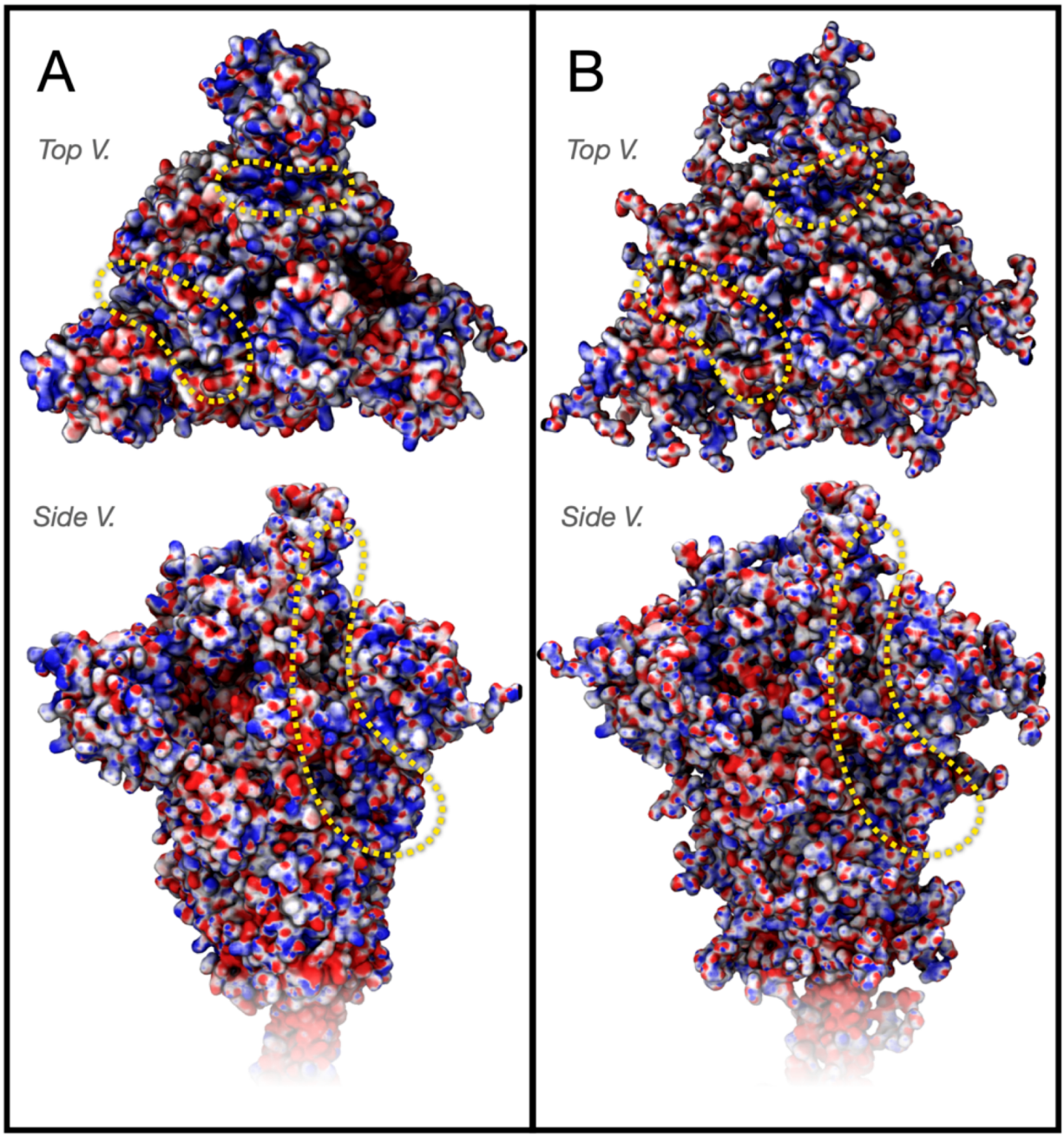


Figure S3.15: ESP maps of WT spike in "1 up," or "open" state

(A) is without glycans and (B) is with glycans. Yellow indicators outline key partial positive surface patches posited to have high affinity for GAGs.

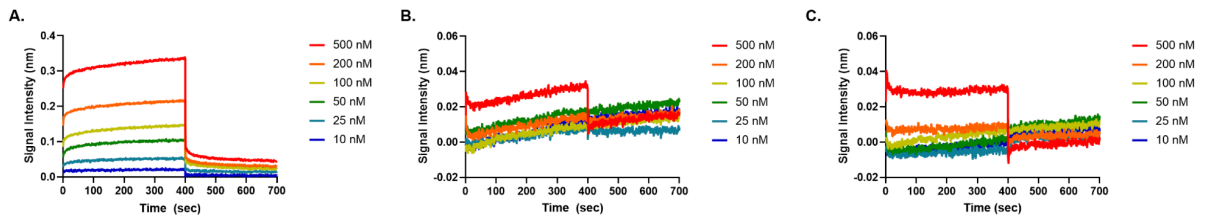


Figure S3.16: BLI curves of HEP15 to SARS-CoV-2 spike in different NaCl concentration
(A) 10 mM NaCl, (B) 75 mM NaCl, and (C) 150 mM NaCl.

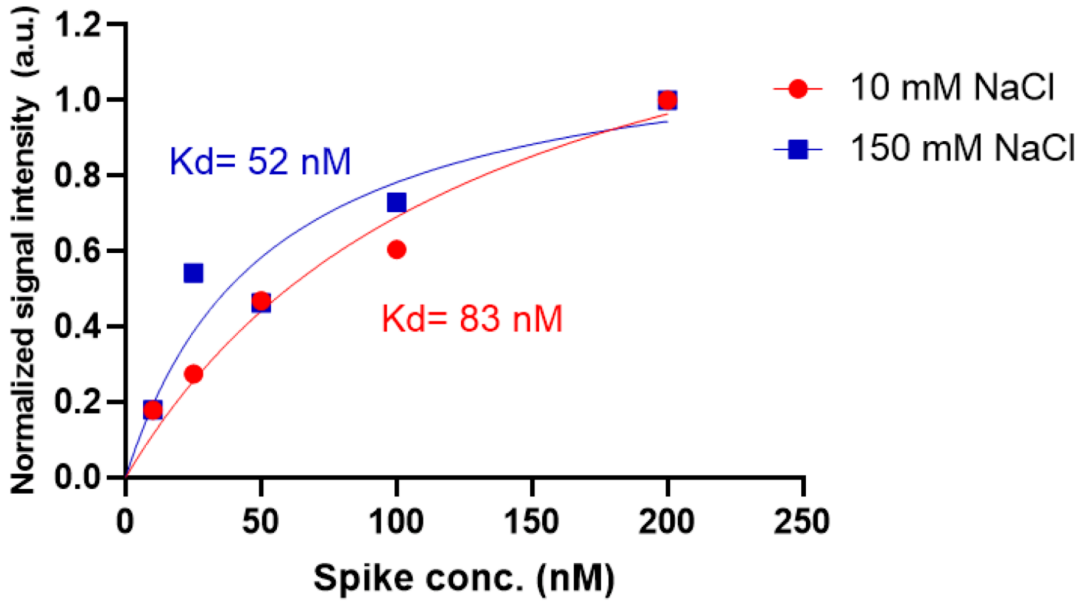


Figure S3.17: Binding curves of NTD antibody to SARS-CoV-2 spike in two different NaCl concentrations (10 mM and 150 mM)

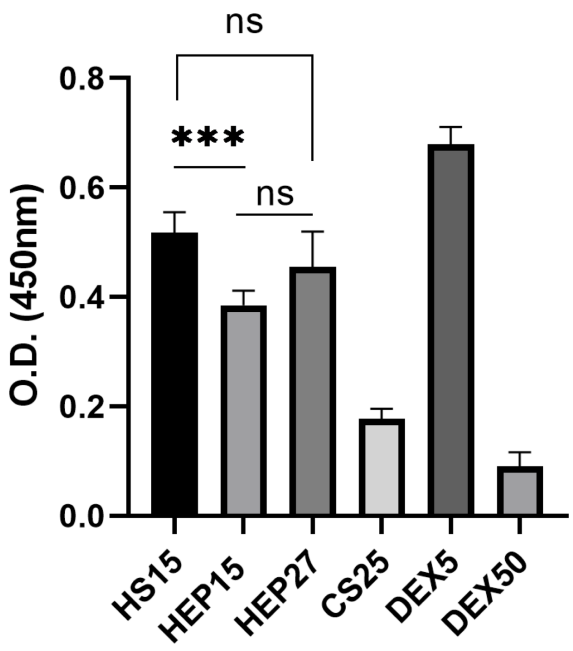


Figure S3.18: Sandwich-type binding of all GAGs and NTD Ab to Wild type (WT) full-length spike protein

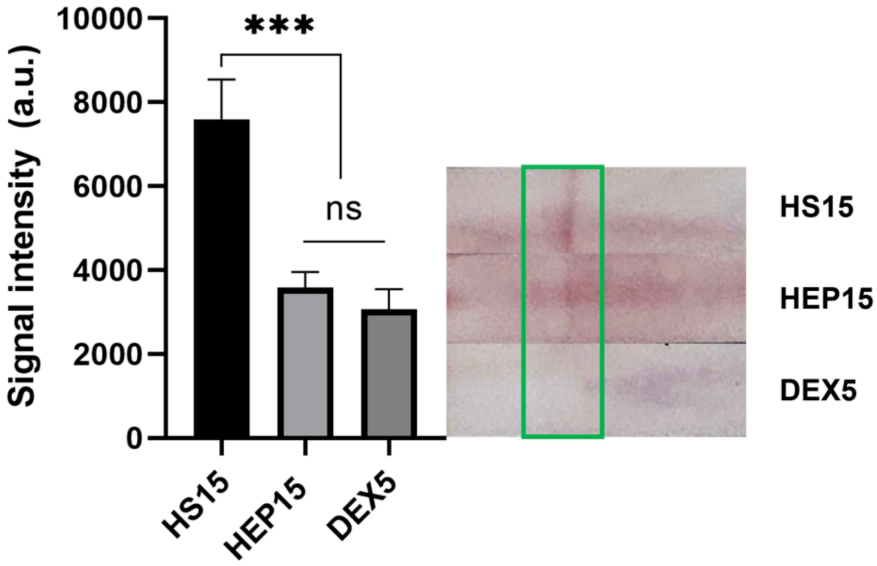


Figure S3.19: Adsorption of GAGs onto the nitrocellulose membrane

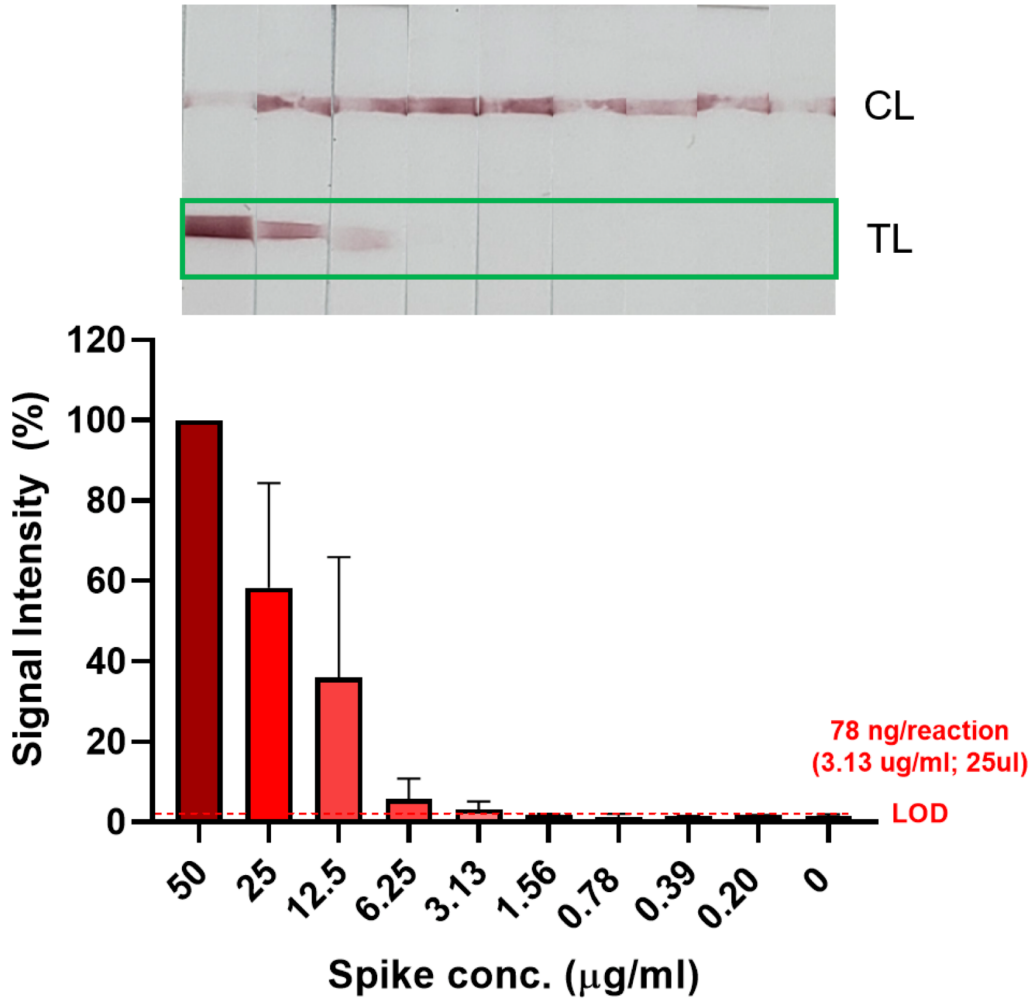


Figure S3.20: Analytical performance of heparan sulfate (HS15) based GlycoGrip LF biosensor

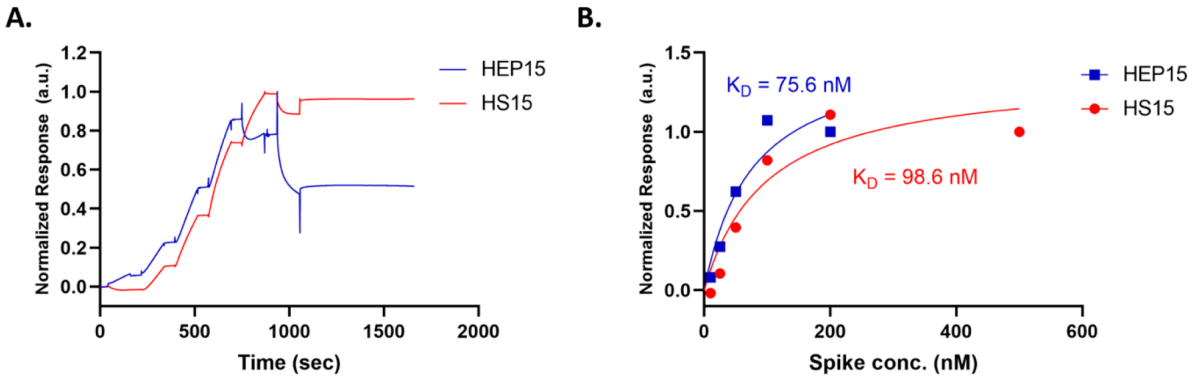


Figure S3.21: Binding of wild-type spike protein to HEP15 and HS15 using SPR

(A) Sensorgram of single cycle kinetic analysis, and (B) corresponding steady-state binding curve.

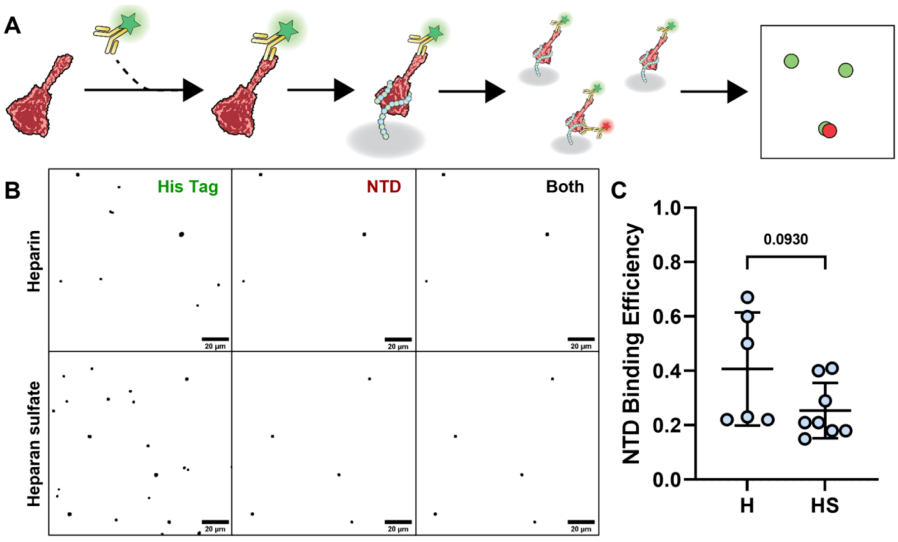


Figure S3.22: NTD Binding Efficiency

(A) Scheme of experimental setup and confocal microscope output. (B) Sample binary confocal images to show location of each protein. Images have been dilated 3x for visibility. (C) NTD Binding efficiency defined as the number of NTD regions divided by the number of points recognized as a his-tag. Statistical analysis was done using a two-tailed student's t-test.

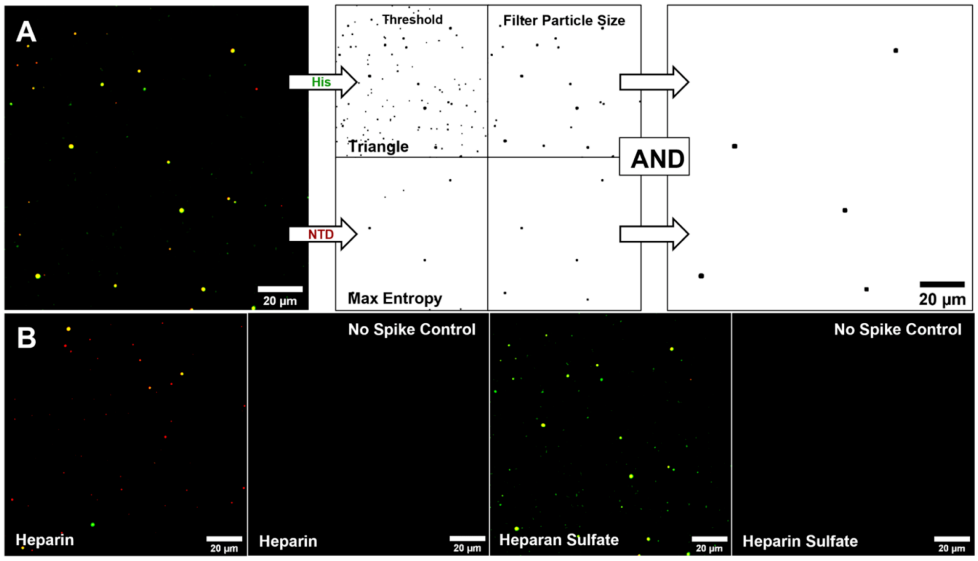


Figure S3.23: Quantification of Confocal Experiment

(A) Method of quantification. Prior to visualization images were blurred (Gaussian blur, 3 pixels) and filtered using a rolling pin filter (50 pixels) to reduce background. Thresholds were created using either triangle or MaxEntropy algorithms. Random noise pixels were then removed by filtering for particle size to only include points above 0.5 um^2. Finally, the binarized images were subjected to an "AND" filter to determine the number of points containing both NTD and His Tag Signal. (B) Images of each condition. Controls were ran with no spike. Instead of pre-incubation, Anti-His Alexa 488 was added during secondary step for controls. All images show the same brightness and contrast to see signal intensities. Control images were not quantified since there was no signal.

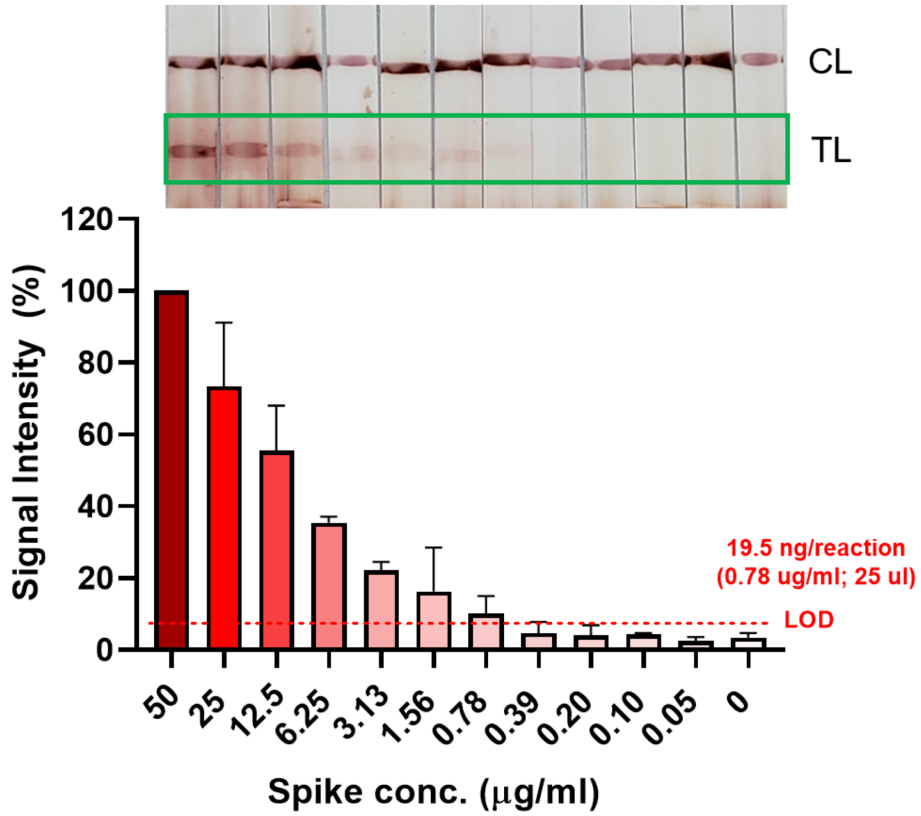


Figure S3.24: Analytical performance of heparin (HEP15) based GlycoGrip LF biosensor with signal enhancement

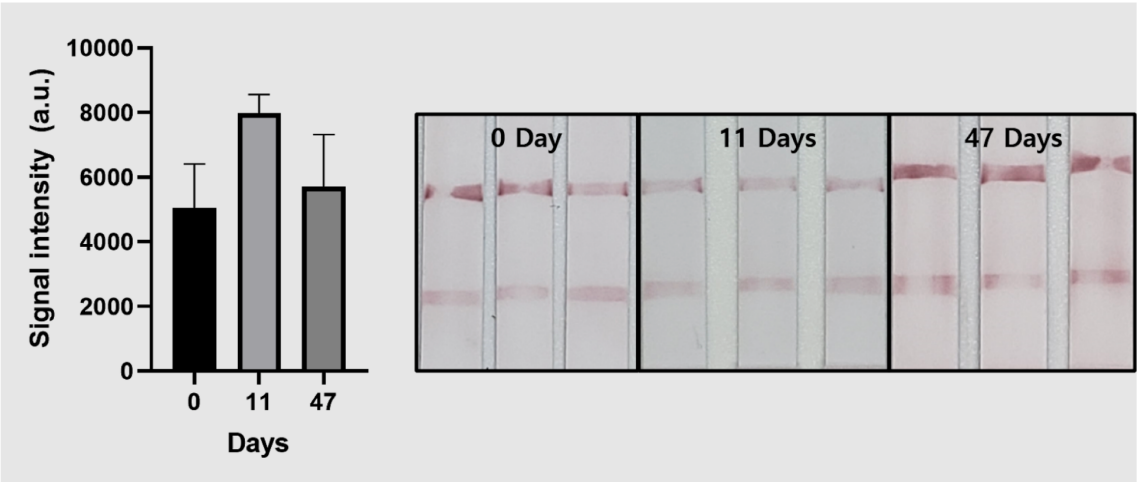


Figure S3.25: Stability testing of the heparin (HEP15) based GlycoGrip LF biosensor

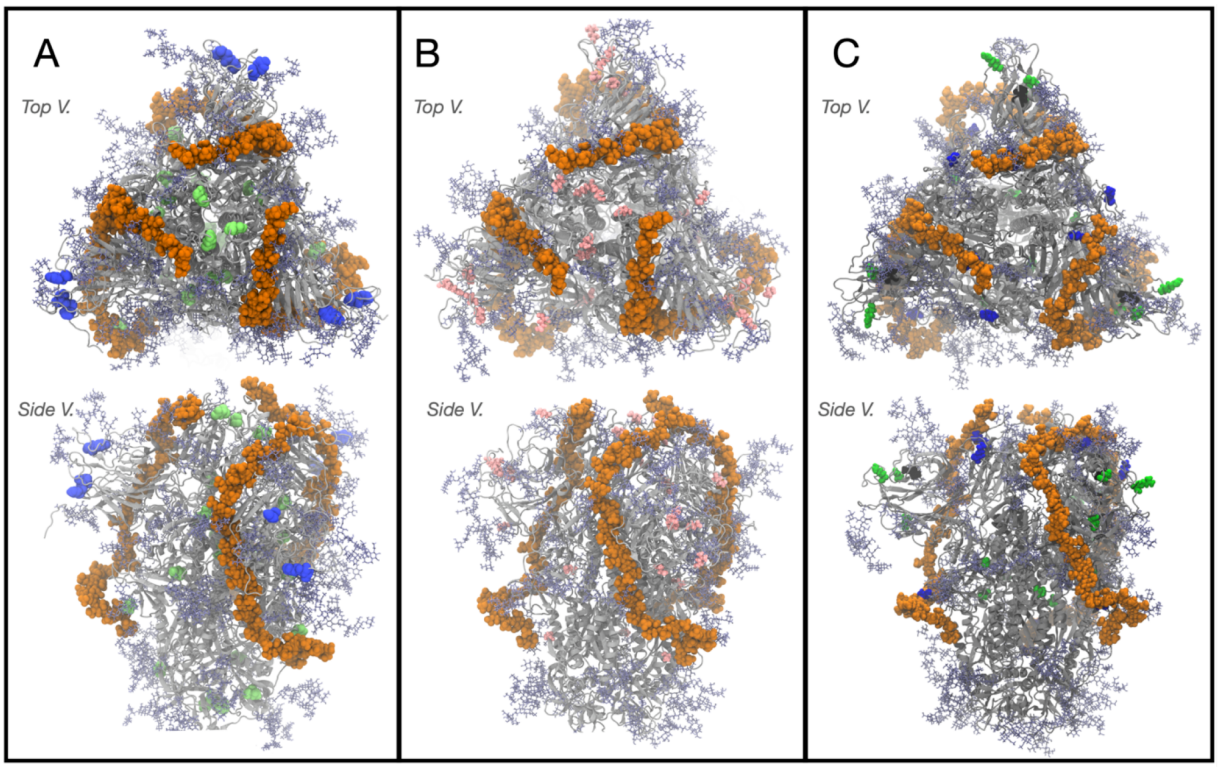


Figure S3.26: Cartoon representations of the Alpha, Beta, and Delta spike constructs examined in this work with trivalent hep40mer aligned onto each variant spike

For all images, the spike structure is shown in grey ribbon, and trivalent hep40mer is represented as orange van der Waals atoms. (A) Cartoon representation of Alpha (B.1.1.7) spike: light green van der Waals atoms represent locations of single point mutations, blue van der Waals atoms represent residues around the NTD deletion regions. (B) Cartoon representation of Beta (B.1.351) spike: pink van der Waals atoms represent locations of single point mutations characterizing the Beta spike. (C) Cartoon representation of Delta (B.1.617.2) spike: green van der Waals atoms represent locations of single point mutations characterizing the Delta spike, blue van der Waals atoms represent locations of single point mutations that lie along the heparin binding site: L452R, T478K, P681R.

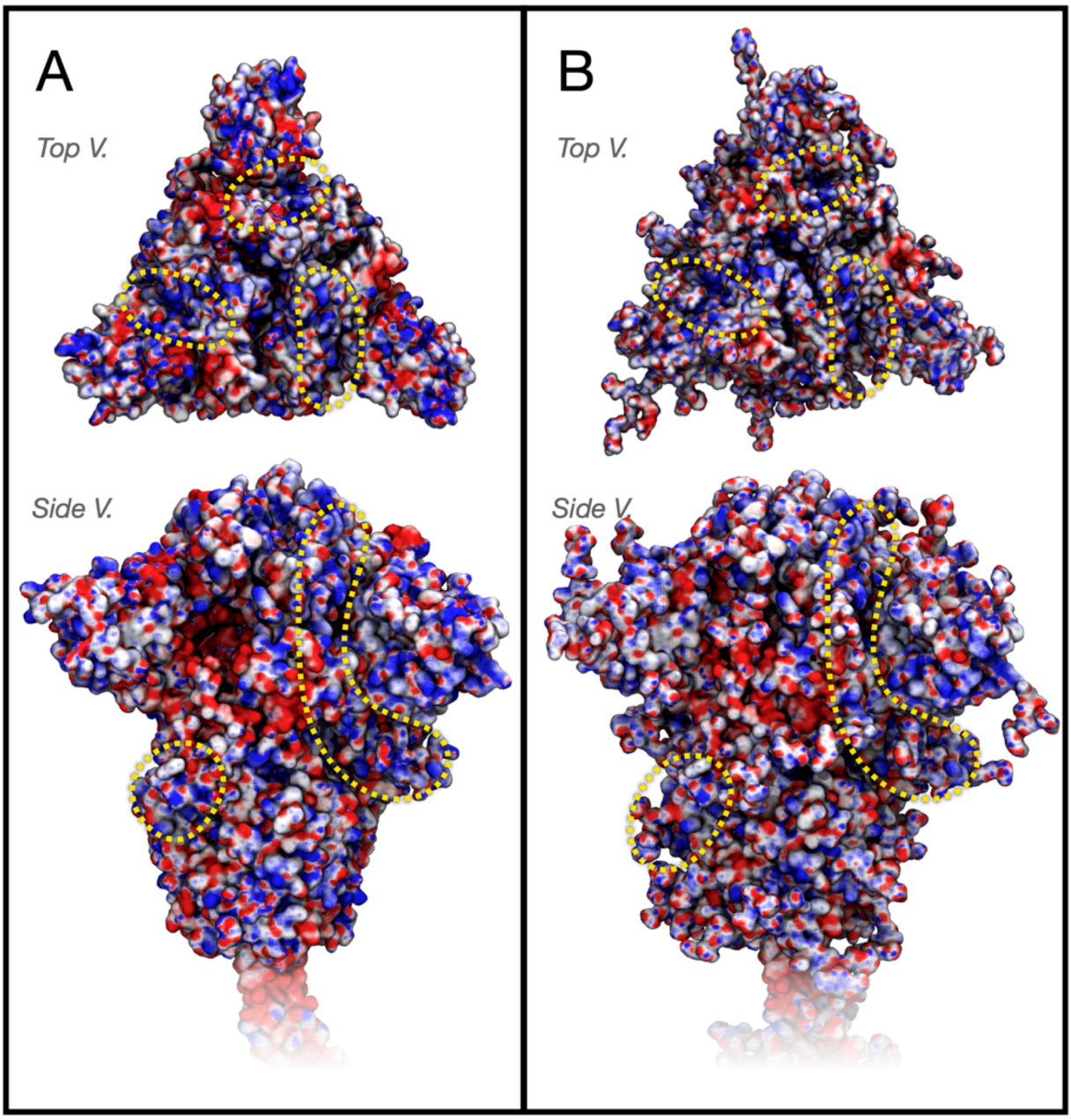


Figure S3.27: ESP Maps of Alpha (B.1.1.7) spike variant

(A) is without glycans and (B) is with glycans. Yellow indicators outline key partial positive surface patches posited to have high affinity for GAGs.

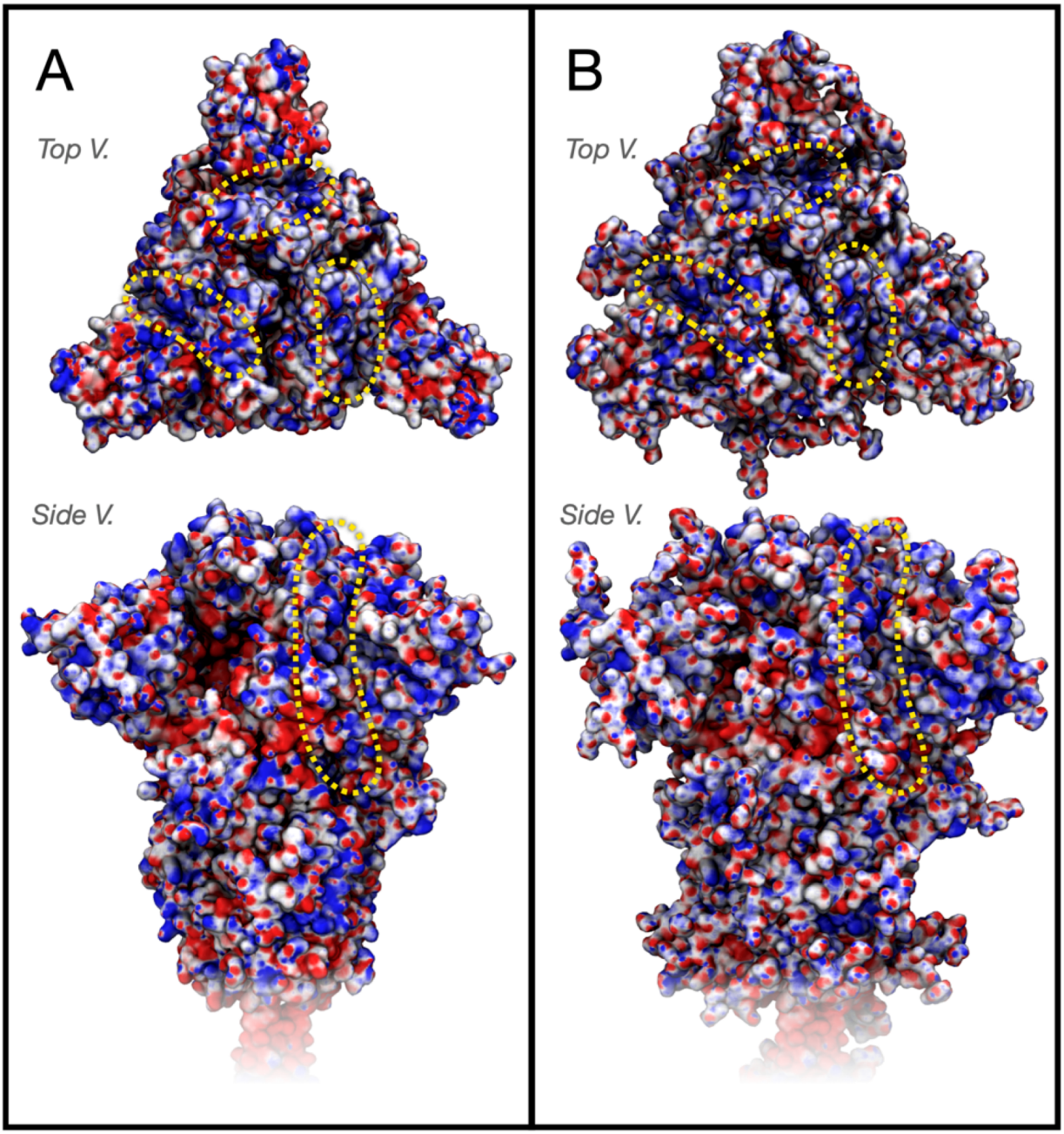


Figure S3.28: ESP maps of Beta (B.1.351) spike variant

(A) is without glycans and (B) is with glycans. Yellow indicators outline key partial positive surface patches posited to have high affinity for GAGs.

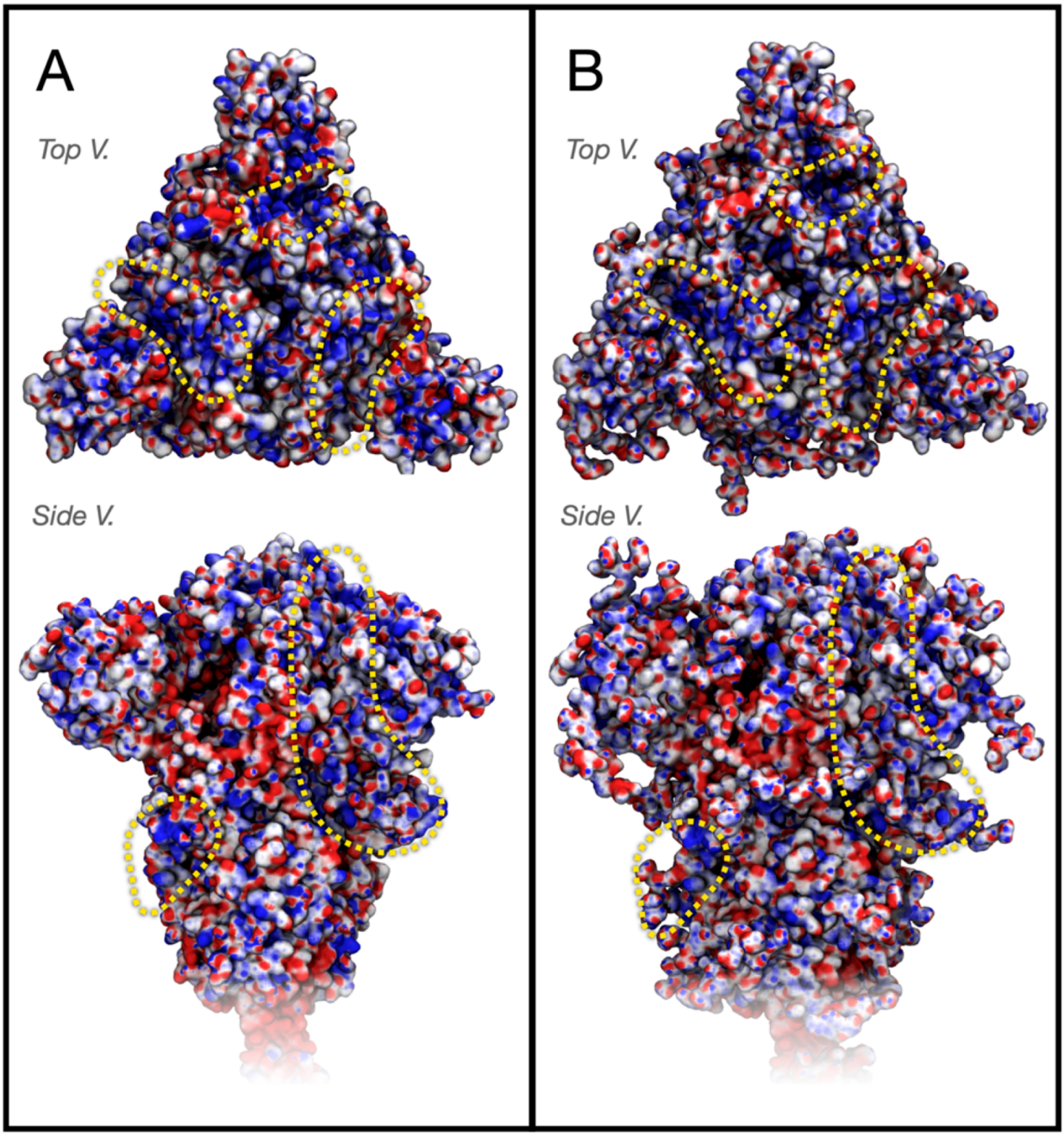


Figure S3.29: ESP maps of Delta (B.1.617.2) spike variant

(A) is without glycans and (B) is with glycans. Yellow indicators outline key partial positive surface patches posited to have high affinity for GAGs.

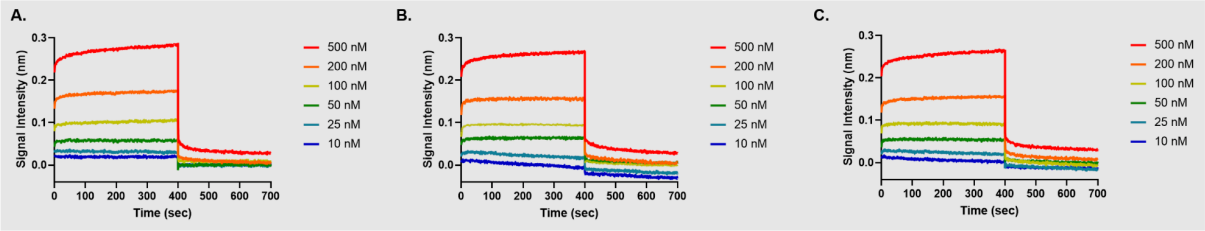


Figure S3.30: BLI curves of HEP15 to variants

(A) Alpha (B.1.1.7), (B) Beta (B.1.351), and (C) Delta (B.1.617.2) variant full-length spike proteins.

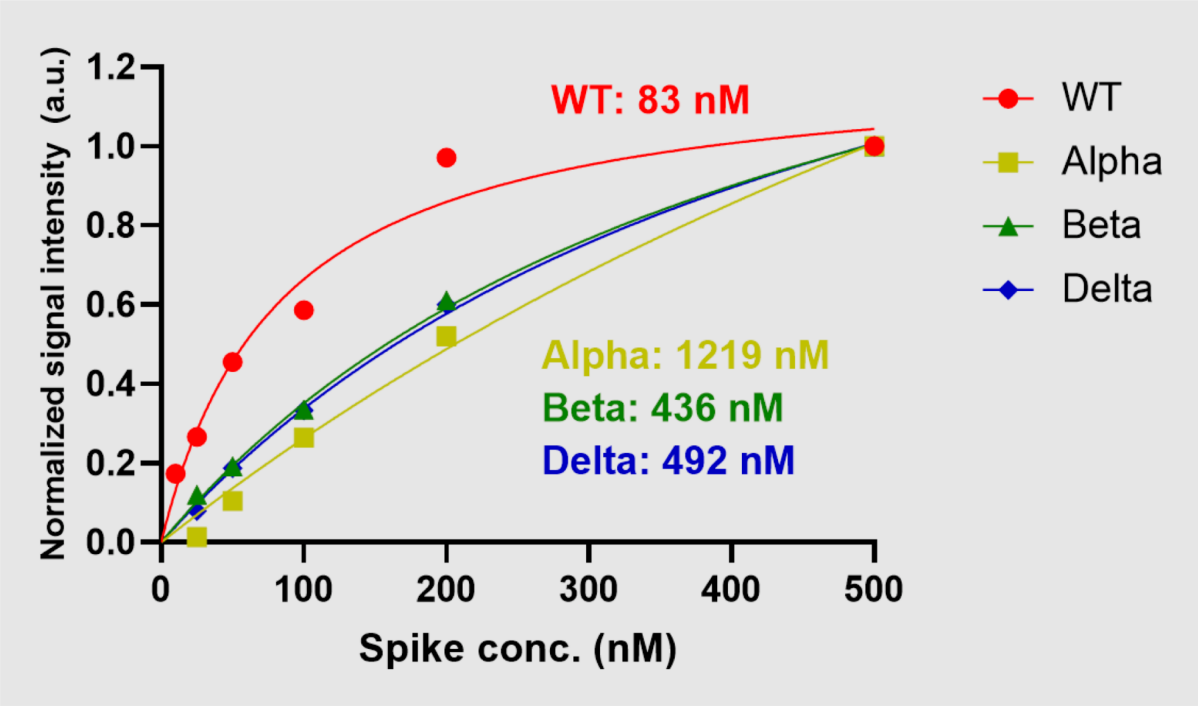


Figure S3.31: Binding curve of the NTD antibody

Binding or NTD shown to Wild type (WT), Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2) variant full-length spike proteins.

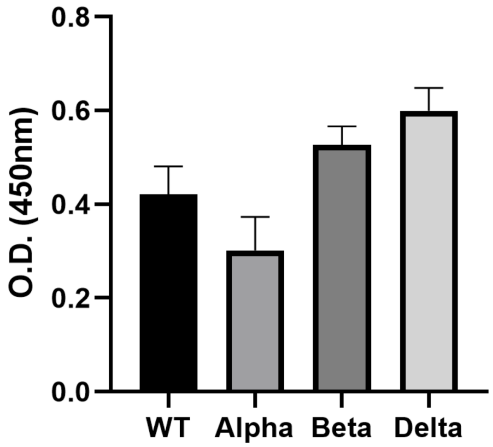


Figure S3.32: Sandwich-type binding of HEP15 and NTD Ab to spike

Binding is shown to Wild type (WT), Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2) variant full-length spike proteins.

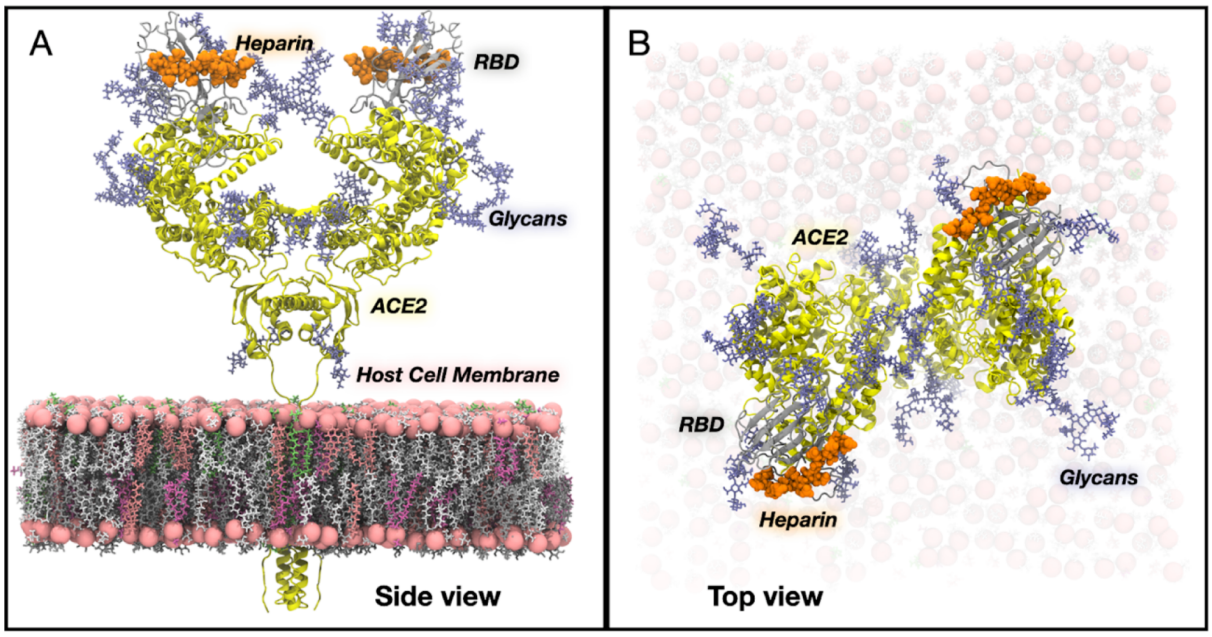


Figure S3.33: Cartoon representation of ACE2-spike RBD-hep8mer structure used in MD simulations

(A) Side view. (B) Top view. Yellow ribbons indicate the ACE2 protein structure, grey ribbons indicate the spike RBD protein structure, light blue licorice atoms indicate ACE2 or spike RBD glycans, orange van der Waals atoms indicate hep8mer structures (one bound to each spike RBD). Pink spheres indicate the Phosphate head groups of the membrane and pink, white, mauve, magenta, and lime green licorice atoms indicate lipid membrane atoms.

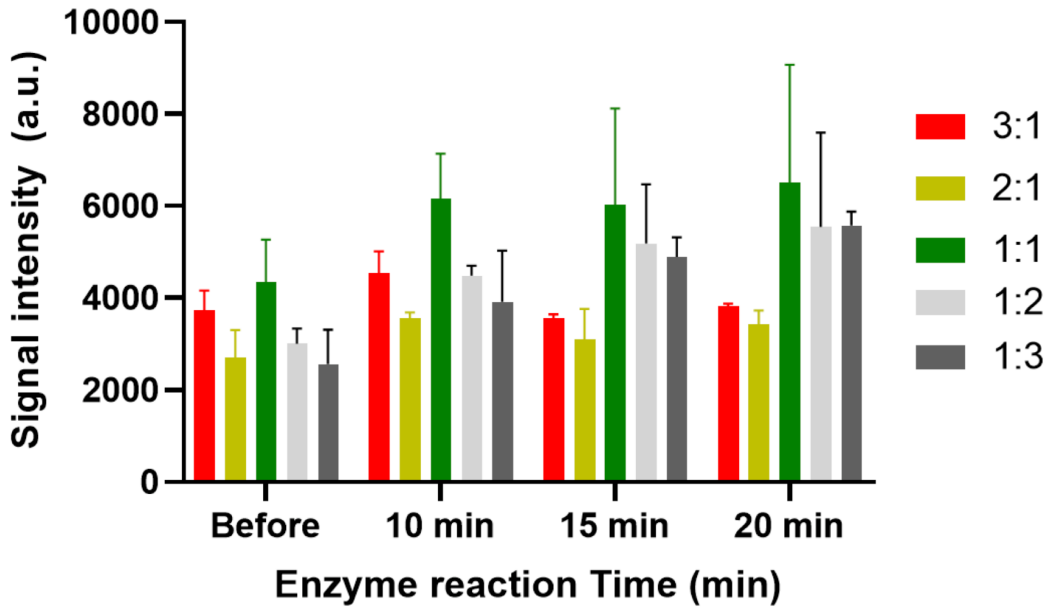


Figure S3.34: Optimization of NTD-AuNP and NTD-HRP-AuNP ratio, and HRP reaction time for LF signal enhancement

3.7 **Acknowledgements**

FLK, MAR, would like to thank Robert Konecny and J. Andrew McCammon for insightful discussions about the use of APBS. FLK, MAR would like to thank Elisa Fadda and Sebastien Vidal for their vital insights in glycobiology. FLK, MAR, LC, REA would like to thank TACC Frontera for their unwavering support of the Amaro Lab's SARS-CoV-2 related efforts and their dedication to answering our many and varied emails. BLI measurements taken at the UNC Macromolecular Interactions Facility are supported in part by P30 CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center funded by the National Cancer Institute. Confocal was done at Microscopy Services Laboratory, Department of Pathology and Laboratory Medicine, which is supported in part by P30 CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National

Institutes of Health. We would like to thank Peter Harris for the graphical art incorporated into this manuscript.

Chapter 3, in full, is a modified reprint of the materials as it appears in “Kim SH†, Kearns FL†, Rosenfeld MA†, Casalino L, Papanikolas MJ, Simmerling C, Amaro RE, Freeman R. *GlycoGrip: Cell Surface-Inspired Universal Sensor for Betacoronaviruses.* ACS Cent Sci. 2022 Jan 26;8(1):22-42.” The dissertation author was a primary co-investigator and co-author of this work.

3.8 **References**

3.7.1 Main Text References

(1) Reitsma, S.; Slaaf, D. W.; Vink, H.; van Zandvoort, M. A. M. J.; oude Egbrink, M. G. A. The Endothelial Glycocalyx: Composition, Functions, and Visualization. Pflügers Arch. - Eur. J. Physiol. 2007, 454 (3), 345–359. https://doi.org/10.1007/s00424-007-0212-8.

(2) Möckl, L. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation . Frontiers in Cell and Developmental Biology . 2020, p 253.

(3) Weinbaum, S.; Tarbell, J. M.; Damiano, E. R. The Structure and Function of the Endothelial Glycocalyx Layer. Annu. Rev. Biomed. Eng. 2007, 9 (1), 121–167. https://doi.org/10.1146/annurev.bioeng.9.060906.151959.

(4) Essentials of Glycobiology, 3rd ed.; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).

(5) Feyzi, E.; Saldeen, T.; Larsson, E.; Lindahl, U.; Salmivirta, M. Age-Dependent Modulation of Heparan Sulfate Structure and Function *. J. Biol. Chem. 1998, 273 (22), 13395–13398. https://doi.org/10.1074/jbc.273.22.13395.

(6) Han, X.; Sanderson, P.; Nesheiwat, S.; Lin, L.; Yu, Y.; Zhang, F.; Amster, I. J.; Linhardt, R. J. Structural Analysis of Urinary Glycosaminoglycans from Healthy Human Subjects. Glycobiology 2020, 30 (3), 143–151. https://doi.org/10.1093/glycob/cwz088.

(7) Vongchan, P.; Warda, M.; Toyoda, H.; Toida, T.; Marks, R. M.; Linhardt, R. J. Structural Characterization of Human Liver Heparan Sulfate. Biochim. Biophys. Acta - Gen. Subj. 2005, 1721 (1), 1–8. https://doi.org/https://doi.org/10.1016/j.bbagen.2004.09.007.

(8) Warda, M.; Toida, T.; Zhang, F.; Sun, P.; Munoz, E.; Xie, J.; Linhardt, R. J. Isolation and Characterization of Heparan Sulfate from Various Murine Tissues. Glycoconj. J. 2006, 23 (7), 555–563. https://doi.org/10.1007/s10719-006-7668-1.

(9) Casu, B. Chapter 1 - Structure and Active Domains of Heparin; Garg, H. G., Linhardt, R. J., Hales, C. A. B. T.-C. and B. of H. and H. S., Eds.; Elsevier Science: Amsterdam, 2005; pp 1–28. https://doi.org/https://doi.org/10.1016/B978-008044859-6/50002-2.

(10) Koehler, M.; Delguste, M.; Sieben, C.; Gillet, L.; Alsteens, D. Initial Step of Virus Entry: Virion Binding to Cell-Surface Glycans. Annu. Rev. Virol. 2020, 7 (1), 143–165. https://doi.org/10.1146/annurev-virology-122019-070025.

(11) Ströh, L. J.; Stehle, T. Glycan Engagement by Viruses: Receptor Switches and Specificity. Annu. Rev. Virol. 2014, 1 (1), 285–306. https://doi.org/10.1146/annurev-virology-031413-085417.

(12) Thompson, A. J.; de Vries, R. P.; Paulson, J. C. Virus Recognition of Glycan Receptors. Curr. Opin. Virol. 2019, 34, 117–129. https://doi.org/10.1016/j.coviro.2019.01.004.

(13) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. Science (80-.). 2020, 369 (6501), 330 LP – 333. https://doi.org/10.1126/science.abb9983.

(14) Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P. Deducing the N- and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. Glycobiology 2020, 30 (12), 981–988. https://doi.org/10.1093/glycob/cwaa042.

(15) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent. Sci. 2020, 6 (10), 1722–1734. https://doi.org/10.1021/acscentsci.0c01056.

(16) Harmer, D.; Gilbert, M.; Borman, R.; Clark, K. L. Quantitative MRNA Expression Profiling of ACE 2, a Novel Homologue of Angiotensin Converting Enzyme. FEBS Lett. 2002, 532 (1–2), 107–110. https://doi.org/10.1016/S0014-5793(02)03640-2.

(17) Tipnis, S. R.; Hooper, N. M.; Hyde, R.; Karran, E.; Christie, G.; Turner, A. J. A Human Homolog of Angiotensin-Converting Enzyme. J. Biol. Chem. 2000, 275 (43), 33238–33243. https://doi.org/10.1074/jbc.M002615200.

(18) Hamming, I.; Timens, W.; Bulthuis, M.; Lely, A.; Navis, G.; van Goor, H. Tissue Distribution of ACE2 Protein, the Functional Receptor for SARS Coronavirus. A First Step in Understanding SARS Pathogenesis. J. Pathol. 2004, 203 (2), 631–637. https://doi.org/10.1002/path.1570.

(19) Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. Nat. Chem. 2021, 2021.02.15.431212. https://doi.org/10.1038/s41557-021-00758-3.

(20) Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection. Biophys. J. 2021, 120 (6), 1072–1084. https://doi.org/https://doi.org/10.1016/j.bpj.2020.10.036.

(21) Verdecchia, P.; Cavallini, C.; Spanevello, A.; Angeli, F. The Pivotal Link between ACE2 Deficiency and SARS-CoV-2 Infection. Eur. J. Intern. Med. 2020, 76, 14–20. https://doi.org/10.1016/j.ejim.2020.04.037.

(22) Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallimadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P.; Hurley, M. F. D.; Harbison, A. M.; Fogarty, C. A.; Coffland, J. E.; Fadda, E.; Voelz, V. A.; Chodera, J. D.; Bowman, G. R. SARS-CoV-2 Simulations Go Exascale to Predict

Dramatic Spike Opening and Cryptic Pockets across the Proteome. Nat. Chem. 2021. https://doi.org/10.1038/s41557-021-00707-0.

(23) Mycroft-West, C. J.; Su, D.; Pagani, I.; Rudd, T. R.; Elli, S.; Gandhi, N. S.; Guimond, S. E.; Miller, G. J.; Meneghetti, M. C. Z.; Nader, H. B.; Li, Y.; Nunes, Q. M.; Procter, P.; Mancini, N.; Clementi, M.; Bisio, A.; Forsyth, N. R.; Ferro, V.; Turnbull, J. E.; Guerrini, M.; Fernig, D. G.; Vicenzi, E.; Yates, E. A.; Lima, M. A.; Skidmore, M. A. Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. Thromb. Haemost. 2020, 120 (12), 1700–1715. https://doi.org/10.1055/s-0040-1721319.

(24) Milewska, A.; Zarebski, M.; Nowak, P.; Stozek, K.; Potempa, J.; Pyrc, K. Human Coronavirus NL63 Utilizes Heparan Sulfate Proteoglycans for Attachment to Target Cells. J. Virol. 2014, 88 (22), 13221 LP – 13230. https://doi.org/10.1128/JVI.02078-14.

(25) Madu, I. G.; Chu, V. C.; Lee, H.; Regan, A. D.; Bauman, B. E.; Whittaker, G. R. Heparan Sulfate Is a Selective Attachment Factor for the Avian Coronavirus Infectious Bronchitis Virus Beaudette (El Heparán Sulfato Es Un Factor de Unión Para La Cepa Beaudette Del Virus de Bronquitis Infecciosa). Avian Dis. 2007, 51 (1), 45–51.

(26) Kim, S. Y.; Jin, W.; Sood, A.; Montgomery, D. W.; Grant, O. C.; Fuster, M. M.; Fu, L.; Dordick, J. S.; Woods, R. J.; Zhang, F.; Linhardt, R. J. Characterization of Heparin and Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Spike Glycoprotein Binding Interactions. Antiviral Res. 2020, 181, 104873. https://doi.org/10.1016/j.antiviral.2020.104873.

(27) Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Perrett, H. R.; Painter, C. D.; Narayanan, A.; Majowicz, S. A.; Kwong, E. M.; McVicar, R. N.; Thacker, B. E.; Glass, C. A.; Yang, Z.; Torres, J. L.; Golden, G. J.; Bartels, P. L.; Porell, R. N.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B. M.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Leibel, S. L.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. SARS-CoV-2 Infection Depends

on Cellular Heparan Sulfate and ACE2. Cell 2020, 183 (4), 1043-1057.e15. https://doi.org/https://doi.org/10.1016/j.cell.2020.09.033.

(28) Zhang, Q.; Chen, C. Z.; Swaroop, M.; Xu, M.; Wang, L.; Lee, J.; Wang, A. Q.; Pradhan, M.; Hagen, N.; Chen, L.; Shen, M.; Luo, Z.; Xu, X.; Xu, Y.; Huang, W.; Zheng, W.; Ye, Y. Heparan Sulfate Assists SARS-CoV-2 in Cell Entry and Can Be Targeted by Approved Drugs in Vitro. Cell Discov. 2020, 6 (1), 80. https://doi.org/10.1038/s41421-020-00222-5.

(29) Liu, L.; Chopra, P.; Li, X.; Bouwman, K. M.; Tompkins, S. M.; Wolfert, M. A.; de Vries, R. P.; Boons, G.-J. Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2. bioRxiv 2021, 2020.05.10.087288. https://doi.org/10.1101/2020.05.10.087288.

(30) Yu, M.; Zhang, T.; Zhang, W.; Sun, Q.; Li, H.; Li, J. Elucidating the Interactions Between Heparin/Heparan Sulfate and SARS-CoV-2-Related Proteins—An Important Strategy for Developing Novel Therapeutics for the COVID-19 Pandemic . Frontiers in Molecular Biosciences . 2021, p 490.

(31) Kalra, R. S.; Kandimalla, R. Engaging the Spikes: Heparan Sulfate Facilitates SARS-CoV-2 Spike Protein Binding to ACE2 and Potentiates Viral Infection. Signal Transduct. Target. Ther. 2021, 6 (1), 39. https://doi.org/10.1038/s41392-021-00470-1.

(32) de Haan, C. A. M.; Li, Z.; te Lintelo, E.; Bosch, B. J.; Haijema, B. J.; Rottier, P. J. M. Murine Coronavirus with an Extended Host Range Uses Heparan Sulfate as an Entry Receptor. J. Virol. 2005, 79 (22), 14451 LP – 14456. https://doi.org/10.1128/JVI.79.22.14451-14456.2005.

(33) Tandon, R.; Sharp, J. S.; Zhang, F.; Pomin, V. H.; Ashpole, N. M.; Mitra, D.; McCandless, M. G.; Jin, W.; Liu, H.; Sharma, P.; Linhardt, R. J. Effective Inhibition of SARS-CoV-2 Entry by Heparin and Enoxaparin Derivatives. J. Virol. 2020, 95 (3). https://doi.org/10.1128/JVI.01987-20.

(34) Harbison, A. M.; Fogarty, C. A.; Phung, T. K.; Satheesan, A.; Schulz, B. L.; Fadda, E. Fine-Tuning the Spike: Role of the Nature and Topology of the Glycan Shield in the Structure and Dynamics of SARS-CoV-2 S. bioRxiv 2021, 2021.04.01.438036. https://doi.org/10.1101/2021.04.01.438036.

(35) Paiardi, G.; Richter, S.; Rusnati, M.; Wade, R. C. Mechanism of Inhibition of SARS-CoV-2 Infection by the Interaction of the Spike Glycoprotein with Heparin. arXiv.org 2021.

(36) Schuurs, Z. P.; Hammond, E.; Elli, S.; Rudd, T. R.; Mycroft-West, C. J.; Lima, M. A.; Skidmore, M. A.; Karlsson, R.; Chen, Y.-H.; Bagdonaite, I.; Yang, Z.; Ahmed, Y. A.; Richard, D. J.; Turnbull, J.; Ferro, V.; Coombe, D. R.; Gandhi, N. S. Evidence of a Putative Glycosaminoglycan Binding Site on the Glycosylated SARS-CoV-2 Spike Protein N-Terminal Domain. Comput. Struct. Biotechnol. J. 2021, 19, 2806–2818. https://doi.org/10.1016/j.csbj.2021.05.002.

(37) de Haan, C. A. M.; Haijema, B. J.; Schellen, P.; Schreur, P. W.; te Lintelo, E.; Vennema, H.; Rottier, P. J. M. Cleavage of Group 1 Coronavirus Spike Proteins: How Furin Cleavage Is Traded Off against Heparan Sulfate Binding upon Cell Culture Adaptation. J. Virol. 2008, 82 (12), 6078 LP – 6083. https://doi.org/10.1128/JVI.00074-08.

(38) Vankadari, N. Structure of Furin Protease Binding to SARS-CoV-2 Spike Glycoprotein and Implications for Potential Targets and Virulence. J. Phys. Chem. Lett. 2020, 11 (16), 6655–6663. https://doi.org/10.1021/acs.jpclett.0c01698.

(39) Mohammad, A.; Alshawaf, E.; Marafie, S. K.; Abu-Farha, M.; Abubaker, J.; Al-Mulla, F. Higher Binding Affinity of Furin for SARS-CoV-2 Spike (S) Protein D614G Mutant Could Be Associated with Higher SARS-CoV-2 Infectivity. Int. J. Infect. Dis. 2021, 103, 611–616. https://doi.org/10.1016/j.ijid.2020.10.033.

(40) Wasik, D.; Mulchandani, A.; Yates, M. V. A Heparin-Functionalized Carbon Nanotube-Based Affinity Biosensor for Dengue Virus. Biosens. Bioelectron. 2017, 91, 811–816. https://doi.org/10.1016/j.bios.2017.01.017.

(41) Jelinek, R.; Kolusheva, S. Carbohydrate Biosensors. Chem. Rev. 2004, 104 (12), 5987–6016. https://doi.org/10.1021/cr0300284.

(42) Griffith, A. R.; Rogers, C. J.; Miller, G. M.; Abrol, R.; Hsieh-Wilson, L. C.; Goddard, W. A. Predicting Glycosaminoglycan Surface Protein Interactions and Implications for Studying Axonal Growth. Proc. Natl. Acad. Sci. 2017, 114 (52), 13697 LP – 13702. https://doi.org/10.1073/pnas.1715093115.

(43) Lee, B. H.; Kim, S. H.; Ko, Y.; Park, J. C.; Ji, S.; Gu, M. B. The Sensitive Detection of ODAM by Using Sandwich-Type Biosensors with a Cognate Pair of Aptamers for the Early Diagnosis of Periodontal Disease. Biosens. Bioelectron. 2019, 126, 122–128. https://doi.org/https://doi.org/10.1016/j.bios.2018.10.040.

(44) Kim, S. H.; Lee, J.; Lee, B. H.; Song, C.-S.; Gu, M. B. Specific Detection of Avian Influenza H5N2 Whole Virus Particles on Lateral Flow Strips Using a Pair of Sandwich-Type Aptamers. Biosens. Bioelectron. 2019, 134, 123–129. https://doi.org/https://doi.org/10.1016/j.bios.2019.03.061.

(45) Posthuma-Trumpie, G. A.; Korf, J.; van Amerongen, A. Lateral Flow (Immuno)Assay: Its Strengths, Weaknesses, Opportunities and Threats. A Literature Survey. Anal. Bioanal. Chem. 2009, 393 (2), 569–582. https://doi.org/10.1007/s00216-008-2287-2.

(46) Mak, W. C.; Beni, V.; Turner, A. P. F. Lateral-Flow Technology: From Visual to Instrumental. TrAC Trends Anal. Chem. 2016, 79, 297–305. https://doi.org/https://doi.org/10.1016/j.trac.2015.10.017.

(47) Ahmad Raston, N. H.; Nguyen, V.-T.; Gu, M. B. A New Lateral Flow Strip Assay (LFSA) Using a Pair of Aptamers for the Detection of Vaspin. Biosens. & Bioelectron. 2017, 93, 21–25. https://doi.org/10.1016/j.bios.2016.11.061.

(48) Li, J.; Jing, L.; Song, Y.; Zhang, J.; Chen, Q.; Wang, B.; Xia, X.; Han, Q. Rapid Detection of Rongalite via a Sandwich Lateral Flow Strip Assay Using a Pair of Aptamers. Nanoscale Res. Lett. 2018, 13 (1), 296. https://doi.org/10.1186/s11671-018-2709-9.

(49) Koczula, K. M.; Gallotta, A. Lateral Flow Assays. Essays Biochem. 2016, 60 (1), 111–120. https://doi.org/10.1042/EBC20150012.

(50) Hanack, K.; Messerschmidt, K.; Listek, M. Antibodies and Selection of Monoclonal Antibodies; 2016; pp 11–22. https://doi.org/10.1007/978-3-319-32805-8_2.

(51) Mercer, T. R.; Salit, M. Testing at Scale during the COVID-19 Pandemic. Nat. Rev. Genet. 2021. https://doi.org/10.1038/s41576-021-00360-w.

(52) Mina, M. J.; Peto, T. E.; García-Fiñana, M.; Semple, M. G.; Buchan, I. E. Clarifying the Evidence on SARS-CoV-2 Antigen Rapid Tests in Public Health Responses to COVID-19. Lancet 2021, 397 (10283), 1425–1427. https://doi.org/10.1016/S0140-6736(21)00425-6.

(53) Crozier, A.; Rajan, S.; Buchan, I.; McKee, M. Put to the Test: Use of Rapid Testing Technologies for Covid-19. BMJ 2021, 372, n208. https://doi.org/10.1136/bmj.n208.

(54) Bermejo-Jambrina, M.; Eder, J.; Kaptein, T. M.; Helgers, L. C.; Brouwer, P. J. M.; van Hamme, J. L.; Vlaar, A. P. J.; van Baarle, F. E. H. P.; de Bree, G. J.; Nijmeijer, B. M.; Koostra, N. A.; van Gils, M. J.; Sanders, R. W.; Geijtenbeek, T. B. H. SARS-CoV-2 Infection and Transmission Depends on Heparan Sulfates and Is Blocked by Low Molecular Weight Heparins. bioRxiv 2020, 2020.08.18.255810. https://doi.org/10.1101/2020.08.18.255810.

(55) Hao, W.; Ma, B.; Li, Z.; Wang, X.; Gao, X.; Li, Y.; Qin, B.; Shang, S.; Cui, S.; Tan, Z. Binding of the SARS-CoV-2 Spike Protein to Glycans. Sci. Bull. 2021, 66 (12), 1205–1214. https://doi.org/10.1016/j.scib.2021.01.010.

(56) Esko, J. D.; Lindahl, U. Molecular Diversity of Heparan Sulfate. J. Clin. Invest. 2001, 108 (2), 169–173. https://doi.org/10.1172/JCI13530.

(57) Xu, D.; Esko, J. D. Demystifying Heparan Sulfate–Protein Interactions. Annu. Rev. Biochem. 2014, 83 (1), 129–157. https://doi.org/10.1146/annurev-biochem-060713-035314.

(58) Yue, J.; Jin, W.; Yang, H.; Faulkner, J.; Song, X.; Qiu, H.; Teng, M.; Azadi, P.; Zhang, F.; Linhardt, R. J.; Wang, L. Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer. Front. Mol. Biosci. 2021, 8. https://doi.org/10.3389/fmolb.2021.649575.

(59) Chautard, E.; Fatoux-Ardore, M.; Ballut, L.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB, the Extracellular Matrix Interaction Database. Nucleic Acids Res. 2011, 39 (Database), D235–D240. https://doi.org/10.1093/nar/gkq830.

(60) Chautard, E.; Ballut, L.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB, a Database Focused on Extracellular Protein–Protein and Protein–Carbohydrate Interactions. Bioinformatics 2009, 25 (5), 690–691. https://doi.org/10.1093/bioinformatics/btp025.

(61) Launay, G.; Salza, R.; Multedo, D.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB, the Extracellular Matrix Interaction Database: Updated Content, a New Navigator and Expanded Functionalities. Nucleic Acids Res. 2015, 43 (D1), D321–D327. https://doi.org/10.1093/nar/gku1091.

(62) Clerc, O.; Deniaud, M.; Vallet, S. D.; Naba, A.; Rivet, A.; Perez, S.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB: Integration of New Data with a Focus on Glycosaminoglycan Interactions. Nucleic Acids Res. 2019, 47 (D1), D376–D381. https://doi.org/10.1093/nar/gky1035.

(63) Park, S.-J.; Lee, J.; Patel, D. S.; Ma, H.; Lee, H. S.; Jo, S.; Im, W. Glycan Reader Is Improved to Recognize Most Sugar Types and Chemical Modifications in the Protein Data Bank. Bioinformatics 2017, 33 (19), 3051–3057. https://doi.org/10.1093/bioinformatics/btx358.

(64) Jo, S.; Song, K. C.; Desaire, H.; MacKerell Jr., A. D.; Im, W. Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins. J. Comput. Chem. 2011, 32 (14), 3135–3141. https://doi.org/https://doi.org/10.1002/jcc.21886.

(65) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. J. Chem. Theory Comput. 2009, 5 (9), 2353–2370. https://doi.org/10.1021/ct900242e.

(66) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. J. Comput. Chem. 2008, 29 (11), 1859–1865. https://doi.org/https://doi.org/10.1002/jcc.20945.

(67) Cladera, J.; Martin, I.; O'Shea, P. The Fusion Domain of HIV Gp41 Interacts Specifically with Heparan Sulfate on the T-Lymphocyte Cell Surface. EMBO J. 2001, 20 (1–2), 19–26. https://doi.org/https://doi.org/10.1093/emboj/20.1.19.

(68) Connell, B. J.; Lortat-Jacob, H. Human Immunodeficiency Virus and Heparan Sulfate: From Attachment to Entry Inhibition. Front. Immunol. 2013, 4. https://doi.org/10.3389/fimmu.2013.00385.

(69) Harrop, H. A.; Rider, C. C. Heparin and Its Derivatives Bind to HIV-1 Recombinant Envelope Glycoproteins, Rather than to Recombinant HIV-1 Receptor, CD4. Glycobiology 1998, 8 (2), 131–137. https://doi.org/10.1093/glycob/8.2.131.

(70) Hansen, J.; Baum, A.; Pascal, K. E.; Russo, V.; Giordano, S.; Wloga, E.; Fulton, B. O.; Yan, Y.; Koon, K.; Patel, K.; Chung, K. M.; Hermann, A.; Ullman, E.; Cruz, J.; Rafique, A.; Huang, T.; Fairhurst, J.; Libertiny, C.; Malbec, M.; Lee, W.; Welsh, R.; Farr, G.; Pennington, S.; Deshpande, D.; Cheng, J.; Watty, A.; Bouffard, P.; Babb, R.; Levenkova, N.; Chen, C.; Zhang, B.; Romero Hernandez, A.; Saotome, K.; Zhou, Y.; Franklin, M.; Sivapalasingam, S.; Lye, D. C.; Weston, S.; Logue, J.; Haupt, R.; Frieman, M.; Chen, G.; Olson, W.; Murphy, A. J.; Stahl, N.; Yancopoulos, G. D.; Kyratsous, C. A. Studies in Humanized Mice and Convalescent Humans Yield a SARS-CoV-2 Antibody Cocktail. Science (80-.). 2020, 369 (6506), 1010–1014. https://doi.org/10.1126/science.abd0827.

(71) 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 (80-.). 2020, 367 (6485), 1444–1448. https://doi.org/10.1126/science.abb2762.

(72) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. Nature 2020, 581 (7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5.

(73) Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y.; Chen, Z.; Guo, Y.; Zhang, J.; Li, Y.; Song, X.; Chen, Y.; Xia, L.; Fu, L.; Hou, L.; Xu, J.; Yu, C.; Li, J.; Zhou, Q.; Chen, W. A Neutralizing Human Antibody Binds to the N-Terminal Domain of the Spike Protein of SARS-CoV-2. Science (80-.). 2020, 369 (6504), 650–655. https://doi.org/10.1126/science.abc6952.

(74) Shrake, A.; Rupley, J. A. Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin. J. Mol. Biol. 1973, 79 (2), 351–371. https://doi.org/https://doi.org/10.1016/0022-2836(73)90011-9.

(75) Yuan, M.; Wu, N. C.; Zhu, X.; Lee, C.-C. D.; So, R. T. Y.; Lv, H.; Mok, C. K. P.; Wilson, I. A. A Highly Conserved Cryptic Epitope in the Receptor Binding Domains of SARS-CoV-2 and SARS-CoV. Science (80-.). 2020, 368 (6491), 630 LP – 633. https://doi.org/10.1126/science.abb7269.

(76) Barnes, C. O.; West Jr., A. P.; Huey-Tubman, K. E.; Hoffmann, M. A. G.; Sharaf, N. G.; Hoffman, P. R.; Koranda, N.; Gristick, H. B.; Gaebler, C.; Muecksch, F.; Lorenzi, J. C. C.; Finkin, S.; Hägglöf, T.; Hurley, A.; Millard, K. G.; Weisblum, Y.; Schmidt, F.; Hatziioannou, T.; Bieniasz, P. D.; Caskey, M.; Robbiani, D. F.; Nussenzweig, M. C.; Bjorkman, P. J. Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. Cell 2020, 182 (4), 828-842.e16. https://doi.org/10.1016/j.cell.2020.06.025.

(77) Yuan, M.; Liu, H.; Wu, N. C.; Wilson, I. A. Recognition of the SARS-CoV-2 Receptor Binding Domain by Neutralizing Antibodies. Biochem. Biophys. Res. Commun. 2021, 538, 192–203. https://doi.org/https://doi.org/10.1016/j.bbrc.2020.10.012.

(78) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. Protein Sci. 2018, 27 (1), 112–128. https://doi.org/https://doi.org/10.1002/pro.3280.

(79) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. Proc. Natl. Acad. Sci. 2001, 98 (18), 10037 LP – 10041. https://doi.org/10.1073/pnas.181342398.

(80) Holst, M.; Saied, F. Multigrid Solution of the Poisson—Boltzmann Equation. J. Comput. Chem. 1993, 14 (1), 105–113. https://doi.org/https://doi.org/10.1002/jcc.540140114.

(81) Holst, M. J.; Saied, F. Numerical Solution of the Nonlinear Poisson–Boltzmann Equation: Developing More Robust and Efficient Methods. J. Comput. Chem. 1995, 16 (3), 337–364. https://doi.org/https://doi.org/10.1002/jcc.540160308.

(82) Gasbarri, M.; V'kovski, P.; Torriani, G.; Thiel, V.; Stellacci, F.; Tapparel, C.; Cagno, V. SARS-CoV-2 Inhibition by Sulfonated Compounds. Microorganisms . 2020. https://doi.org/10.3390/microorganisms8121894.

(83) Gandhi, N. S.; Mancera, R. L. Free Energy Calculations of Glycosaminoglycan–Protein Interactions. Glycobiology 2009, 19 (10), 1103–1115. https://doi.org/10.1093/glycob/cwp101.

(84) Sarkar, A.; Yu, W.; Desai, U. R.; MacKerell, A. D.; Mosier, P. D. Estimating Glycosaminoglycan–Protein Interaction Affinity: Water Dominates the Specific Antithrombin–Heparin Interaction. Glycobiology 2016, 26 (10), 1041–1047. https://doi.org/10.1093/glycob/cww073.

(85) Cramer, J.; Sager, C. P.; Ernst, B. Hydroxyl Groups in Synthetic and Natural-Product-Derived Therapeutics: A Perspective on a Common Functional Group. J. Med. Chem. 2019, 62 (20), 8915–8930. https://doi.org/10.1021/acs.jmedchem.9b00179.

(86) Přistoupil, T. I.; Kramlová, M.; Štěrbíková, J. On the Mechanism of Adsorption of Proteins to Nitrocellulose in Membrane Chromatography. J. Chromatogr. A 1969, 42, 367–375. https://doi.org/https://doi.org/10.1016/S0021-9673(01)80636-1.

(87) Ceron, J.; Lamy, E.; Martinez-Subiela, S.; Lopez-Jornet, P.; Capela-Silva, F.; Eckersall, P.; Tvarijonaviciute, A. Use of Saliva for Diagnosis and Monitoring the SARS-CoV-2: A General Perspective. J. Clin. Med. 2020, 9 (5), 1491. https://doi.org/10.3390/jcm9051491.

(88) Wyllie, A. L.; Fournier, J.; Casanovas-Massana, A.; Campbell, M.; Tokuyama, M.; Vijayakumar, P.; Warren, J. L.; Geng, B.; Muenker, M. C.; Moore, A. J.; Vogels, C. B. F.; Petrone, M. E.; Ott, I. M.; Lu, P.; Venkataraman, A.; Lu-Culligan, A.; Klein, J.; Earnest, R.; Simonov, M.; Datta, R.; Handoko, R.; Naushad, N.; Sewanan, L. R.; Valdez, J.; White, E. B.; Lapidus, S.; Kalinich, C. C.; Jiang, X.; Kim, D. J.; Kudo, E.; Linehan, M.; Mao, T.; Moriyama, M.; Oh, J. E.; Park, A.; Silva, J.; Song, E.; Takahashi, T.; Taura, M.; Weizman, O.-E.; Wong, P.; Yang, Y.; Bermejo, S.; Odio, C. D.; Omer, S. B.; Dela Cruz, C. S.; Farhadian, S.; Martinello, R. A.; Iwasaki, A.; Grubaugh, N. D.; Ko, A. I. Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. N. Engl. J. Med. 2020, 383 (13), 1283–1286. https://doi.org/10.1056/NEJMc2016359.

(89) Zhou, D.; Wu, C. Saliva Glycoproteins Bind to Spike Protein of SARS-CoV-2. Preprints 2020. https://doi.org/10.20944/preprints202005.0192.v1.

(90) Corum, J.; Zimmer, C. Coronavirus Variants and Mutations https://www.nytimes.com/interactive/2021/health/coronavirus-variant-tracker.html (accessed May 25, 2021).

(91) World Health Organization. No Title https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed Aug 8, 2021).

(92) McCallum, M.; Walls, A. C.; Sprouse, K. R.; Bowen, J. E.; Rosen, L.; Dang, H. V; deMarco, A.; Franko, N.; Tilles, S. W.; Logue, J.; Miranda, M. C.; Ahlrichs, M.; Carter, L.; Snell, G.; Pizzuto, M. S.; Chu, H. Y.; Van Voorhis, W. C.; Corti, D.; Veesler, D. Molecular Basis of Immune Evasion by the Delta and Kappa SARS-CoV-2 Variants. bioRxiv 2021, 2021.08.11.455956. https://doi.org/10.1101/2021.08.11.455956.

(93) Gobeil, S. M.-C.; Janowska, K.; McDowell, S.; Mansouri, K.; Parks, R.; Stalls, V.; Kopp, M. F.; Manne, K.; Saunders, K.; Edwards, R. J.; Haynes, B. F.; Henderson, R. C.; Acharya, P. Effect of Natural Mutations of SARS-CoV-2 on Spike Structure, Conformation and Antigenicity. bioRxiv Prepr. Serv. Biol. 2021. https://doi.org/10.1101/2021.03.11.435037.

(94) Shiliaev, N.; Lukash, T.; Palchevska, O.; Crossman, D. K.; Green, T. J.; Crowley, M. R.; Frolova, E. I.; Frolov, I. Natural Isolate and Recombinant SARS-CoV-2 Rapidly Evolve in Vitro to Higher Infectivity through More Efficient Binding to Heparan Sulfate and Reduced S1/S2 Cleavage. bioRxiv 2021, 2021.06.28.450274. https://doi.org/10.1101/2021.06.28.450274.

(95) Parolo, C.; Sena-Torralba, A.; Bergua, J. F.; Calucho, E.; Fuentes-Chust, C.; Hu, L.; Rivas, L.; Álvarez-Diduk, R.; Nguyen, E. P.; Cinti, S.; Quesada-González, D.; Merkoçi, A. Tutorial: Design and Fabrication of Nanoparticle-Based Lateral-Flow Immunoassays. Nat. Protoc. 2020, 15 (12), 3788–3816. https://doi.org/10.1038/s41596-020-0357-x.

(96) Yetisen, A. K.; Akram, M. S.; Lowe, C. R. Paper-Based Microfluidic Point-of-Care Diagnostic Devices. Lab Chip 2013, 13 (12), 2210. https://doi.org/10.1039/c3lc50169h.

(97) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J. Comput. Chem. 2009, 30 (16), 2785–2791. https://doi.org/https://doi.org/10.1002/jcc.21256.

(98) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. J. Comput. Chem. 2010, 31 (2), 455–461. https://doi.org/https://doi.org/10.1002/jcc.21334.

(99) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. J. Mach. Learn. Res. 2011, 12 (85), 2825–2830.

(100) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. J. Mol. Graph. 1996, 14 (1), 33–38. https://doi.org/https://doi.org/10.1016/0263-7855(96)00018-5.

(101) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 2020, 181 (2), 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

(102) Bangaru, S.; Ozorowski, G.; Turner, H. L.; Antanasijevic, A.; Huang, D.; Wang, X.; Torres, J. L.; Diedrich, J. K.; Tian, J.-H.; Portnoff, A. D.; Patel, N.; Massare, M. J.; Yates, J. R.; Nemazee, D.; Paulson, J. C.; Glenn, G.; Smith, G.; Ward, A. B. Structural Analysis of Full-Length SARS-CoV-2 Spike Protein from an Advanced Vaccine Candidate. Science (80-.). 2020, 370 (6520), 1089 LP – 1094. https://doi.org/10.1126/science.abe1502.

(103) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. J. Chem. Theory Comput. 2011, 7 (2), 525–537. https://doi.org/10.1021/ct100578z.

(104) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of PKa Values. J. Chem. Theory Comput. 2011, 7 (7), 2284–2295. https://doi.org/10.1021/ct200133y.

(105) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. J. Comput. Aided. Mol. Des. 2013, 27 (3), 221–234. https://doi.org/10.1007/s10822-013-9644-8.

(106) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. J. Comput. Chem. 2005, 26 (16), 1781–1802. https://doi.org/10.1002/jcc.20289.

(107) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. J. Chem. Phys. 2020, 153 (4), 44130. https://doi.org/10.1063/5.0014475.

(108) Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. Analysis of the SARS-CoV-2 Spike Protein Glycan Shield Reveals Implications for Immune Recognition. Sci. Rep. 2020, 10 (1), 14991. https://doi.org/10.1038/s41598-020-71748-7.

(109) Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. New Insights about HERG Blockade Obtained from Protein Modeling, Potential Energy Mapping, and Docking Studies. Bioorg. Med. Chem. 2006, 14 (9), 3160–3173. https://doi.org/https://doi.org/10.1016/j.bmc.2005.12.032.

(110) Sherman, W.; Beard, H. S.; Farid, R. Use of an Induced Fit Receptor Structure in Virtual Screening. Chem. Biol. Drug Des. 2006, 67 (1), 83–84. https://doi.org/https://doi.org/10.1111/j.1747-0285.2005.00327.x.

(111) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. J. Med. Chem. 2006, 49 (2), 534–553. https://doi.org/10.1021/jm050540c.

(112) van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane Lipids: Where They Are and How They Behave. Nat. Rev. Mol. Cell Biol. 2008, 9 (2), 112–124. https://doi.org/10.1038/nrm2330.

(113) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 1983, 79 (2), 926–935. https://doi.org/10.1063/1.445869.

(114) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. Nat. Methods 2017, 14 (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(115) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. J. Comput. Chem. 2013, 34 (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(116) Amaro, R. E.; Mulholland, A. J. Biomolecular Simulations in the Time of COVID-19, and After. Comput. Sci. Eng. 2020, 22 (6), 30–36. https://doi.org/10.1109/MCSE.2020.3024155.

(117) Mulholland, A. J.; Amaro, R. E. COVID19 - Computational Chemists Meet the Moment. J. Chem. Inf. Model. 2020, 60 (12), 5724–5726. https://doi.org/10.1021/acs.jcim.0c01395.

3.7.2 Supplemental References

(1) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent. Sci. 2020, 6 (10), 1722–1734. https://doi.org/10.1021/acscentsci.0c01056.

(2) Shrake, A.; Rupley, J. A. Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin. J. Mol. Biol. 1973, 79 (2), 351–371. https://doi.org/10.1016/0022-2836(73)90011-9.

(3) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J. Comput. Chem. 2009, 30 (16), 2785–2791. https://doi.org/https://doi.org/10.1002/jcc.21256.

(4) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. J. Comput. Chem. 2010, 31 (2), 455–461. https://doi.org/https://doi.org/10.1002/jcc.21334.

(5) Kim, S. Y.; Jin, W.; Sood, A.; Montgomery, D. W.; Grant, O. C.; Fuster, M. M.; Fu, L.; Dordick, J. S.; Woods, R. J.; Zhang, F.; Linhardt, R. J. Characterization of Heparin and Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Spike Glycoprotein Binding Interactions. Antiviral Res. 2020, 181, 104873. https://doi.org/10.1016/j.antiviral.2020.104873.

(6) Tandon, R.; Sharp, J. S.; Zhang, F.; Pomin, V. H.; Ashpole, N. M.; Mitra, D.; McCandless, M. G.; Jin, W.; Liu, H.; Sharma, P.; Linhardt, R. J. Effective Inhibition of SARS-CoV-2 Entry by Heparin and Enoxaparin Derivatives. J. Virol. 2020, 95 (3). https://doi.org/10.1128/JVI.01987-20.

(7) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. J. Mach. Learn. Res. 2011, 12 (85), 2825–2830.

(8) Essentials of Glycobiology, 3rd ed.; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).

(9) Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Perrett, H. R.; Painter, C. D.; Narayanan, A.; Majowicz, S. A.; Kwong, E. M.; McVicar, R. N.; Thacker, B. E.; Glass, C. A.; Yang, Z.; Torres, J. L.; Golden, G. J.; Bartels, P. L.; Porell, R. N.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B. M.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Leibel, S. L.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. Cell 2020, 183 (4), 1043-1057.e15. https://doi.org/https://doi.org/10.1016/j.cell.2020.09.033.

(10) Liu, L.; Chopra, P.; Li, X.; Bouwman, K. M.; Tompkins, S. M.; Wolfert, M. A.; de Vries, R. P.; Boons, G.-J. Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2. bioRxiv 2021, 2020.05.10.087288. https://doi.org/10.1101/2020.05.10.087288.

(11) Harbison, A. M.; Fogarty, C. A.; Phung, T. K.; Satheesan, A.; Schulz, B. L.; Fadda, E. Fine-Tuning the Spike: Role of the Nature and Topology of the Glycan Shield in the Structure and Dynamics of SARS-CoV-2 S. bioRxiv 2021, 2021.04.01.438036. https://doi.org/10.1101/2021.04.01.438036.

(12) Paiardi, G.; Richter, S.; Rusnati, M.; Wade, R. C. Mechanism of Inhibition of SARS-CoV-2 Infection by the Interaction of the Spike Glycoprotein with Heparin. arXiv.org 2021.

(13) Abraham, M. J.; van der Spoel, D.; Lindahl, E.; Hess, B. GROMACS User Manual version 2019.

(14) Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y.; Chen, Z.; Guo, Y.; Zhang, J.; Li, Y.; Song, X.; Chen, Y.; Xia, L.; Fu, L.; Hou, L.; Xu, J.; Yu, C.; Li, J.; Zhou, Q.; Chen, W. A Neutralizing Human Antibody Binds to the N-Terminal Domain of the Spike Protein of SARS-CoV-2. Science (80-.). 2020, 369 (6504), 650–655. https://doi.org/10.1126/science.abc6952.

(15) Hansen, J.; Baum, A.; Pascal, K. E.; Russo, V.; Giordano, S.; Wloga, E.; Fulton, B. O.; Yan, Y.; Koon, K.; Patel, K.; Chung, K. M.; Hermann, A.; Ullman, E.; Cruz, J.; Rafique, A.; Huang, T.; Fairhurst, J.; Libertiny, C.; Malbec, M.; Lee, W.; Welsh, R.; Farr, G.; Pennington, S.; Deshpande, D.; Cheng, J.; Watty, A.; Bouffard, P.; Babb, R.; Levenkova, N.; Chen, C.; Zhang, B.; Romero Hernandez, A.; Saotome, K.; Zhou, Y.; Franklin, M.; Sivapalasingam, S.; Lye, D. C.; Weston, S.; Logue, J.; Haupt, R.; Frieman, M.; Chen, G.; Olson, W.; Murphy, A. J.; Stahl, N.; Yancopoulos, G. D.; Kyratsous, C. A. Studies in Humanized Mice and Convalescent Humans Yield a SARS-CoV-2 Antibody Cocktail. Science (80-.). 2020, 369 (6506), 1010–1014. https://doi.org/10.1126/science.abd0827.

(16) 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 (80-.). 2020, 367 (6485), 1444–1448. https://doi.org/10.1126/science.abb2762.

(17) Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection. Biophys. J. 2021, 120 (6), 1072–1084. https://doi.org/https://doi.org/10.1016/j.bpj.2020.10.036.

(18) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. J. Chem. Phys. 2020, 153 (4), 44130. https://doi.org/10.1063/5.0014475.

(19) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. J. Comput. Chem. 2005, 26 (16), 1781–1802. https://doi.org/10.1002/jcc.20289.

(20) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. J. Comput. Chem. 2013, 34 (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(21) Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. Glycobiology 2019, 29 (4), 320–331. https://doi.org/10.1093/glycob/cwz003.

(22) Jo, S.; Song, K. C.; Desaire, H.; MacKerell Jr., A. D.; Im, W. Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins. J. Comput. Chem. 2011, 32 (14), 3135–3141. https://doi.org/https://doi.org/10.1002/jcc.21886.

(23) Park, S.-J.; Lee, J.; Patel, D. S.; Ma, H.; Lee, H. S.; Jo, S.; Im, W. Glycan Reader Is Improved to Recognize Most Sugar Types and Chemical Modifications in the Protein Data Bank. Bioinformatics 2017, 33 (19), 3051–3057. https://doi.org/10.1093/bioinformatics/btx358.

(24) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. Nat. Methods 2017, 14 (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(25) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. J. Comput. Chem. 2008, 29 (11), 1859–1865. https://doi.org/https://doi.org/10.1002/jcc.20945.

(26) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. J. Chem. Theory Comput. 2009, 5 (9), 2353–2370. https://doi.org/10.1021/ct900242e.

(27) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. Protein Sci. 2018, 27 (1), 112–128. https://doi.org/https://doi.org/10.1002/pro.3280.

(28) Lee, J.-H.; Choi, M.; Jung, Y.; Lee, S. K.; Lee, C.-S.; Kim, J.; Kim, J.; Kim, N. H.; Kim, B.-T.; Kim, H. G. A Novel Rapid Detection for SARS-CoV-2 Spike 1 Antigens Using Human Angiotensin Converting Enzyme 2 (ACE2). Biosens. Bioelectron. 2021, 171, 112715. https://doi.org/10.1016/j.bios.2020.112715.

(29) Baker, A. N.; Richards, S.-J.; Guy, C. S.; Congdon, T. R.; Hasan, M.; Zwetsloot, A. J.; Gallo, A.; Lewandowski, J. R.; Stansfeld, P. J.; Straube, A.; Walker, M.; Chessa, S.; Pergolizzi, G.; Dedola, S.; Field, R. A.; Gibson, M. I. The SARS-COV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. ACS Cent. Sci. 2020, 6 (11), 2046–2052. https://doi.org/10.1021/acscentsci.0c00855.

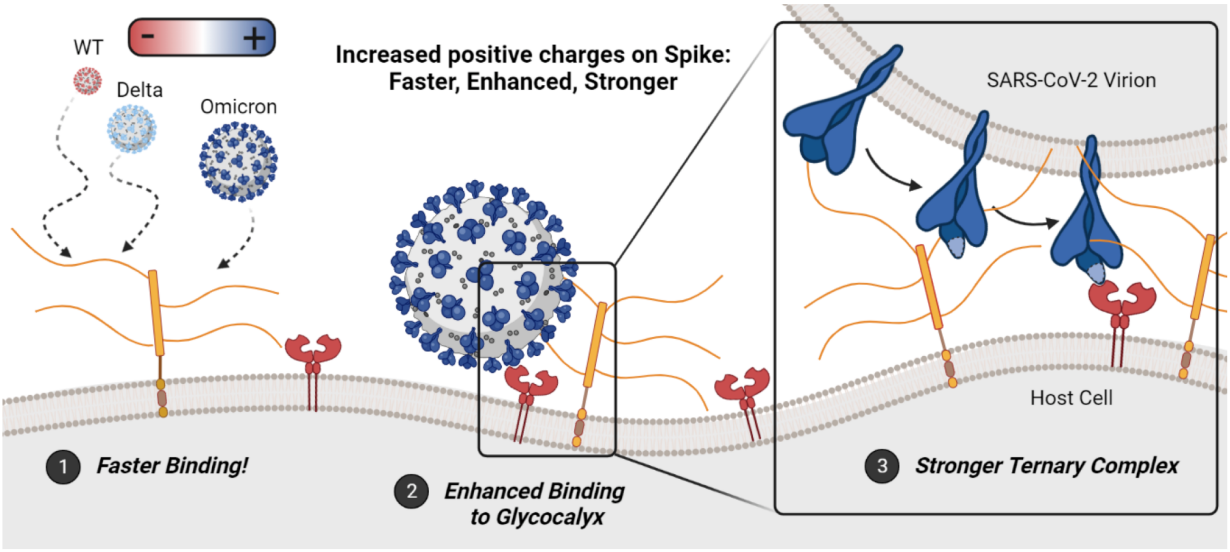
(30) Amaro, R. E.; Mulholland, A. J. Biomolecular Simulations in the Time of COVID-19, and After. Comput. Sci. Eng. 2020, 22 (6), 30–36. https://doi.org/10.1109/MCSE.2020.3024155.

(31)Mulholland, A. J.; Amaro, R. E. COVID19 - Computational Chemists Meet the Moment. J. Chem. Inf. Model. 2020, 60 (12), 5724–5726. https://doi.org/10.1021/acs.jcim.0c01395.

Chapter 4 - Positively bound: SARS-CoV2 Evolved Variants Optimize Binding to Cellular Glycocalyx

4.1 **Abstract**

The SARS-CoV-2 Omicron sub-lineage contains a significant number of mutations in the spike protein relative to earlier SARS-CoV-2 variants. These mutations significantly increased the positive charge of the spike protein that is postulated to confer increased infectivity, potentially through enhanced interactions with cell surface receptors, and an altered host-cell entry mechanism. While heparan sulfate (HS) was shown to be a key co-receptor in the host-cell entry of SARS-CoV-2, the effect of spike charge on its interactions with heparan sulfate has not been clearly elucidated. Here, we investigate the role of evolving spike positive charge in accelerating long-range interactions to heparan sulfate and ACE2 in the glycocalyx. We show that the positively charged Omicron evolved enhanced binding rates to the negatively charged glycocalyx. Moreover, we discovered that while the Omicron spike-ACE2 affinity is comparable to Delta, the Omicron spike interactions with heparan sulfate are significantly enhanced, giving rise to a ternary complex of spike-HS-ACE2 with a larger proportion of double-bound and triple-bound ACE2. Our findings suggest that SARS-CoV-2 variants evolve to be more dependent on heparan sulfate in viral attachment and infection. The evolving enhanced binding pattern between SARS-CoV-2 spikes and the glycocalyx enabled us to engineer a second-generation lateral flow test strip that harnesses heparin and ACE-2 to reliably detect all variants of concern including Omicron.



Scheme 4.1 Table of Contents Image

4.2 **Introduction**

Several SARS-CoV-2 variants of concern (VOCs) have emerged over the course of the COVID-19 pandemic including Alpha, Beta, Gamma, Delta, and Omicron, the latter with its own sub-lineages BA.1, BA.2, BA.3, BA.4, and BA.5.[1–3] Each of these VOCs are characterized by key mutations throughout the genome.[1–3] The SARS-CoV-2 viral envelope is studded with approximately 30 homotrimer glycoproteins, called spike proteins, which play the primary role in initiating host-cell entry via their receptor binding domains (RBDs). Genomic mutations to the spike protein sequence have been implicated in increasing infectivity and/or immune escape.[4–8] The Alpha, Beta, Delta, and Omicron BA.1 genomes, for example, contain 8, 8, 9, and 34 mutations in their spike mRNA sequences relative to the original "wild type" (WT) 2019 strain[9–13] (see Table S4.1 for complete list of mutations considered in this work).

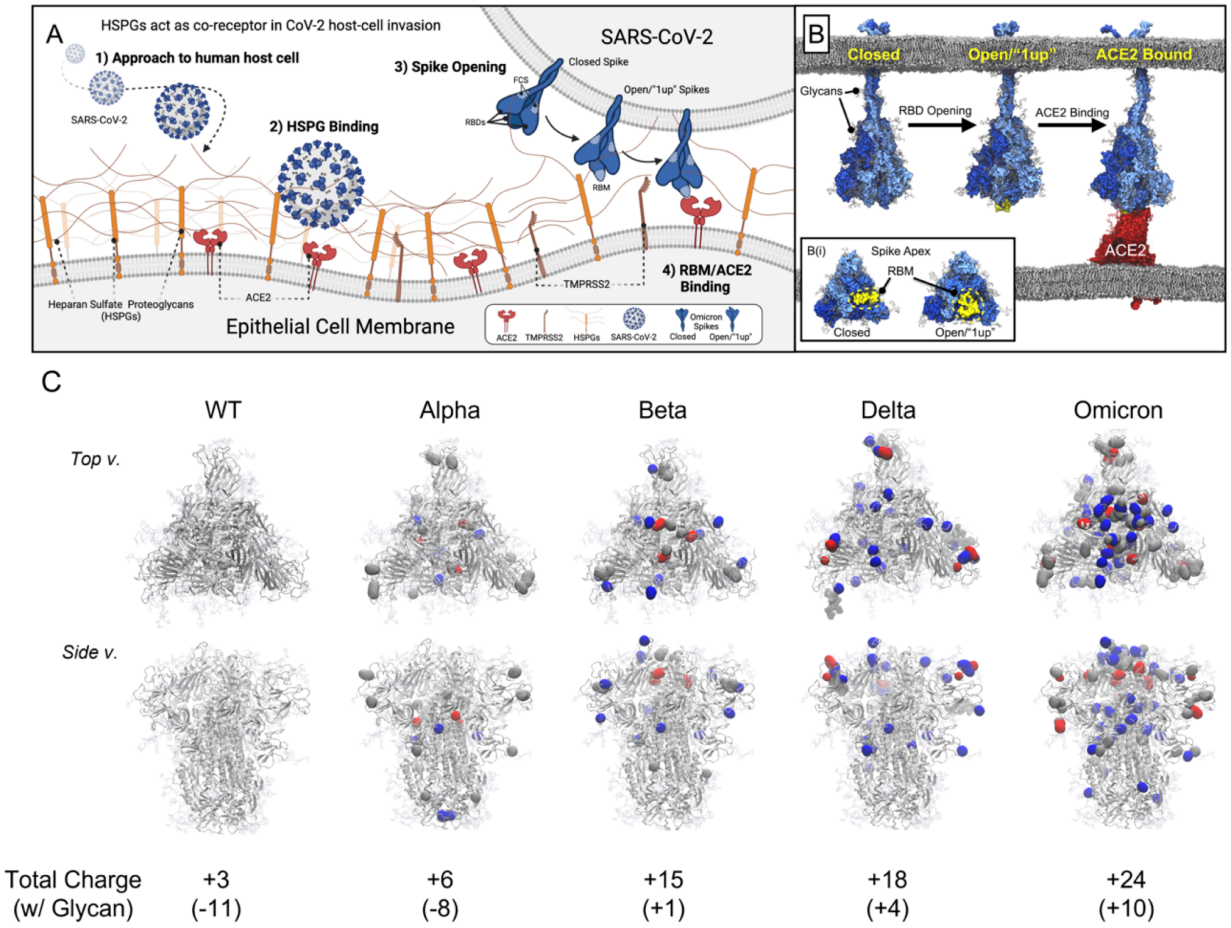
The high number of sequence mutations characteristic of the Omicron variant presented a concern for potential impact on initial PCR detection.[14,15] Rapid antigen detection was also impaired for Omicron variants even though these commercially available kits detect nucleocapsid proteins which incur a lower rate of mutation.[14,15] Furthermore, the newer subvariants of Omicron, particularly BA.5, could completely escape

from detection in current rapid kits.[16] To overcome the reduced sensitivity of the rapid kits, the FDA recommended (August 11, 2022) repeated testing within 48 hours .[17] Rapid antigen detection of variant spike proteins is even more challenging considering the spike genome's high mutation rate, often necessitating re-screening of spike antibodies, a step which can cause significant lag time behind the emergence of new variants.[18–20]

Elucidating the factors driving SARS-CoV-2 evolution affecting spike binding kinetics and stability at the host-cell surface will help predict further mutations or gains of function, as well as aid in developing variant-specific antiviral therapies and better antigen testing platforms. As cellular invasion by SARS-CoV-2 may involve at least 3 distinct cell surface glycoproteins (HSPG, ACE2, TMPRSS2), Scheme 1, the spike mutations might alter its interactions with any or all of them.[21–26] As the SARS-CoV-2 virion approaches the cell surface (Step 1 in Scheme 1A), it encounters the glycocalyx, a dense sugary matrix extending from the epithelial cell membrane.[27,28] Heparan sulfate proteoglycans (HSPGs), key components of the glycocalyx, are known to serve as attachment factors for many viruses and likely to make first contact with SARS-CoV-2.[29–31] HSPGs contain long, intrinsically disordered protein backbones decorated with longer (40-400 monomeric units) poly-sulfated and densely negatively charged glycosaminoglycans (GAGs).[32] Heparan sulfate (HS) itself is biosynthesized natively in repeating dimeric units of N-acetyl-D-glucosamine and D-glucuronic acid; post-processing enzymes then add sulfate groups to certain positions along an HS sequence, and epimerization enzymes may convert some D-glucuronic acid monomers to L-iduronic acid. Neither sulfation nor epimerization reactions go to completion, thus there exist locally controlled regions of high/low sulfation/epimerization proportions further contributing to vast degree of glycocalyx heterogeneity. HS regions with particularly high proportions of sulfation and L-iduronic acid are referred to as "heparin-like" domains, calling upon their similarity to short-chain medicinal heparin (HEP), which is almost completely sulfated and epimerized.[32]

Scheme 4.2: Schematic of host-cell invasion and electrostatic evolution of the SARS-CoV-2 Spike glycoprotein

(A) Scheme depicting initiation of host-cell invasion process through HS/ACE2/spike ternary complex formation via (1) SARS-CoV-2 approach to the human host-cell, (2) SARS-CoV-2 viral binding to Heparan Sulfate Proteoglycans (HSPGs) on the host-cell membrane, (3) conformational change of SARS-CoV-2 spike proteins from a closed to open state, open spike demonstrating 1 Receptor Binding Domain (RBD) in the "up" state with an exposed RBM, i.e. a "1up" spike, and finally (4) spike/ACE2 binding mediated by the Receptor Binding Motif (RBM). (B) Molecular models depicting steps 3 and 4 from panel (A).[33] (B(i)) Inset illustrating an apical view of the spike head highlighting the RBM's relative exposure in closed and open spike conformations. (C) mutations on the VOCs spike protein, introduction of positive charges due to mutation on spike protein marked as blue, negative charges as red, and neutral as grey; Total charge of each spike protein head domain (residues 13 to 1140, residues titrated to pH 7.4 with PROPKA[34]) given in parentheses next to the strain indicator.



Several studies in early 2020 first illustrated that the SARS-CoV-2 spike protein, particularly the spike RBD, can bind to HS and/or HEP,[35–38] and Clausen et al. showed that cellular HS was required for SARS-CoV-2 host-cell invasion.[27] With these prior results, the second step in the viral invasion process

comes into focus, as shown in Scheme 4.2A: Binding of the virion to HSPGs, likely through direct spike-HS interactions as shown by us[39] and others.[27,28,35–37,40–48] This step may increase virion residence time at the host-cell surface, thereby increasing the likelihood of encountering angiotensin converting enzyme 2 (ACE2), SARS-CoV-2's primary host-cell receptor.[44,45,49] The spike protein binds to ACE2 via a highly specialized interface within the Receptor Binding Domain (RBD) called the Receptor Binding Motif (RBM).[50–55] Each of the spike's three RBDs (1 RBD per spike protomer, 3 RBDs per spike protein, 1 RBM per RBD, 3 RBMs per spike protein) may often occupy a "closed" or "down" conformational state wherein these RBMs are largely shielded from recognition.[56–59] Thus, before spike-ACE2 binding can occur, at least one of these RBDs needs to emerge from its down/"shielded" state to an "up"/"open" state, to expose its RBM (Scheme 4.2B(i)).[56,57] While this step (Scheme 4.2B) can occur anytime along the spike lifecycle, Clausen et al report that short chain HEP is capable of inducing spike RBDs to move into the up state, suggesting that RBD opening could be induced after spike binding to HS and/or HEP within the glycocalyx.[27] With the RBM exposed, ACE2 can bind, further stabilizing the virion at the host-cell surface. Finally, with the spike protein immobilized by HS and/or ACE2, TMPRSS2 can cleave the S1 from the S2 spike domain at the Furin Cleavage Site (FCS).[60–64] After cleavage, the freed S1 domain peels off from the S2, revealing the spike's fusion peptide (FP) which then penetrates the host cell membrane and initiates membrane fusion.[65–67]

Recent studies have demonstrated that the Omicron SARS-CoV-2 virion relies less on membrane fusion as catalyzed by TMPRSS2 S1/S2 cleavage than earlier strains and rather more on endocytosis. Indeed, syncytial formation – neighboring SARS-CoV-2 infected cells fusing together, a phenomenon indicative of TMPRSS2 activity -- was reduced for Omicron infected tissues.[68,69] Furthermore, infection of TMPRSS2 knock-out cells by SARS-CoV-2 was increased for Omicron relative to other VOCs.[69,70] This mechanistic change could be due to Omicron sequence mutations near the S2' site causing decreased recognition by TMPRSS2.[68–70] Furthermore, recent research suggests that mutations in the Delta and Omicron RBMs result in altered binding affinities to ACE2 compared to WT.[21–26] Several past works have identified that charge-charge interactions heavily stabilize the WT spike-ACE2 interface,[71] and that increasing spike charge over the course of SARS-CoV-2 evolution potentially increases electrostatic recognition of ACE2 at long

range and increases immune escape,[47] especially for Omicron, as its spike's RBM is more positively charged than other spikes.[48] While the role of evolving positive charge on spike proteins has begun to be unraveled for ACE2,[21–26,47,48] this investigation was done in isolation from its required co-factor, HS.[71] As HS is a long, negatively charged polysaccharide, the growing positive charge on spiked virions, especially for Omicron, is expected to impact the stability of interactions with the HS-rich glycocalyx. Additionally, since HS was previously shown to stabilize spike-ACE2 interactions, an altered affinity to HS may in turn also affect binding to ACE2.

Here, we probe the secondary and ternary complex interactions between HS, ACE2, and the spike proteins of WT, Alpha, Beta, Delta, and Omicron SARS-CoV-2 variants. We propose how the positively charged Omicron spike may unlock a critical HS/ACE2 synergy. By harnessing the power of the primary and secondary cell receptors in the glycocalyx, we show how the performance of HEP anchored test strips can co-evolve with the SARS-CoV-2 genome for robust and rapid sensing of Omicron SARS-CoV-2. Finally, given that our synthetic glycocalyx test strips represent a minimal model for the cell surface, we discuss the potential implications our results may have on understanding SARS-CoV-2 infection dynamics at large.

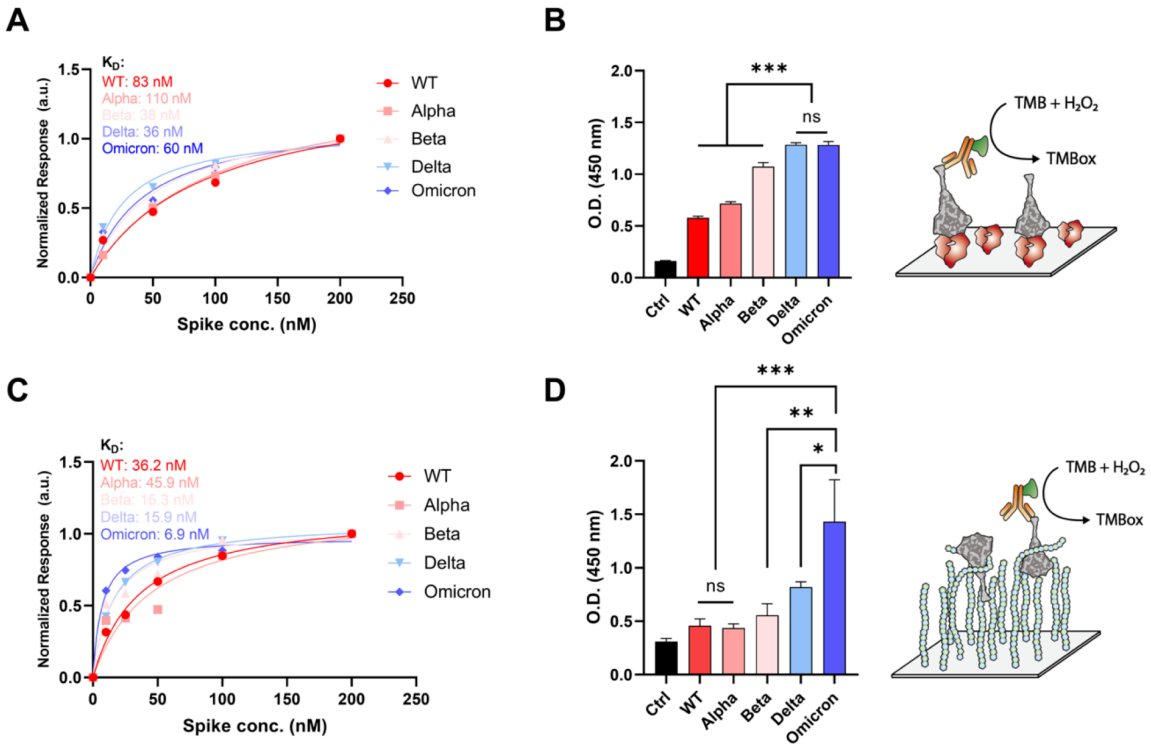


Figure 4.1: Binding affinities and degree of bound complex formation for variants of concerns
(A) Binding affinities of VOC spikes to dimeric ACE2 measured by BLI, K_D values were calculated with steady-state analysis. (B) Degree of bound complex formation for VOC spikes to dimeric ACE2 as measured by ELISA. TMB (3,3',5,5'-tetramethylbenzidine) was used as a chromogen for ELISA. (C) Binding affinities of VOC spikes to HEP measured by BLI, K_D values calculated with steady-state analysis. (D) Degree of bound complex formation for VOC spikes to HEP as measured by ELISA. Three independent tests were performed ($n \geq 3$), and p values <0.05 (*), 0.01 (**) and 0.001 (***) were determined using a one-way ANOVA with Tukey's post hoc test.

4.3 **Results & Discussion**

Spike mutations exhibit increased binding to ACE2 & Heparin

Previous work has demonstrated mixed results with respect to relative binding affinities between ACE2 and different spike protein sequences. Some groups report the Omicron spike protein binds to ACE2 with highest affinity, while others report there is no significant difference in binding affinity between all variants.[21–26] Herein, we have used bio-layer interferometry (BLI), enzyme-linked immunosorbent assays (ELISA), and molecular dynamics (MD) simulations to estimate relative binding affinities between WT, Alpha, Beta, Delta, and Omicron BA.1 spike proteins to ACE2. BLI results indicate similar binding affinities within nanomolar

range for all protein complexes, with Beta and Delta having the highest affinity, followed by Omicron BA.1, WT, and finally Alpha (Figure 4.1A, Figure S4.1). ELISA results show a clearer trend: Delta and Omicron spikes similarly show higher affinity to ACE2 than all other spike proteins (Figure 4.1B, Figure S4.2 for original ELISA data). Our BLI and ELISA results were consistent with previous reports showing affinities of Delta and Omicron similarly increased compared to WT.[21,22,25,26]

We also performed MD simulations of WT, Delta, and Omicron RBDs bound to ACE2 (SI Methods). From these simulations we see interactions at the RBD/ACE2 interface that may explain increased binding affinity between Omicron and WT. Barros et al. discuss the relative contact frequencies of three subregions within the RBM to ACE2: the L3 loop, the central beta-strands, and the right-hand loops (Figure S4.2).[53] They demonstrate that contacts formed between the central beta-strands and ACE2 are the strongest and maintain almost completely during their 1 μs of conformational sampling, whereas contacts formed between the L3 loop and right-hand loops to ACE2 are weaker, measured by decreased contacting frequency. From our MD simulations we see the Omicron RBM (Figure S4.2), with 10 total mutations, both strengthens interactions to ACE2 within the central beta strands (addition of salt bridge between Q493R to ACE2's E35) but also increases stabilization within the weaker right-handed loops through an aromatic interaction (N501Y pi-stacking with ACE2's Y41), and a tight hydrogen bonding network (Y505H tightly supporting a hydrogen bond between RBD's Y495 and ACE2's K353). These interactions, as also elucidated by Han et al.,[72] potentially explain the strengthening of affinity between Omicron spike and ACE2 relative to the WT.

As interactions between the SARS-CoV-2 spike protein and HS in the glycocalyx is a requisite step to binding ACE2,[27] we also investigated binding affinity between spike proteins and HEP. BLI results illustrate an increase in affinity between HEP and Omicron spike over Delta and other VOC spikes (Figure 4.1C, Figure S4.3). This result is also confirmed with ELISA as Omicron spikes bind to HEP with highest affinity, followed by Delta, then Beta, then Alpha and WT at similar affinity (Figure 4.1D). To probe differences in affinity to HEP at the molecular scale, we conducted extensive ensemble-based docking studies with AutoDock Vina[73,74] to predict HEP binding modes to WT, Delta, and Omicron spikes in closed and open conformational states (SI Methods). From ~28,800 binding modes, we clustered to identify 19 total HEP binding "hotspots", 4 of which

were highly buried within the spike and 15 of which were located on the spike surface, thus accessible to long-chain HEP binding as would be seen within the glycocalyx (see SI Scheme S1, SI Methods Section 2.1.2, SI Results Section 2.3.1, and SI Figure S4.9 for complete descriptions of surface versus buried sites). Relative affinities and populations of HEP binding modes at each of these sites were similar across the three spike variants (SI Figures S4.4-11). In past work, we predicted 3 sites of high importance for interaction between the spike protein and HS: an RBD cleft site, an RBD patch site, and the FCS.[39] Current ensemble-based docking simulations have confirmed the presence of these sites on WT and variant spikes and indicate there are no significant differences in binding affinities or number of binding modes in these sites between spike variants (SI Figures S4.4-11, SI Results and Discussion). To determine the degree to which induced-fit effects within the RBD Cleft, RBD Patch, FCS, and potential binding at the RBM could impact affinity, we then conducted targeted flexible protein-flexible ligand docking studies with HEP and HS tetramers at each of these sites across WT, Delta, and Omicron variants with Schrödinger's Induced Fit Docking protocol (Scheme S2, SI Methods Section 2.1.3).[75–79] Again, there were no significant differences between average predicted binding energies for HEP or HS tetramers at each of these sites, across the three spike variants, even with global (MD enabled ensemble docking) and local (Schrödinger IFD) protein flexibility incorporated (SI Figure S4.12). This likely indicates that once a HS/HEP fragment finds a site on the spike surface, it is flexible enough to accommodate sequence mutations and maintain affinity at the surface. These docking results suggest that the increased binding affinity between HEP and SARS-CoV-2 Delta and Omicron spikes relative to WT, as observed with BLI and ELISA, most likely do not originate from site-specific changes.

Evolution of positive charges on Spike protein enhances rate of binding to ACE2 & Heparin

As previously noted by us[49,80] and others[23,47,48] the spike protein is becoming more positively charged with each emerging VOC spike sequence (Scheme 1C). The total formal charge of the trimeric WT spike head domain (residues 13 to 1140) at pH 7.4 is +3, Alpha is +6, Beta +15, Delta +18, and Omicron BA.1 +24;

thus, the Alpha, Beta, Delta, and Omicron BA.1 spike protein sequences follow a trend of +3, +12, +15, and +21 in charge relative to the WT. Several glycans on the spike ectodomain are also shown to be sialylated. While a complete differentiation of glycan sialylation rate per spike sequence is far beyond the scope of this work, it is important to estimate the relative contribution of sialic acids to total spike head domain charge. Assuming a glycoprofile consistent with models from Casalino et al.[56] (14 sialic acid residues on 57 head domain glycans) and described by Watanabe et al.,[81] the total formal charge of the trimeric WT spike head domain with glycans is –11, Alpha is -8, Beta +1, Delta +4, and Omicron BA.1 +10 (Scheme 4.2C, Table S4.1). From this accounting of charge, it's clear to see that due to mutations in spike sequence, the spike protein head domains are increasing in total charge. Additionally, to identify where positive and negative charges are most concentrated on the spike surface, we calculated dynamically averaged electrostatic potential maps of the WT, Delta, and Omicron (Figure 4.2A, Figure S4.13, SI Methods). Interestingly, while the total spike charge is increasing, we also observed a clear redistribution of positive and negative charge across the spike surface between VOCs. Particularly, the Omicron spike surface exhibits a distinct redistribution of charges with a dramatic positively charged “bullseye” at its apex (Figure 4.2A(iii)). Of course, this bullseye is only clear in the closed state, but the remapping of charge across the variant timeline is also demonstrated for 1up spike conformations as well (Figure S4.13). Increasing total charge of the spike protein, along with redistribution of charges on the spike surface have the potential to impact HS/HEP binding to spike within the glycocalyx.

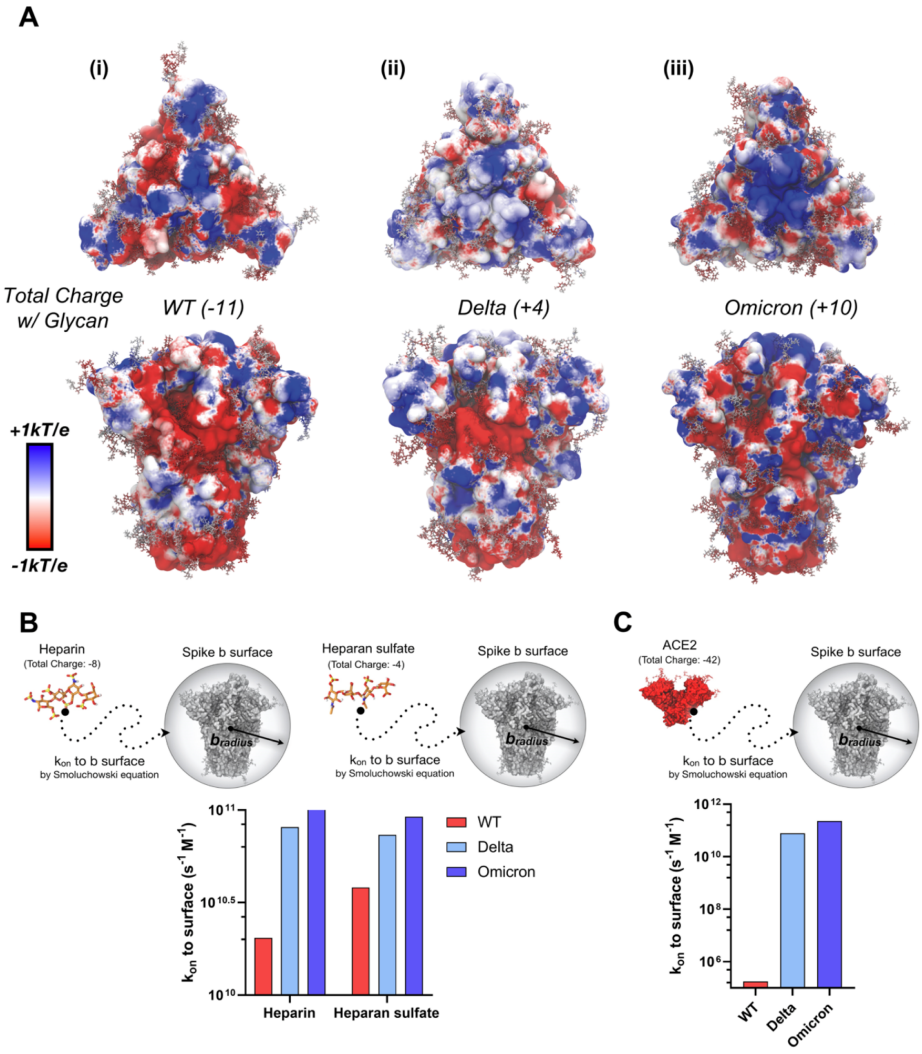


Figure 4.2: Electrostatic potential maps of variants of concern and rate-constants to b-surface between HS to spike variants

(A) Dynamically-averaged electrostatic potential maps collected from 50 ns of MD simulations for (i) WT, (ii) Delta, (iii) and Omicron spike proteins. The protein surfaces are colored according to average electrostatic potential at each site, ranging from –1 k_BT/e (red) to +1 k_BT/e (blue). Total charge of each spike protein head domain (residues 13 to 1140, residues titrated to pH 7.4 with PROPKA[34]) given in parentheses next to the strain indicator; considering including a glycoprofile consistent with Casalino et al.[56] and Watanabe et al.,[81] WT, Delta, and Omicron spike proteins have total charges of –11, +4, and +10, respectively (14 sialic acids). (B) Rate constant (k_{on}) to b-surface calculated between heparin and heparin sulfate tetramer to WT, Delta, and Omicron spike proteins (each titrated to pH 7.4) with a corresponding scheme demonstrating system diffusion to the b-surface. (C) Rate constant (k_{on}) to b-surface calculated between and the ACE2 exodomain (residues 18 to 734, pH 7.4) with a corresponding scheme demonstrating system diffusion to the b-surface. For B and C, when calculating rate constants (k_{on}) to the b-surface, receptor molecules are modeled as spheres defined by a b-radius, and total charge. The k_{on} is calculated analytically according to the Smoluchowski equation (details in SI). Bars are colored according to a red to blue color scale normalized to VOC spike total charge with 14 sialylated glycans: WT (-11) in red, Delta (+4) in light blue, and Omicron (+10) in blue.[56,81] It should be noted that error bars are not necessary for data presented in Figures 4.2B and 4.2C as these values represent exact analytical solutions to the Smoluchowski equation.

To investigate the effects of spike total charge on HEP binding, we used Brownian dynamics (BD)[82,83] simulations with Browndye[82,83] to calculate rate constants (k_{on}) to a “b surface”, wherein the center of mass of a receptor molecule of interest defines the center of a sphere with a "b-radius” (Figure 4.2B). The receptor and ligand molecules, each containing partial atomic charges, approach one another from infinite space. In such a model, a k_{on} between two molecules attaining an intermolecular distance less than the b-radius is largely driven by charge-charge interactions and can thus be solved numerically using the Smoluchowski equation.[84] These results provide insight into long-range electrostatic interactions between molecules. We observe a dramatic increase in k_{on} to the b-surface between a HEP tetramer (charge –8) and WT, Delta, and Omicron spike proteins, 2×10^{10} $M^{-1}s^{-1}$, 8×10^{10} $M^{-1}s^{-1}$, 1×10^{11} $M^{-1}s^{-1}$, respectively, (Figure 4.2B, SI Methods). A similar trend is observed for the k_{on} to the b-surface between a model HS tetramer (charge –4) and WT, Delta, and Omicron spikes (Figure 4.2B). Additionally, seeing as the k_{on} to the b-surface calculated for HS to WT was higher than that for HEP to WT, we predict that HS, owing to its decreased sulfation and charge, is likely to find and bind more quickly to WT spike surface than fully sulfated HEP domains. These results indicate that optimized long-range electrostatic interactions via spike mutations could dramatically impact the rate of SARS-CoV-2 viral approach to the glycocalyx (Step 1 in Scheme 1). Together with our docking results, which predicted very little difference between VOC spikes in HEP binding affinities at HEP-binding hotspots, the BD results illustrate that increased affinities between HEP and Delta/Omicron spikes relative to WT, as seen by BLI and ELISA, may be due to kinetic selection allowing for increased encounters rather than site-specific differences in binding affinity. Furthermore, due to drastic difference in k_{on} to b-surface for WT spike between HEP and HS tetrameric models (Figure 4.2B) -- a trend which is nonexistent, if not reversed, for Delta and Omicron spike proteins – we predict that WT spike proteins have the potential to demonstrate increased selectivity for the less densely sulfated/charged heparin-like domains, while Delta and Omicron spike proteins would demonstrate little to no selectivity, or moderately increased selectivity for more densely sulfated/charged heparin-like domains.

Considering that ACE2's dimeric exodomain has a total charge at pH 7.4 of –42 (total formal charge with glycans and at pH 7.4, –54)[53] rates to ACE2, driven by long-range electrostatic interactions, may also be affected by increasing spike protein charge. Therefore, we also calculated rate constants (k_{on}s) to b-surfaces between the ACE2 ectodomain and WT, Delta, and Omicron spike proteins (SI Methods). Interestingly, we see six orders of magnitude increase in k_{on} between WT and Delta spikes to ACE2, followed by a one order of magnitude increase in k_{on} between Delta and Omicron spikes to ACE2 (Figure 4.2C). As with the HS/HEP results, the increasing total charge of spike proteins may strengthen long-range electrostatic interactions to negatively charged ACE2 (-42). Additionally, recall that binding affinities between SARS-CoV-2 VOC spikes and ACE2 are increasing (decreasing K_D), but only moderately (Figures 4.1A-B). In sum, these BD results for HEP, HS, and ACE2 all point to kinetic fitness for interactions between the spike protein and negatively charged HEP/HS/ACE2 within the charged glycocalyx as a potential underlying evolutionary pressure driving SARS-CoV-2 spike sequence adaptation. Furthermore, many glycoprotein species and glycans within the glycocalyx such as neurophilin, CD147, GRP78, and sialic acids, could interact with spike protein. These interactions could also be altered with the change in the charge on spike should be noted.

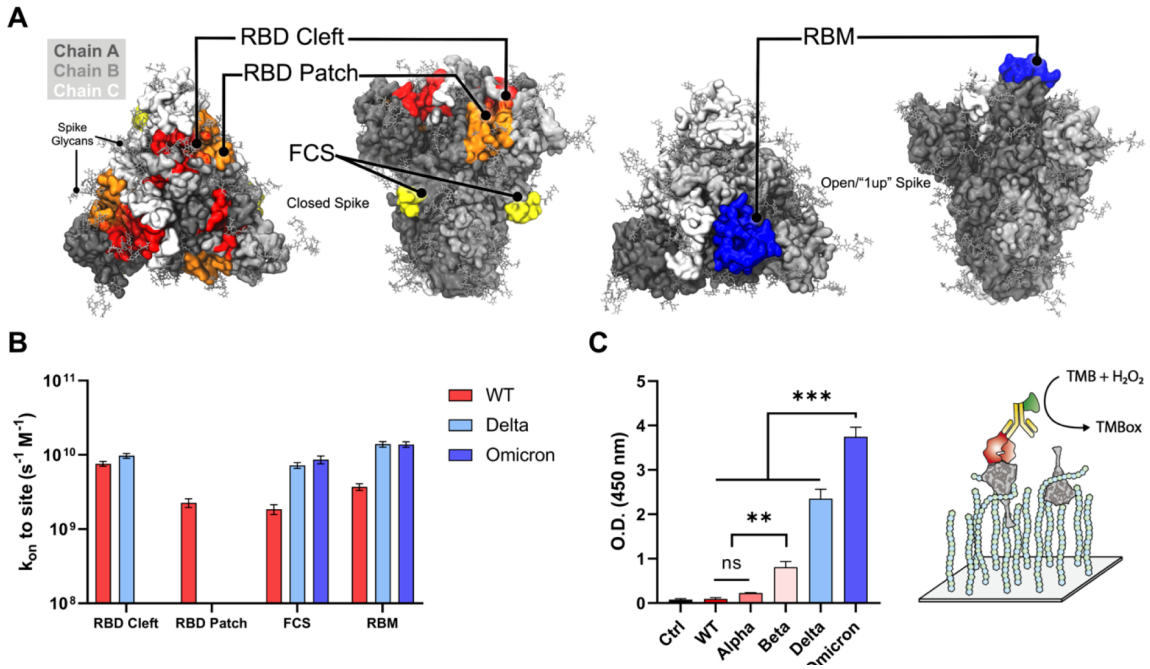


Figure 4.3: Depictions of HS binding hotspots on the spike and corresponding brownian dynamics results

(A) Spike structures illustrating receptor binding domain (RBD) cleft, RBD patch, furin cleavage site (FCS), and the receptor binding motif (RBM) as designated sites targeted with BD simulations and (B) corresponding BD results shown as second-order rate constants between HEP and identified RBD cleft, RBD patch, FCS, and RBM sites. (C) Degree of bound ternary complex formation for VOC spikes to HEP and ACE2 as measured by ELISA. Three independent tests were performed ($n \geq 3$), and p values <0.05 (*), 0.01 (**) and 0.001 (***) were determined using a one-way ANOVA with Tukey's post hoc test.

Remapping of positive charge distribution on Omicron surface maximizes heparin/ACE2 synergy

As can be seen from the dynamically averaged electrostatic potential maps in Figure 4.2A(i-iii), Figure S4.13, spike sequence mutations increase the total spike charge as well as redistribute surface patches of positive and negative charge. As a result, the site of first contact between HS/HEP and spike surface, i.e., a nucleation site for HS/HEP long-chain binding to the spike surface, could be altered on a per-VOC basis. To probe these changes, we again used BD simulations to investigate the rate of HS and HEP tetramer association, this time specifically to the RBM, RBD cleft, RBD patch, and the FCS sites (SI Methods). We find that HEP tetramers associate differentially to spike surface sites due to mutations at each site (Figure 4.3A-B). Upon approaching a WT spike protein, our kinetic experiments indicate that HEP tetramers are most likely to associate with the RBD Cleft site first, followed by the RBM, the RBD Patch, and finally the FCS.

However, upon approaching a Delta spike protein, HEP tetramers are most likely to find the RBM first, followed closely by the RBD Cleft, and finally the FCS, with no observed transitions into the RBD Patch. Similarly, when encountering Omicron spike proteins, HEP tetramers are most likely to first find and bind the RBM, followed by the FCS, with no observed transitions to the RBD Cleft or Patch sites. ESP maps of the Omicron RBM reveal that the Omicron RBM is strongly positively charged, which likely supports the kinetic advantage for binding heparin at this site (Figure S4.13C(i)). These results indicate that redistribution of positive charges, especially for Omicron spikes, might cause a competition between HEP and ACE2 binding on RBM site of spike protein. However, at the cell surface, ternary complex formation between HS, spike, and ACE2 is potentially required for stabilization of the spike-ACE2 interface.[27] Thus, we conducted ternary complex ELISA to identify whether HEP and ACE2 compete with one another for spike binding on a per-VOC basis (Figure 4.3C). While this ternary complex ELISA relies on both spike-HEP and spike-ACE2 binding, and thus could be hindered by decreased affinity at either interface. Strikingly, we observed a significant increase in the affinity of the Omicron spike-HEP-ACE2 ternary complex over all other variants, including Delta. It should be noted that when comparing Figure 4.3C to Figures 4.1B and 4.1D, the trend in ternary complex formation affinities compares more similarly to that seen in spike-HEP binary complex formation (Figure 4.1D) than spike-ACE2 complex formation (Figure 4.1B). Furthermore, the relatively similar affinities for Delta and Omicron spike proteins to ACE2 binding, does not translate to a similar affinity in ternary complex formation for these two spike variants. These results indicate that remapping of positive charges, especially for Omicron, did not hinder the binding of HEP and ACE2 on spike protein. Instead, HEP and ACE2 can co-bind to the spike protein with enhanced binding with the timelines of mutant (Figure S4.13D).

As heparan sulfate has been reported to induce the open conformation of spike protein and enhance the ACE2 binding. We further investigated the synergistic formation of ternary spike-HS-ACE2 complexes we used Mass Photometer (MP) and compared the effect of HS on formation of ternary complexes for WT, Delta, and Omicron spike proteins in the presence of dimeric ACE2 by measuring the mass distributions (Figure 4.4, Figure S4.14-S4.16 for repetition of mass photometer results, SI Methods). The mass of each

trimeric spike protein was measured to be around 560 kDa, and the mass of dimeric ACE2 was 240 kDa. Given that spike likely first encounters the extended tendrils of HS upon approach to the human host-cell, we sequentially added first HS and then ACE2 to spike protein samples to mimic conditions at the cell surface. Co-incubation of spike with just ACE2 generated mass peaks for WT, Delta, and Omicron spike around 800 kDa, followed by signal density in higher mass ranges indicating that spike and ACE2 are interacting to form complexes at varying stoichiometric ratios (Figure 4.4A (iii)). Interestingly, while incubating spike and ACE2 with HS yielded very little differences in MP spectra for WT and Delta spikes compared to no-HS conditions, Omicron showed a significant increased population around 1200 kDa under these testing conditions (Figure 4.4A (iv)), suggesting it plays a role in stabilizing a ternary spike-HS-dACE2. To assign possible stoichiometries to the emerging 1200 kDa complex, we must understand the structural requirements for such assembly. Spike protein binding to ACE2 requires at least one spike RBD to be in the up conformation and a successful binding event between the two proteins is canonically considered as occurring between one ACE2 and one 1-up spike. However, a spike protein with three RBDs in the up-state could accommodate binding of up to three dACE2 ectodomains and dimeric ACE2 could itself accommodate binding of up to two spike proteins.[53,85] Additionally, while dimeric ACE2 was used in this work, without the presence of B^0AT-1 – B^0AT-1 is a sodium-dependent neutral amino acid transporter commonly found co-expressed and complexed with ACE2 in the gastrointestinal tract, B^0AT-1 is often used to stabilize ACE2 during structural elucidation[51] -- and its corresponding stabilization of the ACE2 interfacial neck domain, ACE2 in solution could exhibit more flexibility and adopt dual RBD binding modes as described by Xiao et al.[86] While all such complexes are likely biologically relevant, the degree to which, and by what mechanism(s), spike and ACE2 form such complexes of “intermediate” stoichiometry is an open question. Thus, to parse our current Mass Photometry results, we have enumerated several configurations of spike-ACE2 complexes (illustration within Figure 4.4A) and divided such complexes into three groups based on their expected mass range: A (650 - 900 kDa), B (900 – 1300 kDa), C (1300 – 2000 kDa). To compare the change in ternary complex distribution with or without HS, the fraction of each group (denoted as A, B, C) was calculated for each spike protein (Table S4.4-S4.6 for tables denoting fractions per group for Omicron, Delta, and WT spike

proteins). As shown in Figure 4.4B, although the addition of HS slightly increased the population of group A type complexes for WT and Delta spikes, there was no significant change in degree of complex formation for types A, B, or C for WT and Delta spikes. However, Omicron spikes showed increased proportions of type B complexes in the presence of HS (Figure 4.4B). Considering that HS may stabilize spike RBDs in their up conformation, as reported by Clausen et al.,[27] binding of multiple HS fragments to the Omicron spike could serve to recruit additional nearby ACE2s for binding, thereby increasing the population of group B type complexes (Spike:ACE2 = 1:2, 1:3). Additionally, an increased proportion of group B could also stem from one dimeric ACE2 binding two trimeric spike proteins, an interaction which could easily be facilitated by long chains of HS either binding one or both spike proteins. ACE2 bridging multiple spike proteins is likely an important factor governing complex formation at the cell surface, and HS has the potential to “hold” or cluster multiple spikes near ACE2 in preparation for multiplex binding.

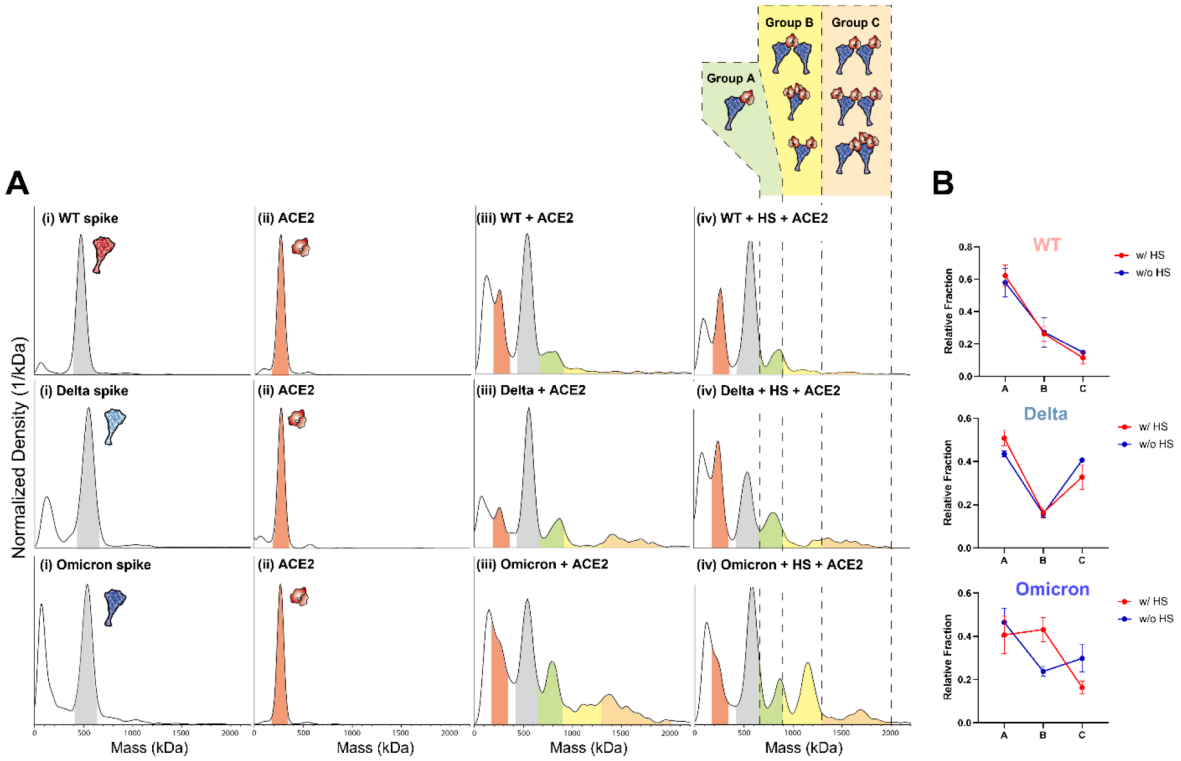


Figure 4.4: Mass photometry results for spike variants of concern and fraction of ternary complex formation

(A) Mass photometer results comparing the WT, Delta, Omicron binding to ACE2 and heparan sulfate (HS). (i) Mass distribution of WT, Delta, and Omicron Spike (mass range highlighted in grey) (ii) Mass distribution of the dACE2 (mass range highlighted in red), (iii) Mass distribution of Spike protein + ACE2, and (iv) Mass distribution of Spike + HS + ACE2. To mimic the viral entry mechanism, HS was incubated with spike proteins first then followed by addition of ACE2. Possible ternary complexes are grouped in A (green), B (yellow), C (orange) based on their expected mass ranges. The molar ratio used in this study was Spike:HS:ACE2 = 1:1:0.5. (B) Fraction of ternary complex with or without HS for WT, Delta, and Omicron obtained by mass photometer. To calculate the fraction of these complexes, count numbers from each group (A, B, C) in Figure 4.4A were obtained. At least three independent experiments were performed, and error bars were calculated by the standard deviation of all experiments. Significance was calculated via multiple t test (unpaired) with Holm-Sidak method (α: 0.05) was performed.

To summarize all results presented thus-far: (1) binding affinity between SARS-CoV-2 VOC spikes to ACE2 are moderately increasing over the variant timeline, (2) site specific affinities between SARS-CoV-2 VOC spikes to HEP dimers and tetramers have not changed significantly over the variant timeline but (3) binding affinity between spike VOCs and long-chain HEP has increased over the variant timeline, (4) increasing total spike charge over the variant timeline may be increasing rates of HEP/HS/ACE2 to spike surfaces, (5) charge redistribution on the spike surface over the variant timeline may be altering HEP/HS

nucleation sites in the context of long-chain binding interactions, and finally (6) Omicron has a particular ability to unlock a key HS/ACE2 synergy by increasing proportions of 1:2 and 1:3 spike:ACE2 complexes. At the cell surface, an individual spike glycoprotein will likely encounter both HS and ACE2. In what order, and by what mechanism(s) does the spike glycoprotein interact with and exploit the native functions of HS and ACE2 to enter the human host-cell, and how do mutations to the spike sequence affect this mechanism remain outstanding questions. As illustrated in Scheme 4.2, for a spike-ACE2 binding event to occur, the spike's RBM needs to be sufficiently exposed, which only occurs when at least one of the spike protein's RBDs moves from a "down"/"shielded" state to the "up"/"exposed" state. Clausen et al. report that short chain HEP can increase proportions of ACE2s bound to the spike protein, suggesting HEP can facilitate RBD opening and ACE2 binding. Based on our results, we hypothesize that as the total formal charge of the spike protein increases so does the spike's fitness for moving through and interacting with the negatively charged glycocalyx and ACE2 as shown in BLI, ELISA, and BD results. Moreover, as the spike approaches the glycocalyx, certain sites (i.e., the RBM and FCS) on VOC spikes may find and bind to HS more quickly than to others due to redistribution of charges on the spike surface. For example, in the case of the Omicron spike protein, given the rate constant for HEP binding is fastest to exposed RBMs, HS/HEP could increase the local concentration of 1up, 2up, and 3up Omicron spike proteins directly at the cell surface. While bound to the Omicron spike protein, HS would thereby stabilize spike in an attack-ready conformation while ACE2 arrives on the cell surface below. ACE2 could eventually displace HS from the RBM, which could in turn shift to one of the many other GAG-hotspots on the spike surface, including the FCS. In this fashion, the Omicron spike protein's RBM could be capitalizing on HS/HEP's capacity for kinetic selection, thereby increasing the localized concentration of ACE2-ready binding partners at the cell surface. Taken together, we suggest a mechanism by which SARS-CoV-2 variants evolve to better bind the co-receptor glycocalyx HS, which indirectly enhances its chances to bind and the stability of its interactions with the primary receptor, ACE2.

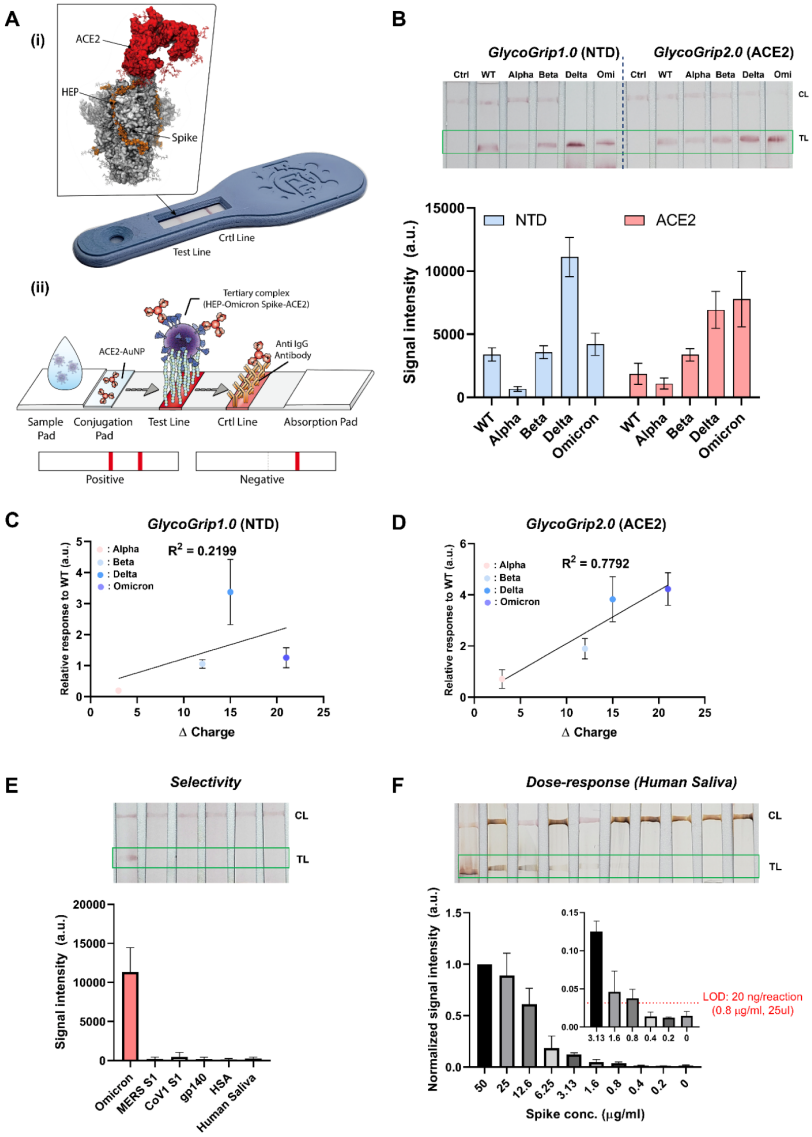


Figure 4.5: Fine-tuning GlycoGrip for variants of concern

A (i)) Image of a GlycoGrip prototype with callout image depicting a spike protein bound to both HEP (the GlycoGrip capture agent) and ACE2 (the signaling probe). (A (ii)) Schematic illustrating GlycoGrip's capture and signaling of SARS-CoV-2 virions. (B) Comparison of the GlycoGrip1.0 and GlycoGrip2.0 to VOCs with two different signaling probes (NTD Ab and ACE2; NTD Ab was previously used in GlycoGrip1.0), and Correlation between relative GlycoGrip1.0 signal intensity versus change in spike charge; charge change calculated as VOC spike charge – WT spike charge, relative signal intensities plotted as ratio with respect to WT (same data as shown in panel B (NTD)). (D) Correlation between relative GlycoGrip2.0 signal intensity versus change in spike charge; charge change calculated as VOC spike charge – WT spike charge, relative signal intensities plotted as ratio with respect to WT (same data as shown in panel B (ACE2)). (E) Selectivity of the GlycoGrip2.0 to relevant proteins including MERS, CoV1 spike, HIV envelope protein (gp140), Humans serum albumin (HSA), and human saliva. (F) Dose-dependency results of Omicron detection using GlycoGrip2.0 with signal enhancement in human saliva condition. The limit of the detection was calculated by the blank + 3x (Standard deviation of blank). At least three independent tests were performed (n ≥ 3) for GlycoGrip.

GlycoGrip is a minimal model for the host-cell surface and signals effectively for evolved variants

As discussed in the introduction, maintaining robust testing via rapid antigen and PCR detection platforms becomes a challenge during an actively progressing public health crisis such as the COVID-19 pandemic. Recently, we showed that the interaction between the host-cell surface glycopolymers and the spike glycoprotein can be exploited in a sandwich-style lateral flow strip-based assay (LFSA) to tackle this exact challenge.[39] Our sensor termed GlycoGrip, was inspired by the interaction between SARS-CoV-2 virions and the glycocalyx. GlycoGrip uses long-chain Heparin (HEP) to capture, and Au-nanoparticle conjugated anti-spike antibodies to signal for the presence of SARS-CoV-2 spike proteins. In our first generation of GlycoGrip, a.k.a. GlycoGrip1.0, we used an N-terminal domain (NTD) based anti-spike antibody to signal for the presence of WT, Alpha, Beta, and Delta spike proteins. In the current work, we investigate the potential of using ACE2 as an ever-relevant signaling probe: if the virus infects, GlycoGrip2.0 will detect, (Figure 4.5A (i,ii)).

For both generations of GlycoGrip, when a sample contains spike protein, a double banded signal will appear on the lateral flow strip: one band at the test line (TL) indicating ternary complex formation between HEP, spike, and the signaling probe, and one band at the control line (CL) indicating binary complex formation between the signaling probe and an anti-signaling probe antibody. We compared GlycoGrip1.0 and GlycoGrip2.0 against all VOC spike proteins (Figure 4.5B). We observed that, while GlycoGrip1.0 still signaled well for Omicron spike, signal intensity dropped significantly relative to Delta. However, the trend observed for WT, Alpha, Beta, and Delta spikes on GlycoGrip1.0 was similar to that reported previously.[39] Notably, GlycoGrip2.0 elicited the strongest signal for Omicron spikes compared to GlycoGrip1.0. Moreover, we observed a clear trend: signal intensity on our GlycoGrip2.0 increases along with the variant timeline (Figure 4.5B) in a manner strikingly similar to the increase in total spike charge. Plotting this change in spike charge (i.e., total charge changes relative to WT spike) against relative signal intensity on our reconfigured GlycoGrip2.0, we see these two quantities correlate with one another: $R^2 = 0.7792$ (Figure 4.5E) while there

was no clear trend for GlycoGrip1.0 (Figure 4.5C): R^2 = 0.2199. This correlation is striking given that, with HEP as the capture agent and ACE2 as the signaling probe, GlycoGrip2.0 can be seen as a simplified model for the cell surface environment. These results indicate that, in contrast to antibody-based detection, our cell-surface mimetic sensor easily and effectively adapts to viral mutations, suggesting a novel paradigm shift in designing LFSA platforms to sensing viral antigens with high mutation rates.

Finally, we investigated the selectivity and sensitivity of our GlycoGrip2.0 specifically for detection of Omicron spike proteins. To determine the selectivity, we interrogated our sensor with related coronavirus (MERS, CoV1) and HIV (gp140) envelope proteins along with relevant complex proteins such as human serum albumin (HSA). To illustrate GlycoGrip's feasibility when used against complex biologically relevant media, we also tested our sensor against a non-infected human saliva sample to check the false-positive signal from the complex biological samples. As shown in Figure 4.5E, GlycoGrip2.0 selectively captures and signals for Omicron spike proteins while not binding to related viral proteins, HSA, or other saliva matrix elements. Finally, Omicron was detectable as low as 40 ng/reaction (1.6 μg/ml, 25 μl, Figure S4.17) with ACE2, and 78 ng/reaction with NTD Abs. We then adopted a silver staining method to further enhance detection 4-fold for Omicron spikes: down to 10 ng/reaction (0.4 μg/ml, Figure S4.18). Lastly, we validated our sensor performance in human saliva samples. Human saliva contains various glycoproteins including neutrophil elastase and histone H2A which could interfere the binding of either ACE2 or heparin to spike protein.[87] The LOD of our sensor was estimated to be 20 ng/reaction (Figure 4.5F) which was comparable to the buffer condition and reported LFSA sensor for the spike protein detection (Table S4.7 for summary of analytical performance with reported LFSA sensors). Indicating the feasibility of our sensor for more complexed patient samples application. In all, these results indicate GlycoGrip is selective for SARS-CoV-2 spike proteins, signals strongly in the presence of Omicron spikes, and is rapidly adaptable and deployable within the context of the ever-evolving COVID-19 public health crisis. Furthermore, as SARS-CoV-2 continues to adapt to niche evolutionary pressures within glycocalyx, GlycoGrip's detection capacities will likely maintain and even strengthen without the need to change any of the sensor components.

4.4 **Conclusions**

As the COVID-19 pandemic now progresses into its third year, public health experts continue to scan the epidemic-horizon for new variants. Delineating the environmental and immunological pressures driving SARS-CoV-2 genomic adaptation can help predict the likely range of future mutations, and the potential impacts of those mutations on infection, re-infection, hospitalization, and mortality rates. In this work, we revealed that increased total charge on the spike proteins of SARS-CoV-2 variants, due to the progressive addition of positively charged mutations, strengthens long-range electrostatic interactions with the negatively-charged host cell surface. Furthermore, we showed that the redistribution of positive and negative charges on evolving spike protein variants, particularly for Omicron, which adopts a striking 'bullseye' like patch of positive charge near the RBM, selectively enhances the rate and strength of HS binding to exposed RBMs. We thus hypothesize that Omicron SARS-CoV-2 kinetically increases the local concentration of ACE2-binding-ready spikes at the cell surface and unlocks a key synergy between HS and ACE2. We believe this remapping of positive charge on the SARS-CoV-2 spike protein is an evolutionary driver for the optimization of electrostatic interactions of spike proteins with both HS and ACE2, thereby increasing the rate of viral entry. With these conclusions in hand, one could predict that emerging variants will exhibit additional charge redistribution to further fine-tune these interactions and in-turn increase SARS-CoV-2 infectivity.

Finally, we leveraged our findings of a favorable tertiary complex formation between the Omicron spike protein, HS, and ACE2 to develop the GlycoGrip2.0 sensor. Additionally, we demonstrated GlycoGrip's ability to "co-evolve" alongside the SARS-CoV-2 genome and we improved its detection of all variants of concern. By harnessing the primary (ACE2) and secondary (HS) cellular receptors in one sensor, GlycoGrip2.0 essentially serves as a minimal model of the glycocalyx environment, which may also serve as a useful platform for viral surveillance. This highlights the advantage of glycocalyx-inspired sensing in a rapidly adapting public health crisis, as it is quickly reconfigurable and employable against evolving variants. As the COVID-19 pandemic is still ongoing, due to continuous evolution of the virus, glycocalyx-inspired LFSAs are

likely to be a great benefit for global health monitoring power, not only for SARS-CoV-2 but other rapidly mutating respiratory viral antigens.

4.5 **Materials & Methods**

4.5.1 Computational Methods

4.5.1.1 WT, Delta, Omicron Spike System Construction, MD Simulation, and Clustering:

Fully glycosylated, all-atom models of WT, Delta, and Omicron SARS-CoV-2 spike glycoprotein head domains (residues 13 to 1140) were constructed according to the following protocols.

WT: For construction of our WT "closed"/all RBD down system, a cryo-EM structure with 2.8 Å resolution was used (PDB ID 6VXX).[1] To improve the accuracy of our model, we incorporated fully resolved NTD, RBD, and pre-fusion loops from another closed spike structure (PDB ID 7JJI).[2] For construction of our WT "open"/1 RBD up system, a cryo-EM structure with 3.46 Å resolution was used (PDB ID 6VSB).[3] To improve the accuracy of our open model, we incorporated fully resolved RBD in an up-state bound to ACE2 (PDB ID 6M17)[4] and fully resolved NTD and pre-fusion loops (PDB ID 7JJI).[2]

Delta: We used the WT open and closed structures as described above as the basis for construction of our Delta variant closed and open spike glycoprotein systems. To account for the mutation profile in the Delta variant, we induced single point mutations using the "mutate" command in psfgen. Experimental data showed significant rearrangement/structural remodeling of the NTD on Delta, so we incorporated a cryo-EM structure of the remodeled Delta variant NTD (PDB ID 7SO9).[5]

Omicron: For construction of our Omicron variant closed system, a cryo-EM structure with 3.36 Å resolution was used (PDB ID 7TF8)[6] as the base structure. Missing loops from the furin cleavage site in the Omicron PDB were grafted in from PDB ID 6VSB.[3] For construction of our Omicron variant open system, a cryo-EM structure with 3.40 A resolution was used (PDB ID 7TEI)[6] as the base structure. Missing loops in the furin cleavage site, fusion peptide, and RBD were grafted in from PDB ID 6VSB.[3] For both open and closed omicron spike glycoprotein systems, we incorporated a cryo-EM structure of a fully resolved Omicron NTD (PDB ID 7K4N).[7]

Glycosylation/ Protonation/ Solvation/ Neutralization: All spike models were then glycosylated following the same glycoprofile as used by Casalino et al.,[8] consistent with Watanabe et al.[9] Protonation states were assigned by performing stand-alone PROPKA3[10] so that the glycan atoms were considered in the calculation, however protonation states (HSE vs HSD) for histidines were assigned by use of PROPKA through Schrödinger's Protein Preparation Wizard tool.[11] AutoIMD, a VMD tool,[12] was used to resolve any glycan/protein clashes or ring penetrations in our glycoprotein systems. Glycoprotein models were then each solvated in explicit water boxes of 215 x 215 x 215 $Å^3$ and neutralized with 0.15 M NaCl.

Molecular Dynamics Simulations: All structures (6 models in total) were then subjected to the following Molecular Dynamics (MD) simulation protocol (1 replica each) with NAMD2.14,[13,14] all atoms described according to CHARMM36m all-atom force field:[15–17] 20,000 steps of Steepest Descent minimization for TIP3 water molecules and NaCl ions. Protein and glycan atoms were held fixed with a Lagrangian constraint. Heating of the solvated system from 10K to 310K by increments of 25K with protein and glycan atoms held in light restraint according to a force constant of 1 kcal/mol/Å. With each increase in temperature, 10080 steps (1fs/step) of MD simulation were performed within the NpT ensemble. Once the temperature reached 310K, 0.5 ns of NVT equilibration was performed with restraints maintained. NpT equilibration (310K, 1.01325 bar) for 0.5 ns (2fs/step) with restraint (force constant = 1 kcal/mol/Å) applied to all protein backbone atoms. Pressure was maintained a Langevin barostat. Box cell dimensions were set to flexible during pressure equilibration. NVT free (no restraints) production (310K, 1.01325 bar) simulations for 50 ns (2fs/step). 50ns NVT production runs were performed on TACC Frontera. As system pressure was equilibrated in the prior step, box cell flexibility was turned off in this step (useFlexibleCell = no).

Clustering: In preparation for ensemble-based spike/GAG docking studies with AutoDock Vina,[18,19] we selected the final frame of each 50ns MD simulation to serve as a rigid-docking receptor. We also clustered the 50ns trajectories to generate 5 other receptor structures per spike glycoprotein model. The 50ns trajectories were clustered in Python (with MDAnalysis[20,21] and Scikit-learn[22]) according to the following protocol: To remove global rotational and translational degrees of freedom before clustering, VMD[12] was used to align trajectories according to minimum Root Mean Square Deviation (RMSD) distance of all C_a atoms from

their first frame positions. Water and ion atoms were stripped from resultant aligned trajectories. Aligned trajectories were then opened as universes in MDAnalysis[20,21] wherein the RMSDs of C_α atoms and glycan carbon atoms were calculated for each frame. Python Scikit-learn's[22] Kmeans clustering package was then used to cluster all frames according to C_α atom and glycan carbon atom RMSDs, and the knee locator algorithm was used to select the optimal number of clusters per simulation. Representative structures -- i.e., those simulation frames closest in RMSD space from the true cluster center – for the 5 most populated clusters were then selected for ensemble-based docking with AutoDock Vina.[18,19] PSF/PDB pairs were generated for each structure selected herein (i.e., for the final frame and for all clustered frames) and have been made available with our shared data sets on the AmaroLab website (https://amarolab.ucsd.edu/covid19.php).

4.5.1.2 Ensemble-based Docking with AutoDock Vina:

As described above, we selected 6 total structures per spike conformation (the final frame from 50ns and 5 representative structures from most populated clusters) to serve as receptors in ensemble based rigid docking studies with AutoDock Vina.[18,19] The following docking methods are detailed graphically in Scheme S1 below. Each chosen spike receptor structure (3 variants x 2 conformational states x 6 selected frames = 36 total receptor structures) was subjected to the following protocol. To ensure receptor grids generated with AutoDockTools[19] would be similarly applied to each receptor structure, first all receptor PDB structures were aligned to one another according to the S2 domain C_α atoms (protein residues 686 to 1140). After alignment, all receptor structures were converted to pdbqt filetype with AutoDockTools.[19] Heparin (HEP) dimer and tetramer .pdbqt structures were used in previous work[19] and thus the same files were used in this work (see Figure S4.19 for images denoting exact molecular structures modeled in this work). Per protein structure, the center of the AutoDock[14,15] receptor grid was defined as the geometric center of the central helix atoms (protein residues 985 to 1000); this was a choice made to ensure relative consistency of the box center for all structures, regardless of closed/1up conformation. Grid box size was set to 150 x 150 x 150 $Å^3$ for all spike

structures, this was chosen to ensure RBDs in the 1up state would still be encompassed within the resultant grid. All docking input files can be found in the datasets shared on the AmaroLab website (https://amarolab.ucsd.edu/covid19.php). AutoDock Vina[18,19] settings were applied as follows: energy_range = 30, exhaustiveness = 80, num_modes = 100, a combination which gave 20 binding modes per docking study. To thoroughly sample binding sites and modes on the spike surface, we conducted 20 runs of each docking procedure. A "docking procedure" being defined as one GAG model (HEP dimer or tetramer) docked into one spike receptor structure (e.g., a "docking procedure" could be described as dimeric HEP docked to WT spike in 1up state, clustered frame #1). Thus, with 20 replicas per docking procedure and 20 resultant binding modes per procedure, we obtained 400 binding modes per docking procedure. With 72 total docking procedures (2 GAG models, 3 spike variants, 2 spike conformational states, 6 receptor structures per spike variant/conformational state), we obtained 28,800 total binding modes from ensemble-based docking in this work.

We then used Scikit-learn's KMeans[22] clustering algorithm to cluster the geometric centers of all 28,800 resultant binding modes, and kneed, an inflection point calculation algorithm, to find the optimal number of clusters. From this clustering, we identified 19 distinctive GAG hotspots (Figure S4.8). To determine which of these 19 sites were accessible to long-chain HEP or HS, as would be encountered on the cell surface, we scanned all selected receptor structures (3 spike variants x 2 spike conformational states x 6 selected frames per variant/conformational state) to identify all residues within 10 Å of each hotspot centroid (for a full list of residues per site Table S4.3). As done in previous work,[23] we then used the Shrake-Rupley algorithm[24] to calculate the Accessible Surface Area (ASA) of each of these sites from the ~2 μs of freely available MD trajectories provided by Casalino et al.[8] We used these simulations to estimate ASA of defined sites, as opposed to our own 50ns trajectories, because Casalino et al.'s are much longer, and therefore likely to be more representative of conformational variability, especially with respect to glycan degrees of freedom. We calculated ASA for all sites at probe radii ranging from r=1.4 Å (reflective of water molecule probe), 7.2 Å (reflective of small molecule binding or an antibody hypervariable loop), and 18.6 Å (reflective of a small protein binding partner or antibody's variable fragment domain). We compared ASA results calculated at

r=7.2 Å between all sites and saw that in the closed state (Figure S4.9), sites K, M, N, and R, were highly buried and likely not accessible to ligand binding. However, from the 1up state, site M becomes moderately exposed. Upon visualizing these sites on the spike structure with VMD we determined sites K, M, N, and R were indeed buried sites, however further investigation will be necessary to determine if site M does indeed become sufficiently exposed after spikes move into the 1up conformation. We then conducted further statistical analyses of these sites with MDAnalysis[20,21] as are described in the Supporting Information Results and Discussion below.

4.5.1.3 Schrödinger IFD:

In past work,[23] through ensemble-based docking, we identified 3 sites on the spike surface with high affinity to GAGs which could be important for anchoring the spike to long-chain GAG binding modes within the glycocalyx: the RBD cleft, the RBD patch, and the FCS. Ensemble-based docking studies in this current work (described above) reconfirmed the presence of these sites. Additionally, Brownian Dynamics simulations show the importance of the RBM as a potentially adapting kinetic discriminator for HS within the glycocalyx. While we have already incorporated a degree of protein and glycan motion with our ensemble-based docking studies -- by selecting clustered spike structures from 50ns of MD simulation -- local binding site conformations can adapt to ligand binding. To assess the degree to which local rearrangements at key binding sites could contribute to GAG binding and to see how these rearrangements do or don't change with the introduction of spike mutations, we conducted site specific flexible ligand-flexible receptor docking simulations with Schrödinger IFD (Scheme S2).[25] For docking into the RBD Cleft, RBD Patch, and FCS sites, the final frame from 50ns simulations of each variant closed spike structure was taken. Since the spike is a trimeric protein, there exist three RBD Cleft, RBD Patch, and FCS sites on the spike structure. To avoid complications and confounding variability due to glycan positioning during flexible docking simulations, we specifically then selected the specific RBD Cleft, RBD Patch, and FCS sites for docking based on which sites were not occupied by glycans in the final frame. This selection was particularly

important as Schrödinger IFD does not handle or treat glycan atoms and therefore would not have been able to appropriately include glycan atoms during these studies. For docking into the RBM, we selected the final frame of 50 ns simulations of each variant 1up spike structure. No special care was needed for selecting frames without glycans in the case of the RBM as the RBM, when in the up/open state, is practically unreachable by spike glycans.

To prepare all spike structures for docking with Schrödinger IFD,[25] all protein structures were first titrated to pH 7.4 with PROPKA3.[10] Then all glycan atoms were removed from protein structures as Schrödinger does not properly treat glycans. HSD, HSE, and HSP residue names were converted to the Schrödinger compatible names HID, HIE, and HIP. Structures were then converted to .mae format and prepared according to OPLS4[26] force field using Schrödinger Protein Preparation Wizard[11] according to the following settings: missing hydrogen atoms were added (necessary after glycan deletion), bond orders were assigned, disulfide bonds were added (necessary as Schrödinger cannot also take topology files), hydrogen bonds were optimized with PROPKA3[10] at pH 7.4, and a restrained minimization of hydrogen atoms was performed according to energies and forces described by the OPLS4[22] force field (again necessary after glycan deletion and addition of missing hydrogen atoms). All prepared protein structures will be shared with this work. Heparin and heparan sulfate tetrameric structures (specific chemical structures of which are shown in Figure S4.19 below) were prepared for flexible docking with Schrödinger's LigPrep.[27]

Binding sites for flexible docking were then defined as the center of mass of the following residues, again with care taken to ensure no site was occupied by a glycan in the selected frame from MD simulations:

- WT:

– RBD Cleft: (chain B and resid 346 348 349 351 352 354 355 356 357 450 454 466 467 469 489 472 490) or (chain C and resid 113 114 115 132 165 167)

– RBD Patch: (chain B and resid 337 356 357 359 360 393 394 516 520 521 523 561 562 577 579 580 582) or (chain C and resid 41 170 172 173 226 227 228)

– FCS: (chain B 675 676 677 678 679 680 681 682 683 684 685)

– RBM: (chain A and resid 438 to 508)

● Delta:

– RBD Cleft: (chain A and resid 346 348 349 351 352 354 355 356 357 450 454 466 467 469 489 472 490) or (chain B and resid 113 114 115 132 165 167)

– RBD Patch: (chain B and resid 337 356 357 359 360 393 394 516 520 521 523 561 562 577 579 580 582) or (chain C and resid 41 170 172 173 226 227 228)

– FCS: (chain B 675 676 677 678 679 680 681 682 683 684 685)

– RBM: (chain A and resid 438 to 508)

● Omicron:

– RBD Cleft: (chain A and resid 346 348 349 351 352 354 355 356 357 450 454 466 467 469 489 472 490) or (chain B and resid 113 114 115 132 165 167)

– RBD Patch: (chain B and resid 337 356 357 359 360 393 394 516 520 521 523 561 562 577 579 580 582) or (chain C and resid 41 170 172 173 226 227 228)

– FCS: (chain B 675 676 677 678 679 680 681 682 683 684 685)

– RBM: (chain A and resid 438 to 508)

From each docking procedure, Glide scores were collected and analyzed holistically as well as individual binding modes were inspected to determine interactions of interest within each binding site.

4.5.1.4 Brownian Dynamics Simulations:

Following the preparation and docking of glycoproteins and ligands, all structures were submitted to the PDB2PQR program[28,29] to assign atomic partial charges and radii according to the CHARMM36m forcefield.[15–17] Protonation states for all systems were assigned using PROPKA3[10] at pH of 7.4. Then the "make_apbs_input" and "run_apbs_input" programs in the Browndye2 package[30,31] were used to prepare input files and run APBS 1.5[32–37] to solve the linear Poisson-Boltzmann equation for the creation of electrostatic potential grids for each molecule. Electrostatics calculations, as well as BD simulations, were

performed at a temperature of 298.15K, with a NaCl electrolyte concentration of 10mM, a solvent dielectric of 78, and a solute interior dielectric of 4, and with a grid spacing of 0.5Å.

BD simulations to study the association kinetics of bimolecular reactions require definitions of reaction criteria. Following the docking procedure, key interacting residues for each of the sites on the glycoproteins were identified. For each site on each monomer of each glycoprotein, the center of mass of these residues was computed. Separately, the center of mass for each ligand was also computed. The distance between the glycoprotein site center of mass and the ligand center of mass was used as the reaction coordinate, and if this distance ever fell below a defined threshold of 14Å, a "reaction" was assumed to have occurred.

The Browndye2 package[30,31] as used to prepare and run all BD simulations. Hydrodynamics were enabled. Upon independent investigation, we observed anomalous behavior for these systems when desolvation forces were enabled, most likely due to the high magnitude of molecular charges involved. For this reason, we chose not to enable desolvation forces for these simulations. A total of 24 separate systems were simulated on the TACC Frontera supercomputer. For each system, the BD simulations were spread onto a 56-core node and ran for 24 hours. The total number of BD simulations varied between systems, and anywhere from a few hundred to a hundred thousand separate trajectories were completed per system. Following the simulations, the obtained reaction statistics may be used to estimate k_{on}s for each system. BD simulations were performed on TACC Frontera.

To compute the association rate constants to the b-surface, we use the following equation, which is derived from the Smoluchowski equation:[38,39]

$$k(r) = -\frac{DQ_cQ_s}{\left[1 - exp\left\{\frac{Q_cQ_s}{4\pi\varepsilon_0\varepsilon_r k_B T r}\right\}\right]\varepsilon_0\varepsilon_r k_B T}$$

Where k(r) is the association rate constant to the spherical b-surface of radius r, Q_s is the charge of the substrate, Q_c is the charge of the receptor, D is the radial relative diffusion coefficient of the two molecules, ε_0 is the vacuum permittivity, ε_r is the dielectric constant of the solvent, k_B is the Boltzmann constant, and T is the system temperature.

4.5.1.5 Spike RBD+ACE2 MD Simulations:

To investigate the stability of the ACE2/RBD interface over the course of the variant timeline, RBD+ACE2 systems were constructed for WT, Delta, and Omicron variants. RBDs were extracted from our full spike WT, Delta, and Omicron models and then aligned to a 2.90 Å cryo-EM structure of the WT ACE2/RBD complex (PDB ID 6M17).[4] ACE2 and aligned RBD complex were extracted for each variant and full glycosylation profile of ACE2 and RBD were replicated from Barros et al.[40] PROPKA3[10] was used to ensure all protonation states for ACE2 and the RBDs were still appropriate, and they were. Special attention was paid to ensure the Zn^{2+} atoms from ACE2 were retained in RBD/ACE2 system model building. Additionally, special care was taken to make sure there were no residue clashes along the RBD/ACE2 interface as the Delta and Omicron interfaces were constructed from alignment to the WT RBD structure and not resolved experimentally. All systems were solvated in water boxes of ~130 x 140 x 180 $Å^3$ and ionized with 0.15M NaCl. For each RBD/ACE2 system, we then performed 3 replicas of the following MD simulation protocol with NAMD2.14[13,14] and CHARMM36m all atom force field:[15–17] 10,000 steps of Steepest Descent minimization for all atoms (no restraints nor constraints). Heating of the solvated system from 10K to 310K by increments of 25K with protein and glycan atoms held in light restraint according to a force constant of 1 kcal/mol/Å. With each increase in temperature, 10080 steps (1fs/step) of MD simulation were performed within the NVT ensemble. Once the temperature reached 310K, 0.5 ns of NVT equilibration was performed with restraints maintained. NpT equilibration (310K, 1.01325 bar) for 0.5 ns (1fs/step) with restraint (force constant = 1 kcal/mol/Å) applied to all protein backbone atoms. Pressure was maintained a Langevin barostat. Box cell dimensions were set to flexible during pressure equilibration. GPU accelerated NVT free (no restraints) production (310K, 1.01325 bar) simulations for 25 ns (1fs/step) conducted with NAMD3.0.[14] As system pressure was equilibrated in the prior step, box cell flexibility was turned off in this step (useFlexibleCell = no). GPU accelerated NVT production runs were performed on the Hopper GPU cluster at SDSC TSCC. To prepare for analysis, VMD[12] was used to align trajectories according to protein C_α atomic positions in the first frame, and

water and ion atoms were stripped from trajectories. Trajectories were then ported into MDAnalysis[20,21] as universes where native contacts analysis was performed.

4.5.1.6 Dynamical Electrostatic Potential Map Calculations:

To confirm the presence of large, positively charged regions on the spike surface we used a time-averaged implementation of Adaptive Poisson Boltzmann Solver (APBS)[32–37] to calculate the electrostatic potential at equally spaced grid points along the spike surface over our aligned 50ns classical MD simulation trajectories (described in Section 2.1.1 above). We calculated electrostatic potential maps for the WT spike in closed and open states as well as for Delta and Omicron BA.1 variants full-length spike structures in the closed and open states. All resulting ESP volume (.dx) files for each frame of the 50ns trajectories (1260 frames per simulation) were averaged using the APBS's dxmath functionality. For each structure and each frame, we calculated electrostatic potential maps using the following options, and example input scripts can be found in the downloadable tar.gz file associated with this supporting information:

```
elec name frame
    mg-auto
    dime 321 321 321
    cglen 400 400 400
    fglen 200 200 200
    cgcent mol 1
    fgcent mol 1
    lpbe
    bcfl sdh
    ion charge 1 conc 0.150 radius 1.36375
    ion charge -1 conc 0.150 radius 2.27
    pdie 4.0
    sdie 78.00
    chgm spl2
    srfm smol
```

```
    srad 1.4
    swin 0.3
    sdens 10.0
    temp 298.15
    gamma 0.105
    calcenergy total
    write pot dx frame
end
```

4.5.2 Experimental Materials & Methods

4.5.1 Materials

Heparin (HEP001) was purchased from Galen laboratory supplies. Heparan sulfate from bovine kidney (H7640), Human serum albumin (A3782), and sucrose (S0389), Silver lactate (359750), Hydroquinone (H9003) were purchased from Sigma-Aldrich. Biotin-PEG3-amine (BG-17) was purchased from G-Biosciences. Tween 20 (J20605-AP) was purchased from Thermo Fisher Scientific. Sodium phosphate monobasic (389872500) and Sodium phosphate dibasic (204851000) were purchased from ACROS Organics. Bovine serum albumin (105033) was purchased from MP biomedicals. Gold nanoparticles 10nm (15703-20) were purchased from Ted Pella Inc. Human ACE2, Fc Tag (AC2-H5257) was purchased from Acro Biosystems. Rabbit anti-human IgG (31143) and Horseradish peroxidase conjugated goat anti-human IgG with Horseradish peroxidase (A18811) was purchased from Invitrogen. Horseradish peroxidase conjugated anti-His Tag antibody (652504) was purchased from Biolegend. Nitrocellulose membrane (FF120HP), sample pad (Whatman CF4 dipstick pad), and absorbent pad (Whatman standard 17) were purchased from Cytiva. SARS-CoV-2 Wild type Spike (40589-V08H4), Delta (40589-V08H10) Spike, and Omicron (40589-V08H26) Spike, and HIV gp140 envelope protein (11677-V08H) were purchased from Sino biological. SARS-CoV-2 Alpha (B.1.1.7) Spike (10796-CV-100) and Beta (B.1.351) spike (10786-CV-100) were purchased from R&D systems. Streptavidin modified BLI biosensor tips (18-5019) and anti-human IgG Fc Capture (AHC) BLI biosensor tips (18-5060) were purchased from Sartorius. N-Terminal domain binding

antibody (LT-2000) was purchased from Leinco Technologies. Human saliva pooled from human donors (991-05-P) was purchased from LEE Biosolutions.

4.5.2 Biotin conjugation to heparin

For BLI, ELISA, and LFSA preparation, biotin modified heparin was prepared. Specifically, 2 mg of biotin-PEG3-amine and 2 mg of heparin was dissolved in the 10 mM sodium phosphate buffer pH 7.4. Followed by the addition of the sodium cyanoborohydride (5 mg), and the solution was incubated 24 hr at 60°C. After the incubation, same amount of sodium cyanoborohydride was added and incubated for another 24 hr. Resulting solution was purified by centrifugation with 3k filter to remove the unreacted biotins. Finally, the solution was lyophilized and stored in -20°C until further use.

4.5.3 Immobilization and binding of heparin, ACE2 to variant spike proteins

To compare the binding of heparin or ACE2 to variant spike proteins, enzyme-linked immunoassay (ELISA) was utilized. Firstly, streptavidin (200 nM; 50 µL) was added to the Nunc maxisorp flat bottom 96 well plate and incubated overnight at 4°C. The plates were washed with 200 µl of 1xPBST (0.05% tween-20) three times to remove unbound streptavidin. Then the plate was blocked with 100 µl of 2% BSA for 1 hr at room temperature and washed with 1xPBST. Biotinylated heparin (800 nM; 50 µL) was incubated for 1hr and washed thoroughly to remove unbound heparin. SARS-CoV-2 Spike proteins (WT, Alpha, Beta, Delta, and Omicron; 100 nM; 50 µL) in LFSA buffer were incubated for 1 hr, and LFSA buffer without any spike protein was also incubated for 1 hr as a control. After 1 hr incubation, wells were washed with 200 µl of 1xPBST three times. For the signal generation, anti-his tag-HRP (1 µg/ml; 50 µl) was incubated for 30 min. To remove the unbound HRP, plates were thoroughly washed 5 times with 1xPBST. Finally, 50 µl of TMB substrate (Thermo Fisher Scientific, 34028) was added to each well to develop color. The reaction was stopped by adding 50 µl of stop solution (Thermo Fisher Scientific, N600), and absorbance was measured at 450 nm. In case of ACE2 binding, same procedure was used as heparin binding except that ACE2 functionalized plates

were prepared by directly absorbing ACE2 (10 µg/mL; 50 µL) to Nunc maxisorp flat bottom 96 well plates at 4 °C overnight instead of streptavidin functionalization.

4.5.4 Biolayer interferometry (BLI)

For measuring the binding affinities of heparin or ACE2 to variants of SARS-CoV-2 Spike, biolayer interferometry (BLI) was used. For heparin binding affinity measurement, streptavidin coated BLI tips were functionalized with biotin-heparin (100 µg/mL, 40 µL) for 180 s. Unbound or loosely bound biotin-heparin was washed for 500 s. Heparin functionalized tips were treated with various concentrations of spike protein (10, 25, 50, 100, 200 nM) including control without spike protein for 400 sec and dissociation was measured for 500 s. Binding affinity, Dissociation Constant (K_D), was calculated with steady-state analysis using the HT 11.1 software. ACE2 binding affinity was measured with the same procedure except that anti-human IgG Fc capture (AHC) BLI tips were utilized when ACE2 (1 µg/mL; 40 µL) was loaded. For all measurements, LFSA buffer (10 mM Sodium phosphate, 0.05% tween-20, pH 7.4) was used.

4.5.5 Checking ternary complex formation (Spike-HS-ACE2) using ELISA

To check the synergy of heparin and ACE2 binding in spike protein binding, ELISA was utilized. Firstly, streptavidin (200 nM; 50 µL) was added to the Nunc maxisorp flat bottom 96 well plate and incubated overnight at 4°C. The plates were washed with 1xPBST (0.05% tween-20) three times to remove unbound streptavidin. Then the plate was blocked with 2% BSA (100 µl) for 1hr at room temperature and washed with 1xPBST. Biotinylated heparin (800 nM; 50 µL) was incubated for 1hr and washed thoroughly to remove unbound heparin. SARS-CoV-2 Spike proteins (100 nM; 50 µL) in LFSA buffer were incubated for 1 hr including the control without spike proteins. After incubation wells were washed three times with 1xPBST to remove unbound spike proteins. ACE2 (1 ug/mL; 50 µL) in LFSA buffer was added and incubated for 1 hr. After incubation, each well was washed three times with 1xPBST. Finally, anti-IgG-HRP (1 µg/mL; 50 µL) was incubated for 30 min, and plates were thoroughly washed 5 times with 1xPBST. Finally, 50 µL of TMB

substrate (Thermo Fisher Scientific, 34028) was added to each well to develop color. The reaction was stopped by adding 50 µL of stop solution (Thermo Fisher Scientific, N600), and absorbance was measured at 450 nm.

4.5.6 Checking ternary complex formation (Spike-HS-ACE2) using Mass Photometer

To measure the synergy effect of heparan sulfate and ACE2 in spike protein binding, mass photometry (Refeyn One mass photometry) was utilized. Microscope coverslips (CG15KH1, 24 x 50 mm, Thickness 170 ± 5 µm, Thorlabs) were cleaned by sonication in 2-Propanol (A516-4, Fisher Scientific) followed by DW (10 min for each step). Silicone gasket (3 mm diameter x 1mm depth, Grace Bio-labs) was applied to the cleaned coverslip to form a chamber. After preparing the gasket, 10 µl of LFSA buffer (10 mM sodium phosphate, pH 7.4) without tween-20 was first injected to the chamber to find the focus using the autofocus function called droplet-dilution using Acquire MP Sofware provided by the instrument. Samples (Spike, Spike + ACE2, Spike + HS + ACE2) were prepared by incubating Spike (40 µg/ml; 5 µl), ACE2 (16 µg/ml; 2.5 µl), and HS (2 µg/ml; 2.5 µl) for 30 min at room temperature prior to mass photometer. For each acquisition, 10 µl of mixture of protein solution was introduced to the chamber (total volume: 20 µl) and movies of 60 s duration were recorded. Each sample was measured in new chambers (each chamber was used once). All mass photometer data were analyzed with DiscoverMP (2022 ver. R1). Firstly, using a mass calibration curve, ratiometric data was converted to the mass. Then, to check the fraction of the ternary complexes, count numbers for binding event were obtained for each mass range (denoted as group A, B, and C). Obtained count numbers were used to calculate the relative fraction of each and compare the fraction of the ternary complex with or without HS.

4.5.7 Preparation of signaling probes

To prepare the signaling probes, ACE2 and NTD Ab were conjugated to the gold nanoparticles (AuNP). Firstly, 10 nm AuNP (1 ml) was equilibrated with borate buffer (0.1 M pH 8.0; 100 µl), then ACE2 (0.6 mg/ml; 8.3 µl) or NTD Ab (1 mg/ml; 5 µl) was added to prepare the signaling probes for GlycoGrip2.0 and GlycoGrip1.0, respectively. After incubating the resulting solution for 1 hr at room temperature with continuous rotation, 100 µl of 1% BSA was added and incubated for an additional 30 min. To remove the unbound proteins, ACE2 or NTD Ab conjugated AuNPs were centrifugated at 22000g, 4°C for 45 min and the supernatant was removed. AuNP pellets were resuspended in the 1% BSA (1 ml) solution. This procedure was repeated 3 times. Finally, ACE2 or NTD Ab conjugated AuNPs were resuspended in the LFSA buffer and stored in 4°C until further use.

4.5.8 Preparation of GlycoGrip LFSA

For the immobilization of the heparin to the nitrocellulose membrane, heparin was conjugated to the streptavidin as previously reported. Briefly, biotin-heparin (2 mg/ml, 50 µl) was incubated with streptavidin (1 mg/ml, 100 µl) for 1 hr. The mixture solutions were purified by centrifugation with a 30k amicon filter to remove excess biotin-heparin. The concentration of heparin was measured using Azure A assay. Heparin conjugated with streptavidin (300 µM) and rabbit antihuman IgG (1 mg/mL) were dispensed on the nitrocellulose membrane (FF120HP). Dispensed nitrocellulose membrane was dried at 65 °C for 3 min. After drying, the nitrocellulose membrane was blocked with a blocking buffer (1% BSA, 0.05% Tween 20 in 10 mM PB, pH 7.4). Finally, the sample pad (Whatman CF4 dipstick pad) and the absorbent pad (Whatman standard 17) were assembled onto the nitrocellulose membrane. Assembled strips were stored at room temperature with desiccant before use.

4.5.9 Comparison of the variant detection in GlycoGrip1.0 and GlycoGrip2.0

For the comparison of the SARS-CoV-2 variant detection in GlycoGrip1.0 and GlycoGrip2.0, 25 μL of each of the Spike proteins (25 μg/ml) in LFSA buffer were incubated with 25 μl of signaling probes (i.e., 20 nM of NTD Ab-AuNP for GlycoGrip1.0, and 20 nM of ACE2-AuNP for GlycoGrip2.0) for 30 min at room temperature. Then, the dipstick method was used to compare the detection. Briefly, the resulting solutions were dispensed in the flat bottom 96 well plate and lateral flow strips were dipped for 10 min. After 10 min, the image was taken by smartphone camera and signals were quantitatively analyzed by ImageJ software.

4.5.10 Evaluation of Selectivity and sensitivity of GlycoGrip2.0

To evaluate the selectivity of the GlycoGrip2.0, 25 μL of each samples including SARS-CoV-2 Omicron spike (25 μg/mL), CoV1 S1 (25 μg/mL), MERS S1 (25 μg/mL), HIV gp140 (group M, CRF07_BC) envelope protein (25 μg/mL), Human serum albumin (50 mg/mL), and Human saliva (25x diluted from the stock solution) was incubated with 25 μL of signaling probe (ACE2-AuNP; 20 nM) for 30 min. Then, signals were generated with the dipstick method and analyzed with ImageJ.

For the sensitivity testing, different concentrations of SARS-CoV-2 Omicron spike (0, 0.4, 0.8, 1.6, 3.13, 6.25, 12.5, 25, 50 μg/ml; 25 μl) were incubated with ACE2-AuNP (20 nM; 25 μl) for 30 min. After 30 min, the same dipstick method was utilized. Signal intensity of the test line was analyzed with ImageJ software and the limit of the detection (LOD) was estimated by blank + 3 standard deviations. At least 3 independent tests were performed to calculate the LOD.

4.5.11 Evaluation of sensitivity of GlycoGrip2.0 in human saliva

For human saliva testing, various concentrations of SARS-CoV-2 Omicron spike (0. 0.8, 1.6, 6.25, 25, 50, 100 ug/ml; 12.5 ul) were introduced into human saliva (25μL; 25x diluted from the stock solution). Prepared spike in human saliva solution were incubated with ACE2-AuNP (20 nM; 25 μL) for 30 min. After 30 min, dipstick method was utilized, and signal was enhanced with silver enhancement methods as described in 2.2.11.

4.5.12 Signal enhancement analysis

To enhance the sensitivity of the GlycoGrip2.0, silver enhancement methods[37] was adopted. Specifically, different concentrations of SARS-CoV-2 Omicron spike (0, 0.1, 0.2, 0.4, 0.8, 1.6, 3.13, 6.25, 12.5, 25, 50 ug/mL; 25 µL) were incubated with ACE2-AuNP (20 nM; 25 µL) for 30 min. After 30 min, each solution was dispensed to 96 well plates and LFSA strips were dipped for 10 min. After 10 min, Strips were sequentially washed with LFSA buffer (100 µL) and DW (100 µL) for 3 min. Finally, Strips were dipped into the mixture of 0.3% silver lactate and 3% hydroquinone (100 µL each) for 3 min and the enhancement reaction was stopped by washing with DW. Signal intensity of the test line was analyzed with ImageJ software and the limit of the detection (LOD) was estimated by blank + 3 standard deviations. At least 3 independent tests were performed to calculate the LOD.

4.6 **Supplemental Information**

4.6.1 Supplemental Results and Discussion:

4.6.1.1 Summary of AutoDock Vina results:

28,800 GAG binding modes were collected from ensemble-based docking studies as laid out in Scheme S1 below. All binding modes for WT, Delta, and Omicron spike structures in closed and 1up conformations can be seen in Figures S4.4, S4.5, and S4.6 respectively. The centers of mass of each of these resultant binding modes were collected passed through kmeans clustering to identify GAG binding “hotspots” on the spike surface. Figure S4.7 shows the sum of squared Euclidean distances (SSE) for the generated 1-100 clusters; kneed elbow locator algorithm was used to determine 19 to be the optimal number of clusters of all 28,800 binding modes. These 19 clusters were then given names A-S and their locations on the spike structure were identified, Figure S4.8. To determine which of these binding sites were “surface exposed”, I.e. sites accessible to long chain GAGs within the glycocalyx, we calculated the Accessible Surface Area (ASA) for each of these sites using the Shrake-Rupley algorithm[20] through VMD tools,[12] Figure S4.9, with a probe

radius of r=7.2 Å from WT trajectories shared for closed and 1up state spike proteins shared by Casalino et al.[8] Each site was defined as any residue (from all structures considered, i.e., all conformations of WT, Delta, and Omicron structures) within 10 Å of the centroid of said site as defined by kmeans clustering. A probe radius of 7.2 Å was chosen to be consistent with an approaching GAG fragment. All residue numbers per site can be found in Table S3. From ASA results we see that in the closed spike conformation, sites K, M, N, and R are largely inaccessible to GAG fragments within the glycocalyx. However, in the 1up spike conformation site M becomes marginally exposed due to the lifting up of the neighboring RBD.

To identify any sites of particular importance on a per spike variant basis we then analyzed the distributions of predicted binding energies within each site as given by AutoDock Vina,[18,19] Figure S4.10. As can be seen from these results, both heparin dimers and heparin tetramers bind with relatively similar predicted energies to all sites on a per variant basis, save for a few instances. We also compared the distribution of predicted binding energies at each site as a function of spike conformation, i.e., closed versus 1up states, Figure S4.11. Again, very little difference could be seen between closed and 1up spike structures to suggest heparin dimers or tetramers favor binding to either state. Given the similarity in predicted energies, and the broad distribution of predicted energies at each site in all resultant binding modes, we cannot predict, at this time, any significant differences in binding affinity at each site resulting from changes in spike sequence. As such we predict binding affinity differences as observed with BLI between heparin and variant spike structures are likely due to effects only captured by use of long-chain heparins or due to kinetic effects such as rate of encounter complex formation.

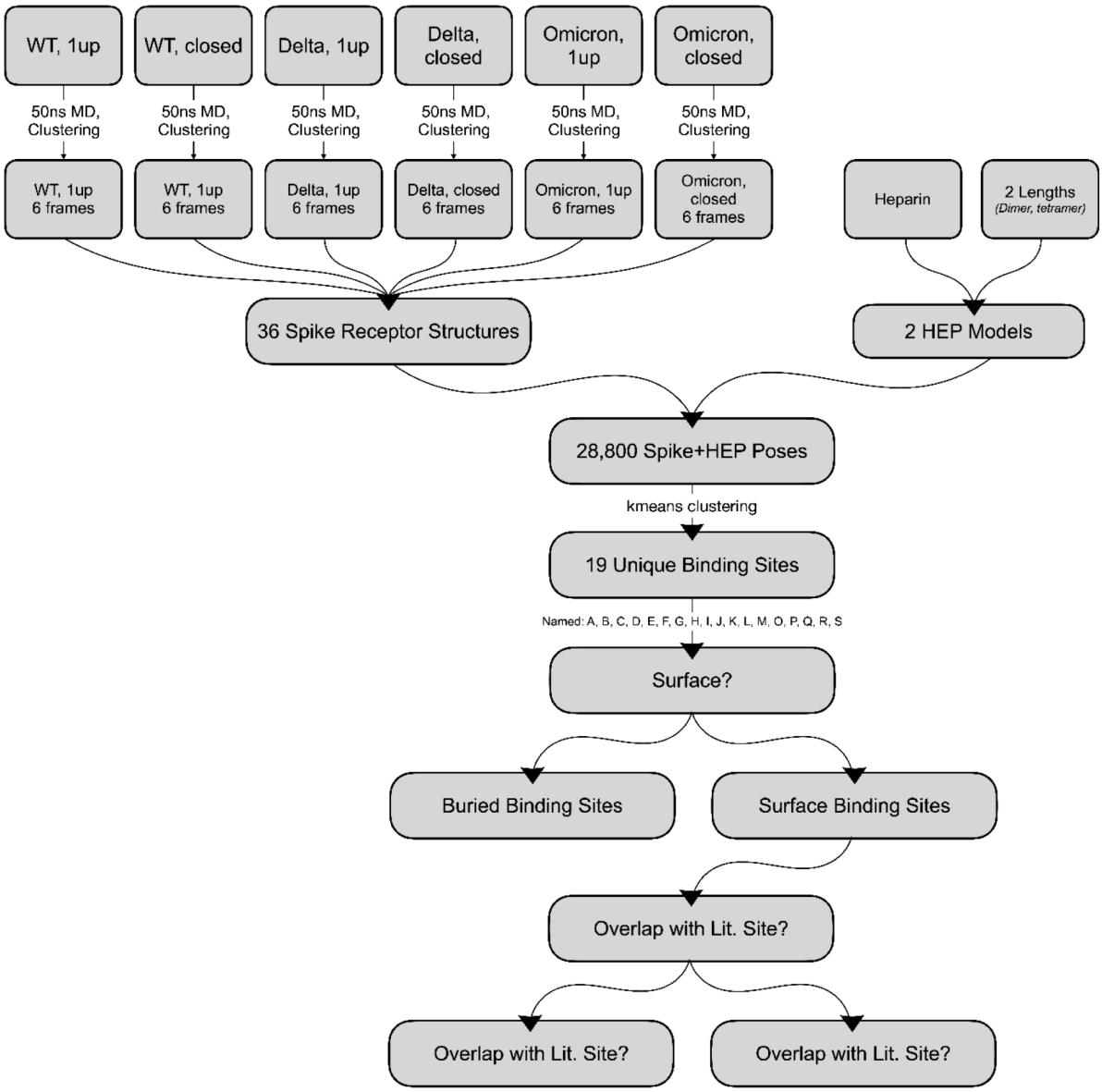
4.6.1.2 Summary of Schrödinger IFD Results:

While we have already incorporated a degree of protein flexibility in our docking studies by conducting extensive ensemble-based sampling (i.e., 3 spike proteins (WT, Delta, Omicron) x 2 spike states (closed, 1up) x 6 conformations each = 36 total spike receptor structures) we were interested in identifying any potential induced fit effects and how such effects may adapt/change over the variant timeline. As such we conducted

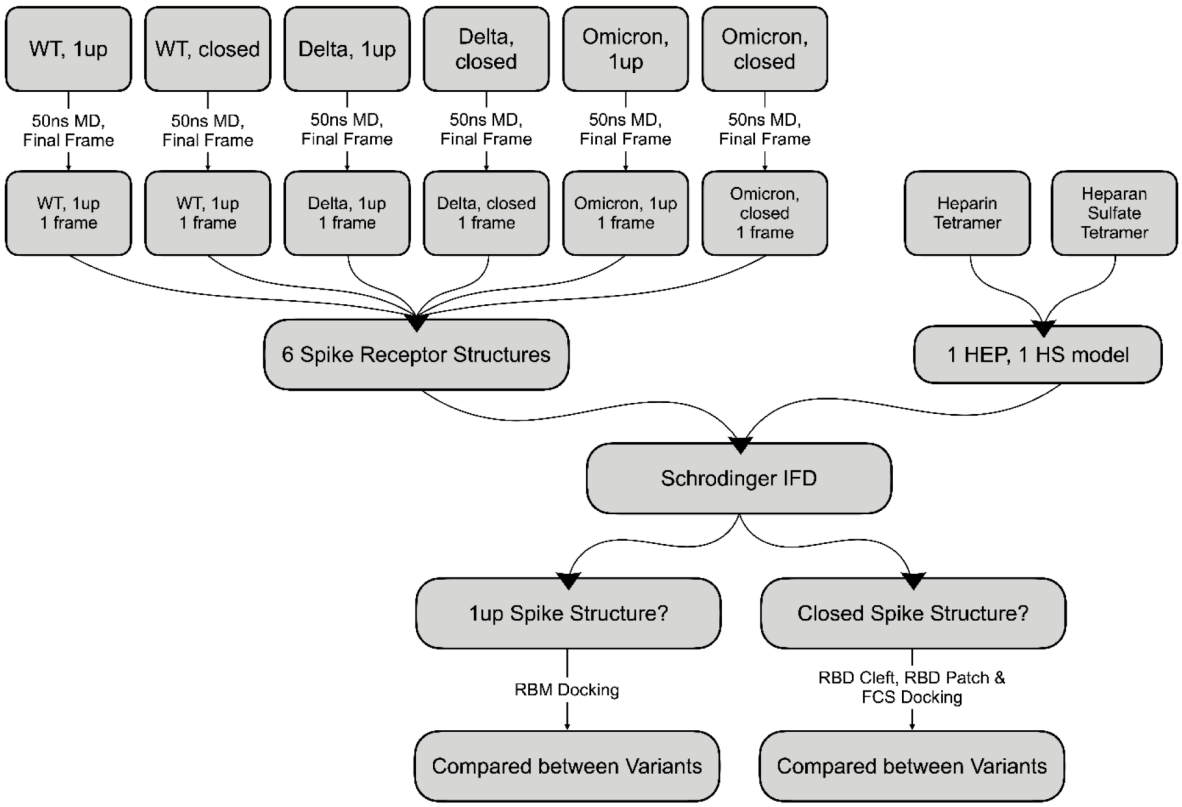
flexible ligand-flexible receptor docking studies with Schrödinger IFD on targeted sites within the spike protein using heparin and heparan sulfate tetrameric models. From these results, Figure S4.12, we again see predicted binding energies (Glide Scores) are broad at each site, and there is virtually no difference in these distributions across the variant timeline despite mutations within each site. We predict that this is due to the innate flexibility of GAG molecules like heparin and heparan sulfate. These ligands can adapt to mutations within these binding sites and thus still bind at each site. Thus, we predict differences in binding affinity between heparin/heparan sulfate and spike proteins as seen from BLI and ELISA results are likely due to effects that can only be seen at the long-chain binding mode scale or due to kinetic effects not captured in docking studies.

4.6.2 Supplemental Schemes

Scheme S4.1: Schematic outlining variant spike simulation, clustering, and AutoDock Vina procedures
(as discussed in Computational Methods Sections 2.1.1 to 2.1.2)



Scheme S4.2: Schrödinger IFD Methodology at pH 7.4 outlining variant spike simulation, clustering, and Schrödinger IFD procedures
(as discussed in Computational Methods 2.1.1 and 2.1.3.)



4.6.3 Supplemental Tables

Table S4.1: Complete list of all mutations per variant considered for modeling and charge calculations in this work

Glycan contribution calculated according to Watanabe et al.[6] and with glycans chosen consistent to Casalino et al.[8] with 14 sialic acid residues included (total glycan charge of –14).

Variant	Mutations from WT/2019 (charge change relative to WT due to mutation)	ΔT_Q	T_Q
WT			-11
Alpha	ΔH69-V70 (0), ΔY144 (0), A570D (-1), D614G (+1), P681H (0), T716I (0), S982A (0), D1118H (+1)	(+1)*3 = +3	-8
Beta	D80A (+1), D215G (+1), ΔL241-L242-A243 (0), K417N (-1), E484K (+2), N501Y (0), D614G (+1), A701V (0)	(+4)*3 = +12	+1
Delta	T19R (+1), G142D (-1), ΔE156-F157 (+1), R158G (-1), L452R (+1), T478K (+1), D614G (+1), P681R (+1), D950N (+1), ΔN17-Glycan (0)	(+5)*3 = +15	+4
Omicron (BA.1)	A67V (0), ΔH69-V70 (0), T95I (0), G142D (-1), ΔV143-Y144-Y145 (0), ΔN211 (0), L212I, ins214EPE (-2), G338D (-1), S371L (0), S373P (0), S375F (0), K417N (-1), N440K (+1), G446S (0), S466N (0), T478K (+1), E484A (+1), Q493R (+1), G498R (+1), N501Y (0), Y505H (0), T547K (+1), D614G (+1), H655Y (0), N679K (+1), P681H (0), N764K (+1), D796Y (+1), N856K (+1), Q954H (0), N969K (+1), L981F (0)	(+7)*3 = +21	+10

Table S4.2: Complete list of all titratable residues and their selected protonation states each spike/ACE2 structure to pH = 7.4, as calculated by PROPKA

Full pK_a calculation data can be found in the shared files associated with this supporting information.

pH	Prot	Cnf	P.S.	Residue IDs
7.4	WT	Clo.	ASP	40 53 80 88 111 138 178 198 215 228 253 287 290 294 364 389 398 405 420 427 428 442 467 568 571 574 578 586 614 627 663 767 745 775 796 808 820 830 839 843 848 867 936 950 979 985 994 1041 1084 1118 1127 1139
			GLU	96 132 154 156 169 180 191 224 281 598 309 324 340 406 465 471 484 516 554 583 619 654 661 702 725 748 773 780 819 868 918 988 990 1017 1031 1072 1092 1111
			GLUP	none
			HSD	146 207 245 519 1058
			HSE	49 66 69 625 655 1048 1064 1083 1101
			HSP	none
7.4	WT	1up	ASP	40 53 80 88 111 138 178 198 215 228 253 287 290 294 364 389 398 405 420 427 428 442 467 568 571 574 578 586 614 627 663 767 745 775 796 808 820 830 839 843 848 867 936 950 979 985 994 1041 1084 1118 1127 1139
			ASPP	none
			GLU	96 132 154 156 169 180 191 224 281 598 309 324 340 406 465 471 484 516 554 583 619 654 661 702 725 748 773 780 819 868 918 988 990 1017 1031 1072 1092 1111
			GLUP	none
			HSD	146 207 245 519 1058
			HSE	49 66 69 625 655 1048 1064 1083 1101
			HSP	none
7.4	Del	Clo.	ASP	40 53 80 88 111 138 142 178 198 215 228 253 287 290 294 364 389 398 405 420 427 428 442 467 568 571 574 578 586 627 663 737 745 775 796 808 820 830 839 843 848 867 936 979 985 994 1041 1084 1118 1127 1139

Table S4.2: Complete list of all titratable residues and their selected protonation states each spike/ACE2 structure to pH = 7.4, as calculated by PROPKA (continued)

			ASPP	none
			GLU	96 132 154 169 180 191 224 281 298 309 324 340 406 465 471 484 516 554 583 619 654 661 702 725 748 773 780 819 868 918 988 990 1017 1031 1072 1092 1111
			GLUP	none
			HSD	519 625 655 1058 1083 1088
			HSE	49 66 69 146 207 245 1048 1064 1101
			HSP	None
7.4	Del	1up	ASP	40 53 80 88 111 138 142 178 198 215 228 253 287 290 294 364 389 398 405 420 427 428 442 467 568 571 574 578 586 627 663 737 745 775 796 808 820 830 839 843 848 867 936 979 985 994 1041 1084 1118 1127 1139
			ASPP	none
			GLU	96 132 154 169 180 191 224 281 298 309 324 340 406 465 471 484 516 554 583 619 654 661 702 725 748 773 780 819 868 918 988 990 1017 1031 1072 1092 1111
			GLUP	none
			HSD	519 625 655 1058 1083 1088
			HSE	49 66 69 146 207 245 1048 1064 1101
			HSP	None
7.4	Omi	Clo.	ASP	40 53 80 88 111 138 142 178 198 215 228 253 287 290 294 339 364 389 398 405 420 427 428 442 467 568 571 574 578 586 627 663 737 745 775 808 820 830 839 843 848 867 936 950 979 985 994 1041 1084 1118 1127 1139
			ASPP	none

Table S4.2: Complete list of all titratable residues and their selected protonation states each spike/ACE2 structure to pH = 7.4, as calculated by PROPKA (continued)

			GLU	96 132 154 156 169 180 191 2141 (inserted E) 2143 (inserted E) 224 281 298 309 324 340 406 465 471 516 554 583 619 654 661 702 725 748 773 780 819 868 918 988 990 1017 1031 1072 1092 1111
			GLUP	None
			HSD	146 207 245 519 681 1058 1088
			HSE	49 66 505 625 954 1048 1064 1083 1101
			HSP	none
7.4	Omi	1up	ASP	40 53 80 88 111 138 142 178 198 215 228 253 287 290 294 339 364 389 398 405 420 427 428 442 467 568 571 574 578 586 627 663 737 745 775 808 820 830 839 843 848 867 936 950 979 985 994 1041 1084 1118 1127 1139
			ASPP	none
			GLU	96 132 154 156 169 180 191 2141 (inserted E) 2143 (inserted E) 224 281 298 309 324 340 406 465 471 516 554 583 619 654 661 702 725 748 773 780 819 868 918 988 990 1017 1031 1072 1092 1111
			GLUP	None
			HSD	146 207 245 519 681 1058 1088
			HSE	49 66 505 625 954 1048 1064 1083 1101
			HSP	none
7.4	ACE2	--	ASP	30 38 67 111 136 157 198 201 206 213 216 225 269 292 295 299 303 335 350 355 367 368 382 427 431 471 494 499 509 543 597 609 615 630 637 669 693 713 719
			ASPP	none
			GLU	22 23 35 37 56 57 75 87 110 140 145 150 160 166 171 181 182 189 197 208 224 227 231 232 238 310 312 329 375 398 402 406 430 433 435 457 467 479 483 489 495 527 536 549 564 571 589 634 639 667 668 699 701 723

Table S4.2: Complete list of all titratable residues and their selected protonation states each spike/ACE2 structure to pH = 7.4, as calculated by PROPKA (continued)

			GLUP	none
			HSD	34 195 239 373 374 378 417 493
			HSE	228 241 265 345 401 505 535 540
			HSP	none

Prot. = protein. Cnf = conformational state, relevant to spike protein structures only. P.S. = protonation state. Clo. = spike in closed conformational state. 1up = spike in 1up/open conformational state. WT = Wild Type. Del. = Delta. Omi. = Omicron. ACE2 = angiotensin converting enzyme 2. ASP = deprotonated/negatively charged aspartate. ASPP = protonated/neutral aspartate. GLU = deprotonated/negatively charged glutamate. GLUP = protonated/neutral glutamate. HSD = singly protonated/neutral histidine with protonation on the N_δ atom. HSE = singly protonated/neutral histidine with protonation on the N_ε atom. HSP = double protonated/positively charged histidine with protonation on N_δ and Nε atoms.

Table S4.3: Complete list of protein residues per GAG binding hotspot as predicted by ensemble-based docking with AutoDock and Kmeans clustering

Site	Protein Residues: Residue Name, Residue Number, Chain ID Glycans: Glycan Protein Residue Name, Protein Residue Number, (Protein Chain ID)
A	Arg403A, Asp405A, Glu406A, Asn417A, Lys417A, Tyr421A, Ser443A, Lys444A, Val445A, Gly446A, Asn448A, Tyr449A, Tyr451A, Leu452A, Tyr453A, Leu455A, Phe456A, Lys458A, Ser459A, Asn460A, Lys462A, Ser477A, Phe490A, Leu492A, Gln493A, Arg493A, Ser494A, Tyr495A, Ser496A, Phe497A, Gly496A, Gln498A, Arg498A, Pro499A, Thr500A, Asn501A, Tyr501A, Gly502A, Val503A, Gly504A, Tyr505A, His505A, Gln506A, Pro507A, Phe342B, Asn343B, Tyr369B, Asn370B, Leu371B, Ser371B, Ala372B, Ser373B, Pro373B, Ser375B, Phe375B, Phe374B, Trp436B, Asn437B, Ser438B, Asn439B, Asn440B, Lys440B, Leu441B, Gly447B, Glycan N343(B)
B	Ser49B, Arg346B Phe347B, Ala348B, Val350B, Tyr351B, Ala352B, Trp353B, Leu371C, Ala372C, Phe374C, Phe400B, Val401B, Ile402B, Arg403B, Glu406B, Gly416B, Lys417B, Asn417B, Ile418B, Ala419B, Tyr421B, Asn422B, Tyr423B, Leu441B, Asp442B, Ser443B, Lys444B, Gly447B, Asn448B, Tyr449B, Asn450B, Tyr451B, Leu452B, Arg452B, Tyr453B, Arg454B, Leu455B, Phe456B, Arg457B, Leu461B, Thr470B, Glu471B, Tyr473B, Tyr489B, Phe490B, Pro491B, Leu492B, Gln493B, Arg493B, Ser494B, Tyr495B, Gly496B, Ser496B, Phe497B, Gln498B, Asn501B, Tyr501B, Tyr505B, His505B, Pro507B, Glycan N165(C), Glycan N343(C)
C	Asn370A, Ser371A, Ser373A, Phe377A, Lys378A, Cys379A, Tyr380A, Gly381A, Val382A, Ser383A, Pro384A, Thr385A, Lys386A, Ala411A, Pro412A, Gln414A, Asp427A, Asp428A, Phe429A, Thr430A, Gly431A, Leu517A, Tyr51C, Ser349C, Val350C, Ala352C, Ile402C, Arg403C, Glu406C, Gly416C, Asn417C, Lys417C, Ile418C, Ala419C, Asp420C, Tyr421C, Asn422C, Tyr423C, Tyr449C, Asn450C, Tyr451C, Leu452C, Arg452C, Tyr453C, Arg454C, Leu455C, Phe456C, Lys458C, Ser459C, Leu461C, Asp467C, Ile468C, Arg469C, Thr470C, Glu471C, Ile472C, Tyr473C, Gln474C, Ala475C, Glu484C, Cys488C, Tyr489C, Phe490C, Pro491C, Leu492C, Gln493C, Arg493C, Ser494C, Tyr495C, Gly496C, Ser496C, Phe497C, Tyr501C, Glycan N165(A), Glycan N343(A)
D	Ser13C, Cys15C, Val16C, Leu18C, Ala67C, Val67C, Ile68C, His69C, Ser71C, Phe79C, Arg78C, Asn81C, Glu96C, Lys97C, Ser98C, Asn99C, Ile100C, Ile101C, Arg102C, Gly103C, Leu118C, Val120C, Asn121C, Asn122C, Pro139C, Ala123C, Thr124C, Asn125C, Val127C, Asp138C, Pro139C, Phe140C, Leu141C, Asp142C, Gly142C, Val143C, Tyr144C, Tyr145C, His146C, Lys147C, Asn148C, Asn149C, Lys150C, Ser151C, Trp152C, Met153C, Glu154C, Gly156C, Ser155C, Glu156C, Phe157C, Arg158C, Val159C, Phe175C, Leu176C, Met177C, Asp178C, Leu179C, Glu180C, Gly181C, Lys182C, Gln183C, Arg190C, Thr240C, Leu241C, Leu242C, Ala243C, Leu244C, His245C, His245C, Arg246C, Ser247C, Tyr248C, Leu249C, Thr250C, Ser254C, Gly257C, Trp258C, Thr259C, Ala260C, Gly261C, Ala262C, Ala263C, Glycan N122(C), Glycan N149(C)
E	Asp420A, Tyr421A, Phe456A, Arg457A, Lys458A, Ser459A, Asn460A, Thr385B, Lys386B, Asn388B, Asp389B, Pro527B, Lys528B, Thr415A, Gly416A, Tyr473A, Asp364B, Ser366B, Gly526B, Lys529B, Tyr369B, Asn370B, Val367B, Leu455A, Leu461A, Leu387B, Tyr453A, Arg454A, Pro384B, Lys417A, Ile418A, Tyr489A, Asp985C, Tyr365B, Lys462A, Val327B, Lys424A, Pro463A, Leu371B, Glycan N234(B), Glycan N343(B), Glycan T323(B)

Table S4.3: Complete list of protein residues per GAG binding hotspot as predicted by ensemble-based docking with AutoDock and Kmeans clustering (continued)

F	Pro330B, Ile332B, Arg357B, Ile358B, Ser359B, Asn360B, Cys361B, Phe392B, Thr393B, Asn394B, Val395B, Tyr396B, Glu516B, Leu518B, His519B, Ala520B, Pro521B, Ala522B, Thr523B, Val524B, Asn544B, Leu560B, Pro561B, Phe562B, Gln563B, Gln564B, Arg577B, Leu582B, Lys41C, Ile128C, Phe168C, Glu169C, Tyr170C, Val171C, Ser172C, Gln173C, Pro174C, Phe175C, Tyr200C, Lys202C, Pro225C, Leu226C, Val227C, Asp228C, Leu229C, Pro230C, Glycan N331(B), Glycan N122(C)
G	Ile119A, Thr124A, Asn125A, Val126A, Val127A, Ile128A, Phe168A, Glu169A, Tyr170A, Val171A, Ser172A, Gln173A, Pro174A, Phe175A, Leu179A, Ile203A, Leu226A, Val227A, Asp228A, Leu229A, Pro230A, Arg357C, Ser359C, Asn360C, Asn394C, Tyr396C, Thr523C, Glycan N122(A), Glycan N149(A), Glycan N331(C),
H	Arg357A, Ser359A, Asn360A, Thr393A, Asn394A, Ala520A, Pro521A, Ala522A, Thr523A, Pro561A, Phe562A, Lys41B, Leu117B, Ile119B, Val120B, Asn121B, Thr124B, Asn125B, Val126B, Val127B, Ile128B, Lys129B, Val130B, Phe168B, Glu169B, Tyr170B, Val171B, Ser172B, Gln173B, Pro174B, Phe175B, Leu176B, Met177B, Leu179B, Phe192B, Tyr200B, Phe201B, Lys202B, Ile203B, Tyr204B, Ser205B, Glu224B, Pro225B, Leu226B, Val227B, Asp228B, Leu229B, Pro230B, Ile231B, Gly232B, Glycan N331(A), Glycan N122(B), Glycan N282(B)
I	Ser325A, Val327A, Val382A, Ser383A, Thr385A, Lys386A, Leu387A, Asn388A, Asp389A, Leu390A, Leu518A, Lys528A, Asn540A, Phe541A, Asn542A, Gly545A, Leu546A, Thr547A, Gly548A, Glu748B, Asn978B, Ile980B, Leu981B, Ser982B, Arg983B, Leu984B, Asp985B, Lys986B, Ala989B, Thr747B, Asp979B, Ser746B, Phe329A, Leu977B, Phe543A, Asn544A, Thr549A, Val976B, Ile326A, Val987B, Glu988B, Ile993B, Pro322A, Glu324A, Asp745B, Lys417C, Gln321A, Thr323A, Val539A, Lys547A, Phe981B, Arg328A, Gly550A, Glucan T323(A), Glycan S325(A), Glycan N234(A)
J	Gln52B, Thr274B, Gln271B, Arg273B, Asp290B, Cys291B, Ala292B, Leu293B, Glu298B, Ser316B, Asn317B, Phe318B, Arg319B, Val320B, Gln321B, Pro322B, Thr323B, Glu324B, Lys537B, Cys538B, Asn540B, Thr549B, Val551B, Cys590B, Ser591B, Phe592B, Gly593B, Val595B, Tyr612B, Glu619B, Val620B, Pro621B Val622B, Ala623B, Ile624B, His625B, Ala626B, Asp627B, Gln628B, Leu629B, Thr630B, Pro631B, Thr632B, Trp633B, Arg634B, Val635B, Met740C, Asp745C, Thr747C, Glycan N234(B), Glycan T323(B)
K	Asp737A, Cys738A, Thr739A, Met740A, Ile742A, Cys743A, Gly744A, Asp745A, Ser746A, Thr747C, Glu748A, Cys749A, Ser750A, Asn751A, Leu752A, Leu753A, Leu754A, Gln755A, Tyr756A, Phe759A, Gly757A, Ser758A, Cys760A, Thr761A, Lys764A, Ile993A, Ile997A, Leu1001A, Ser50C, Gln52C, Asp53C, Pro272C, Thr274C, Glu298C, Lys304C, Ser316C, Asn317C, Phe318C, Arg319C, Thr302C, Phe592C, Thr630C, Glycan N234(C)

Table S4.3: Complete list of protein residues per GAG binding hotspot as predicted by ensemble-based docking with AutoDock and Kmeans clustering (continued)

L	Leu24B, Pro25B, Pro26B, Ala27B, Tyr28B, Thr29B, Asn30B, Asn61B, Val62B, Thr63B, Trp64B, Phe65B, His66B, Ile68B, Val70B, Arg78B, Val213B, Arg214B, Asp215B, Tyr266B, Leu212B, ins-Glu2141B, ins-Pro2142B, ins-Glu2143B, Glycan N61(B), Glycan N74(B), Glycan N603(B)
M	Val47B, Leu48B, His49B, Ser50B, Thr51B, Gln52B, Thr274B, Leu276B, Val289B, Cys291B, Pro295B, Ser297B, Glu298B, Thr299B, Lys300B, Cys301B, Thr302B, Leu303B, Lys304B, Ser305B, Phe306B, Thr307B, Val308B, Tyr313B, Gln314B, Thr315B, Ser316B, Asn317B, Phe318B, Ile569A, Trp633B, Arg634B, Cys738C, Thr739C, Leu753C, Leu754C, Tyr756C, Gly757C, Ser758C, Phe759C, Cys760C, Thr761C, Gln762C, Leu763C, Lys764C, Asn764C, Arg765C, Gln957B, Asn960B, Thr961B, Lys964B, Gln965B, Ser968B, Ser967B,
N	Leu48A, His49A, Ser50A, Thr51A, Thr274A, Phe275A, Leu276A, Leu277A, Val289A, Cys291A, Ser297A, Glu298A, Thr299A, Lys300A, Cys301A, Thr302A, Leu303A, Lys304A, Ser305A, Phe306A, Thr315A, Thr961A, Lys964A, Leu754B, Gly757B, Ser758B, Cys760B, Thr761B, Asn764B, Gln52A, Leu296A, Ser316A, Arg765B, Asp290A, Ala292A, Phe759B, Gln762B, Thr739B, Leu753B, Val308A, Asp294A, Pro295A, Gln314A, Thr307A, Gln957A, Gln755B, Tyr756B, Tyr313A, Val597A, Gln965A, Cys738B, Lys764B, Asn317A, Phe318A, Ile569C Leu849A, Glycan N234(A), Glycan T323(A)
O	Ile834A, Ile312C, Val595C, Ser596C, Ile598C, Val608C, Ala609C, Val610C, Leu611C, Tyr612C, Gln613C, Gly614C, Val615C, Asn616C, Cys617C, Thr618C, Glu619C, Val620C, Pro621C, Val622C, Ala623C, Ile624C, Leu629C, Thr630C, Pro631C, Thr632C, Trp633C, Arg634C, Val635C, Tyr636C, Ser637C, Thr638C, Gly639C, Ser640C, Asn641C, Val642C, Phe643C, Gln644C, Thr645C, Arg646C, Ala647C, Gly648C, Cys649C, Leu650C, Ile651C, Gly652C, Ala653, Glu654C, His655C, Tyr655C, Val656C, Ile666C, Ile670C, Cys671C, Ala672C, Gln675C, Arg682C, Arg683C, Ser686C, Ala688C, Ser689C, Gln690C, Ser691C, Ile692C, Ile693C, Ala694C, Tyr695C, Glycan N616(C), Glycan N657(C)
P	Asn616B, Gln644B, Thr645B, Arg646B, Ala647B, Gly648B, Ala668B, Gly669B, Ile670B, Pro812C, Ser813C, Lys814C, Arg815C, Phe823C, Leu828C, Ala829C, Asp830C, Ala831C, Gly832C, Phe833C, Ile834C, Lys835C, Gln836C, Tyr837C, Pro862C, Pro863C, Leu865C, Thr866C, Asp867C, Glu868C, Met869C, His1058C, Glycan N616(B), Glycan N657(B), Glycan N282(C)
Q	Ala609A, Val610A, Leu611A, Tyr612A, Val615A, Asn616A, Cys617A, Thr618A, Glu619A, Val620A, Pro621A, Leu629A, Thr632A, Gly639A, Ser640A, Asn641A, Val642A, Phe643A, Gln644A, Thr645A, Gly648A, Cys649A, Leu650A, Ile651A, Gly652A, Ala653A, Glu654A, His655A, Val656A, Ile670A, Arg682A, Ala694A, Thr630A, Arg646A, Gln628A, Ile666A, His655A, Arg634A, Asn657A, Arg681A, Arg683A, Ile693A, Tyr695A, Thr638A, His625A, Thr696A, Ala623A, Ser686A, Ser691A, Ile692A, Tyr655A, Glycan N616(A), Glycan N657(A)

Table S4.3: Complete list of protein residues per GAG binding hotspot as predicted by ensemble-based docking with AutoDock and Kmeans clustering (continued)

R	Leu1024A, Thr1027A, Lys1028A, Glu1031A, Ser1037A, Lys1038A, Arg1039A, Val1040A, Asp1041A, Phe1042A, Cys1043A, Glu780B, Val781B, Gln784B, Ala1020B, Ser1021B, Ala1022B, Asn1023B, Leu1024B, Ala1025B, Ala1026B, Thr1027B, Lys1028B, Met1029B, Ser1030B, Glu1031B, Cys1032B, Val1033B, Leu1034B, Ser1037B, Arg1039B, Phe1042B, Thr1027C, Glu1031C, Arg1039C, Cys1032A, Leu727B, Cys1043B, Gly1035B, Phe1062B, Trp886B, Asn1023C, Ser1030C, Gln1036B, Lys1038B, His1064B, Phe1042C, Gly1044A, Lys1045A, Glu725A
S	Val722C, Gly799C, Phe800C, Asn801C, Phe802C, Ser803C, Gln804C, Ile805C, Gln920C, Lys921C, Leu922C, Ile923C, Ala924C, Asn925C, Gln926C, Phe927C, Asn928C, Ser929C, Ala930C, Ile931C, Gly932C, Lys933C, Ile934C, Gln935C, Asp936C, Glycan N709 (B), Glycan N717(C), Glycan N801(C)

Table S4.4: Fraction of each group measured by mass photometer for Omicron ternary complex with or without heparan sulfate

Group / Sample	A (fraction)	B (fraction)	C (fraction)
Omicron + ACE2	0.448	0.215	0.337
Omicron + ACE2	0.453	0.240	0.307
Omicron + ACE2	0.428	0.236	0.336
Average (Standard Deviation)	0.443 (0.014)	0.230 (0.013)	0.327 (0.017)
Group / Sample	A (fraction)	B (fraction)	C (fraction)
Omicron + HS + ACE2	0.468	0.395	0.137
Omicron + HS + ACE2	0.349	0.469	0.181
Omicron + HS + ACE2	0.316	0.488	0.195
Average (Standard Deviation)	0.378 (0.080)	0.451 (0.050)	0.171 (0.030)

Table S4.5: Fraction of each group measured by mass photometer for Delta ternary complex with or without heparan sulfate

Group / Sample	A (fraction)	B (fraction)	C (fraction)
Delta + ACE2	0.447	0.147	0.406
Delta + ACE2	0.438	0.152	0.409
Delta + ACE2	0.421	0.174	0.405
Average (Standard Deviation)	0.435 (0.013)	0.158 (0.014)	0.407 (0.002)
Group / Sample	A (fraction)	B (fraction)	C (fraction)
Delta + HS + ACE2	0.507	0.145	0.348
Delta + HS + ACE2	0.473	0.155	0.372
Delta + HS + ACE2	0.545	0.191	0.264
Average (Standard Deviation)	0.508 (0.036)	0.163 (0.024)	0.328 (0.057)

Table S4.6: Fraction of each group measured by mass photometer for WT ternary complex with or without heparan sulfate

Group / Sample	A (fraction)	B (fraction)	C (fraction)
WT + HS + ACE2	0.573	0.312	0.115
WT + HS + ACE2	0.697	0.224	0.079
WT + HS + ACE2	0.597	0.254	0.150
Average (Standard Deviation)	0.622 (0.066)	0.263 (0.045)	0.115 (0.035)
Group / Sample	A (fraction)	B (fraction)	C (fraction)
WT + HS + ACE2	0.517	0.336	0.147
WT + HS + ACE2	0.641	0.209	0.150
WT + HS + ACE2	0.553	0.276	0.171
Average (Standard Deviation)	0.570 (0.064)	0.274 (0.064)	0.156 (0.013)

Table S4.7: Comparison of the limit of detection for spike protein detection in lateral flow assay

Target	Bioreceptor	Detection type	Limit of Detection (ng/reaction)	Ref.
Spike	Primary: ACE2 Secondary: Antibody	sandwich-type	5 ng/reaction (buffer)	Lee2020[41]
Spike	Primary: NeuNAc[a] Secondary: NueNAc-PHEA$_{50}$[b]	sandwich-type	200 ng/reaction	Baker2020[42]
Spike	Primary: ACE2 Secondary: Antibody	sandwich-type	1 ng/reaction (buffer)	Lee2022[43]
Spike	Primary: Heparin Secondary: NTD Ab	sandwich-type	19.5 ng/reaction (buffer) 78 ng/reaction (saliva)	Kim2022[23]
Spike	Primary: Heparin Secondary: ACE2	Sandwich-type	10 ng/reaction (buffer) 20 ng/reaction (saliva)	In this study

4.6.4 Supplemental Figures

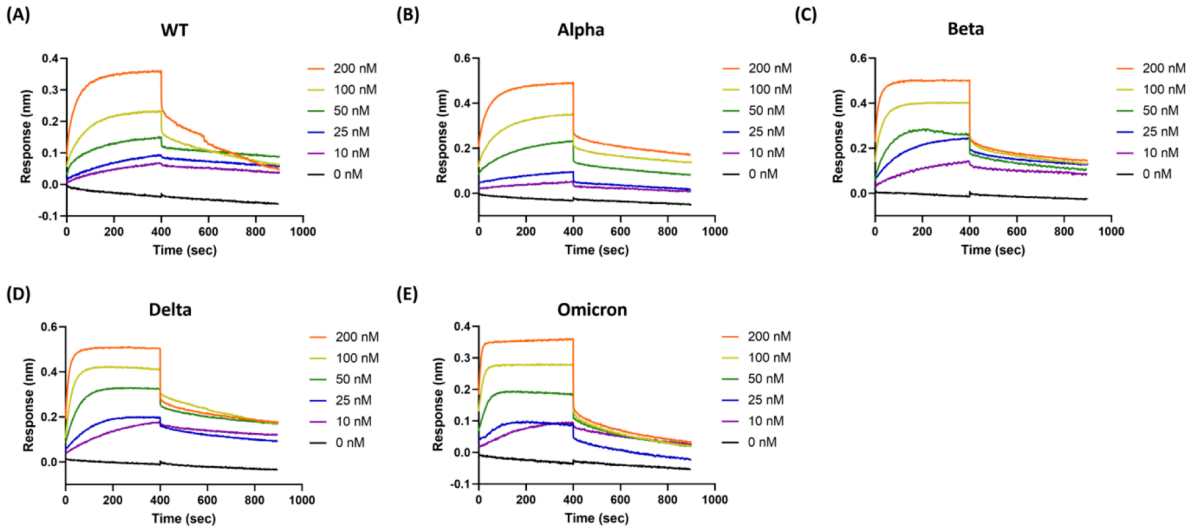


Figure S4.1: BLI sensogram of the ACE2 binding to variant of SARS-CoV-2 spike proteins

(A) Wild type, (B) Alpha, (C) Beta, (D) Delta, and (E) Omicron. ACE2 binding to 0, 10, 25, 50, 100, 200 nM of spike proteins for each variant were measured. Control sample (0 nM) was subtracted to analyze the binding affinity using the steady-state analysis.

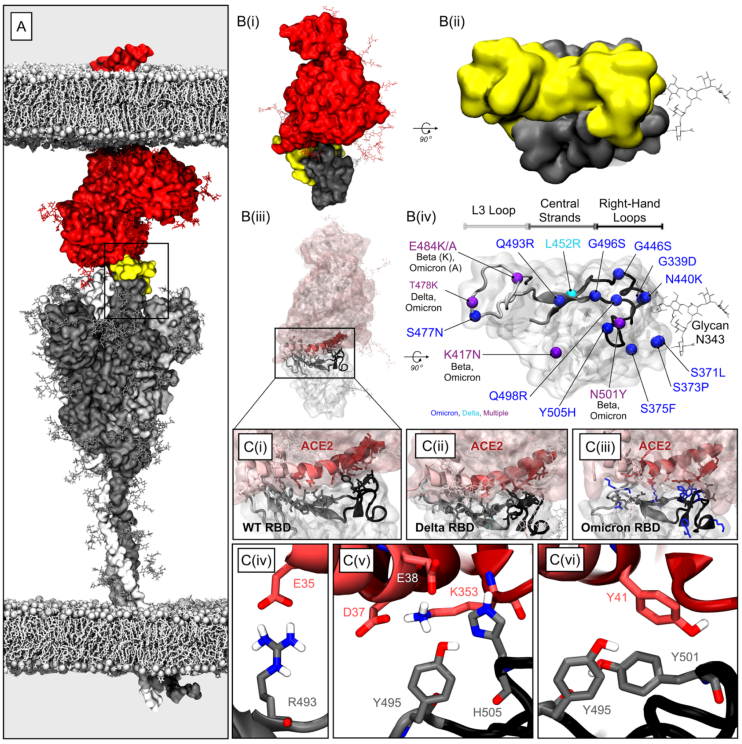


Figure S4.2: Image of (A) the spike/ACE2 double membrane complex.[44]

ACE2 shown in dark red surface representation, ACE2 glycans shown in dark red licorice; SARS-CoV-2 spike protein in 1up conformation is shown in a grey surface representation (chain A as dark grey, chain B as light grey, chain C as white), the spike's exposed RBM is highlighted with a yellow surface, spike glycans shown in grey licorice; human and SARS-CoV-2 membranes are shown in grey licorice representation with phosphorous atoms represented in VDW. A rectangular box outlines the ACE2+RBM interface. (B(i)) Image showing the ACE2+RBD system constructed in this work for MD simulations. All molecules colored and represented as described in panel A. (B(ii)) Spike RBD with RBM highlighted. Spike RBD in grey surface representation, RBM highlighted in yellow. N-linked glycan at N343 shown in grey licorice representation. (B(iii)) Same system as shown in B(i) but with transparent surfaces to highlight the secondary structure at the ACE2+RBD interface. ACE2 surface is shown in transparent red, with ACE2 glycans shown in pink licorice, spike RBD surface is shown in transparent grey with glycan N343 shown in grey licorice. The RBD side of the ACE2+RBD interface is represented in opaque ribbons and colored in shades of grey; the L3 Loop (residues 470-489), central beta strands (residues 450-456, 490-496), and right-handed loops (residues 438-450, 497-508) are colored as light grey, grey, and black ribbons, respectively. The ACE2 side of the ACE2+RBD interface is composed largely of one N-terminal alpha helix (residues 18-50), a neighboring helix (residues 82-83), and a neighboring beta-strand (residues 353-357). ACE2 residues within the ACE2+RBD interface that interact primarily with the RBD L3 loops, RBD central beta strands, and RBD right-handed loops are represented as pink, red, and dark red ribbons, respectively. A callout box highlights the ACE2+RBD interface. (B(iv)) Spike RBD shown in same orientation as shown in B(ii) but with transparent surface to reveal the positions of mutations in Beta, Delta, and Omicron BA.1 SARS-CoV-2 variant spike sequences. The L3 Loop, central beta strands, and right-handed loops are represented as in B(iii). Positions of mutations seen in the Delta and Omicron spike sequences are shown as cyan and blue beads, respectively, while positions of mutations seen in multiple spike sequences are shown as purple beads with denotation of sequence. Glycan N343 is represented in grey licorice. (C(i-iii)) Visualization of WT, Delta, and Omicron RBDs bound to ACE2. Secondary structural elements at the ACE2+RBD interface are represented in panels B(iii) and B(iv). Cyan and blue residues indicate positions of mutation within the Delta and Omicron spike sequences, respectively. (C(iv-vi)) Zoom-in images highlighting specific interactions of high interest observed in ACE2+Omicron spike RBD simulations as discussed in the main text. Red carbon atoms denote ACE2 residues, grey carbon atoms denote Omicron RBD residues.

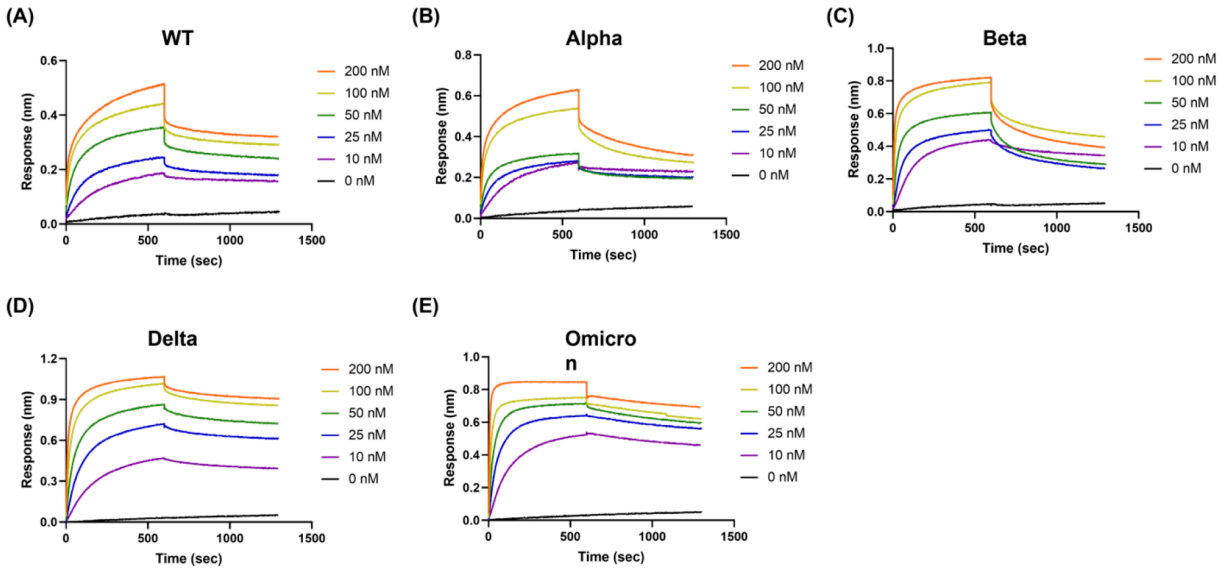


Figure S4.3: BLI sensogram of the heparin (HEP) binding to variant of SARS-CoV-2 spike proteins

(A) Wild type, (B) Alpha, (C) Beta, (D) Delta, and (E) Omicron. HEP binding to 0, 10, 25, 50, 100, 200 nM of spike proteins for each variant were measured. Control sample (0 nM) was subtracted to analyze the binding affinity using the steady-state analysis.

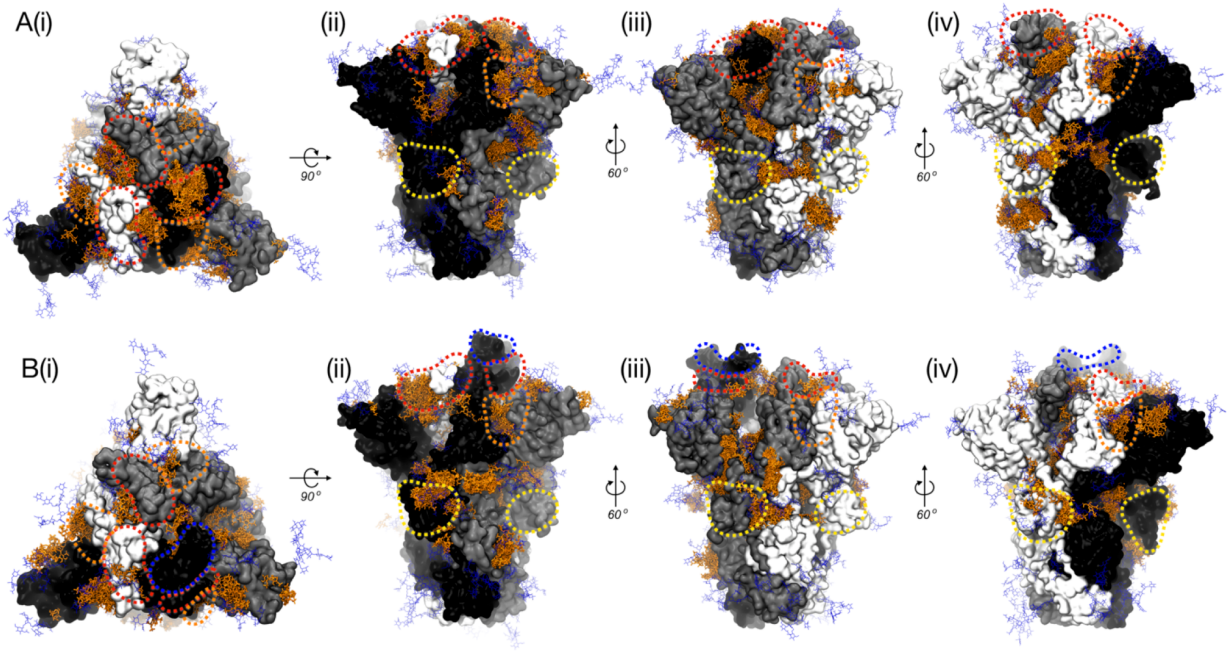


Figure S4.4: AutoDock Vina docking results illustrating heparin dimeric and tetrameric models bound to WT

(A) closed and (B) 1-up structures. In all structures, spike chains A, B, and C are shown in black, grey, and white surface representations. Spike glycans are shown in thin blue licorice. Heparin dimeric and tetrameric binding modes are shown in orange licorice. Subpanels (i-iv) each represent the same structure in a different orientation. In all images, red, orange, yellow, and blue dotted lines highlight the RBD patch, RBD cleft, FCS, and RBM sites respectively.

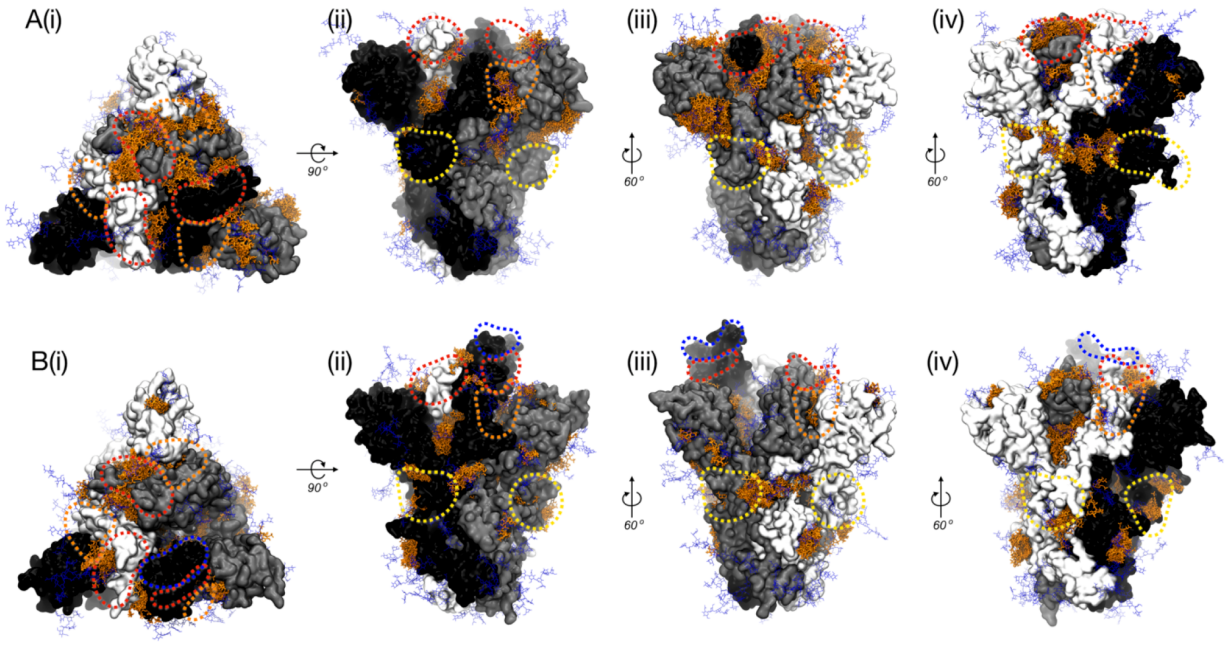


Figure S4.5: AutoDock Vina docking results illustrating heparin dimeric and tetrameric models bound to Delta

(A) is closed and (B) is 1-up. In all structures, spike chains A, B, and C are shown in black, grey, and white surface representations. Spike glycans are shown in thin blue licorice. Heparin dimeric and tetrameric binding modes are shown in orange licorice. In all images, red, orange, yellow, and blue dotted lines highlight the RBD patch, RBD cleft, FCS, and RBM sites respectively.

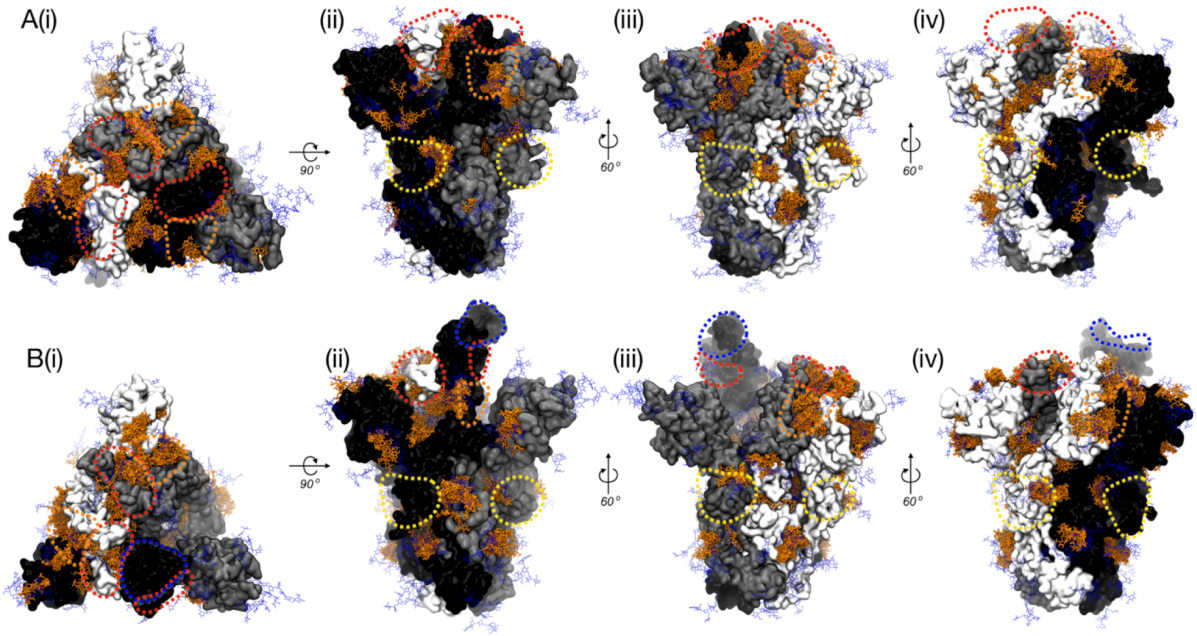


Figure S4.6: AutoDock Vina docking results illustrating heparin dimeric and tetrameric models bound to Omicron

(A) is closed and (B) is 1-up. In all structures, spike chains A, B, and C are shown in black, grey, and white surface representations. Spike glycans are shown in thin blue licorice. Heparin dimeric and tetrameric binding modes are shown in orange licorice. In all images, red, orange, yellow, and blue dotted lines highlight the RBD patch, RBD cleft, FCS, and RBM sites respectively.

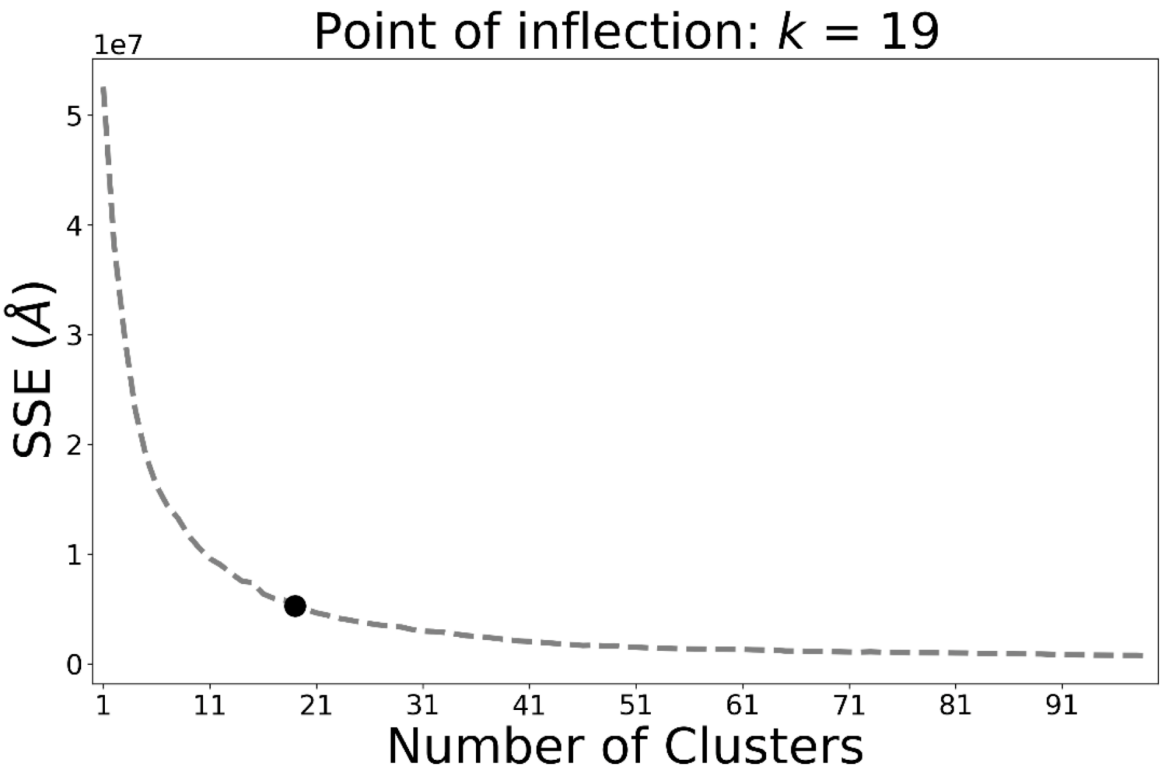


Figure S4.7: K-means clustering results determining the optimal number of clusters from 28,800 AutoDock vina binding modes

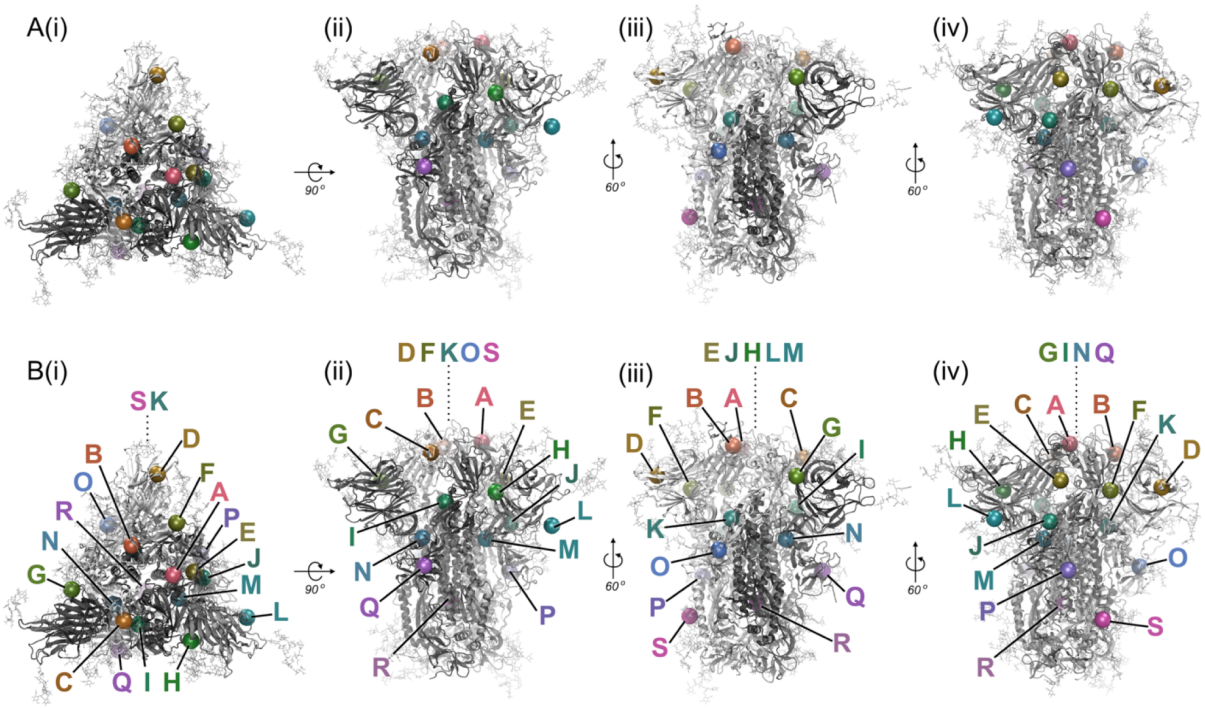


Figure S4.8: All 19 heparin hotspots found via ensemble-based docking with AutoDock and kmeans clustering.

(A) Panels (i-iv) show the WT spike protein in closed conformation; chains A, B, and C are represented in grey, light grey, and white ribbons, respectively, and spike glycans are represented in light grey licorice. The centroid of each identified "hotspot" is represented with a 5Å sphere. The hotspots were ordered by height on the spike protein (i.e., according to z-coordinate) and colored according to a rainbow color palette. All images are based on an identical structure and VMD scene but rotated to give complete viewing of spike apex (i) and three sides (ii-iv). (B) Panels (i-iv) show the same WT spike protein in closed conformation with the same representation scheme as in (A), but with each identified binding hotspot labeled according to its name A-R. Hotspots were named as described in Supporting Information methods. Sites not visible in a given orientation, due to rotation or fog in the image, are denoted with a dashed line.

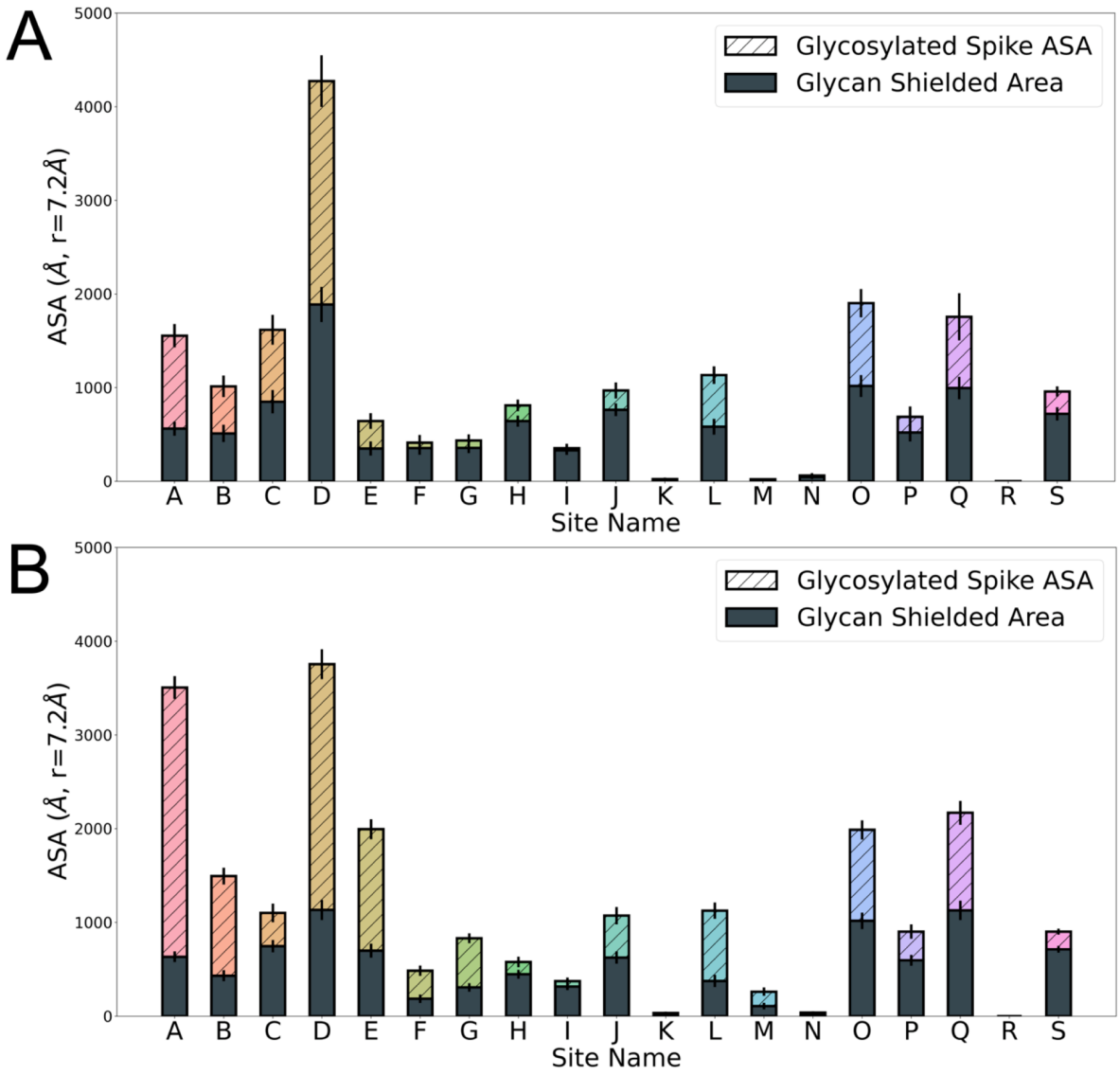


Figure S4.9: Accessible Surface Area plotted for each site calculated with a probe radius of 7.2 Å, calculated according to the Shrake-Rupley algorithm

(A) ASAs calculated for each binding site from wild type spike simulations in the all-RBDs-down/closed conformation. (B) ASAs calculated for each binding site from wild type spike simulations in the 1-up RBD conformation.

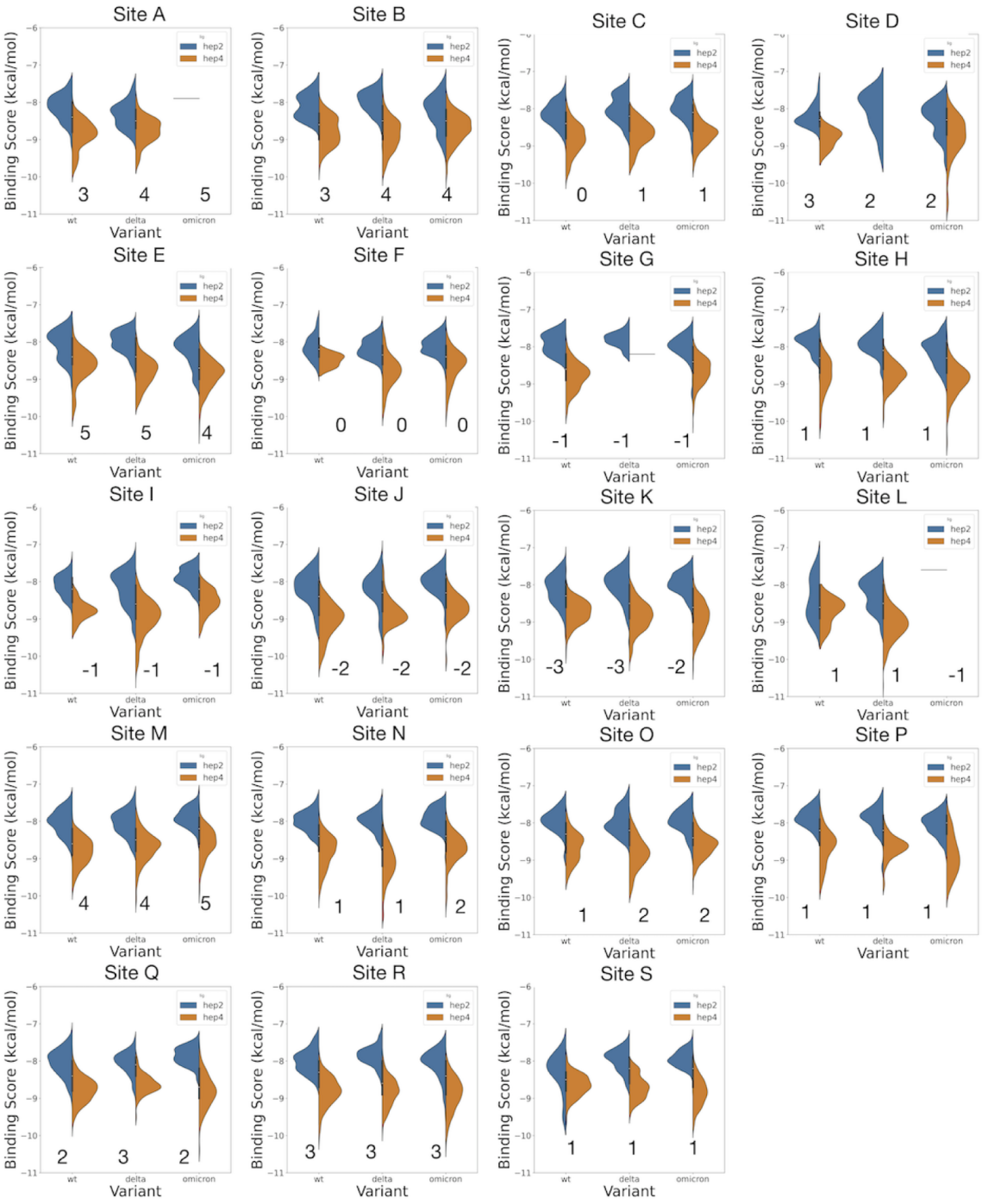


Figure S4.10: Violin plots demonstrating distribution of AutoDock Vina binding scores for heparin dimer (hep2) and tetramer (hep4) models each binding hotspot across all variants

Total charge per site per variant is denoted next to each violin within the plot. Site based distributions of AutoDock Vina binding scores were taken from both closed and 1up spike conformations per variant within each site.

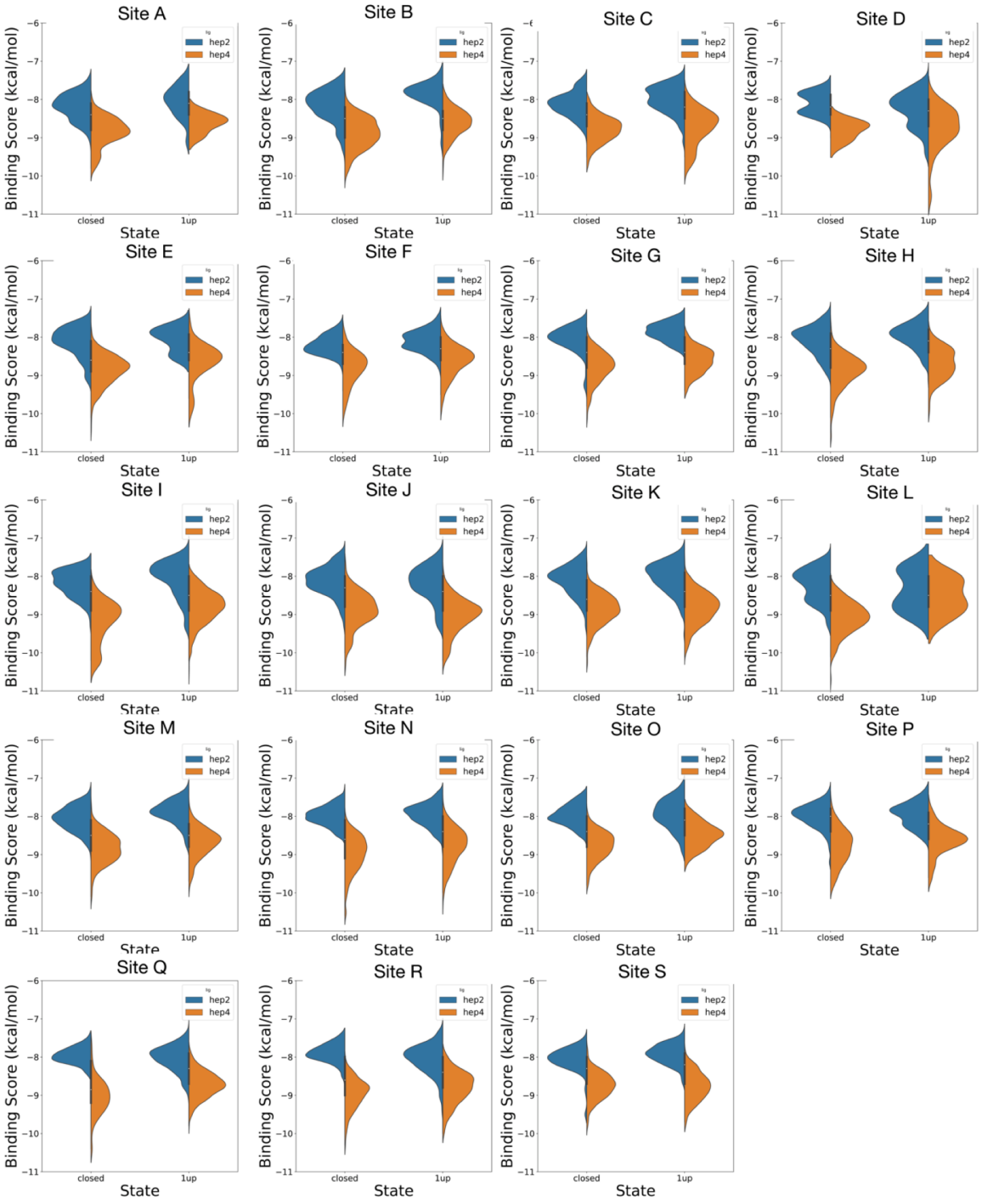


Figure S4.11: Violin plots demonstrating distribution of AutoDock Vina binding scores for heparin dimer (hep2) and tetramer (hep4) models each binding hotspot across 1up and closed spike structures

Site based distributions of AutoDock Vina binding scores were taken from all three spike variants (WT, Delta, and Omicron) per spike conformation.

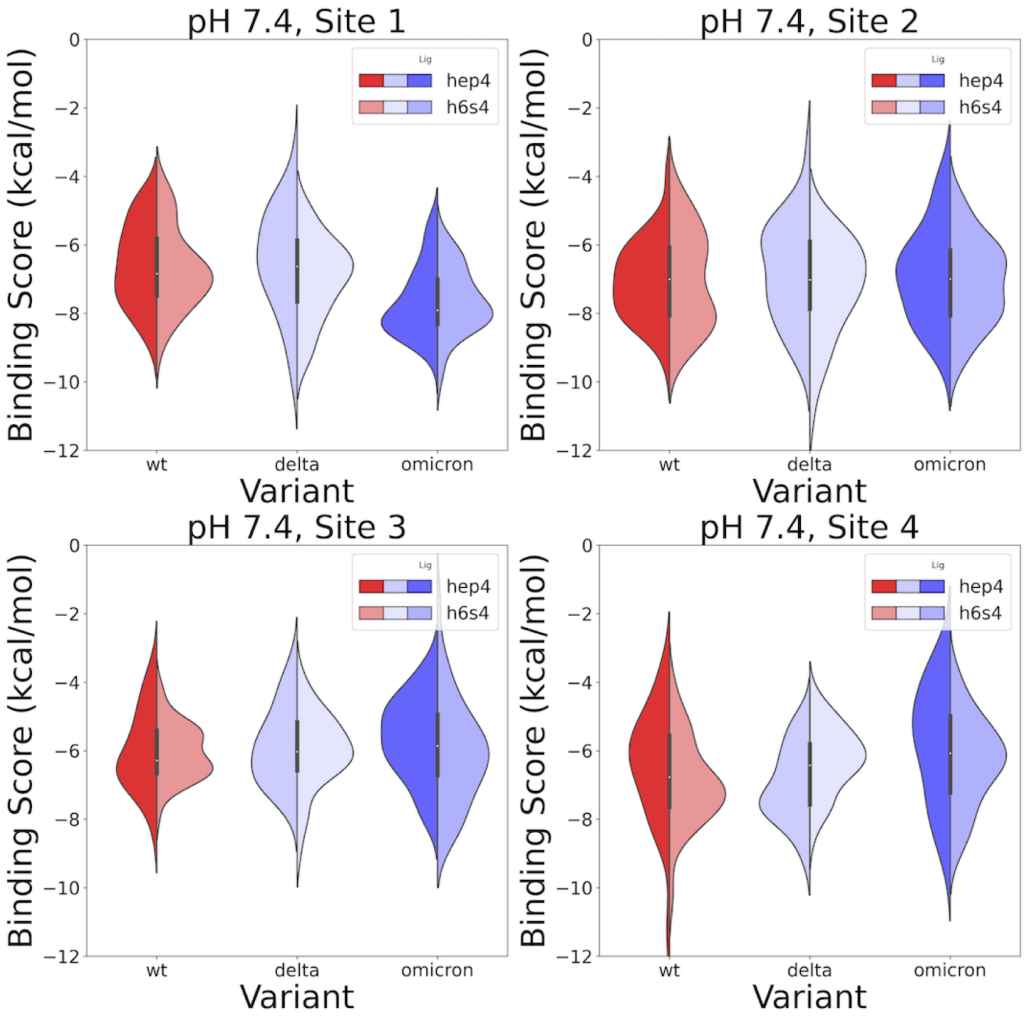


Figure S4.12: Violin plots illustrating the distribution of binding scores predicted by Schrodinger's Induced Fit Docking protocol in each of the probed sites, for each of the Variants, at pH's 7.4

Site 1 is the RBD Cleft Site, Site 2 is the RBD Patch Site, Site 3 is the Furin Cleavage Site, and Site 4 is the Receptor Binding Motif.

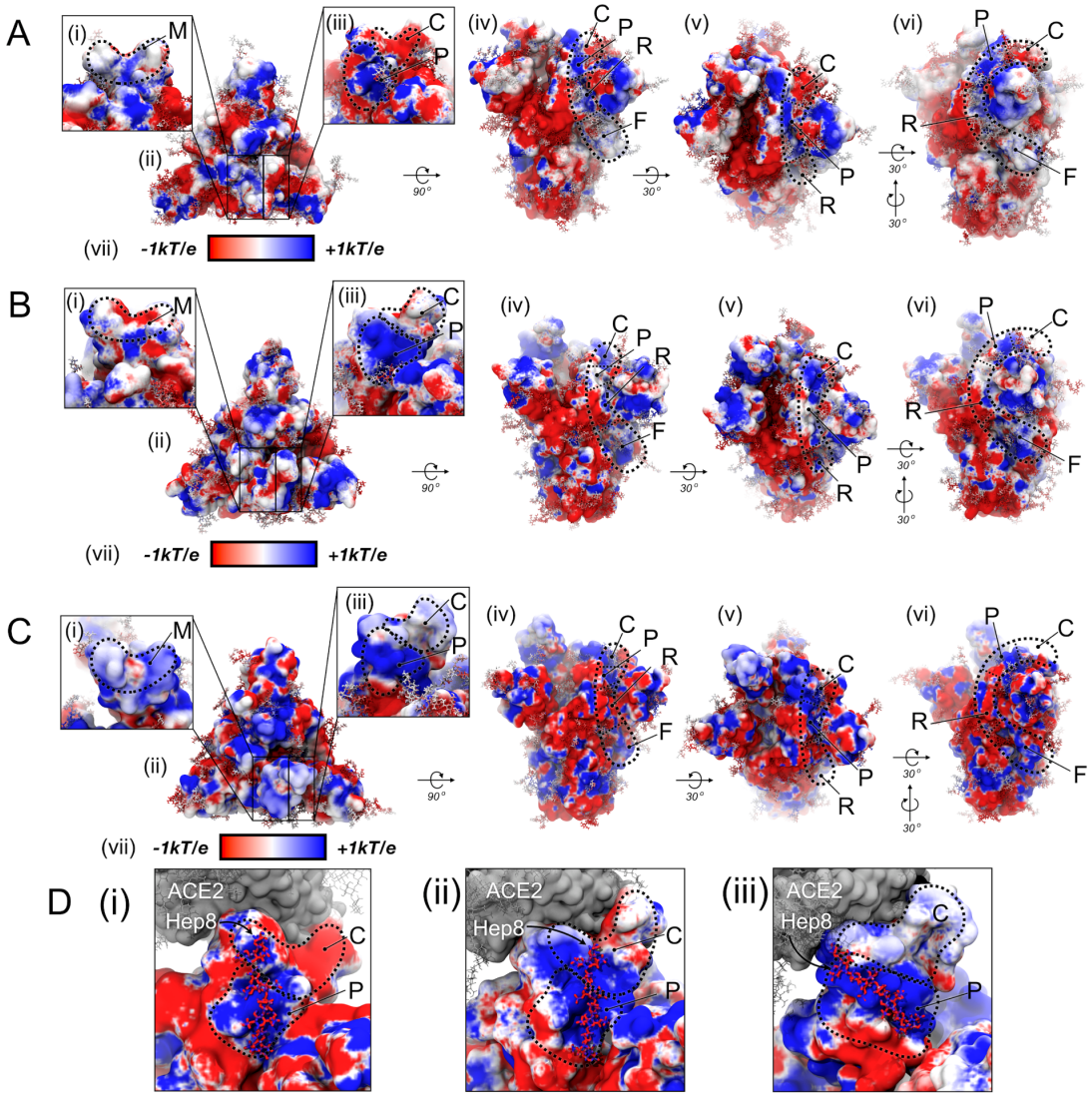


Figure S4.13: Dynamically-averaged electrostatic potential maps collected from 50 ns of MD simulations for (A) WT, (B) Delta, and (C) Omicron spike proteins in the 1up RBD conformation

For images A-C, labels (i-vii) indicate the following images: a close-up view of the RBM site (denoted as M), a top-down view of the spike protein, a close-up view of the RBD Cleft and RBD Patch sites (denoted as C and P, respectively), a side view of the spike protein with RBD Cleft, RBD Patch, Connecting Ridge, and FCS sites (denoted as C, P, R, and F, respectively) highlighted, a tilted-top-down view of the spike protein with RBD Cleft, RBD Patch, and Connecting Ridge sites (denoted as C, P, and R, respectively) highlighted, a rotated-side view of the spike protein with RBD Cleft, RBD Patch, Connecting Ridge, and FCS sites (denoted as C, P, R, and F, respectively) highlighted, and a color bar demonstrating the color ranges for each image and the corresponding calculated electrostatic potential. In all panels, protein surfaces are colored according to average electrostatic potential at each site, ranging from –1 k_BT/e (red) to +1 k_BT/e (blue). (D) Images demonstrating how the SARS-CoV-2 spike can, in the 1up state, accommodate binding of ACE2 (grey surface) at the RBM and a heparin octamer (Hep8) bound to the positively charged RBD Patch site for (i) WT, (ii) Delta, and (iii) Omicron spike proteins. These images were generated by alignment with VMDtools.

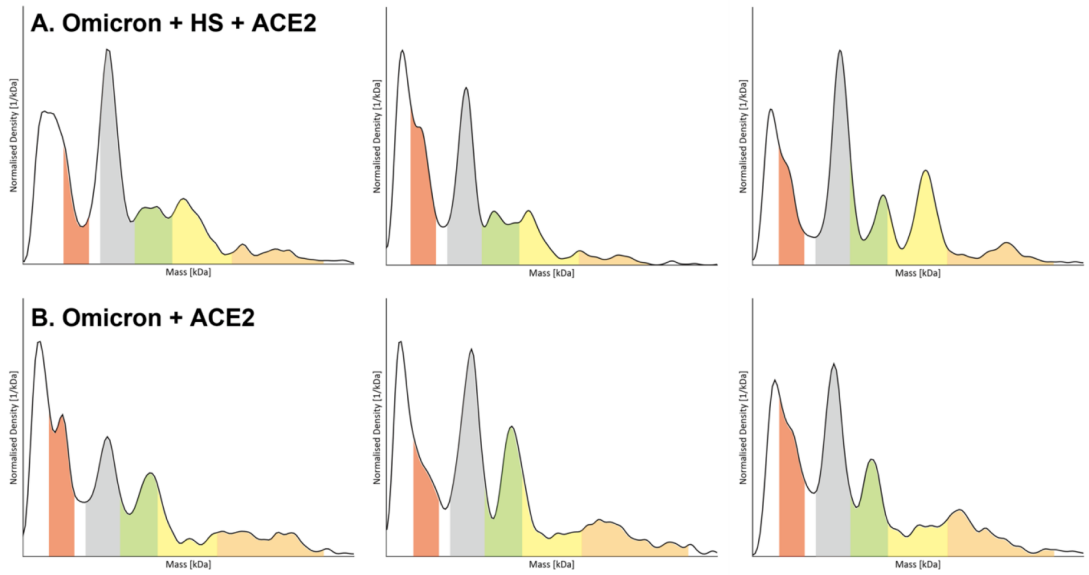


Figure S4.14: Repetition results of Omicron variant mass photometer

Row (A) Omicron spike + HS + ACE2, and row (B) Omicron spike + ACE2. Mass distribution of Omicron spike (mass range highlighted in grey), dACE2 (mass range highlighted in red), and Possible ternary complexes are grouped in A (green), B (yellow), C (orange) based on their expected mass ranges.

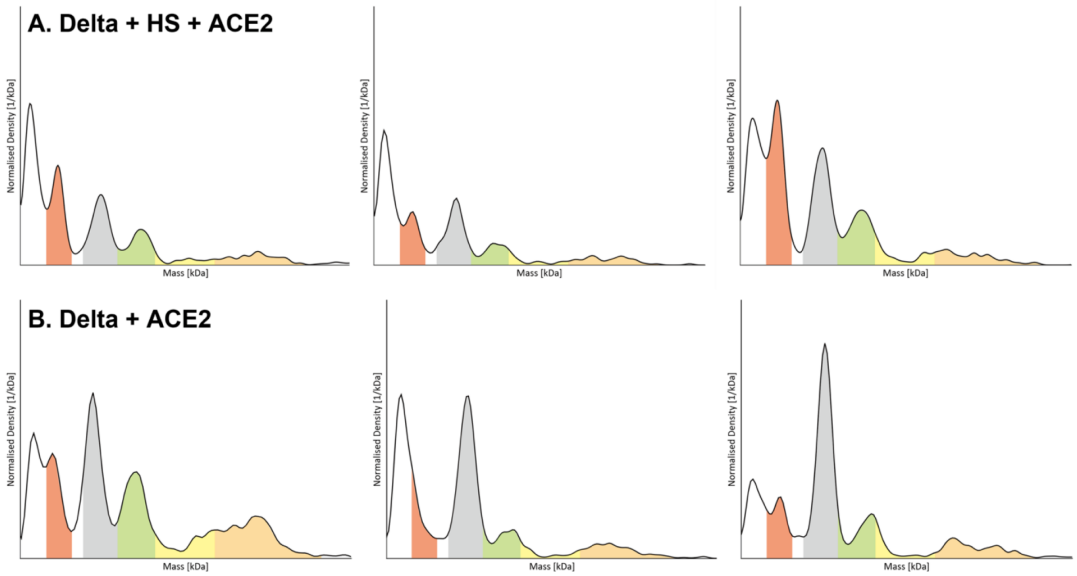


Figure S4.15: Repetition results of Delta variant mass photometer

Row (A) Delta spike + HS + ACE2, and row (B) Delta spike + ACE2. Mass distribution of Delta spike (mass range highlighted in grey), dACE2 (mass range highlighted in red), and Possible ternary complexes are grouped in A (green), B (yellow), C (orange) based on their expected mass ranges.

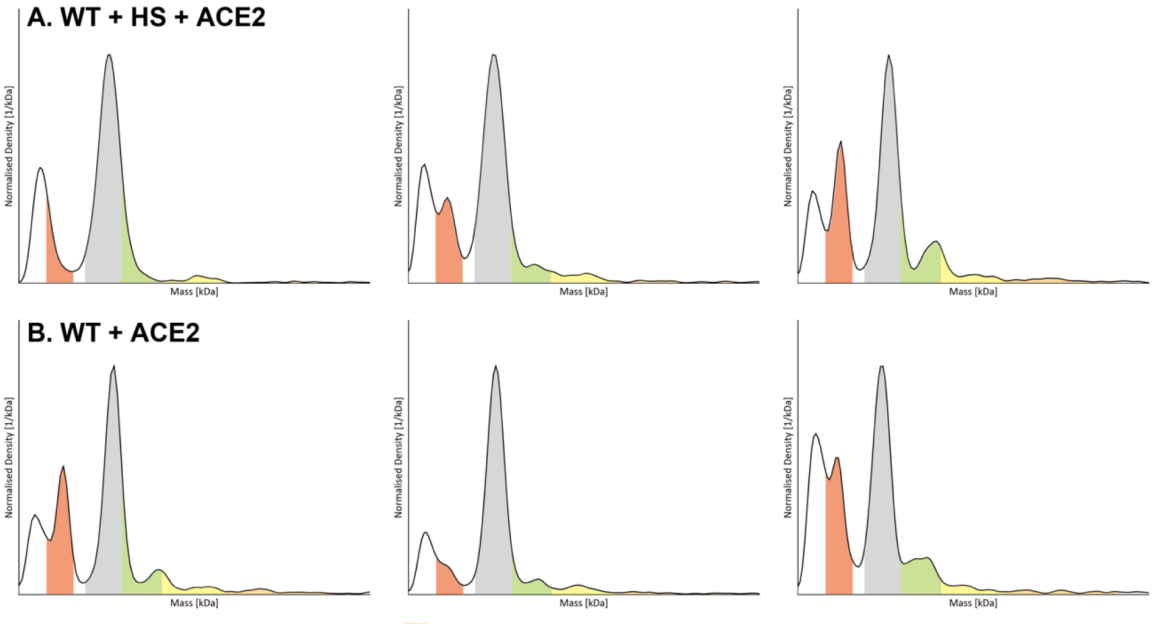


Figure S4.16: Repetition results of Wild type mass photometer

Row (A) WT spike + HS + ACE2, and row (B) WT spike + ACE2. Mass distribution of WT spike (mass range highlighted in grey), dACE2 (mass range highlighted in red), and Possible ternary complexes are grouped in A (green), B (yellow), C (orange) based on their expected mass ranges.

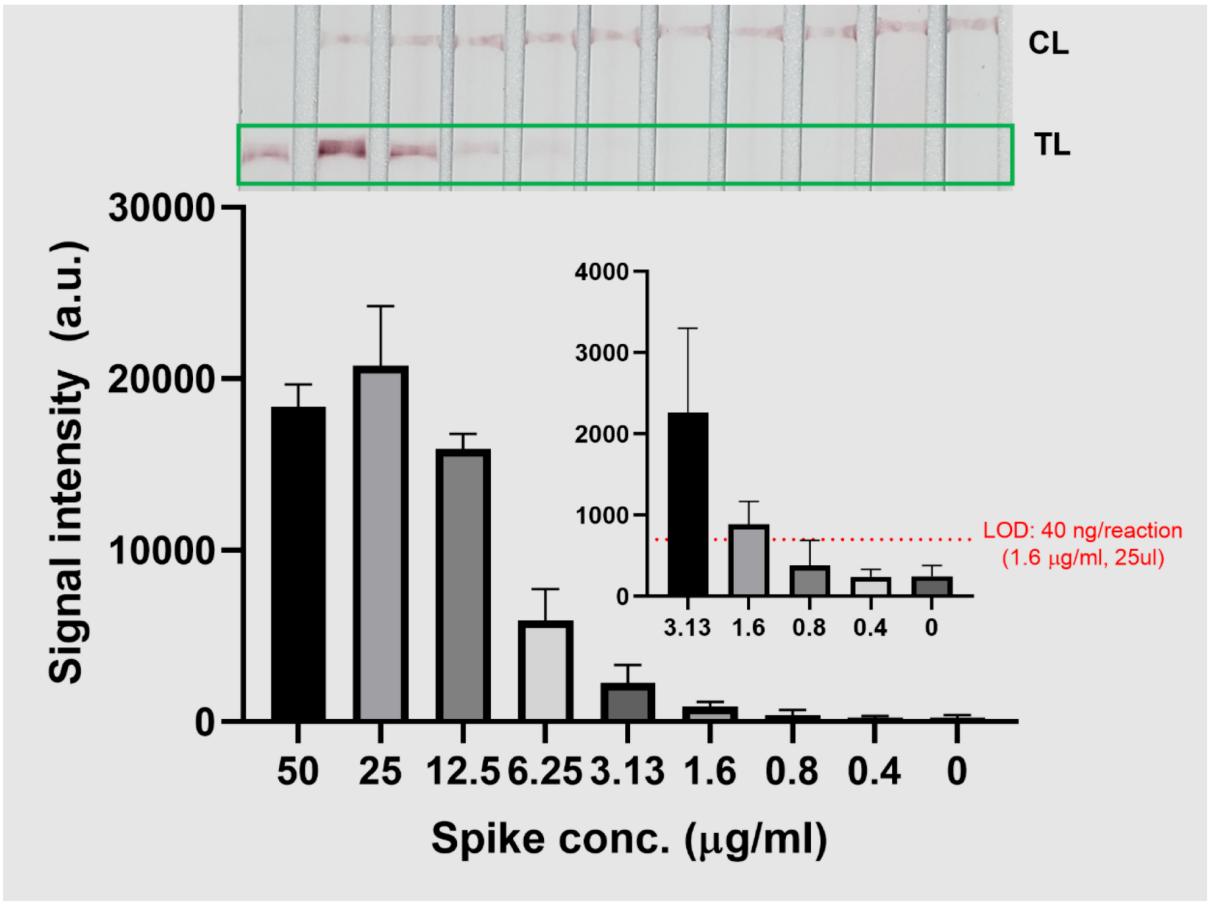


Figure S4.17: Dose-dependency results of Omicron detection using *GlycoGrip2.0* without signal enhancement in buffer condition

The limit of the detection was calculated by the blank + 3x (Standard deviation of blank). At least three independent tests were performed (n ≥ 3) for *Glycogrip.*

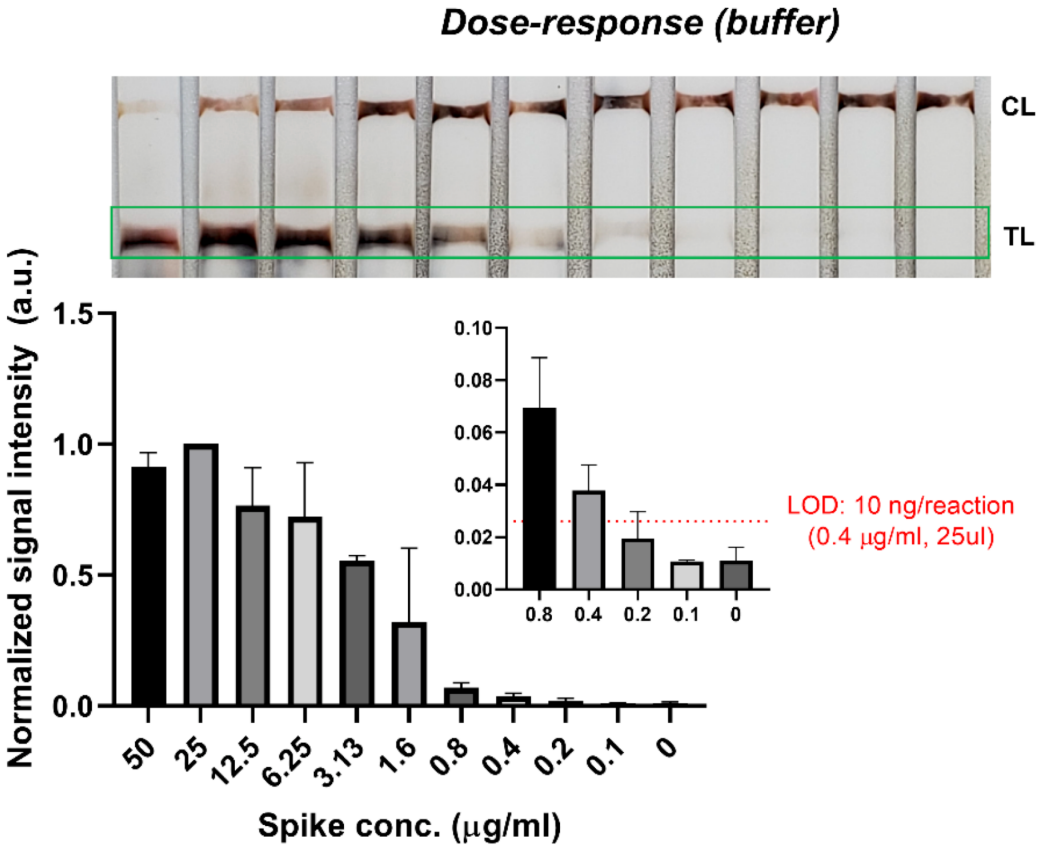


Figure S4.18: Dose-dependency results of Omicron detection using *GlycoGrip2.0* with signal enhancement in buffer condition

The limit of the detection was calculated by the blank + 3x (Standard deviation of blank). At least three independent tests were performed (n ≥ 3).

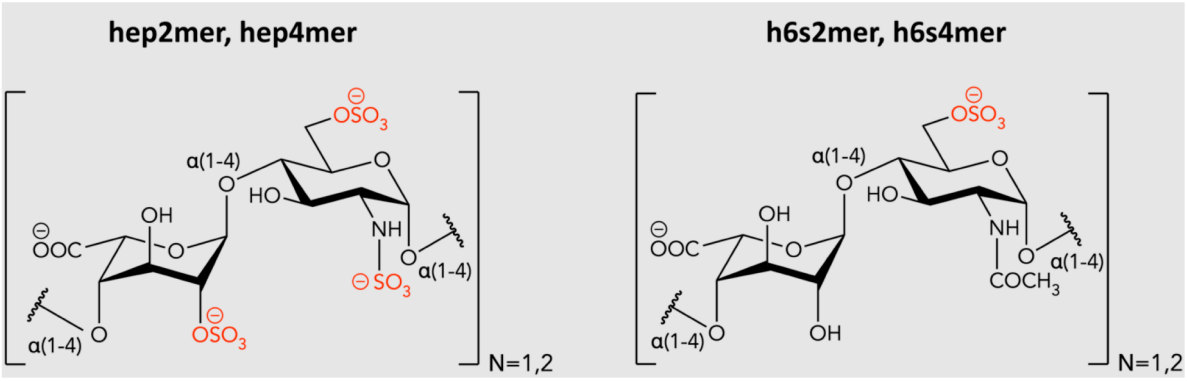


Figure S4.19: ChemDraws of all molecules modeled and docked in this work. Hep2mer, hep4mer, h6s2mer, h6s4mer.

4.7 **Acknowledgements**

FLK, MAR, LV, LC, and REA would like to acknowledge TACC Frontera and SDSC TSCC for their continued, unwavering support of our work over the past several years.

Chapter 4, in full, is a modified reprint of the materials as it has been accepted for publication in Cell Reports Physical Science as "Kim SH†, Kearns FL†, Rosenfeld MA†, Votapka L, Casalino L, Papanikolas MJ, Amaro RE, Freeman R. Positively bound: *SARS-CoV-2 Evolved Variants Optimize Binding to Cellular Glycocalyx.* Cell Reports Physical Science. 2023. (accepted)." The dissertation author was a primary co-investigator and co-author of this work.

4.8 **References**

4.7.1 Main Text References

(1) WHO Coronavirus (COVID-19) Dashboard. World Health Organization

(2) World Health Organization. Tracking SARS-CoV-2 Variants of Concern. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed 2021-12-12).

(3) Center for Disease Control and Prevention. SARS-CoV-2 Variant Classifications and Definitions. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html (accessed 2022-08-08).

(4) Evolutionary Insight into the Emergence of SARS-CoV-2 Variants of Concern. Nat Med 2022, 28 (7), 1357–1358. https://doi.org/10.1038/s41591-022-01892-2.

(5) Bashor, L.; Gagne, R. B.; Bosco-Lauth, A. M.; Bowen, R. A.; Stenglein, M.; VandeWoude, S. SARS-CoV-2 Evolution in Animals Suggests Mechanisms for Rapid Variant Selection. Proceedings of the National Academy of Sciences 2021, 118 (44). https://doi.org/10.1073/pnas.2105253118.

(6) Otto, S. P.; Day, T.; Arino, J.; Colijn, C.; Dushoff, J.; Li, M.; Mechai, S.; van Domselaar, G.; Wu, J.; Earn, D. J. D.; Ogden, N. H. The Origins and Potential Future of SARS-CoV-2 Variants of Concern in the Evolving COVID-19 Pandemic. Current Biology 2021, 31 (14), R918–R929. https://doi.org/10.1016/j.cub.2021.06.049.

(7) Maher, M. C.; Bartha, I.; Weaver, S.; di Iulio, J.; Ferri, E.; Soriaga, L.; Lempp, F. A.; Hie, B. L.; Bryson, B.; Berger, B.; Robertson, D. L.; Snell, G.; Corti, D.; Virgin, H. W.; Kosakovsky Pond, S. L.; Telenti, A. Predicting the Mutational Drivers of Future SARS-CoV-2 Variants of Concern. Sci Transl Med 2022, 14 (633). https://doi.org/10.1126/scitranslmed.abk3445.

(8) Harvey, W. T.; Carabelli, A. M.; Jackson, B.; Gupta, R. K.; Thomson, E. C.; Harrison, E. M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S. J.; Robertson, D. L. SARS-CoV-2 Variants, Spike Mutations and Immune Escape. Nat Rev Microbiol 2021, 19 (7), 409–424. https://doi.org/10.1038/s41579-021-00573-0.

(9) McCallum, M.; Walls, A. C.; Sprouse, K. R.; Bowen, J. E.; Rosen, L.; Dang, H. v; DeMarco, A.; Franko, N.; Tilles, S. W.; Logue, J.; Miranda, M. C.; Ahlrichs, M.; Carter, L.; Snell, G.; Pizzuto, M. S.; Chu, H. Y.; van Voorhis, W. C.; Corti, D.; Veesler, D. Molecular Basis of Immune Evasion by the Delta and Kappa SARS-CoV-2 Variants. Science (1979) 2021, 0 (0), eabl8506. https://doi.org/10.1126/science.abl8506.

(10) Public Health England. Public Health England Investigation of Novel SARS-COV-2 Variant 202012/01: Technical Briefing; 2020.

(11) Zhou, D.; Dejnirattisai, W.; Supasa, P.; Liu, C.; Mentzer, A. J.; Ginn, H. M.; Zhao, Y.; Duyvesteyn, H. M. E.; Tuekprakhon, A.; Nutalai, R.; Wang, B.; Paesen, G. C.; Lopez-Camacho, C.; Slon-Campos, J.; Hallis, B.; Coombes, N.; Bewley, K.; Charlton, S.; Walter, T. S.; Skelly, D.; Lumley, S. F.; Dold, C.; Levin, R.; Dong, T.; Pollard, A. J.; Knight, J. C.; Crook, D.; Lambe, T.; Clutterbuck, E.; Bibi, S.; Flaxman, A.; Bittaye, M.; Belij-Rammerstorfer, S.; Gilbert, S.; James, W.; Carroll, M. W.; Klenerman, P.; Barnes, E.; Dunachie, S. J.; Fry, E. E.; Mongkolsapaya, J.; Ren, J.; Stuart, D. I.; Screaton, G. R. Evidence of Escape of SARS-CoV-2 Variant B.1.351 from Natural and Vaccine-Induced Sera. Cell 2021, 184 (9), 2348-2361.e6. https://doi.org/10.1016/j.cell.2021.02.037.

(12) Quandt, J.; Muik, A.; Salisch, N.; Lui, B. G.; Lutz, S.; Krüger, K.; Wallisch, A.-K.; Adams-Quack, P.; Bacher, M.; Finlayson, A.; Ozhelvaci, O.; Vogler, I.; Grikscheit, K.; Hoehl, S.; Goetsch, U.; Ciesek, S.; Türeci, Ö.; Sahin, U. Omicron BA.1 Breakthrough Infection Drives Cross-Variant Neutralization and Memory B Cell Formation against Conserved Epitopes. Sci Immunol 2022. https://doi.org/10.1126/sciimmunol.abq2427.

(13) Shrestha, L. B.; Foster, C.; Rawlinson, W.; Tedla, N.; Bull, R. A. Evolution of the SARS-CoV-2 Omicron Variants BA.1 to BA.5: Implications for Immune Escape and Transmission. Rev Med Virol 2022. https://doi.org/10.1002/rmv.2381.

(14) Mohapatra, R. K.; Kandi, V.; Verma, S.; Dhama, K. Challenges of the Omicron (B.1.1.529) Variant and Its Lineages: A Global Perspective. ChemBioChem 2022, 23 (9). https://doi.org/10.1002/cbic.202200059.

(15) Shao, W.; Zhang, W.; Fang, X.; Yu, D.; Wang, X. Challenges of SARS-CoV-2 Omicron Variant and Appropriate Countermeasures. Journal of Microbiology, Immunology and Infection 2022, 55 (3), 387–394. https://doi.org/10.1016/j.jmii.2022.03.007.

(16) Leuzinger, K.; Roloff, T.; Egli, A.; Hirsch, H. H. Impact of SARS-CoV-2 Omicron on Rapid Antigen Testing Developed for Early-Pandemic SARS-CoV-2 Variants. Microbiol Spectr 2022, 10 (4). https://doi.org/10.1128/spectrum.02006-22.

(17) United States Food and Drug Administration. At-Home COVID-19 Antigen Tests-Take Steps to Reduce Your Risk of False Negative: FDA Safety Communication.

(18) VanBlargan, L. A.; Errico, J. M.; Halfmann, P. J.; Zost, S. J.; Crowe, J. E.; Purcell, L. A.; Kawaoka, Y.; Corti, D.; Fremont, D. H.; Diamond, M. S. An Infectious SARS-CoV-2 B.1.1.529 Omicron Virus Escapes Neutralization by Therapeutic Monoclonal Antibodies. Nat Med 2022, 28 (3), 490–495. https://doi.org/10.1038/s41591-021-01678-y.

(19) Tuekprakhon, A.; Nutalai, R.; Dijokaite-Guraliuc, A.; Zhou, D.; Ginn, H. M.; Selvaraj, M.; Liu, C.; Mentzer, A. J.; Supasa, P.; Duyvesteyn, H. M. E.; Das, R.; Skelly, D.; Ritter, T. G.; Amini, A.; Bibi, S.; Adele, S.; Johnson, S. A.; Constantinides, B.; Webster, H.; Temperton, N.; Klenerman, P.; Barnes, E.; Dunachie, S. J.; Crook, D.; Pollard, A. J.; Lambe, T.; Goulder, P.; Paterson, N. G.; Williams, M. A.; Hall, D. R.; Fry, E. E.; Huo, J.; Mongkolsapaya, J.; Ren, J.; Stuart, D. I.; Screaton, G. R.; Conlon, C.; Deeks, A.; Frater, J.; Frending, L.; Gardiner, S.; Jämsén, A.; Jeffery, K.; Malone, T.; Phillips, E.; Rothwell, L.; Stafford, L. Antibody Escape of SARS-CoV-2 Omicron BA.4 and BA.5 from Vaccine and BA.1 Serum. Cell 2022, 185 (14), 2422-2433.e13. https://doi.org/10.1016/j.cell.2022.06.005.

(20) Planas, D.; Saunders, N.; Maes, P.; Guivel-Benhassine, F.; Planchais, C.; Buchrieser, J.; Bolland, W.-H.; Porrot, F.; Staropoli, I.; Lemoine, F.; Péré, H.; Veyer, D.; Puech, J.; Rodary, J.; Baele, G.; Dellicour, S.; Raymenants, J.; Gorissen, S.; Geenen, C.; Vanmechelen, B.; Wawina-Bokalanga, T.; Martí-Carreras, J.; Cuypers, L.; Sève, A.; Hocqueloux, L.; Prazuck, T.; Rey, F. A.; Simon-Loriere, E.; Bruel, T.; Mouquet, H.; André, E.; Schwartz, O. Considerable Escape of SARS-CoV-2 Omicron to Antibody Neutralization. Nature 2022, 602 (7898), 671–675. https://doi.org/10.1038/s41586-021-04389-z.

(21) Hoffmann, M.; Zhang, L.; Pöhlmann, S. Omicron: Master of Immune Evasion Maintains Robust ACE2 Binding. Signal Transduct Target Ther 2022, 7 (1), 118. https://doi.org/10.1038/s41392-022-00965-5.

(22) Socher, E.; Heger, L.; Paulsen, F.; Zunke, F.; Arnold, P. Molecular Dynamics Simulations of the Delta and Omicron SARS-CoV-2 Spike – ACE2 Complexes Reveal Distinct Changes between Both Variants. Comput Struct Biotechnol J 2022, 20, 1168–1176. https://doi.org/10.1016/j.csbj.2022.02.015.

(23) Cotten, M.; Phan, M. V. T. Evolution to Increased Positive Charge on the Viral Spike Protein May Be Part of the Adaptation of SARS-CoV-2 to Human Transmission. bioRxiv 2022, 2022.07.30.502143. https://doi.org/10.1101/2022.07.30.502143.

(24) Mehta, P.; Ravi, V.; Devi, P.; Maurya, R.; Parveen, S.; Mishra, P.; Yadav, A.; Swaminathan, A.; Saifi, S.; Khare, K.; Chattopadhyay, P.; Yadav, M.; Chauhan, N. S.; Tarai, B.; Budhiraja, S.; Shamim, U.; Pandey, R. Mutational Dynamics across VOCs in International Travellers and Community Transmission Underscores Importance of Spike-ACE2 Interaction. Microbiol Res 2022, 262, 127099. https://doi.org/10.1016/j.micres.2022.127099.

(25) da Costa, C. H. S.; de Freitas, C. A. B.; Alves, C. N.; Lameira, J. Assessment of Mutations on RBD in the Spike Protein of SARS-CoV-2 Alpha, Delta and Omicron Variants. Sci Rep 2022, 12 (1), 8540. https://doi.org/10.1038/s41598-022-12479-9.

(26) Kim, S.; Liu, Y.; Ziarnik, M.; Cao, Y.; Zhang, X. F.; Im, W. Binding of Human ACE2 and RBD of Omicron Enhanced by Unique Interaction Patterns Among SARS-CoV-2 Variants of Concern. bioRxiv 2022, 2022.01.24.477633. https://doi.org/10.1101/2022.01.24.477633.

(27) Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Perrett, H. R.; Painter, C. D.; Narayanan, A.; Majowicz, S. A.; Kwong, E. M.; McVicar, R. N.; Thacker, B. E.; Glass, C. A.; Yang, Z.; Torres, J. L.; Golden, G. J.; Bartels, P. L.; Porell, R. N.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B. M.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Leibel, S. L.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. Cell 2020, 183 (4), 1043-1057.e15. https://doi.org/https://doi.org/10.1016/j.cell.2020.09.033.

(28) Yue, J.; Jin, W.; Yang, H.; Faulkner, J.; Song, X.; Qiu, H.; Teng, M.; Azadi, P.; Zhang, F.; Linhardt, R. J.; Wang, L. Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer. Front Mol Biosci 2021, 8. https://doi.org/10.3389/fmolb.2021.649575.

(29) Cagno; Tseligka; Jones; Tapparel. Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias? Viruses 2019, 11 (7), 596. https://doi.org/10.3390/v11070596.

(30) Stencel-Baerenwald, J. E.; Reiss, K.; Reiter, D. M.; Stehle, T.; Dermody, T. S. The Sweet Spot: Defining Virus–Sialic Acid Interactions. Nat Rev Microbiol 2014, 12 (11), 739–749. https://doi.org/10.1038/nrmicro3346.

(31) Connell, B. J.; Lortat-Jacob, H. Human Immunodeficiency Virus and Heparan Sulfate: From Attachment to Entry Inhibition. Front Immunol 2013, 4. https://doi.org/10.3389/fimmu.2013.00385.

(32) Xu, D.; Esko, J. D. Demystifying Heparan Sulfate–Protein Interactions. Annu Rev Biochem 2014, 83 (1), 129–157. https://doi.org/10.1146/annurev-biochem-060713-035314.

(33) Casalino, L.; Dommer, A. C.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S.-H.; Trifan, A.; Brace, A.; Bogetti, A. T.; Clyde, A.; Ma, H.; Lee, H.; Turilli, M.; Khalid, S.; Chong, L. T.; Simmerling, C.; Hardy, D. J.; Maia, J. D.; Phillips, J. C.; Kurth, T.; Stern, A. C.; Huang, L.; McCalpin, J. D.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. Int J High Perform Comput Appl 2021, 35 (5), 432–451. https://doi.org/10.1177/10943420211006452.

(34) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. J Chem Theory Comput 2011, 7 (2), 525–537. https://doi.org/10.1021/ct100578z.

(35) Mycroft-West, C. J.; Su, D.; Pagani, I.; Rudd, T. R.; Elli, S.; Gandhi, N. S.; Guimond, S. E.; Miller, G. J.; Meneghetti, M. C. Z.; Nader, H. B.; Li, Y.; Nunes, Q. M.; Procter, P.; Mancini, N.; Clementi, M.; Bisio, A.; Forsyth, N. R.; Ferro, V.; Turnbull, J. E.; Guerrini, M.; Fernig, D. G.; Vicenzi, E.; Yates, E. A.; Lima, M. A.; Skidmore, M. A. Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. Thromb Haemost 2020, 120 (12), 1700–1715. https://doi.org/10.1055/s-0040-1721319.

(36) Liu, L.; Chopra, P.; Li, X.; Bouwman, K. M.; Tompkins, S. M.; Wolfert, M. A.; de Vries, R. P.; Boons, G.-J. Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2. ACS Cent Sci 2021, 7 (6), 1009–1018. https://doi.org/10.1021/acscentsci.1c00010.

(37) Kim, S. Y.; Jin, W.; Sood, A.; Montgomery, D. W.; Grant, O. C.; Fuster, M. M.; Fu, L.; Dordick, J. S.; Woods, R. J.; Zhang, F.; Linhardt, R. J. Characterization of Heparin and Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Spike Glycoprotein Binding Interactions. Antiviral Res 2020, 181, 104873. https://doi.org/10.1016/j.antiviral.2020.104873.

(38) Kalra, R. S.; Kandimalla, R. Engaging the Spikes: Heparan Sulfate Facilitates SARS-CoV-2 Spike Protein Binding to ACE2 and Potentiates Viral Infection. Signal Transduct Target Ther 2021, 6 (1), 39. https://doi.org/10.1038/s41392-021-00470-1.

(39) Kim, S. H.; Kearns, F. L.; Rosenfeld, M. A.; Casalino, L.; Papanikolas, M. J.; Simmerling, C.; Amaro, R. E.; Freeman, R. GlycoGrip: Cell Surface-Inspired Universal Sensor for Betacoronaviruses. ACS Cent Sci 2021.

(40) Schuurs, Z. P.; Hammond, E.; Elli, S.; Rudd, T. R.; Mycroft-West, C. J.; Lima, M. A.; Skidmore, M. A.; Karlsson, R.; Chen, Y.-H.; Bagdonaite, I.; Yang, Z.; Ahmed, Y. A.; Richard, D. J.; Turnbull, J.; Ferro, V.; Coombe, D. R.; Gandhi, N. S. Evidence of a Putative Glycosaminoglycan Binding Site on the Glycosylated SARS-CoV-2 Spike Protein N-Terminal Domain. Comput Struct Biotechnol J 2021, 19, 2806–2818. https://doi.org/10.1016/j.csbj.2021.05.002.

(41) Tandon, R.; Sharp, J. S.; Zhang, F.; Pomin, V. H.; Ashpole, N. M.; Mitra, D.; McCandless, M. G.; Jin, W.; Liu, H.; Sharma, P.; Linhardt, R. J. Effective Inhibition of SARS-CoV-2 Entry by Heparin and Enoxaparin Derivatives. J Virol 2020, 95 (3). https://doi.org/10.1128/JVI.01987-20.

(42) Zhang, Q.; Chen, C. Z.; Swaroop, M.; Xu, M.; Wang, L.; Lee, J.; Wang, A. Q.; Pradhan, M.; Hagen, N.; Chen, L.; Shen, M.; Luo, Z.; Xu, X.; Xu, Y.; Huang, W.; Zheng, W.; Ye, Y. Heparan Sulfate Assists SARS-CoV-2 in Cell Entry and Can Be Targeted by Approved Drugs in Vitro. Cell Discov 2020, 6 (1), 80. https://doi.org/10.1038/s41421-020-00222-5.

(43) Paiardi, G.; Richter, S.; Oreste, P.; Urbinati, C.; Rusnati, M.; Wade, R. C. The Binding of Heparin to Spike Glycoprotein Inhibits SARS-CoV-2 Infection by Three Mechanisms. Journal of Biological Chemistry 2022, 298 (2), 101507. https://doi.org/10.1016/j.jbc.2021.101507.

(44) Milewska, A.; Nowak, P.; Owczarek, K.; Szczepanski, A.; Zarebski, M.; Hoang, A.; Berniak, K.; Wojarski, J.; Zeglen, S.; Baster, Z.; Rajfur, Z.; Pyrc, K. Entry of Human Coronavirus NL63 into the Cell. J Virol 2018, 92 (3). https://doi.org/10.1128/JVI.01933-17.

(45) Lang, J.; Yang, N.; Deng, J.; Liu, K.; Yang, P.; Zhang, G.; Jiang, C. Inhibition of SARS Pseudovirus Cell Entry by Lactoferrin Binding to Heparan Sulfate Proteoglycans. PLoS One 2011, 6 (8), e23710. https://doi.org/10.1371/journal.pone.0023710.

(46) Pascarella, S.; Ciccozzi, M.; Bianchi, M.; Benvenuto, D.; Cauda, R.; Cassone, A. The Electrostatic Potential of the Omicron Variant Spike Is Higher than in Delta and Delta-plus Variants: A Hint to Higher Transmissibility? J Med Virol 2022, 94 (4), 1277–1280. https://doi.org/10.1002/jmv.27528.

(47) Gan, H. H.; Zinno, J.; Piano, F.; Gunsalus, K. C. Omicron Spike Protein Has a Positive Electrostatic Surface That Promotes ACE2 Recognition and Antibody Escape. Frontiers in Virology 2022, 2. https://doi.org/10.3389/fviro.2022.894531.

(48) Nie, C.; Sahoo, A. K.; Netz, R. R.; Herrmann, A.; Ballauff, M.; Haag, R. Charge Matters: Mutations in Omicron Variant Favor Binding to Cells. ChemBioChem 2022, 23 (6). https://doi.org/10.1002/cbic.202100681.

(49) Kearns, F. L.; Sandoval, D. R.; Casalino, L.; Clausen, T. M.; Rosenfeld, M. A.; Spliid, C. B.; Amaro, R. E.; Esko, J. D. Spike-Heparan Sulfate Interactions in SARS-CoV-2 Infection. Curr Opin Struct Biol 2022, 76, 102439. https://doi.org/10.1016/j.sbi.2022.102439.

(50) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. Nature 2020, 581 (7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5.

(51) 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 (1979) 2020, 367 (6485), 1444–1448. https://doi.org/10.1126/science.abb2762.

(52) Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; Brindley, M. A.; Lewis, N. E.; Tiemeyer, M.; Chen, B.; Woods, R. J.; Wells, L. Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. Cell Host Microbe 2020, 28 (4), 586-601.e6. https://doi.org/10.1016/j.chom.2020.08.004.

(53) Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection. Biophys J 2021, 120 (6), 1072–1084. https://doi.org/10.1016/j.bpj.2020.10.036.

(54) Ozono, S.; Zhang, Y.; Ode, H.; Sano, K.; Tan, T. S.; Imai, K.; Miyoshi, K.; Kishigami, S.; Ueno, T.; Iwatani, Y.; Suzuki, T.; Tokunaga, K. SARS-CoV-2 D614G Spike Mutation Increases Entry Efficiency with Enhanced ACE2-Binding Affinity. Nat Commun 2021, 12 (1), 848. https://doi.org/10.1038/s41467-021-21118-2.

(55) Verdecchia, P.; Cavallini, C.; Spanevello, A.; Angeli, F. The Pivotal Link between ACE2 Deficiency and SARS-CoV-2 Infection. Eur J Intern Med 2020, 76, 14–20. https://doi.org/10.1016/j.ejim.2020.04.037.

(56) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent Sci 2020, 6 (10), 1722–1734. https://doi.org/10.1021/acscentsci.0c01056.

(57) Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. Nat Chem 2021, 2021.02.15.431212. https://doi.org/10.1038/s41557-021-00758-3.

(58) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 2020, 181 (2), 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

(59) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. Science (1979) 2020, 367 (6483), 1260 LP – 1263. https://doi.org/10.1126/science.abb2507.

(60) Rahbar Saadat, Y.; Hosseiniyan Khatibi, S. M.; Zununi Vahed, S.; Ardalan, M. Host Serine Proteases: A Potential Targeted Therapy for COVID-19 and Influenza. Front Mol Biosci 2021, 8. https://doi.org/10.3389/fmolb.2021.725528.

(61) Sasaki, M.; Uemura, K.; Sato, A.; Toba, S.; Sanaki, T.; Maenaka, K.; Hall, W. W.; Orba, Y.; Sawa, H. SARS-CoV-2 Variants with Mutations at the S1/S2 Cleavage Site Are Generated in Vitro during Propagation in TMPRSS2-Deficient Cells. PLoS Pathog 2021, 17 (1), e1009233. https://doi.org/10.1371/journal.ppat.1009233.

(62) Papa, G.; Mallery, D. L.; Albecka, A.; Welch, L. G.; Cattin-Ortolá, J.; Luptak, J.; Paul, D.; McMahon, H. T.; Goodfellow, I. G.; Carter, A.; Munro, S.; James, L. C. Furin Cleavage of SARS-CoV-2 Spike Promotes but Is Not Essential for Infection and Cell-Cell Fusion. PLoS Pathog 2021, 17 (1), e1009246. https://doi.org/10.1371/journal.ppat.1009246.

(63) Essalmani, R.; Jain, J.; Susan-Resiga, D.; Andréo, U.; Evagelidis, A.; Derbali, R. M.; Huynh, D. N.; Dallaire, F.; Laporte, M.; Delpal, A.; Sutto-Ortiz, P.; Coutard, B.; Mapa, C.; Wilcoxen, K.; Decroly, E.; NQ Pham, T.; Cohen, É. A.; Seidah, N. G. Distinctive Roles of Furin and TMPRSS2 in SARS-CoV-2 Infectivity. J Virol 2022, 96 (8). https://doi.org/10.1128/jvi.00128-22.

(64) Bestle, D.; Heindl, M. R.; Limburg, H.; van Lam van, T.; Pilgram, O.; Moulton, H.; Stein, D. A.; Hardes, K.; Eickmann, M.; Dolnik, O.; Rohde, C.; Klenk, H.-D.; Garten, W.; Steinmetzer, T.; Böttcher-Friebertshäuser, E. TMPRSS2 and Furin Are Both Essential for Proteolytic Activation of SARS-CoV-2 in Human Airway Cells. Life Sci Alliance 2020, 3 (9), e202000786. https://doi.org/10.26508/lsa.202000786.

(65) Koppisetti, R. K.; Fulcher, Y. G.; van Doren, S. R. Fusion Peptide of SARS-CoV-2 Spike Rearranges into a Wedge Inserted in Bilayered Micelles. J Am Chem Soc 2021, 143 (33), 13205–13211. https://doi.org/10.1021/jacs.1c05435.

(66) Jackson, C. B.; Farzan, M.; Chen, B.; Choe, H. Mechanisms of SARS-CoV-2 Entry into Cells. Nat Rev Mol Cell Biol 2022, 23 (1), 3–20. https://doi.org/10.1038/s41580-021-00418-x.

(67) Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell Entry Mechanisms of SARS-CoV-2. Proceedings of the National Academy of Sciences 2020, 117 (21), 11727–11734. https://doi.org/10.1073/pnas.2003138117.

(68) Meng, B.; Abdullahi, A.; Ferreira, I. A. T. M.; Goonawardane, N.; Saito, A.; Kimura, I.; Yamasoba, D.; Gerber, P. P.; Fatihi, S.; Rathore, S.; Zepeda, S. K.; Papa, G.; Kemp, S. A.; Ikeda, T.; Toyoda, M.; Tan, T. S.; Kuramochi, J.; Mitsunaga, S.; Ueno, T.; Shirakawa, K.; Takaori-Kondo, A.; Brevini, T.; Mallery, D. L.; Charles, O. J.; Baker, S.; Dougan, G.; Hess, C.; Kingston, N.; Lehner, P. J.; Lyons, P. A.; Matheson, N. J.; Ouwehand, W. H.; Saunders, C.; Summers, C.; Thaventhiran, J. E. D.; Toshner, M.; Weekes, M. P.; Maxwell, P.; Shaw, A.; Bucke, A.; Calder, J.; Canna, L.; Domingo, J.; Elmer, A.; Fuller, S.; Harris, J.; Hewitt, S.; Kennet, J.; Jose, S.; Kourampa, J.; Meadows, A.; O'Brien, C.; Price, J.; Publico, C.; Rastall, R.; Ribeiro, C.; Rowlands, J.; Ruffolo, V.; Tordesillas, H.; Bullman, B.; Dunmore, B. J.; Gräf, S.; Hodgson, J.; Huang, C.; Hunter, K.; Jones, E.; Legchenko, E.; Matara, C.; Martin, J.; Mescia, F.; O'Donnell, C.; Pointon, L.; Shih, J.; Sutcliffe, R.; Tilly, T.; Treacy, C.; Tong, Z.; Wood, J.; Wylot, M.; Betancourt, A.; Bower, G.; Cossetti, C.; de Sa, A.; Epping, M.; Fawke, S.; Gleadall, N.; Grenfell, R.; Hinch, A.; Jackson, S.; Jarvis, I.; Krishna, B.; Nice, F.; Omarjee, O.; Perera, M.; Potts, M.; Richoz, N.; Romashova, V.; Stefanucci, L.; Strezlecki, M.; Turner, L.; de Bie, E. M. D. D.; Bunclark, K.; Josipovic, M.; Mackay, M.; Butcher, H.; Caputo, D.; Chandler, M.; Chinnery, P.; Clapham-Riley, D.; Dewhurst, E.; Fernandez, C.; Furlong, A.; Graves, B.; Gray, J.; Hein, S.; Ivers, T.; le Gresley, E.; Linger, R.; Kasanicki, M.; King, R.; Kingston, N.; Meloy, S.; Moulton, A.; Muldoon, F.; Ovington, N.; Papadia, S.; Penkett, C. J.; Phelan, I.; Ranganath, V.; Paraschiv, R.; Sage, A.; Sambrook, J.; Scholtes, I.; Schon, K.; Stark, H.; Stirrups, K. E.; Townsend, P.; Walker, N.; Webster, J.; Butlertanaka, E. P.; Tanaka, Y. L.;

Ito, J.; Uriu, K.; Kosugi, Y.; Suganami, M.; Oide, A.; Yokoyama, M.; Chiba, M.; Motozono, C.; Nasser, H.; Shimizu, R.; Kitazato, K.; Hasebe, H.; Irie, T.; Nakagawa, S.; Wu, J.; Takahashi, M.; Fukuhara, T.; Shimizu, K.; Tsushima, K.; Kubo, H.; Kazuma, Y.; Nomura, R.; Horisawa, Y.; Nagata, K.; Kawai, Y.; Yanagida, Y.; Tashiro, Y.; Tokunaga, K.; Ozono, S.; Kawabata, R.; Morizako, N.; Sadamasu, K.; Asakura, H.; Nagashima, M.; Yoshimura, K.; Cárdenas, P.; Muñoz, E.; Barragan, V.; Márquez, S.; Prado-Vivar, B.; Becerra-Wong, M.; Caravajal, M.; Trueba, G.; Rojas-Silva, P.; Grunauer, M.; Gutierrez, B.; Guadalupe, J. J.; Fernández-Cadena, J. C.; Andrade-Molina, D.; Baldeon, M.; Pinos, A.; Bowen, J. E.; Joshi, A.; Walls, A. C.; Jackson, L.; Martin, D.; Smith, K. G. C.; Bradley, J.; Briggs, J. A. G.; Choi, J.; Madissoon, E.; Meyer, K. B.; Mlcochova, P.; Ceron-Gutierrez, L.; Doffinger, R.; Teichmann, S. A.; Fisher, A. J.; Pizzuto, M. S.; de Marco, A.; Corti, D.; Hosmillo, M.; Lee, J. H.; James, L. C.; Thukral, L.; Veesler, D.; Sigal, A.; Sampaziotis, F.; Goodfellow, I. G.; Matheson, N. J.; Sato, K.; Gupta, R. K. Altered TMPRSS2 Usage by SARS-CoV-2 Omicron Impacts Infectivity and Fusogenicity. Nature 2022, 603 (7902), 706–714. https://doi.org/10.1038/s41586-022-04474-x.

(69) Willett, B. J.; Grove, J.; MacLean, O. A.; Wilkie, C.; Logan, N.; de Lorenzo, G.; Furnon, W.; Scott, S.; Manali, M.; Szemiel, A.; Ashraf, S.; Vink, E.; Harvey, W. T.; Davis, C.; Orton, R.; Hughes, J.; Holland, P.; Silva, V.; Pascall, D.; Puxty, K.; da Silva Filipe, A.; Yebra, G.; Murcia, P. R.; Patel, A. H.; The COVID-19 Genomics UK (COG-UK) Consortium; Haughney, J.; Robertson, D. L.; Palmarini, M.; Ray, S.; Thomson, E. C. The Hyper-Transmissible SARS-CoV-2 Omicron Variant Exhibits Significant Antigenic Change, Vaccine Escape and a Switch in Cell Entry Mechanism. medRxiv 2022.

(70) Peacock, T. P.; Brown, J. C.; Zhou, J.; Thakur, N.; Sukhova, K.; Newman, J.; Kugathasan, R.; Yan, A. W. C.; Furnon, W.; de Lorenzo, G.; Cowton, V. M.; Reuss, D.; Moshe, M.; Quantrill, J. L.; Platt, O. K.; Kaforou, M.; Patel, A. H.; Palmarini, M.; Bailey, D.; Barclay, W. S. The Altered Entry Pathway and Antigenic Distance of the SARS-CoV-2 Omicron Variant Map to Separate Domains of Spike Protein. bioRxiv 2022, 2021.12.31.474653. https://doi.org/10.1101/2021.12.31.474653.

(71) Wang, X.; Bie, L.; Gao, J. Structural Insights into the Cofactor Role of Heparin/Heparan Sulfate in Binding between the SARS-CoV-2 Spike Protein and Host Angiotensin-Converting Enzyme II. J Chem Inf Model 2022, 62 (3), 656–667. https://doi.org/10.1021/acs.jcim.1c01484.

(72) Han, P.; Li, L.; Liu, S.; Wang, Q.; Zhang, D.; Xu, Z.; Han, P.; Li, X.; Peng, Q.; Su, C.; Huang, B.; Li, D.; Zhang, R.; Tian, M.; Fu, L.; Gao, Y.; Zhao, X.; Liu, K.; Qi, J.; Gao, G. F.; Wang, P. Receptor Binding and Complex Structures of Human ACE2 to Spike RBD from Omicron and Delta SARS-CoV-2. Cell 2022, 185 (4), 630-640.e10. https://doi.org/10.1016/j.cell.2022.01.001.

(73) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. J Comput Chem 2010, 31 (2), 455–461. https://doi.org/https://doi.org/10.1002/jcc.21334.

(74) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J Comput Chem 2009, 30 (16), 2785–2791. https://doi.org/https://doi.org/10.1002/jcc.21256.

(75) Schrodinger Release 2022-3. Induced Fit Docking Protocol. Schr{\"o}dinger, LLC: New York, NY 2022.

(76) Sherman, W.; Beard, H. S.; Farid, R. Use of an Induced Fit Receptor Structure in Virtual Screening. Chem Biol Drug Des 2006, 67 (1), 83–84. https://doi.org/https://doi.org/10.1111/j.1747-0285.2005.00327.x.

(77) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. J Med Chem 2006, 49 (2), 534–553. https://doi.org/10.1021/jm050540c.

(78) Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. New Insights about HERG Blockade Obtained from Protein Modeling, Potential Energy Mapping, and Docking Studies. Bioorg Med Chem 2006, 14 (9), 3160–3173. https://doi.org/10.1016/j.bmc.2005.12.032.

(79) Schrödinger Release 2021-3. Glide. Schrödinger, LLC: New York, NY 2021.

(80) Zimmer, C.; Corum, J. Coronavirus in a Tiny Drop. New York Times. 2021.

(81) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. Science (1979) 2020, 369 (6501), 330–333. https://doi.org/10.1126/science.abb9983.

(82) Huber, G. A.; McCammon, J. A. Brownian Dynamics Simulations of Biological Molecules. Trends Chem 2019, 1 (8), 727–738. https://doi.org/10.1016/j.trechm.2019.07.008.

(83) Huber, G. A.; McCammon, J. A. Browndye: A Software Package for Brownian Dynamics. Comput Phys Commun 2010, 181 (11), 1896–1905. https://doi.org/10.1016/j.cpc.2010.07.022.

(84) Chavanis, P.-H. The Generalized Stochastic Smoluchowski Equation. Entropy 2019, 21 (10), 1006. https://doi.org/10.3390/e21101006.

(85) Mannar, D.; Saville, J. W.; Zhu, X.; Srivastava, S. S.; Berezuk, A. M.; Tuttle, K. S.; Marquez, A. C.; Sekirov, I.; Subramaniam, S. SARS-CoV-2 Omicron Variant: Antibody Evasion and Cryo-EM Structure of Spike Protein–ACE2 Complex. Science (1979) 2022, 375 (6582), 760–764. https://doi.org/10.1126/science.abn7760.

(86) Xiao, T.; Lu, J.; Zhang, J.; Johnson, R. I.; McKay, L. G. A.; Storm, N.; Lavine, C. L.; Peng, H.; Cai, Y.; Rits-Volloch, S.; Lu, S.; Quinlan, B. D.; Farzan, M.; Seaman, M. S.; Griffiths, A.; Chen, B. A Trimeric Human Angiotensin-Converting Enzyme 2 as an Anti-SARS-CoV-2 Agent. Nat Struct Mol Biol 2021, 28 (2), 202–209. https://doi.org/10.1038/s41594-020-00549-3.

(87) Yoshizato, K.; Taira, T.; Sato-Matsubara, M.; Sekiguchi, S.; Yabunaka, Y.; Kira, Y.; Ohashi, T.; Daikoku, A.; Ofusa, K.; Kadono, C.; Oikawa, D.; Matsubara, T.; Nakagama, Y.; Kido, Y.; Tokunaga, F.; Ikeda, K.; Kaneko, A.; Kawada, N. Cloaking the ACE2 Receptor with Salivary Cationic Proteins Inhibits SARS-CoV-2 Entry. The Journal of Biochemistry 2022, 172 (4), 205–216. https://doi.org/10.1093/jb/mvac054.

(88) Amaro, R. E.; Mulholland, A. J. A Community Letter Regarding Sharing Biomolecular Simulation Data for COVID-19. J Chem Inf Model 2020, 60 (6), 2653–2656. https://doi.org/10.1021/acs.jcim.0c00319.

(89) Amaro, R. E.; Mulholland, A. J. Biomolecular Simulations in the Time of COVID-19, and After. Comput Sci Eng 2020, 22 (6), 30–36. https://doi.org/10.1109/MCSE.2020.3024155.

(90) Mulholland, A. J.; Amaro, R. E. COVID19 - Computational Chemists Meet the Moment. J Chem Inf Model 2020, 60 (12), 5724–5726. https://doi.org/10.1021/acs.jcim.0c01395.

4.7.2 Supplementary Text References

(1) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 2020, 181 (2), 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

(2) Bangaru, S.; Ozorowski, G.; Turner, H. L.; Antanasijevic, A.; Huang, D.; Wang, X.; Torres, J. L.; Diedrich, J. K.; Tian, J.-H.; Portnoff, A. D.; Patel, N.; Massare, M. J.; Yates, J. R.; Nemazee, D.; Paulson, J. C.; Glenn, G.; Smith, G.; Ward, A. B. Structural Analysis of Full-Length SARS-CoV-2 Spike Protein from an Advanced Vaccine Candidate. Science (1979) 2020, 370 (6520), 1089 LP – 1094. https://doi.org/10.1126/science.abe1502.

(3) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. Science (1979) 2020, 367 (6483), 1260 LP – 1263. https://doi.org/10.1126/science.abb2507.

(4) 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 (1979) 2020, 367 (6485), 1444–1448. https://doi.org/10.1126/science.abb2762.

(5) McCallum, M.; Walls, A. C.; Sprouse, K. R.; Bowen, J. E.; Rosen, L.; Dang, H. v; DeMarco, A.; Franko, N.; Tilles, S. W.; Logue, J.; Miranda, M. C.; Ahlrichs, M.; Carter, L.; Snell, G.; Pizzuto, M. S.; Chu, H. Y.; van Voorhis, W. C.; Corti, D.; Veesler, D. Molecular Basis of Immune Evasion by the Delta and Kappa SARS-CoV-2 Variants. Science (1979) 2021, 0 (0), eabl8506. https://doi.org/10.1126/science.abl8506.

(6) Gobeil, S. M.-C.; Henderson, R.; Stalls, V.; Janowska, K.; Huang, X.; May, A.; Speakman, M.; Beaudoin, E.; Manne, K.; Li, D.; Parks, R.; Barr, M.; Deyton, M.; Martin, M.; Mansouri, K.; Edwards, R. J.; Eaton, A.; Montefiori, D. C.; Sempowski, G. D.; Saunders, K. O.; Wiehe, K.; Williams, W.; Korber, B.; Haynes, B. F.;

Acharya, P. Structural Diversity of the SARS-CoV-2 Omicron Spike. Mol Cell 2022, 82 (11), 2050-2068.e6. https://doi.org/10.1016/j.molcel.2022.03.028.

(7) Tortorici, M. A.; Beltramello, M.; Lempp, F. A.; Pinto, D.; Dang, H. v.; Rosen, L. E.; McCallum, M.; Bowen, J.; Minola, A.; Jaconi, S.; Zatta, F.; de Marco, A.; Guarino, B.; Bianchi, S.; Lauron, E. J.; Tucker, H.; Zhou, J.; Peter, A.; Havenar-Daughton, C.; Wojcechowskyj, J. A.; Case, J. B.; Chen, R. E.; Kaiser, H.; Montiel-Ruiz, M.; Meury, M.; Czudnochowski, N.; Spreafico, R.; Dillen, J.; Ng, C.; Sprugasci, N.; Culap, K.; Benigni, F.; Abdelnabi, R.; Foo, S.-Y. C.; Schmid, M. A.; Cameroni, E.; Riva, A.; Gabrieli, A.; Galli, M.; Pizzuto, M. S.; Neyts, J.; Diamond, M. S.; Virgin, H. W.; Snell, G.; Corti, D.; Fink, K.; Veesler, D. Ultrapotent Human Antibodies Protect against SARS-CoV-2 Challenge via Multiple Mechanisms. Science (1979) 2020, 370 (6519), 950–957. https://doi.org/10.1126/science.abe3354.

(8) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent Sci 2020, 6 (10), 1722–1734. https://doi.org/10.1021/acscentsci.0c01056.

(9) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. Science (1979) 2020, 369 (6501), 330–333. https://doi.org/10.1126/science.abb9983.

(10) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. J Chem Theory Comput 2011, 7 (2), 525–537. https://doi.org/10.1021/ct100578z.

(11) Schrödinger Release 2021-3. Protein Preparation Wizard. Schrödinger, LLC: New York, NY 2021.

(12) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. J Mol Graph 1996, 14 (1), 33–38. https://doi.org/https://doi.org/10.1016/0263-7855(96)00018-5.

(13) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. J Comput Chem 2005, 26 (16), 1781–1802. https://doi.org/10.1002/jcc.20289.

(14) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. v; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. v; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. J Chem Phys 2020, 153 (4), 44130. https://doi.org/10.1063/5.0014475.

(15) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. J Comput Chem 2013, 34 (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(16) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. Nat Methods 2017, 14 (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(17) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. J Chem Theory Comput 2009, 5 (9), 2353–2370. https://doi.org/10.1021/ct900242e.

(18) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. J Comput Chem 2010, 31 (2), 455–461. https://doi.org/https://doi.org/10.1002/jcc.21334.

(19) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J Comput Chem 2009, 30 (16), 2785–2791. https://doi.org/https://doi.org/10.1002/jcc.21256.

(20) 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 (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

(21) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations; 2016; pp 98–105. https://doi.org/10.25080/Majora-629e541a-00e.

(22) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research 2011, 12 (85), 2825–2830.

(23) Kim, S. H.; Kearns, F. L.; Rosenfeld, M. A.; Casalino, L.; Papanikolas, M. J.; Simmerling, C.; Amaro, R. E.; Freeman, R. GlycoGrip: Cell Surface-Inspired Universal Sensor for Betacoronaviruses. ACS Cent Sci 2021.

(24) Shrake, A.; Rupley, J. A. Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin. J Mol Biol 1973, 79 (2), 351–371. https://doi.org/10.1016/0022-2836(73)90011-9.

(25) Schrodinger Release 2022-3. Induced Fit Docking Protocol. Schr{\"o}dinger, LLC: New York, NY 2022.

(26) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. J Chem Theory Comput 2021, 17 (7), 4291–4300. https://doi.org/10.1021/acs.jctc.1c00302.

(27) Schrödinger Release 2021-3. LigPrep. Schrödinger, LLC: New York, NY 2021.

(28) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. Nucleic Acids Res 2004, 32 (Web Server), W665–W667. https://doi.org/10.1093/nar/gkh381.

(29) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. Nucleic Acids Res 2007, 35 (Web Server), W522–W525. https://doi.org/10.1093/nar/gkm276.

(30) Huber, G. A.; McCammon, J. A. Browndye: A Software Package for Brownian Dynamics. Comput Phys Commun 2010, 181 (11), 1896–1905. https://doi.org/10.1016/j.cpc.2010.07.022.

(31) Huber, G. A.; McCammon, J. A. Brownian Dynamics Simulations of Biological Molecules. Trends Chem 2019, 1 (8), 727–738. https://doi.org/10.1016/j.trechm.2019.07.008.

(32) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. Protein Science 2018, 27 (1), 112–128. https://doi.org/https://doi.org/10.1002/pro.3280.

(33) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. Proceedings of the National Academy of Sciences 2001, 98 (18), 10037 LP – 10041. https://doi.org/10.1073/pnas.181342398.

(34) Holst, M. Adaptive Numerical Treatment of Elliptic Systems on Manifolds. Adv Comput Math 2001, 15 (1), 139–191. https://doi.org/10.1023/A:1014246117321.

(35) Holst, M. J.; Saied, F. Numerical Solution of the Nonlinear Poisson–Boltzmann Equation: Developing More Robust and Efficient Methods. J Comput Chem 1995, 16 (3), 337–364. https://doi.org/https://doi.org/10.1002/jcc.540160308.

(36) Bank, R. E.; Holst, M. A New Paradigm for Parallel Adaptive Meshing Algorithms. SIAM Review 2003, 45 (2), 291–323. https://doi.org/10.1137/S003614450342061.

(37) Holst, M.; Saied, F. Multigrid Solution of the Poisson—Boltzmann Equation. J Comput Chem 1993, 14 (1), 105–113. https://doi.org/https://doi.org/10.1002/jcc.540140114.

(38) Chavanis, P.-H. The Generalized Stochastic Smoluchowski Equation. Entropy 2019, 21 (10), 1006. https://doi.org/10.3390/e21101006.

(39) Votapka, L. W.; Amaro, R. E. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. PLoS Comput Biol 2015, 11 (10), e1004381. https://doi.org/10.1371/journal.pcbi.1004381.

(40) Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection. Biophys J 2021, 120 (6), 1072–1084. https://doi.org/10.1016/j.bpj.2020.10.036.

(41) Lee, J.-H.; Choi, M.; Jung, Y.; Lee, S. K.; Lee, C.-S.; Kim, J.; Kim, J.; Kim, N. H.; Kim, B.-T.; Kim, H. G. A Novel Rapid Detection for SARS-CoV-2 Spike 1 Antigens Using Human Angiotensin Converting Enzyme 2 (ACE2). Biosens Bioelectron 2021, 171, 112715. https://doi.org/10.1016/j.bios.2020.112715.

(42) Baker, A. N.; Richards, S.-J.; Guy, C. S.; Congdon, T. R.; Hasan, M.; Zwetsloot, A. J.; Gallo, A.; Lewandowski, J. R.; Stansfeld, P. J.; Straube, A.; Walker, M.; Chessa, S.; Pergolizzi, G.; Dedola, S.; Field, R. A.; Gibson, M. I. The SARS-COV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. ACS Cent Sci 2020, 6 (11), 2046–2052. https://doi.org/10.1021/acscentsci.0c00855.

(43) Lee, J.-H.; Lee, Y.; Lee, S. K.; Kim, J.; Lee, C.-S.; Kim, N. H.; Kim, H. G. Versatile Role of ACE2-Based Biosensors for Detection of SARS-CoV-2 Variants and Neutralizing Antibodies. Biosens Bioelectron 2022, 203, 114034. https://doi.org/10.1016/j.bios.2022.114034.

(44) Casalino, L.; Dommer, A. C.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S.-H.; Trifan, A.; Brace, A.; Bogetti, A. T.; Clyde, A.; Ma, H.; Lee, H.; Turilli, M.; Khalid, S.; Chong, L. T.; Simmerling, C.; Hardy, D. J.; Maia, J. D.; Phillips, J. C.; Kurth, T.; Stern, A. C.; Huang, L.; McCalpin, J. D.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. Int J High Perform Comput Appl 2021, 35 (5), 432–451. https://doi.org/10.1177/10943420211006452.

Chapter 5 - #COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol

5.1 **Abstract**

We seek to completely revise current models of airborne transmission of respiratory viruses by providing never-before-seen atomic- level views of the SARS-CoV-2 virus within a respiratory aerosol. Our work dramatically extends the capabilities of multiscale computational microscopy to address the significant gaps that exist in current experimental methods, which are limited in their ability to interrogate aerosols at the atomic/molecular level and thus obscure our understanding of airborne transmission. We demonstrate how our integrated data-driven platform provides a new way of exploring the composition, structure, and dynamics of aerosols and aerosolized viruses, while driving simulation method development along several important axes. We present a series of initial scientific discoveries for the SARS-CoV-2 Delta variant, noting that the full scientific impact of this work has yet to be realized.

5.2 **Justification**

We develop a novel HPC-enabled multiscale research framework to study aerosolized viruses and the full complexity of species that comprise them. We present technological and methodological advances that bridge time and length scales from electronic structure through whole aerosol particle morphology and dynamics.

5.3 **Performance Attributes**

Table 5.1 Performance Attributes

Performance Attribute	Our Submission
Category of achievement	Scalability, Time-to-solution
Type of method used	Explicit, Deep Learning
Results reported on the basis of	Whole application including I/O
Precision reported	Mixed Precision
System scale	Measured on full system
Measurement Mechanism	Hardware performance counters, Application timers, Performance Modeling

5.4 **Overview of the Problem**

Respiratory pathogens, such as SARS-CoV-2 and influenza, are the cause of significant morbidity and mortality worldwide. These respiratory pathogens are spread by virus-laden aerosols and droplets that are produced in an infected person, exhaled, and transported through the environment[1] (Figure 5.1). Medical dogma has long focused on droplets as the main transmission route for respiratory viruses, where either a person has contact with an infected surface (fomites) or direct droplet transmission by close contact with an infected individual. However, as we continue to observe with SARS-CoV-2, airborne transmission also plays a significant role in spreading disease. We know this from various super spreader events, e.g., during a choir rehearsal.[2] Intervention and mitigation decisions, such as the relative importance of surface cleaning or whether and when to wear a mask, have unfortunately hinged on a weak understanding of aerosol transmission, to the detriment of public health.

A central challenge to understanding airborne transmission has been the inability of experimental science to reliably probe the structure and dynamics of viruses once they are inside respiratory aerosol particles. Single particle experimental methods have poor resolution for smaller particles (<1 micron) and are prone to sample destruction during collection. Airborne viruses are present in low concentrations in the air and are similarly prone to viral inactivation during sampling. In addition, studies of the initial infection event, for example in the deep lung, are limited in their ability to provide a detailed understanding of the myriad of molecular interactions and dynamics taking place in situ. Altogether, these knowledge gaps hamper our

collective ability to understand mechanisms of infection and develop novel effective antivirals, as well as prevent us from developing concrete, science-driven mitigation measures (e.g., masking and ventilation protocols).

Here, we aim to reconceptualize current models of airborne transmission of respiratory viruses by providing never-before-seen views of viruses within aerosols. Our approach relies on the use of all-atom molecular dynamics (MD) simulations as a multiscale 'computational microscope.' MD simulations can synthesize multiple types of biological data (e.g., multiresolution structural datasets, glycomics, lipidomics, etc.) into cohesive, biologically 'accurate' structural models. Once created, we then approximate the model down to its many atoms, creating trajectories of its time dependent dynamics under cell-like (or in this case, aerosol-like) conditions. Critically, MD simulations are more than just 'pretty movies.' MD equations are solved in a theoretically rigorous manner, allowing us to compute experimentally testable macroscopic observables from time-averaged microscopic properties. What this means is that we can directly connect MD simulations with experiments, each vali- dating and providing testable hypotheses to the other, which is the real power of the approach. An ongoing challenge to the successful application of such methods, however, is the need for technological and methodological advances that make it possible to access length scales relevant to the study of large, biologically complex systems (spanning nanometers to one micron in size) and, correspondingly, longer timescales (microseconds to seconds).

Such challenges and opportunities manifest in the study of aerosolized viruses. Aerosols are generally defined as being less than 5 microns in diameter, able to float in the air for hours, travel significant distances (e.g., can fill a room, like cigarette smoke), and be inhaled. Fine aerosols < 1 micron in size can stay in the air for over 12 hours and are enriched with viral particles.[3,4] Our work focuses on these finer aerosols that travel deeper into the respiratory tract. Several studies provide the molecular recipes necessary to reconstitute respiratory aerosols according to their actual biologically-relevant composition.[5,6] These aerosols can contain lipids, cholesterol, albumin (protein), various mono- and di-valent salts, mucins, other surfactants, and water (Figure 5.1). Simulations of aerosolized viruses embody a novel framework for the study of aerosols: they will allow us and others to tune different species, relative humidity, ion concentrations,

etc. to match experiments that can directly and indirectly connect to and inform our simulations, as well as test hypotheses. Some of the species under study here, e.g., mucins, have not yet been structurally characterized or explored with simulations and thus the models we generate are expected to have impact beyond their roles in aerosols.

In addition to varying aerosol composition and size, the viruses themselves can be modified to reflect new variants of concern, where such mutations may affect interactions with particular species in the aerosol that might affect its structural dynamics and/or vi- ability. The virion developed here is the Delta variant (B.1.617.2 lineage) of SARS-CoV-2 (Figure 5.2) which presents a careful integration of multiple biological datasets: (1) a complete viral envelope with realistic membrane composition, (2) fully glycosylated full-length spike proteins integrating 3D structural coordinates from multiple cryoelectron microscopy (cryo-EM) studies[7–10] (3) all biologically known features (post-translational modifications, palmitoylation, etc.), (4) any other known membrane proteins (e.g., the envelope (E) and membrane (M) proteins), and (5) virion size and patterning taken directly from cryoelectron tomography (cryo-ET). Each of the individual components of the virus are built up before being integrated into the composite virion, and thus represent useful molecular-scale scientific contributions in their own right.[11,12]

Altogether in this work, we dramatically extend the capabilities of data-driven, multiscale computational microscopy to provide a new way of exploring the composition, structure, and dynamics of respiratory aerosols. While a seemingly limitless number of putative hypotheses could result from these investigations, the first set of questions we expect to answer are: How does the virus exist within a droplet of the same order of magnitude in size, without being affected by the air-water interface, which is known to destroy molecular structure[13] ? How does the biochemical composition of the droplet, including pH, affect the structural dynamics of the virus? Are there species within the aerosols that “buffer” the viral structure from damage, and are there particular conditions under which the impact of those species changes? Our simulations can also provide specific parameters that can be included in physical models of aerosols, which still assume a simple water or water-salt composition even though it is well known that such models, e.g., using kappa-Kohler theory, break down significantly as the molecular species diversify.[14]

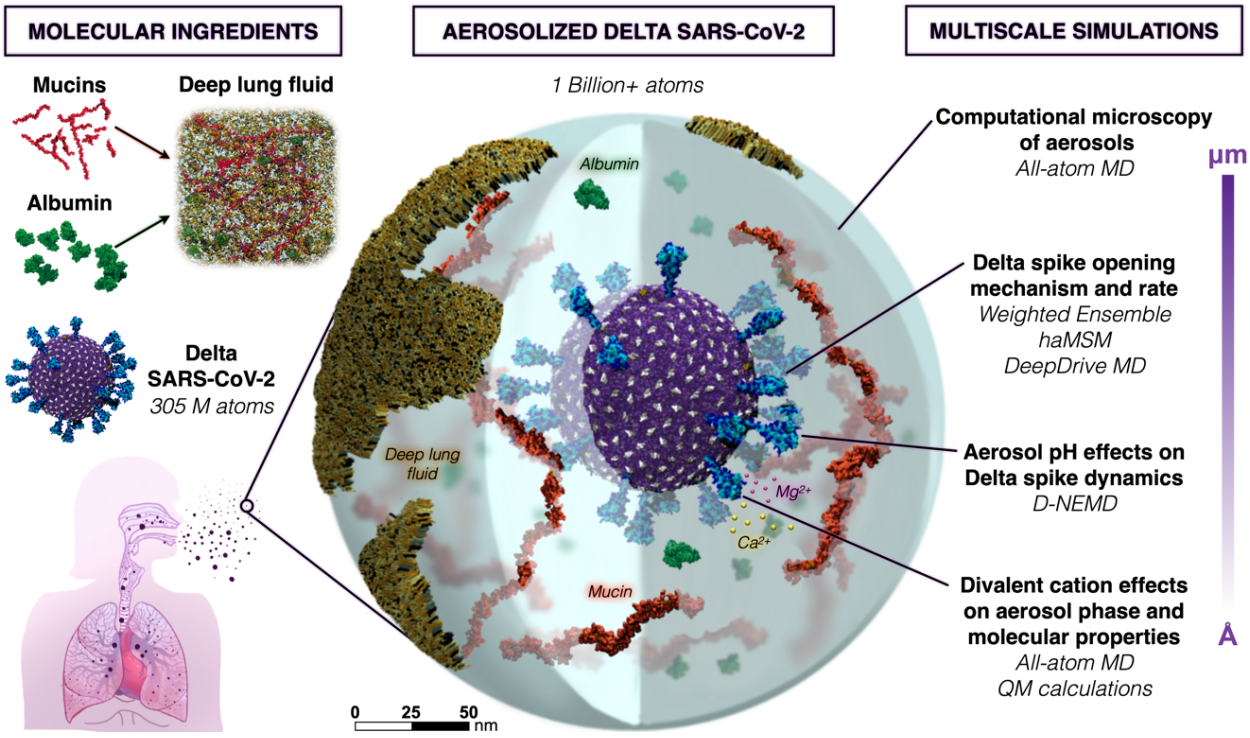


Figure 5.1: Overall schematic depicting the construction and multiscale simulations of Delta SARS-CoV-2 in a respiratory aerosol
(N.B.: The size of divalent cations has been increased for visibility.)

5.5 **Current State of the Art**

Current experimental methods are unable to directly interrogate the atomic-level structure and dynamics of viruses and other molecules within aerosols. Here we showcase computational microscopy as a powerful tool capable to overcome these significant experimental limitations. We present the major elements of our multiscale computational microscope and how they come together in an integrated manner to enable the study of aerosols across multiple scales of resolution. We demonstrate the impact such methods can bring to bear on scientific challenges that until now have been intractable and present a series of new scientific discoveries for SARS-CoV-2.

5.5.1 Parallel molecular dynamics

All-atom molecular dynamics simulation has emerged as an increasingly powerful tool for understanding the molecular mechanisms underlying biophysical behaviors in complex systems. Leading

simulation engines, NAMD[15], AMBER[16], and GROMACS[17], are broadly useful, with each providing unique strengths in terms of specific methods or capabilities as required to address a particular biological question, and in terms of their support for particular HPC hardware platforms. Within the multiscale computational microscopy platform developed here, we show how each of these different codes contributes different elements to the overall framework, oftentimes utilizing different computing modalities/architectures, while simultaneously extending on state-of-the-art for each. Structure building, simulation preparation, visualization, and post-hoc trajectory analysis are performed using VMD on both local workstations and remote HPC resources, enabling modeling of the molecular systems studied herein.[18–22] We show how further development of each of these codes, considered together within the larger-scale collective framework, enables the study of SARS-CoV-2 in a wholly novel manner, with extension to numerous other complex systems and diseases.

5.5.2 AI-enhanced WE simulations

Because the virulence of the Delta variant of SARS-CoV-2 may be partly attributable to spike protein (S) opening, it is of pressing interest to characterize the mechanism and kinetics of the process. Although S-opening in principle can be studied via conventional MD simulations, in practice the system complexity and timescales make this wholly intractable. Splitting strategies that periodically replicate promising MD trajectories, among them the weighted ensemble (WE) method[23,24], have enabled simulations of the spike opening of WT SARS-CoV-2.[12,25] WE simulations can be orders of magnitude more efficient than conventional MD in generating pathways and rate constants for rare events (e.g., protein folding[26] and binding[27]). The WESTPA software for running WE[28] is well-suited for high-performance computing with nearly perfect CPU/GPU scaling. The software is interoperable with any dynamics engine, including the GPU-accelerated AMBER dynamics engine[29] that is used here. As shown below, major upgrades to WESTPA (v. 2.0) have enabled a dramatic demonstration of spike opening in the Delta variant (Figures 5.5 and 7.6) and exponentially improved analysis of spike-opening kinetics.

The integration of AI techniques with WE can further enhance the efficiency of sampling rare events.[11,30,31] One frontier area couples unsupervised linear and non-linear dimensionality reduction methods to identify collective variables/progress coordinates in high-dimensional molecular systems.[32,33] Such methods may be well suited for analyzing the aerosolized virus. Integrating these approaches with WE simulations is advantageous in sampling the closed open transitions in the Delta S landscape (Figure 5.5) as these unsupervised AI approaches automatically stratify progress coordinates (Figure 5.5D).

5.5.3 Dynamical Non-Equilibrium MD

Aerosols rapidly acidify during flight via reactive uptake of atmospheric gases, which is likely to impact the opening/closing of the S protein.[5,34] Here, we describe the extension of dynamical non-equilibrium MD (D- NEMD)[35] to investigate pH effects on the Delta S. D-NEMD simulations[35] are emerging as a useful technique for identifying allosteric effects and communication pathways in proteins[36,37], including recently identifying effects of linoleic acid in the WT spike.[38] This approach complements equilibrium MD simulations, which provide a distribution of configurations as starting points for an ensemble of short non-equilibrium trajectories under the effect of the external perturbation. The response of the protein to the perturbation introduced can then be determined using the Kubo-Onsager by directly tracking the change in atomic positions between the equilibrium and non-equilibrium simulations at equivalent points in time.

5.5.4 OrbNet

Ca^{2+} ions are known to play a key role in mucin aggregation in epithelial tissues.[39] Our RAV simulations would be an ideal case-study to probe such complex interactions between Ca^{2+}, mucins, and the SARS-CoV-2 virion in aerosols. However, Ca^{2+} binding energies can be difficult to capture accurately due to electronic dispersion and polarization, terms which are not typically modeled in classical mechanical force fields. Quantum mechanical (QM) methods are uniquely suited to capture these subtle interactions. Thus, we set out to estimate the correlation in Ca^{2+} binding energies between CHARMM36m and quantum mechanical estimates enabled via AI with OrbNet. Calculation of energies with sufficient accuracy in biological systems can, in many cases, be adequately described with density functional theory (DFT). However, its high cost

limits the applicability of DFT in comparison to fixed charge force-fields. To capture quantum quality energetics at a fraction of the computational expense, we employ a novel approach (OrbNet) based on the featurization of molecules in terms of symmetry-adapted atomic orbitals and the use of graph neural network methods for deep-learning quantum-mechanical properties.[40] Our method outperforms existing methods in terms of its training efficiency and transferable accuracy across diverse molecular systems, opening a new pathway for replacing DFT in large-scale scientific applications such as those explored here.[41]

5.6 **Innovations Realized**

5.6.1 Construction and simulation of SARS-CoV-2 in a respiratory aerosol

Our approach to simulating the entire aerosol follows a composite framework wherein each of the individual molecular pieces is refined and simulated on its own before it is incorporated into the composite model. Simulations of each of the components are useful in their own right, and often serve as the basis for biochemical and biophysical validation and experiments.[11]

Throughout, we refer to the original circulating SARS-CoV-2 strain as “WT”, whereas all SARS-CoV-2 proteins constructed in this work represent the Delta variant (Figure 7.2) All simulated membranes reflect mammalian ER-Golgi intermediate compartment (ERGIC) mimetic lipid compositions. VMD[18,22], psfgen[42] and CHARMM- GUI[43] were used for construction and parameterization. Topologies and parameters for simulations were taken from CHARMM36m all-atom additive force fields.[44–50] NAMD was used to perform MD simulations,[15] adopting similar settings and protocols as in Casalino et al.[11] All systems underwent solvation, charge neutralization, minimization, heating, and equilibration prior to production runs. Refer to Table 5.2 for abbreviations, PBC dimensions, total number of atoms, and total equilibration times for each system of interest.

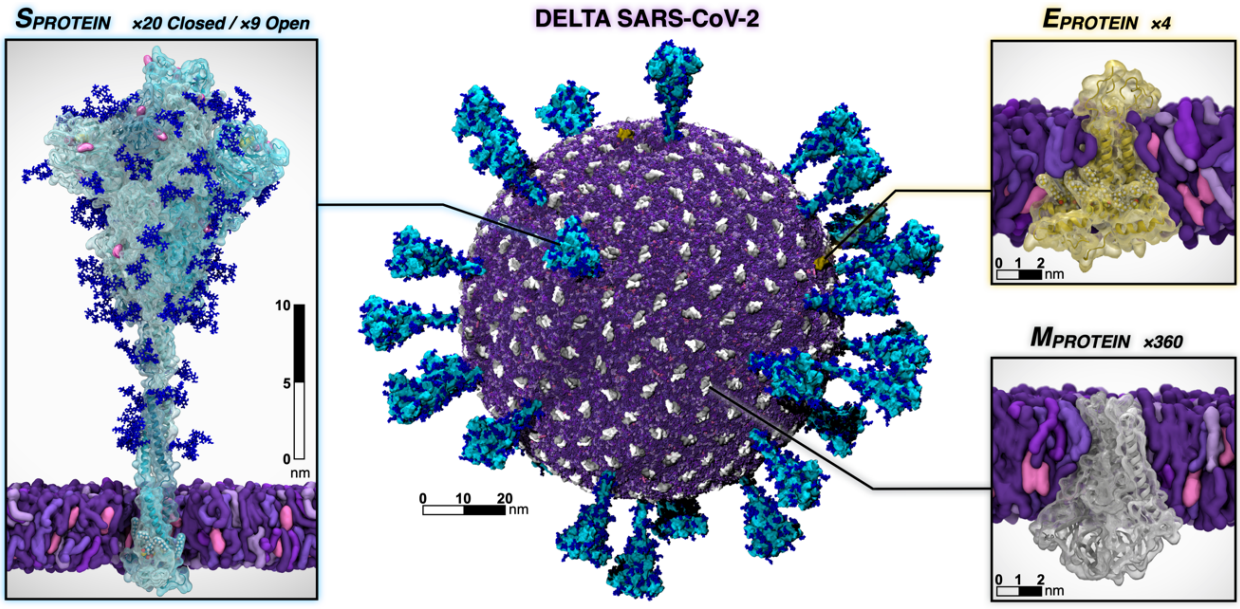


Figure 5.2: Individual protein components of the SARS-CoV-2 Delta virion
The spike is shown with the surface in cyan and with Delta's mutated residues and deletion sites highlighted in pink and yellow, respectively. Glycans attached to the spike are shown in blue. The E protein is shown in yellow and the M protein is shown in silver and white. Visualized with VMD.

5.6.1.1 Simulating the SARS-CoV-2 structural proteins.

Fully glycosylated Delta spike (S) structures in open and closed conformations were built based on WT constructs from Casalino et al.[11] with the following mutations: T19R, T95I, G142D, E156G, Δ157-158, L452R, T478K, D614G, P681R, and D950N.[8,51] Higher resolved regions were grafted from PDB 7JJI.[7] Additionally, coordinates of residues 128-167 accounting for a drastic conformational change seen in the Delta variant S (graciously made available to us by the Veesler Lab) were similarly grafted onto our constructs.[8] Finally, the S proteins were glycosylated following work by Casalino et al.[11] By incorporating the Veesler Lab's bleeding-edge structure and highly resolved regions from 7JJI, our models represent the most complete and accurate structures of the Delta S to date. The S proteins were inserted into membrane patches and equilibrated for 3x110 ns. For nonequilibrium and weighted ensemble simulations, a closed S head (SH, residues 13- 1140) was constructed by removing the stalk from the full-length closed S structure, then re-solvated, neutralized, minimized, and subsequently passed to WE and D-NEMD teams. The M protein was built from a structure graciously provided by the Feig Lab (paper in prep). The model was

inserted into a membrane patch and equilibrated for 700 ns. RMSD-based clustering was used to select a stable starting M protein conformation. From the equilibrated and clustered M structure, VMD's Mutator plugin[18] was used to incorporate the I82T mutation onto each M monomer to arrive at the Delta variant M. To construct the most complete E protein model to-date, the structure was patched together by re- solving incomplete PDBs 5X29,[52] 7K3G,[53] and 7M4R.[54] To do so, the transmembrane domain (residues 8-38) from 7K3G was aligned to the N-terminal domain (residues 1-7) and residues 39 to 68 of 5X29 and residues 69 to 75 of 7M4R by their Cα atoms. E was then inserted into a membrane patch and equilibrated for 40 ns.

5.6.1.2 Constructing the SARS-CoV-2 Delta virion.

The SARS-CoV-2 Delta virion (V) model was constructed following Casalino et al.[55] using CHARMM-GUI,[56] LipidWrapper,[57] and Blender,[58] using a 350 Å lipid bilayer with an equilibrium area per lipid of 63 Å^2 and a 100 nm diameter Blender icospherical surface mesh.[59] The resulting lipid membrane was solvated in a 1100 Å^3 waterbox and subjected to 4 rounds of equilibration and patching.[55] 360 M dimers and 4 E pentamers were then tiled onto the surface, followed by random placement of 29 full-length S proteins (9 open, 20 closed) according to experimentally observed S protein density.[60] M and E proteins were oriented with intravirion C- termini. After solvation in a 1460 Å waterbox, the complete V model tallied >305 million atoms (Table 5.2). V was equilibrated for 41 ns prior to placement in the respiratory aerosol (RA) model. The equilibrated membrane was 90 nm in diameter and remains in close structural agreement with the experimental studies.[60]

Table 5.2: Summary of all systems constructed in this work. See Figure 5.3 for illustration of aerosol construction

[a]systems	[b]Abb.	[c](Å × Å × Å)	[d]N_a	[e](ns)
[f]M dimers	M	125 × 125 × 124	164,741	700
[f]E pentamers	E	123 × 125 × 102	136,775	41
Spikes:				
[f](open)	S	206 × 200 × 410	1,692,444	330
[f](closed)	S	204 × 202 × 400	1,658,224	330
[g](closed, head)	SH	172 × 184 × 206	615,593	73μs
Mucins:				
[f]short mucin 1	m_1	123 × 104 × 72	87,076	25
[f]short mucin 2	m_2	120 × 101 × 72	82,155	25
[f]long mucin 1	m_3	810 × 104 × 115	931,778	23
[f]long mucin 2	m_4	904 × 106 × 109	997,029	15
[f]long mucin 3	m_5	860 × 111 × 113	1,040,215	18
[f]S+m1/m2+ALB	SMA	227 × 229 × 433	2,156,689	840
[f]Virion	V	1460 × 1460 × 1460	305,326,834	41
[f]Resp.Aero.+Vir.	RAV	2834 × 2820 × 2828	1,016,813,441	2.42
TOTAL FLOPS			2.4 ZFLOPS	

[a]M, E, S, SH, and V models represent SARS-CoV-2 Delta strain. [b]Abbreviations used throughout document. [c]Periodic boundary dimensions. [d]Total number of atoms. [e]Total aggregate simulation time, including heating and equilibration runs. [f]Simulated with NAMD. [g]Simulated with NAMD, AMBER, and GROMACS.

5.6.1.3 Building and simulating the respiratory aerosol.

Respiratory aerosols contain a complex mixture of chemical and biological species. We constructed a respiratory aerosol (RA) fluid based on a composition from artificial saliva and surrogate deep lung fluid recipes.[6] This recipe includes 0.7 mM DPPG, 6.5 mM DPPC, 0.3 mM cholesterol, 1.4 mM Ca^{2+}, 0.8 mM Mg^{2+}, and 142 mM Na^{+}, human serum albumin (ALB) protein, and a composition of mucins (Figure 5.3) Mucins are long polymer-like structures that are decorated by dense, heterogeneous, and complex regions of O-glycans. This work represents the first of its kind as, due to their complexity, the O-glycosylated regions of mucins have never before been constructed for molecular simulations. Two short (m1, m2, 5 nm) and three long (m3, m4, m5 55 nm) mucin models were constructed following known experimental compositions of protein and glycosylation sequences[39,61–64] with ROSETTA[65] and CHARMM GUI Glycan Modeler.[43] Mucin models (short and long) were solvated, neutralized by charge matching with Ca^{2+} ions, minimized, and equilibrated for 15-25 ns each (Table 5.2). Human serum albumin (ALB), which is also found in respiratory

aerosols, was constructed from PDB 1AO6.[66] ALB was solvated, neutralized, minimized, and equilibrated for 7ns. Equilibrated structures of ALB and the three long mucins were used in construction of the RAV with m3+m4+m5 added at 6 g/mol and ALB at 4.4 g/mol.

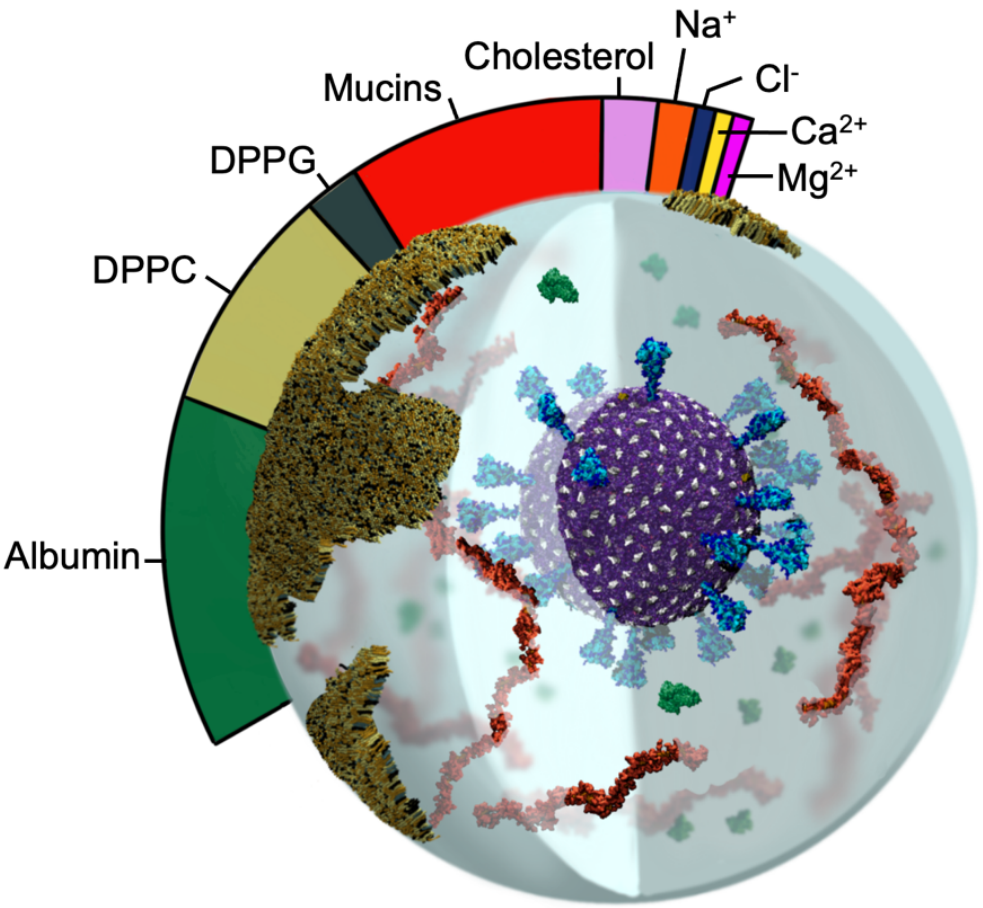


Figure 5.3: Image of RAV with relative mass ratios of RA molecular components represented in the color bar
Water content is dependent on the relative humidity of the environment and is thus omitted from the molecular ratios.

5.6.1.4 Constructing the respiratory aerosolized virion model.

A 100 nm cubic box with the RA fluid recipe specified above was built with PACKMOL,[67] minimized, equilibrated briefly on TACC Frontera, then replicated to form a 300 nm cube. The RA box was then carved into a 270 nm diameter sphere. To make space for the placement of V within the RA, a spherical selection with volume corresponding to that of the V membrane + S crown (radius 734 Å) was deleted from the center of the RA. The final equilibrated V model, including surrounding equilibrated waters and ions (733 Å radius),

was translated into the RA. Atom clashes were resolved using a 1.2 Å cutoff. Hydrogen mass repartitioning[68] was applied to the structure to improve performance. The simulation box was increased to 2800 Å per side to provide a 100 Å vacuum atmospheric buffer. The RAV simulation was conducted in an NVT ensemble with a 4 fs timestep. After minimizing, the RAV was heated to 298 K with 0.1 kcal/mol Å2 restraints on the viral lipid headgroups, then equilibrated for 1.5 ns. Finally, a cross-section of the RAV model – including and open S, m1/m2, and ALB (called the SMA system) – was constructed with PACKMOL to closely observe atomic scale interactions within the RAV model (Figure 5.4).

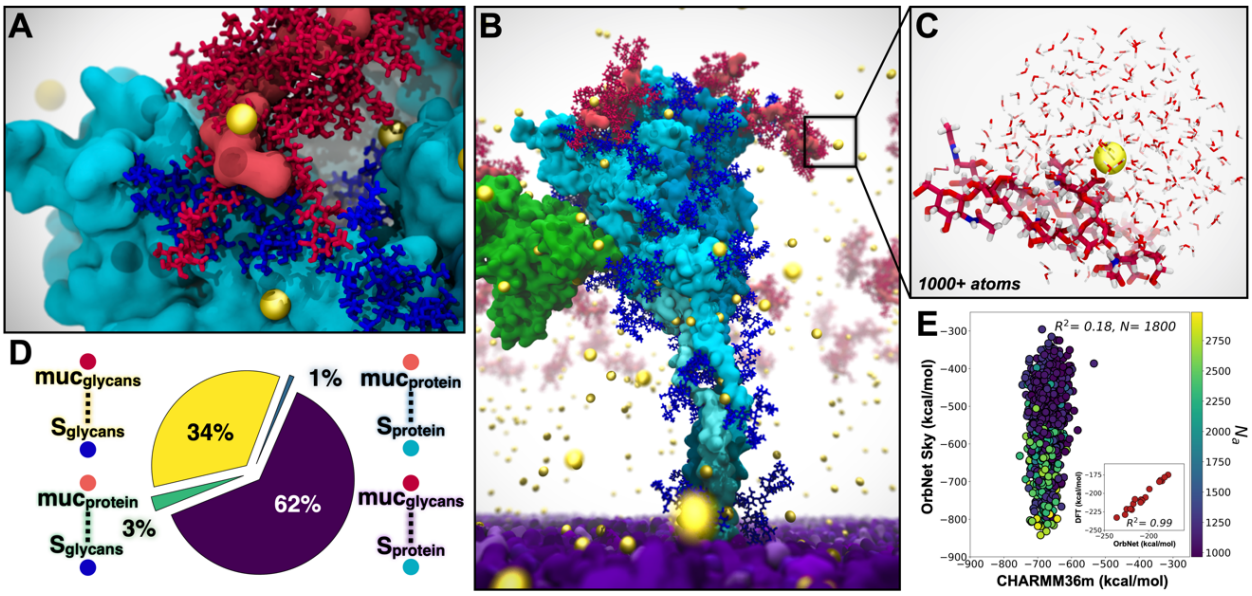


Figure 5.4: SMA system captured with multiscale modeling from classical MD to AI-enabled quantum mechanics

For all panels: S protein shown in cyan, S glycans in blue, m1/m2 shown in red, ALB in orange, Ca2+ in yellow spheres, viral membrane in purple. A) Interactions between mucins and S facilitated by glycans and Ca2+. B) Snapshot from SMA simulations. C) Ex- ample Ca2+ binding site from SMA simulations (1800 sites, each 1000+ atoms) used for AI-enabled quantum mechanical estimates from OrbNet Sky. D) Quantification of contacts between S and mucin from SMA simulations. E) OrbNet Sky energies vs CHARMM36m energies for each sub-selected system, colored by total number of atoms. Performance of OrbNet Sky vs. DFT in subplot (ωB97x-D3/def-TZVP, R2=0.99, for 17 systems of peptides chelating Ca2+. Visualized with VMD.

5.6.2 Parameter evaluation with OrbNet

Comparison to quantum methods reveals significant polarization effects and shows that there is opportunity to improve the accuracy of fixed charge force fields. For the large system sizes associated with

solvated Ca^{2+}-protein interaction motifs (over 1000 atoms, even in aggressively truncated systems) conventional quantum mechanics methods like density functional theory (DFT) are im- practical for analyzing a statistically significant ensemble of distinct configurations (see discussion in Performance Results). In contrast, OrbNet allows for DFT accuracy with over 1000-fold speed-up, providing a useful method for benchmarking and refining the force- field simulation parameters with quantum accuracy.[41] To confirm the accuracy of OrbNet versus DFT (ωB97X- D/def2-TZVP), the inset of Figure 5.4E correlates the two methods for the Ca2+-binding energy in a benchmark dataset of small Ca^{2+}- peptide complexes.[69] The excellent correlation of OrbNet and DFT for the present use case is clear from the inset figure; six datapoints were removed from this plot on the basis of a diagnostic applied to the semi-empirical GFN-xTB solution used for feature generation of OrbNet.

Figure 5.4E presents a comparison of the validated OrbNet method with the CHARMM36m force field for 1800 snapshots taken from the SMA MD simulations. At each snapshot, a subsystem containing a solvated Ca^{2+}-protein complex was extracted (Figure 5.4E) with protein bonds capped by hydrogens. For both OrbNet and the force field, the Ca^{2+}-binding energy was computed and shown in the correlation plot. Lack of correlation between OrbNet and the force field identifies important polarization effects, absent in a fixed charge description. Similarly, the steep slope of the best-fit line in Figure 5.4E reflects the fact that some of the configurations sampled using MD with the CHARMM36m force field are relatively high in energy according to the more accurate OrbNet potential. This approach allows us to test and quantify limitations of empirical force fields, such as lack of electronic polarization.

The practicality of OrbNet for these simulation snapshots with 1000+ atoms offers a straightforward multiscale strategy for refining the accuracy of the CHARMM36m force field. By optimizing the partial charges and other force field parameters, improved correlation with OrbNet for the subtle Ca^{2+}-protein interactions could be achieved, leading to near-quantum accuracy simulations with improved configurational sampling. The calculations presented here present a proof-of-concept of this iterative strategy.

5.6.3 AI-WE simulations of Delta spike opening

While our previous WE simulations of the WT SARS-CoV-2 S- opening[12] were notable in generating pathways for a seconds-timescale process of a massive system, we have made two critical technological advancements in the WESTPA software that greatly enhance the efficiency and analysis of WE simulations. These advances enabled striking observations of Delta-variant S opening (Figures 5.5 and 5.6). First, in contrast to prior manual bins for controlling trajectory replication, we have developed automated and adaptive binning that enables more efficient surmounting of large barriers via early identification of "bottleneck" regions.[70] Second, we have parallelized, memory-optimized, and implemented data streaming for the history-augmented Markov state model (haMSM) analysis scheme[71] to enable application to the TB-scale S-opening datasets. The haMSM approach estimates rate constants from simulations that have not yet reached a steady state.[72]

Our WE simulations generated >800 atomically detailed, Delta- variant S-opening pathways (Figures 5.5B and 5.6) of the receptor binding domain (RBD) switching from a glycan-shielded 'down' to an ex- posed 'up' state using 72 μs of total simulation time within 14 days using 192 NVIDIA V100 GPUs at a time on TACC's Longhorn supercomputer. Among these pathways, 83 reach an 'open' state that aligns with the structure of the human ACE2-bound WT S protein[73] and 18 reach a dramatically open state (Figure 5.6). Our haMSM analysis of WT WE simulations successfully provided long-timescale (steady-state) rate constants for S-opening based on highly transient information (Figure 5.5C).

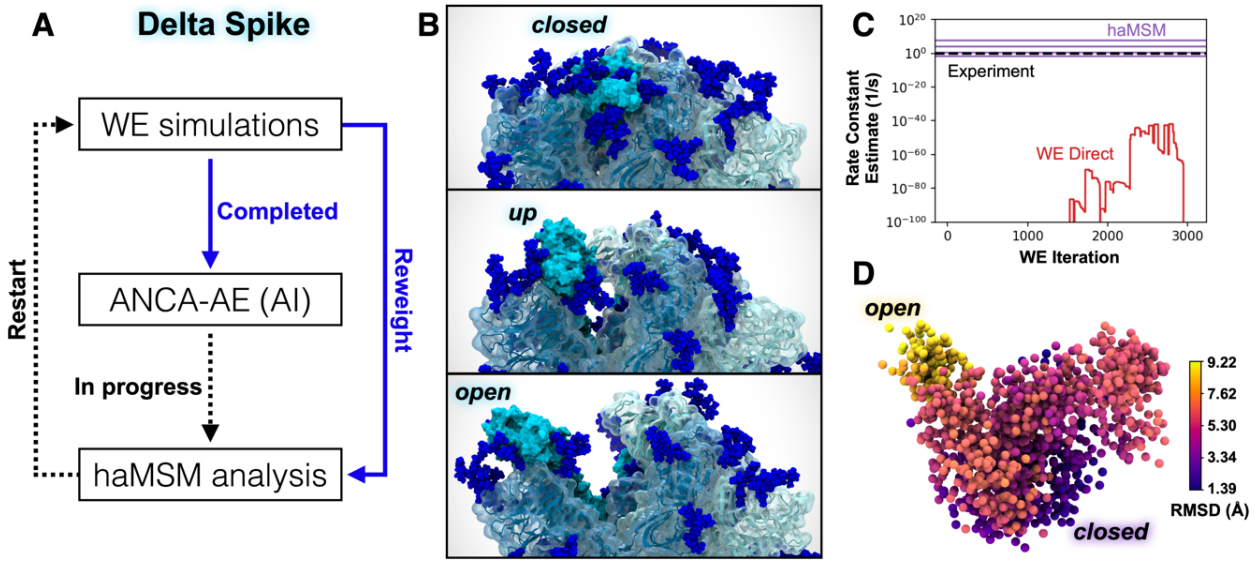


Figure 5.5: Delta-variant spike opening from WE simulations, and AI/haMSM analysis
A) The integrated workflow. B) Snapshots of the 'down', 'up', and 'open' states for Delta S-opening from a representative pathway generated by WE simulation, which represents 105 speedup compared to conventional MD. C) Rate-constant estimation with haMSM analysis of WE data (pur- ple lines) significantly improves direct WE computation (red), by comparison to experimental measurement (black dashed). Varying haMSM estimates result from different featurizations which will be individually cross-validated. D) The first three dimensions of the ANCA-AE embeddings depict a clear separation between the closed (darker purple) and open (yellow) con- formations of the Delta spike. A sub-sampled landscape is shown here where each sphere represents a conformation from the WE simulations and colored with the root-mean squared deviations (Å) with respect to the closed state. Visualized with VMD.

We also leveraged a simple, yet powerful unsupervised deep learning method called Anharmonic Conformational Analysis enabled Autoencoders (ANCA-AE)[33] to extract conformational states from our long-timescale WE simulations of Delta spike opening (Figure 5.5A,D). ANCA-AE first minimizes the fourth order correlations in atomistic fluctuations from MD simulation datasets and projects the data onto a low dimensional space where one can visualize the anharmonic conformational fluctuations. These projections are then input to an autoencoder that further minimizes non-linear correlations in the atomistic fluctuations to learn an embedding where conformations are automatically clustered based on their structural and energetic similarity. A visualization of the first three dimensions from the latent space articulates the RBD opening motion from its closed state (Figure 5.5D). It is notable that while other deep learning techniques need special

purpose hardware (such as GPUs), the ANCA-AE approach can be run with relatively modest CPU resources and can therefore scale to much larger systems (e.g., the virion within aerosol) when optimized.

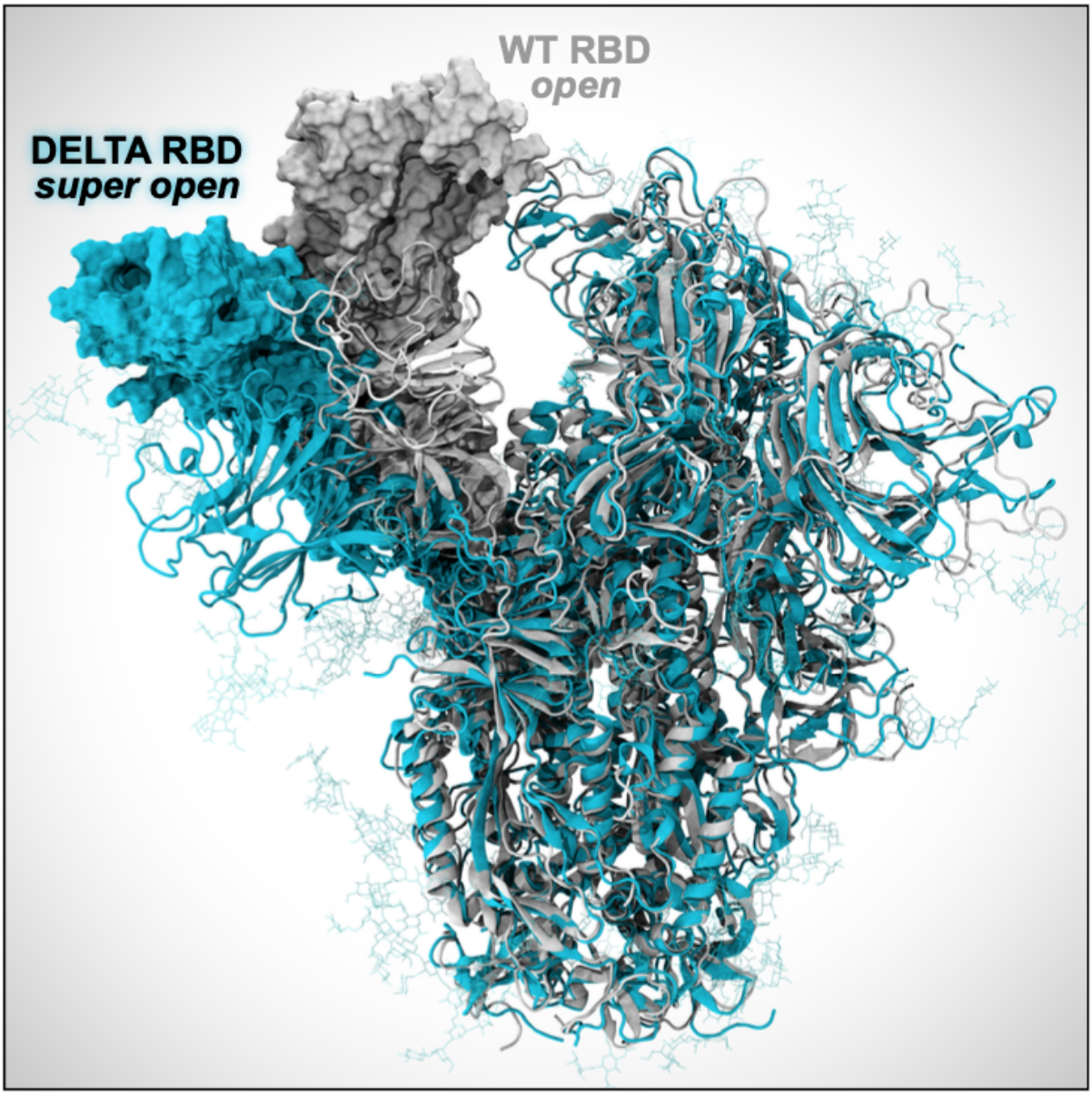


Figure 5.6: WE simulations reveal a dramatic opening of the Delta S (cyan), compared to WT S (white)

While further investigation is needed, this super open state seen in the Delta S may indicate increased capacity for binding to human host-cell receptors.

5.6.4 D-NEMD explores pH effects on Delta spike

We performed D-NEMD simulations of the SH system with GROMACS[74] using a ΔpH=2.0 (from 7.0 to 5.0) as the external perturbation. We ran 3 200-ns equilibrium MD simulations of SH to generate 87 configurations (29 configurations per replicate) that were used as the starting points for multiple short (10 ns) D-NEMD trajectories under the effect of the external perturbation (ΔpH=2.0). The effect of a ΔpH was modelled by changing the protonation state of histidines 66, 69, 146, 245, 625, 655, 1064, 1083, 1088, and 1101 (we note that other residues may also become protonated[75] ; the D-NEMD approach can also be applied to examine those). The structural response of the S to the pH decrease was investigated by measuring the difference in the position for each Cα atom between the equilibrium and corresponding D-NEMD simulation at equivalent points in time[38], namely after 0, 0.1, 1, 5 and 10 ns of simulation. The D-NEMD simulations reveal that pH changes, of the type expected in aerosols, affect the dynamics of functionally important regions of the spike, with potential implications for viral behavior (Figure 5.7). As this approach involves multiple short independent non-equilibrium trajectories, it is well suited for cloud computing. All D-NEMD simulations were performed using Oracle Cloud.

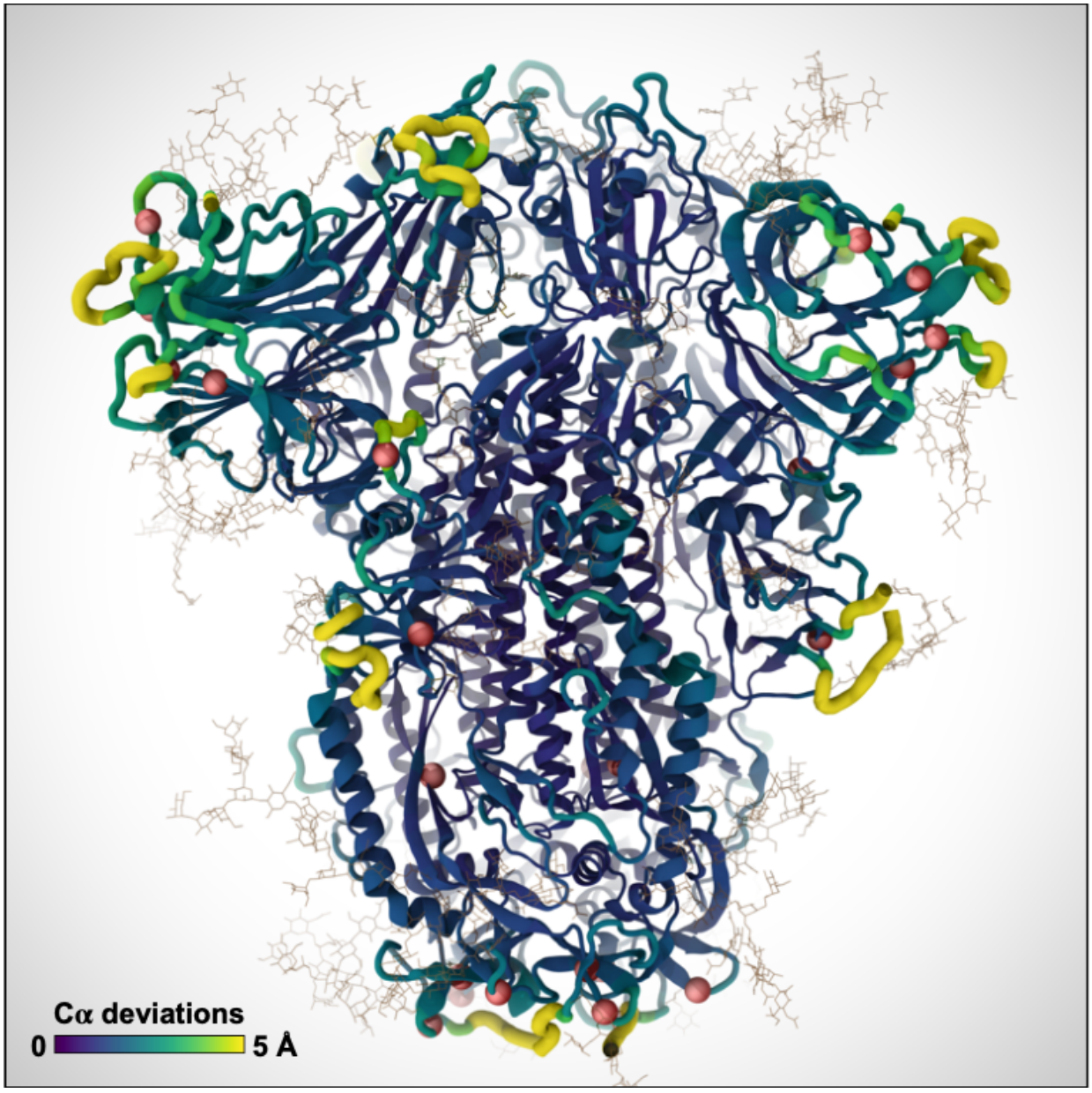


Figure 5.7: D-NEMD simulations reveal changes in key functional regions of the S protein, including the receptor binding domain, as the result of a pH decrease

Color scale and ribbon thickness indicate the degree of deviation of Cα atoms from their equilibrium position. Red spheres indicate the location of positively charged histidines.

5.7 How performance was measured

5.7.1 WESTPA

For the WE simulations of spike opening using WESTPA, we defined the time to solution as the total simulation time required to generate the first spike opening event. Spike opening is essentially impossible to observe via conventional MD. WESTPA simulations were run using the AMBER20 dynamics engine and 192 NVIDIA V100 GPUs at a time on TACC's Longhorn supercomputer.

5.7.2 NAMD

NAMD performance metrics were collected using hardware performance counters for FLOPs/step measurements, and application- internal timers for overall simulation rates achieved by production runs including all I/O for simulation trajectory and checkpoint output. NAMD FLOPs/step measurements were conducted on TACC Frontera, by querying hardware performance counters with the rdmsr utility from Intel msr-tools and the "TACC stats" system programs. For each simulation, FLOP counts were measured for NAMD simulation runs of two different step counts. The results of the two simulation lengths were subtracted to eliminate NAMD startup operations, yielding an accurate estimate of the marginal FLOPs per step for a continuing simulation.[76] Using the FLOPs/step values computed for each simulation, overall FLOP rates were computed by dividing the FLOPs/step value by seconds/step performance data reported by NAMD internal application timers during production runs.

Table 5.3: MD simulation floating point ops per timestep

MD Simulation	Code	Atoms	[a]FLOPs/step
Spike, head	AMBER, GROMACS	0.6M	62.14 GFLOPs/step
Spike	NAMD	1.7M	43.05 GFLOPs/step
S+m_1/m_2+ALB	NAMD	2.1M	54.86 GFLOPs/step
Resp. Aero.+Vir.	NAMD	1B	25.81 TFLOPs/step

[a] FLOPs/step data were computed by direct FLOP measurements from hardware performance counters for NAMD simulations, or by using the application-reported FLOP rates and ns/day simulation performance in the case of GROMACS.

5.7.3 GROMACS

GROMACS 2020.4 benchmarking was performed on Oracle Cloud Infrastructure (OCI) compute shape BM.GPU4.8 consisting of 8 NVIDIA A100 tensor core GPUs, and 64 AMD Rome CPU cores. The

simulation used for benchmarking contained 615,563 atoms and was run for 500,000 steps with 2 fs time steps. The simulations were run on increasing numbers of GPUs, from 1 to 8, using 8 CPU cores per GPU, running for both the production (Nose-Hoover) and GPU- accelerated (velocity rescaling) thermostats. Particle–mesh Ewald (PME) calculations were pinned to a single GPU, with additional GPUs for multi-GPU jobs used for particle–particle calculations. Performance data (ns/day and average single-precision TFLOPS, calculated as total number of TFLOPs divided by total job walltime) were reported by GROMACS itself. Each simulation was repeated four times and average performance figures reported.

5.8 **Performance**

5.8.1 NAMD performance

NAMD was used to perform all of the simulations listed in Table 5.2, except for the closed spike (SH) simulations described further below. With the exception of the aerosol and virion simulation, the other NAMD simulations used conventional protocols and have performance and parallel scaling characteristics that closely match the results reported in our previous SARS-CoV-2 research.[55] NAMD 2.14 scaling performance for the one billion- atom respiratory aerosol and virion simulation run on ORNL Summit is summarized in Tables 7.4 and 8.5. A significant performance challenge associated with the aerosol virion simulation relates to the roughly 50% reduction in particle density as compared with a more conventional simulation with a fully populated periodic cell. The reduced particle density results in large regions of empty space that nevertheless incur additional overheads associated with both force calculations and integration, and creates problems for the standard NAMD load balancing scheme that estimates the work associated with the cubic “patches” used for parallel domain decomposition. The PME electrostatics algorithm and associated 3-D FFT and transpose operations encompass the entire simulation unit cell and associated patches, requiring involvement in communication and reduction operations despite the inclusion of empty space. Enabling NAMD diagnostic output on a 512-node 1B-atom aerosol and virion simulation revealed that ranks assigned empty regions of the periodic cell had 66 times the number of fixed-size patches as ranks assigned dense regions. The initial load estimate for an empty patch was changed from a fixed 10 atoms to a runtime

parameter with a default of 40 atoms, which reduced the patch ratio from 66 to 19 and doubled performance on 512 nodes.

Table 5.4: NAMD performance: Respiratory Aerosol + Virion, 1B atoms, 4 fs timestep w/ HMR, and PME every 3 steps

Nodes	Summit CPU + GPU	Speedup	Efficiency
256	4.18 ns/day	~1.0×	~100%
512	7.68 ns/day	1.84×	92%
1024	13.64 ns/day	3.27×	81%
2048	23.10 ns/day	5.53×	69%
4096	34.21 ns/day	8.19×	51%

Table 5.5: Peak NAMD FLOP rates, ORNL Summit

NAMD Simulation	Atoms	Nodes	Sim rate	Performance
Resp. Aero.+Vir.	1B	4096	34.21 ns/day	2.55 PFLOPS

5.8.2 WESTPA performance

Our time to solution for WE simulations of spike opening (to the “up” state) (Figure 5.5) using the WESTPA software and AMBER20 was 14 μs of total simulation time, which was completed in 4 days using 192 NVIDIA V100 GPUs at a time on TACC’s Longhorn supercomputer. For reference, conventional MD would require an expected 5 orders of magnitude more computing. The WESTPA software is highly scalable, with nearly perfect scaling out to >1000 NVIDIA V100 GPUs and this scaling is expected to continue until the filesystem is saturated. Thus, WESTPA makes optimal use of large supercomputers and is limited by filesystem I/O due to the periodic restarting of trajectories after short time intervals.

5.8.3 AI-enhanced WE simulations

DeepDriveMD is a framework to coordinate the concurrent execution of ensemble simulations and drive them using AI models.[77,78] DeepDriveMD has been shown to improve the scientific performance of diverse problems: from protein folding to conformation of protein-ligand complexes. We coupled WESTPA to DeepDriveMD, which is responsible for resource dynamism and concurrent heterogeneous task execution (ML and AMBER). The coupled workflow was executed on 1024 nodes on Summit (OLCF), and, in spite of

the spatio-temporal heterogeneity of tasks involved, the resource utilization was in the high 90%. Consistent with earlier studies, the coupling of WESTPA to DeepDriveMD results in a 100x improvement in the exploration of phase space.

5.8.4 GROMACS performance

Figure 5.8 shows GROMACS parallelizes well across the 8 NVIDIA A100 GPUs available on each BM.GPU4.8 instance used in the *Cluster in the Cloud* running on OCI. There is a performance drop for two GPUs due to inefficient division of the PME and particle-particle tasks. Methods to address this exist for the two GPU case[17], but were not adopted as we were targeting maximum raw performance across all 8 GPUs. Production simulations achieved 27% of the peak TFLOPS available from the GPUs. Multiple simulations were run across 10 such compute nodes, enabling the ensemble to run at an average combined speed of 425 TFLOPS and sampling up to 1μs/day. We note that the calculations will be able to run 20%–40% faster once the Nose-Hoover thermostat that is required for the simulation is ported to run on the GPU. Benchmarking using a velocity rescaling thermostat that has been ported to GPU shows that this would enable the simulation to extract 34% of the peak TFLOPS from the cards, enabling each node to achieve an average speed of 53.4 TFLOPS, and 125 ns/day. A cluster of 10 nodes would enable GROMACS to run at an average combined speed of over 0.5 PFLOPs, simulating over 1.2 μs/day.

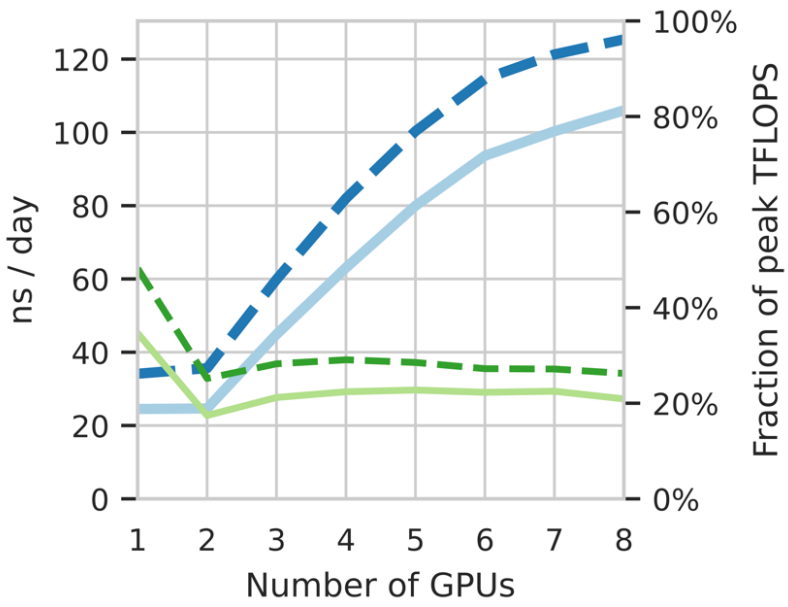


Figure 5.8: GROMACS performance across 1–8 A100 GPUs in ns/day (thicker, blue lines) and the fraction of maximum theoretical TFLOPS (thinner, green lines)
Production setup shown with solid line and runs with the GPU-accelerated thermostat in dashed.

A significant innovation is that this power is available on demand: Cluster in the Cloud with GPU-optimized GROMACS was provisioned and benchmarked within one day of inception of the project. This was handed to the researcher, who submitted the simulations. Automatically, up to ten BM.GPU4.8 compute nodes were provisioned on-demand based on requests from the Slurm scheduler. These simulations were performed on OCI, using *Cluster in the Cloud* to manage automatic scaling.

Cluster in the Cloud was configured to dynamically provision and terminate computing nodes based on the workload. Simulations were conducted using GROMACS 2020.4 compiled with CUDA sup- port. Multiple simultaneous simulations were conducted, with each simulation utilizing a single BM.GPU4.8 node without multi-node parallelism.

This allowed all production simulations to be completed within 2 days. The actual compute cost of the project was less than $6125 USD (on-demand OCI list price). The huge reduction in “time to science” that low-cost cloud enables changes the way that researchers can access and use HPC facilities. In our

opinion, such a setup enables “exclusive on-demand” HPC capabilities for the scientific community for rapid advancement in science.

5.8.5 OrbNet performance

Prior benchmarking reveals that OrbNet provides over 1000-fold speedup compared to DFT. For the calculations presented here, the cost of corresponding high-quality range-separated DFT calculations (wB97X-D/def2-TZVP) can be estimated. In Figure 5.4E, we consider system sizes which would require 14,000–47,000 atomic orbitals for wB97X-D/def2-TZVP, exceeding the range of typical DFT evaluations. Estimation of the DFT computational cost of the 1811 configurations studied in Figure 5.4E suggests a total of 115M core-hours on NERSC Cori Haswell nodes; in contrast, the OrbNet calculations for the current study require only 100k core-hours on the same nodes. DFT cost estimates were based on extrapolation from a dataset of over 1M ChEMBL molecules ranging in size from 40 to 107 atom systems considering only the cubic cost component of DFT.[41]

5.9 **Implications**

Our major scientific achievements are:

(1) We showcase an extensible AI-enabled multiscale computational framework that bridges time and length scales from electronic structure through whole aerosol particle morphology and dynamics.

(2) We develop all-atom simulations of respiratory mucins and use these to understand the structural basis of interaction with the SARS-CoV-2 spike protein. This has implications for viral binding in the deep lung, which is coated with mucins. We expect the impact of our mucin simulations to be far reaching, as malfunctions in mucin secretion and folding have been implicated in progression of severe diseases such as cancer and cystic fibrosis.

(3) We present a significantly enhanced all-atom model and simulation of the SARS-CoV-2 Delta virion, which includes the hundreds of tiled M-protein dimers and the E-protein ion channels. This model can be used as a basis to understand why the Delta virus is so much more infectious than the WT or alpha variants.

(4) We develop an ultra-large (1 billion+) all-atom simulation capturing massive chemical and biological complexity within a respiratory aerosol. This simulation provides the first atomic level views of virus-laden aerosols and is already serving as a basis to develop an untold number of experimentally testable hypotheses. An immediate example suggests a mechanism through which mucins and other species, e.g., lipids, which are present in the aerosol, arrange to protect the molecular structure of the virus, which otherwise would be exposed to the air-water interface. This work also opens the door for developing simulations of other aerosols, e.g., sea spray aerosols, that are involved in regulating climate.

(5) We evidence how changes in pH, which are expected in the aerosol environment, may alter dynamics and allosteric communication pathways in key functional regions of the Delta spike protein.

(6) We characterize atomically detailed pathways for the spike- opening process of the Delta variant using WE simulations, revealing a dramatically open state that may facilitate binding to human host cells.

(7) We demonstrate how parallelized haMSM analysis of WE data can provide physical rate estimates of spike opening, improving prior estimates by many orders of magnitude. The pipeline can readily be applied to the any variant spike protein or other complex systems of interest.

(8) We show how HPC and cloud resources can be used to significantly drive down time-to-solution for major scientific efforts as well as connect researchers and greatly enable complex collaborative interactions.

(9) We demonstrate how AI coupled to HPC at multiple levels can result in significantly improved effective performance, e.g., with AI-driven WESTPA, and extend the reach and do- main of applicability of tools ordinarily restricted to smaller, less complex systems, e.g., with OrbNet.

(10) While our work provides a successful use case, it also exposes weaknesses in the HPC ecosystem in terms of support for key steps in large/complex computational science campaigns. We find lack of widespread support for high performance remote visualization and interactive graphical sessions for

system preparation, debugging, and analysis with diverse science tools to be a limiting factor in such efforts.

5.10 **Acknowledgments**

We thank Prof. Kim Prather for inspiring and informative discussions about aerosols and for her commitment to convey the air- borne nature of SARS-CoV-2. We thank D. Veesler for sharing the Delta spike NTD coordinates in advance of publication. We thank B. Messer, D. Maxwell, and the Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory supported by the DOE under Contract DE-AC05-00OR22725. We thank the Texas Advanced Computing Center Frontera team, especially D. Stanzione and T. Cockerill, and for compute time made available through a Director's Discretionary Allocation (NSF OAC-1818253). We thank the Argonne Leadership Computing Facility supported by the DOE under DE-AC02-06CH11357. We thank the Pittsburgh Supercomputer Center for providing priority queues on Bridges-2 through the XSEDE allocation NSF TG-CHE060063. We thank N. Kern and J. Lee of the CHARMM-GUI support team for help converting topologies between NAMD and GROMACS. We thank J. Copperman, G. Simpson, D. Aristoff, and J. Leung for valuable discussions and sup- port from NIH grant GM115805. NAMD and VMD are funded by NIH P41-GM104601. This work was supported by the NSF Center for Aerosol Impacts on Chemistry of the Environment (CAICE), National Science Foundation Center for Chemical Innovation (NSF CHE-1801971), as well as NIH GM132826, NSF RAPID MCB-2032054, an award from the RCSA Research Corp., a UC San Diego Moore's Cancer Center 2020 SARS-CoV-2 seed grant, to R.E.A. This work was also supported by Oracle Cloud credits and related resources provided by the Oracle for Research program.

Chapter 5, in full, is a modified reprint of the materials as it appears in "Dommer, A., † Casalino, L.†, and Kearns, F. †; Rosenfeld, M.; Wauer, N.; Ahn, S-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A.; Trifan, A.; Brace, A.; Sztain, T.; Clyde, A.; Ma, H.; Chennubhotla, C.; Lee, H.; Turilli, M.; Khalid, S.; Tamayo-Mendoza, T.; Welborn, M.; Christensen, A.; Smith, D.; Qiao, Z.; Sirumalla, S.; O'Connor, M.; Manby, F.; Anandkumar, A.; Hardy, D.; Phillips, J.; Stern, A.; Romero, J.; Clark, D.; Dorrell, M.; Maiden, T.; Huang, L.; McCalpin, J.;

Woods, C.; Grey, A.; Williams, M.; Barker, B.; Rajapaksha, H.; Pitts, R.; Gibbs, T.; Stone, J.; Zuckerman, D.; Mulholland, A.; Miller, T.; Jha, S.; Ramanathan, A.; Chong, L.; Amaro, R. *#COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol*. Int J High Perform Comput Appl. 2023 Jan;37(1):28-44." The dissertation author was a co-investigator and co-author of this work. **Notably, this work was selected as a Finalist for the 2021 ACM Gordon Bell Special Prize for COVID-19 Research.**

5.11 **References**

(1) Wang, C. C.; Prather, K. A.; Sznitman, J.; Jimenez, J. L.; Lakdawala, S. S.; Tufekci, Z.; Marr, L. C. Airborne Transmission of Respiratory Viruses. *Science* 2021, *373* (6558), eabd9149. https://doi.org/10.1126/science.abd9149.

(2) Miller, S. L.; Nazaroff, W. W.; Jimenez, J. L.; Boerstra, A.; Buonanno, G.; Dancer, S. J.; Kurnitski, J.; Marr, L. C.; Morawska, L.; Noakes, C. Transmission of SARS-CoV-2 by Inhalation of Respiratory Aerosol in the Skagit Valley Chorale Superspreading Event. *Indoor Air* 2021, *31* (2), 314–323. https://doi.org/10.1111/ina.12751.

(3) Coleman, K. K.; Tay, D. J. W.; Sen Tan, K.; Ong, S. W. X.; Son, T. T.; Koh, M. H.; Chin, Y. Q.; Nasir, H.; Mak, T. M.; Chu, J. J. H.; Milton, D. K.; Chow, V. T. K.; Tambyah, P. A.; Chen, M.; Wai, T. K. Viral Load of SARS-CoV-2 in Respiratory Aerosols Emitted by COVID-19 Patients While Breathing, Talking, and Singing. *Clinical Infectious Diseases* 2021. https://doi.org/10.1093/cid/ciab691.

(4) Fennelly, K. P. Particle Sizes of Infectious Aerosols: Implications for Infection Control. *The Lancet Respiratory Medicine* 2020, *8* (9), 914–924. https://doi.org/10.1016/S2213-2600(20)30323-4.

(5) Vejerano, E. P.; Marr, L. C. Physico-Chemical Characteristics of Evaporating Respiratory Fluid Droplets. *Journal of the Royal Society Interface* 2018, *15* (139), 1–10. https://doi.org/10.1098/rsif.2017.0939.

(6) Walker, J. S.; Archer, J.; Gregson, F. K. A.; Michel, S. E. S.; Bzdek, B. R.; Reid, J. P. Accurate Representations of the Microphysical Processes Occurring during the Transport of Exhaled Aerosols and Droplets. *ACS Central Science* 2021. https://doi.org/10.1021/acscentsci.0c01522.

(7) Bangaru, S.; Ozorowski, G.; Turner, H. L.; Antanasijevic, A.; Huang, D.; Wang, X.; Torres, J. L.; Diedrich, J. K.; Tian, J.-H.; Portnoff, A. D.; Patel, N.; Massare, M. J.; Yates, J. R.; Nemazee, D.; Paulson, J. C.; Glenn, G.; Smith, G.; Ward, A. B. Structural Analysis of Full-Length SARS-CoV-2 Spike Protein from an Advanced Vaccine Candidate. *Science* 2020, *370* (6520), 1089–1094. https://doi.org/10.1126/science.abe1502.

(8) McCallum, M.; Walls, A. C.; Sprouse, K. R.; Bowen, J. E.; Rosen, L.; Dang, H. V.; deMarco, A.; Franko, N.; Tilles, S. W.; Logue, J.; Miranda, M. C.; Ahlrichs, M.; Carter, L.; Snell, G.; Pizzuto, M. S.; Chu, H. Y.; Van Voorhis, W. C.; Corti, D.; Veesler, D. Molecular Basis of Immune Evasion by the Delta and Kappa SARS-CoV-2 Variants. *bioRxiv* 2021. https://doi.org/10.1101/2021.08.11.455956.

(9) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 2020, *181* (2), 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

(10) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science* 2020, *367* (6483), 1260–1263. https://doi.org/10.1126/science.abb2507.

(11) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; Mclellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* 2020. https://doi.org/10.1021/acscentsci.0c01056.

(12) Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nature Chemistry 2021 13:10* 2021, *13* (10), 963–968. https://doi.org/10.1038/s41557-021-00758-3.

(13) D'Imprima, E.; Floris, D.; Joppe, M.; Sánchez, R.; Grininger, M.; Kühlbrandt, W. Protein Denaturation at the Air-Water Interface and How to Prevent It. *eLife* 2019, *8*, e42747. https://doi.org/10.7554/eLife.42747.

(14) Petters, M. D.; Kreidenweis, S. M. A Single Parameter Representation of Hygroscopic Growth and Cloud Condensation Nucleus Activity. *Atmospheric Chemistry and Physics* 2007, *7* (8), 1961–1971. https://doi.org/10.5194/acp-7-1961-2007.

(15) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem.\ Phys.* 2020, *153*, 044130. https://doi.org/10.1063/5.0014475.

(16) Case, D.; Ben-Shalom, I.; Brozell, S.; article is an open \ldots, D. C.-\ldots Switzerland. T. AMBER 2018; University of California: San Francisco, CA, USA, 2018.© 2019 by the Authors.

(17) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous Parallelization and Acceleration of Molecular Dynamics Simulations in GROMACS. *The Journal of Chemical Physics* 2020, *153* (13), 134110. https://doi.org/10.1063/5.0018516.

(18) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* 1996. https://doi.org/10.1016/0263-7855(96)00018-5.

(19) Sener, M.; Levy, S.; Stone, J. E.; Christensen, A. J.; Isralewitz, B.; Patterson, R.; Borkiewicz, K.; Carpenter, J.; Hunter, C. N.; Luthey-Schulten, Z.; Cox, D. Multiscale Modeling and Cinematic Visualization of Photosynthetic Energy Conversion Processes from Electronic to Cell Scales. *Parallel Computing* 2021, 102698.

(20) Stone, J. E.; Isralewitz, B.; Schulten, K. Early Experiences Scaling VMD Molecular Visualization and Analysis Jobs on Blue Waters. In *Extreme Scaling Workshop (XSW), 2013*; 2013; pp 43–50. https://doi.org/10.1109/XSW.2013.10.

(21) Stone, J. E.; Vandivort, K. L.; Schulten, K. GPU-Accelerated Molecular Visualization on Petascale Supercomputing Platforms. In *Proceedings of the 8th International Workshop on Ultrascale Visualization*; UltraVis '13; ACM: New York, NY, USA, 2013; p 6:1-6:8.

(22) Stone, J. E.; Sener, M.; Vandivort, K. L.; Barragan, A.; Singharoy, A.; Teo, I.; Ribeiro, J. V.; Isralewitz, B.; Liu, B.; Goh, B. C.; Phillips, J. C.; MacGregor-Chatwin, C.; Johnson, M. P.; Kourkoutis, L. F.; Hunter, C. N.; Schulten, K. Atomic Detail Visualization of Photosynthetic Membranes with GPU-Accelerated Ray Tracing. *Parallel Computing* 2016, *55*, 17–27. https://doi.org/10.1016/j.parco.2015.10.015.

(23) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophysical Journal* 1996, *70* (1), 97–110.

(24) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annual Review of Biophysics* 2017, *46*, 43–57. https://doi.org/10.1146/annurev-biophys-070816-033834.

(25) Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallimadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P.; al, et. SARS-CoV-2 Simulations Go Exascale to Predict Dramatic Spike Opening and Cryptic Pockets across the Proteome. *Nature Chemistry* 2021, *13* (7), 651–659. https://doi.org/10.1038/s41557-021-00707-0.

(26) Adhikari, U.; Mostofian, B.; Copperman, J.; Petersen, A.; Zuckerman, D. M. Computational Estimation of Ms-Sec Atomistic Folding Times. *Journal of the American Chemical Society* 2019, 6519–6526. https://doi.org/10.1101/427393.

(27) Saglam, A. S.; Chong, L. T. Protein–Protein Binding Pathways and Calculations of Rate Constants Using Fully-Continuous, Explicit-Solvent Simulations. *Chemical science* 2019, *10* (8), 2360–2372.

(28) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; Zuckerman, D. M.; Chong, L. T. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *Journal of Chemical Theory and Computation* 2015, *11* (2), 800–809. https://doi.org/10.1021/ct5010615.

(29) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* 2013, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.

(30) Brace, A.; Lee, H.; Ma, H.; Trifan, A.; Turilli, M.; Yakushin, I.; Munson, T.; Foster, I.; Jha, S.; Ramanathan, A. Achieving 100X Faster Simulations of Complex Biological Phenomena by Coupling ML to HPC Ensembles. 2021.

(31) Noe, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annual Review of Physical Chemistry* 2020, *71* (1), 361–390. https://doi.org/10.1146/annurev-physchem-042018-052331.

(32) Bhowmik, D.; Gao, S.; Young, M. T.; Ramanathan, A. Deep Clustering of Protein Folding Simulations. *BMC Bioinformatics* 2018, *19* (18), 484. https://doi.org/10.1186/s12859-018-2507-5.

(33) Clyde, A.; Galanie, S.; Kneller, D. W.; Ma, H.; Babuji, Y.; Blaiszik, B.; Brace, A.; Brettin, T.; Chard, K.; Chard, R.; Coates, L.; Foster, I.; Hauner, D.; Kertesz, V.; Kumar, N.; Lee, H.; Li, Z.; Merzky, A.; Schmidt, J. G.; Tan, L.; Titov, M.; Trifan, A.; Turilli, M.; Van Dam, H.; Chennubhotla, S. C.; Jha, S.; Kovalevsky, A.; Ramanathan, A.; Head, M. S.; Stevens, R. High Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Non-Covalent Inhibitor. *bioRxiv* 2021. https://doi.org/10.1101/2021.03.27.437323.

(34) Warwicker, J. A Model for PH Coupling of the SARS-CoV-2 Spike Protein Open/Closed Equilibrium. *Briefings in Bioinformatics* 2021, *22* (2), 1499–1507. https://doi.org/10.1093/bib/bbab056.

(35) Ciccotti, G.; Ferrario, M. Non-Equilibrium by Molecular Dynamics: A Dynamical Approach. *Molecular Simulation* 2016, *42* (16), 1385–1400. https://doi.org/10.1080/08927022.2015.1121543.

(36) Galdadas, I.; Qu, S.; Oliveira, A. S. F.; Olehnovics, E.; Mack, A. R.; Mojica, M. F.; Agarwal, P. K.; Tooke, C. L.; Gervasio, F. L.; Spencer, J.; Bonomo, R. A.; Mulholland, A. J.; Haider, S. Allosteric Communication in Class A β-Lactamases Occurs via Cooperative Coupling of Loop Dynamics. *eLife* 2021, *10*, e66567. https://doi.org/10.7554/eLife.66567.

(37) Oliveira, A. S. F.; Shoemark, D. K.; Campello, H. R.; Wonnacott, S.; Gallagher, T.; Sessions, R. B.; Mulholland, A. J. Identification of the Initial Steps in Signal Transduction in the A4β2 Nicotinic Receptor: Insights from Equilibrium and Nonequilibrium Simulations. *Structure* 2019, *27* (7), 1171-1183.e3. https://doi.org/10.1016/j.str.2019.04.008.

(38) Oliveira, A. S. F.; Ciccotti, G.; Haider, S.; Mulholland, A. J. Dynamical Nonequilibrium Molecular Dynamics Reveals the Structural Basis for Allostery and Signal Propagation in Biomolecular Systems. *The European Physical Journal B* 2021, *94* (7), 144. https://doi.org/10.1140/epjb/s10051-021-00157-0.

(39) Hughes, G. W.; Ridley, C.; Collins, R.; Roseman, A.; Ford, R.; Thornton, D. J. The MUC5B Mucin Polymer Is Dominated by Repeating Structural Motifs and Its Topology Is Regulated by Calcium and PH. *Scientific Reports* 2019, *9* (1), 17350. https://doi.org/10.1038/s41598-019-53768-0.

(40) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. OrbNet: Deep Learning for Quantum Chemistry Using Symmetry-Adapted Atomic-Orbital Features. *The Journal of Chemical Physics* 2020, *153* (12), 124111. https://doi.org/10.1063/5.0021955.

(41) Christensen, A. S.; Sirumalla, S. K.; Qiao, Z.; O'Connor, M. B.; Smith, D. G. A.; Ding, F.; Bygrave, P. J.; Anandkumar, A.; Welborn, M.; Manby, F. R.; au2, T. F. M. I. OrbNet Denali: A Machine Learning Potential for Biological and Organic Chemistry with Semi-Empirical Cost and DFT Accuracy. 2021.

(42) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. 2005. https://doi.org/10.1002/jcc.20289.

(43) Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* 2019. https://doi.org/10.1093/glycob/cwz003.

(44) Beglov, D.; Roux, B. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. *The Journal of Chemical Physics* 1994. https://doi.org/10.1063/1.466711.

(45) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *Journal of Chemical Theory and Computation* 2009, *5* (9), 2353–2370. https://doi.org/10.1021/ct900242e.

(46) Han, K.; Venable, R. M.; Bryant, A.-M.; Legacy, C. J.; Shen, R.; Li, H.; Roux, B.; Gericke, A.; Pastor, R. W. Graph–Theoretic Analysis of Monomethyl Phosphate Clustering in Ionic Solutions. *The Journal of Physical Chemistry B* 2018, *122* (4), 1484–1494. https://doi.org/10.1021/acs.jpcb.7b10730.

(47) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *Journal of Computational Chemistry* 2013, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(48) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature methods* 2017, *14* (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(49) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* 2010, *114* (23), 7830–7843. https://doi.org/10.1021/jp101759q.

(50) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. Simulations of Anionic Lipid Membranes: Development of Interaction-Specific Ion Parameters and Validation Using NMR Data. *The Journal of Physical Chemistry B* 2013, *117* (35), 10183–10192. https://doi.org/10.1021/jp401512z.

(51) Kannan, S. R.; Spratt, A. N.; Cohen, A. R.; Naqvi, S. H.; Chand, H. S.; Quinn, T. P.; Lorson, C. L.; Byrareddy, S. N.; Singh, K. Evolutionary Analysis of the Delta and Delta Plus Variants of the SARS-CoV-2 Viruses. *Journal of Autoimmunity* 2021, *124*, 102715. https://doi.org/10.1016/j.jaut.2021.102715.

(52) Surya, W.; Li, Y.; Torres, J. Structural Model of the SARS Coronavirus E Channel in LMPG Micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 2018, *1860* (6), 1309–1317. https://doi.org/10.1016/j.bbamem.2018.02.017.

(53) Mandala, V. S.; McKay, M. J.; Shcherbakov, A. A.; Dregni, A. J.; Kolocouris, A.; Hong, M. Structure and Drug Binding of the SARS-CoV-2 Envelope Protein Transmembrane Domain in Lipid Bilayers. *Nature Structural and Molecular Biology* 2020, *27* (12), 1202–1208.

(54) Chai, J.; Cai, Y.; Pang, C.; Wang, L.; McSweeney, S.; Shanklin, J.; Liu, Q. Structural Basis for SARS-CoV-2 Envelope Protein Recognition of Human Cell Junction Protein PALS1. *Nature Communications* 2021, *12* (1), 3433. https://doi.org/10.1038/s41467-021-23533-x.

(55) Casalino, L.; Dommer, A. C.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S.-H.; Trifan, A.; Brace, A.; Bogetti, A. T.; Clyde, A.; Ma, H.; Lee, H.; Turilli, M.; Khalid, S.; Chong, L. T.; Simmerling, C.; Hardy, D. J.; Maia, J. D.; Phillips, J. C.; Kurth, T.; Stern, A. C.; Huang, L.; McCalpin, J. D.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. *The International ournal of High*

Performance Computing Applications 2021, *35* (5), 432–451. https://doi.org/10.1177/10943420211006452.

(56) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of computational chemistry* 2008, *29*, 2967–2970. https://doi.org/10.1002/jcc.

(57) Durrant, J. D.; Amaro, R. E. LipidWrapper: An Algorithm for Generating Large-Scale Membrane Models of Arbitrary Geometry. *PLoS Computational Biology* 2014, *10* (7). https://doi.org/10.1371/journal.pcbi.1003720.

(58) Blender Online Community. Blender - a 3D Modelling and Rendering Package. Blender Foundation: Blender Institute, Amsterdam 2020.

(59) Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W. J. H.; Welsch, S.; Blanc, F. E. C.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C.; van Zandbergen, G.; Landry, J.; de Azevedo, N. T. D.; Mosalaganti, S.; Schwarz, A.; Covino, R.; Mühlebach, M. D.; Hummer, G.; Krijnse Locker, J.; Beck, M. In Situ Structural Analysis of SARS-CoV-2 Spike Reveals Flexibility Mediated by Three Hinges. *Science* 2020, eabd5223. https://doi.org/10.1126/science.abd5223.

(60) Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B.; Lu, J. M.; Peukes, J.; Xiong, X.; Kräusslich, H. G.; Scheres, S. H. W.; Bartenschlager, R.; Briggs, J. A. G. Structures and Distributions of SARS-CoV-2 Spike Proteins on Intact Virions. *Nature* 2020, 1–7. https://doi.org/10.1038/s41586-020-2665-2.

(61) Mariethoz, J.; Alocci, D.; Gastaldello, A.; Horlacher, O.; Gasteiger, E.; Rojas-Macias, M.; Karlsson, N. G.; Packer, N. H.; Lisacek, F. Glycomics@ExPASy: Bridging the Gap*. *Molecular and Cellular Proteomics* 2018, *17* (11), 2164–2176. https://doi.org/10.1074/mcp.RA118.000799.

(62) Symmes, B. A.; Stefanski, A. L.; Magin, C. M.; Evans, C. M. Role of Mucins in Lung Homeostasis: Regulated Expression and Biosynthesis in Health and Disease. *Biochemical Society Transactions* 2018, *46* (3), 707–719. https://doi.org/10.1042/BST20170455.

(63) Thomsson, K. A.; Schulz, B. L.; Packer, N. H.; Karlsson, N. G. MUC5B Glycosylation in Human Saliva Reflects Blood Group and Secretor Status. *Glycobiology* 2005, *15* (8), 791–804. https://doi.org/10.1093/glycob/cwi059.

(64) Markovetz, M. R.; Subramani, D. B.; Kissner, W. J.; Morrison, C. B.; Garbarine, I. C.; Ghio, A.; Ramsey, K. A.; Arora, H.; Kumar, P.; Nix, D. B.; Kumagai, T.; Krunkosky, T. M.; Krause, D. C.; Radicioni, G.; Alexis, N. E.; Kesimer, M.; Tiemeyer, M.; Boucher, R. C.; Ehre, C.; Hill, D. B. Endotracheal Tube Mucus as a Source of Airway Mucus for Rheological Study. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2019, *317* (4), L498–L509. https://doi.org/10.1152/ajplung.00238.2019.

(65) Raveh, B.; London, N.; Schueler-Furman, O. Sub-Angstrom Modeling of Complexes between Flexible Peptides and Globular Proteins. *Proteins: Structure, Function, and Bioinformatics* 2010, *78* (9), 2029–2040. https://doi.org/10.1002/prot.22716.

(66) Sugio, S.; Kashima, A.; Mochizuki, S.; Noda, M.; Kobayashi, K. Crystal Structure of Human Serum Albumin at 2.5 Å Resolution. *Protein Engineering, Design and Selection* 1999, *12* (6), 439–446. https://doi.org/10.1093/protein/12.6.439.

(67) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* 2009, *30* (13), 2157–2164. https://doi.org/10.1002/jcc.21224.

(68) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* 2015, *11* (4), 1864–1874. https://doi.org/10.1021/ct5010406.

(69) Hu, X.; Lenz-Himmer, M.-O.; Baldauf, C. Better Force Fields Start with Better Data – A Data Set of Cation Dipeptide Interactions. 2021.

(70) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *The Journal of Physical Chemistry A* 2021, *125* (7), 1642–1649. https://doi.org/10.1021/acs.jpca.0c10724.

(71) Copperman, J.; Zuckerman, D. M. Accelerated Estimation of Long-Timescale Kinetics from Weighted Ensemble Simulation via Non-Markovian "Microbin" Analysis. *Journal of Chemical Theory and Computation* 2020, *16* (11), 6763–6775.

(72) Suarez, E.; Lettieri, S.; Zwier, M. C.; Stringer, C. A.; Subramanian, S. R.; Chong, L. T.; Zuckerman, D. M. Simultaneous Computation of Dynamical and Equilibrium Information Using a Weighted Ensemble of Trajectories. *Journal of chemical theory and computation* 2014, *10* (7), 2658–2667.

(73) Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Receptor Binding and Priming of the Spike Protein of SARS-CoV-2 for Membrane Fusion. *Nature* 2020, *588* (7837), 327–330. https://doi.org/10.1038/s41586-020-2772-0.

(74) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* 2015, *1–2*, 19–25. https://doi.org/10.1016/J.SOFTX.2015.06.001.

(75) Lobo, V. R.; Warwicker, J. Predicted PH-Dependent Stability of SARS-CoV-2 Spike Protein Trimer from Interfacial Acidic Groups. *Computational and Structural Biotechnology Journal* 2021, *19*, 5140–5148. https://doi.org/10.1016/j.csbj.2021.08.049.

(76) Phillips, J.; Zheng, G.; Kumar, S.; Kale, L. NAMD: Biomolecular Simulation on Thousands of Processors. In *Proceedings of the IEEE/ACM SC2002 Conference, Technical Paper 277*; IEEE Press: Baltimore, Maryland, 2002; pp 1–18.

(77) Brace, A.; Salim, M.; Subbiah, V.; Ma, H.; Emani, M.; Trifa, A.; Clyde, A. R.; Adams, C.; Uram, T.; Yoo, H.; Hock, A.; Liu, J.; Vishwanath, V.; Ramanathan, A. Stream-AI-MD: Streaming AI-Driven Adaptive Molecular Simulations for Heterogeneous Computing Platforms. In *Proceedings of the Platform for Advanced Scientific Computing Conference*; Association for Computing Machinery: New York, NY, USA, 2021.

(78) Lee, H.; Turilli, M.; Jha, S.; Bhowmik, D.; Ma, H.; Ramanathan, A. DeepDriveMD: Deep-Learning Driven Adaptive Molecular Simulations for Protein Folding. In *2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS)*; 2019; pp 12–19.

Part II: Mucin domain-containing Immunoprotein Modeling

Chapter 6 - Glycoproteomic landscape and structural dynamics of TIM family immune checkpoints enabled by mucinase SmE

6.1 **Abstract**

Mucin-domain glycoproteins are densely O-glycosylated and play critical roles in a host of biological functions. In particular, the T cell immunoglobulin and mucin-domain containing family of proteins (TIM-1, -3, -4) decorate immune cells and act as key checkpoint inhibitors in cancer. However, their dense O-glycosylation remains enigmatic both in terms of glycoproteomic landscape and structural dynamics, primarily due to the challenges associated with studying mucin domains. Here, we present a mucinase (SmE) and demonstrate its ability to selectively cleave along the mucin glycoprotein backbone, similar to others of its kind. Unlike other mucinases, though, SmE harbors the unique ability to cleave at residues bearing extremely complex glycans which enabled improved mass spectrometric analysis of several mucins, including the entire TIM family. With this information in-hand, we performed molecular dynamics (MD) simulations of TIM-3 and -4 to demonstrate how glycosylation affects structural features of these proteins. Overall, we present a powerful workflow to better understand the detailed molecular structures of the mucinome.

6.2 **Introduction**

Mucin-domain glycoproteins are characterized by extremely dense O-glycosylation that contributes to a unique, bottle-brush secondary structure which can extend away from the cell surface or form extracellular gel-like secretions.[1–3] Mucin-type O-glycans are characterized by an initiating α-N-acetylgalactosamine (GalNAc) that can be further elaborated into several core structures which can contain sialic acid, fucose, and/or ABO blood group antigens.[4,5] As a result, mucin domains serve as highly heterogeneous stretches of glycosylation that exert both biophysical and biochemical influence on the cellular milieu.[6,7] The canonical

family of mucins, e.g. MUC2 and MUC16, bear massive mucin domains that can reach 5-10 MDa in size and are heavily implicated in various diseases.[8,9] That said, many other proteins contain mucin domains that do not necessarily reach that size or complexity. Indeed, we recently introduced the human "mucinome", which comprises hundreds of proteins thought to contain the dense O-glycosylation that is characteristic of mucin domains.[10] For instance, platelet glycoprotein 1bα (GP1bα) interacts with Von Willebrand Factor to mediate platelet adhesion, and mutations in GP1bα are involved in platelet-type Von Willebrand disease.[11] C1 esterase inhibitor (C1-Inh) is a serine protease inhibitor and its deficiency is associated with hereditary angioedema.[12] Finally, the T cell immunoglobulin and mucin domain containing protein family (TIM-1, TIM-3 and TIM-4) are critical regulators of immune responses and are highly implicated in various cancers.[13,14] Though considerable progress has been made in the biological and analytical analyses of these and other mucin-domain glycoproteins, much remains unknown regarding their glycan structures, glycosylation site-specificity, and functional roles within the cellular environment.

This gap in knowledge is due, in part, to the challenges associated with studying mucins by mass spectrometry (MS).[15,16] A typical MS workflow involves digesting proteins with workhorse proteases like trypsin, subjecting the peptides to separation via reverse phase HPLC, then analyzing them by high-resolution MS.[17] Mucins present unique challenges at each stage of this process, but one of the most well-documented issues is the resistance of densely O-glycosylated domains to trypsin digestion.[3,5,18] To address this challenge, several proteases have been introduced that selectively cleave at or near O-glycosylated residues thereby revolutionizing the field of O- glycoproteomics.[19–22] These enzymes are aptly named O-glycoproteases; those that prefer mucin-domain glycoproteins are often termed mucinases. The first of these enzymes, OgpA (Genovis OpeRATOR), was characterized as an O-glycoprotease that cleaves N-terminally to glycosylated Ser or Thr residues but is hindered by the presence of sialic acid.[21,23] Shortly thereafter, we introduced StcE as a mucinase that selectively digests mucin domains with a cleavage motif of T/S*_X_T/S, wherein the asterisk indicates a mandatory glycosylation site; we followed this work with a mucinase toolkit displaying a wide range of cleavage specificities.[19,20] More recently, ImpA (NEB O-glycoprotease) was commercialized and, like OgpA, cleaves N- terminally to glycosylated Ser or Thr

residues but, unlike OgpA, is less restricted by the glycans present.[24] While these enzymes have aided in the analysis of many O-glycoproteins and mucins, each enzyme is accompanied by drawbacks: OgpA is limited by its resistance to sialic acid; the StcE cleavage motif is relatively restrictive; ImpA demonstrates preference for small amino acids adjacent to cleavage. Thus, an ideal, broad-specificity O- glycoprotease conducive to MS has not yet been characterized.

Another issue surrounding characterization of mucin-domain glycoproteins is that typical structural biology techniques are not well suited for glycoproteins, let alone the dense glycosylation characteristic of mucin domains. As covered in our recent review, current knowledge regarding mucin secondary structure originates from various low-resolution images generated by atomic force microscopy (AFM), scanning electron microscopy (SEM), and cryogenic electron microscopy (cryoEM).[3,25–27] While this has allowed us to definitively visualize the linearity of the mucin protein backbones, by nature of the techniques, we are unable to (a) discern the individual glycans and how they contribute to changes in protein structure, or (b) observe the mucin protein dynamics. Additionally, these methods often require large (>50 kDa), pure, and concentrated protein samples, which are difficult to obtain with most native mucin-domain glycoproteins.[3] More recently, and eloquently reviewed in reference 28, many advances have been made in computational modeling of glycoproteins.[28] These molecular dynamics (MD) simulations have revealed some of the many roles that glycans play in the structure, stability, dynamics, and function of glycoproteins. Most notably, Amaro and colleagues revealed, for the first time, the functional role of the glycan shield in the activation mechanism of the SARS-CoV-2 spike protein.[29,30] That said, while mucins have been subjected to MD simulations previously, they are often modeled with a static glycan structure, lack precise glycosylation information, and are simulated with coarse grained MD. Taken together, for mucin-domain glycoproteins, we generally do not know the glycoproteomic landscape nor how the glycans work in concert to control protein and cellular dynamics.

Here, we present a powerful technique to map the complex glycosylation within mucin domains and pair this information with MD simulations in order to better understand how glycans affect glycoprotein secondary structure and dynamics. We first introduce a new mucinase, Serratia marcescens Enhancin (SmE), and demonstrate that its unique ability to cleave at glycosites decorated by a myriad of glycans enabled enhanced glycosite and glycoform analysis by MS. We next showed that SmE outperforms the commercial glycoproteases OgpA and ImpA for these purposes, and we used molecular modeling to understand its uniquely broad tolerance for dense glycosylation. With SmE in-hand, we then obtained complete O-glycoproteomic information for all TIM family proteins and demonstrated that TIM-3 has markedly fewer O-glycosites when compared to TIM-1 and -4. To better understand how these glycans affect overall protein structure, we then employed MD simulations and showed that TIM-3 has a much shorter persistence length and higher flexibility than TIM-4, primarily attributed to the dense glycosylation in the latter. Overall, this workflow aids in unraveling the complex molecular mechanisms behind mucin domains, their glycan patterns, and their contribution to cellular biology.

6.3 **Results**

Characterization of SmE cleavage motif and glycan specificity

Various microorganisms found within mucosal environments secrete proteolytic enzymes that have been shown to be advantageous tools for MS analysis of mucins. We and others have mined the microbiota to generate a toolkit of O-glycoproteases, each having unique peptide and glycan specificities.[19,20,22,24] In particular, Serratia marcescens is a pervasive opportunistic pathogen in humans. This organism secretes a mucinase, SmE, which is a viral enhancin protein shown to promote arboviral infection of mosquitoes by degrading gut membrane-bound mucins.[31] Like previously characterized O-glycoproteases BT4244 and ImpA, SmE contains a catalytic domain belonging to the Pfam family PF13402 (peptidase M60, enhancin, and enhancin-like or M60-like family) that is defined by a conserved HEXXH metallopeptidase motif.[32] We expressed SmE as a 94-kDa soluble, N- terminal His-tagged protein in E. coli at a high-yield expression level

of 65 mg/L (Figure S6.1). To determine SmE's mucin selectivity, we digested glycoproteins with and without mucin domains at a 1:20 enzyme to substrate (E:S) ratio. SmE preferentially cleaved the mucin proteins C1 esterase inhibitor (C1-Inh), CD43, and TIM-1, whereas it did not significantly cleave the non-mucin glycoproteins fibronectin and fetuin (Figure S6.2).

As in our previous work, we then characterized SmE's cleavage motif using biologically relevant mucin-domain glycoproteins (C1-Inh, TIM-1, TIM-3, TIM-4, and GP1bα). These proteins were digested with SmE and subjected to MS analysis (Figure 6.1A). Manually validated glycopeptides were mapped to protein sequences to identify sequence windows, which were input into weblogo.berkeley.edu to determine minimum sequence motifs. As demonstrated in Figure 6.1B, SmE cleaved N-terminally to a glycosylated Ser or Thr. SmE accommodated a variety of O-linked glycans at the P1' position (pie chart, right), including sialylated core 1 and core 2 O-linked glycans; surprisingly, it also tolerated fucosylated ABO blood group antigens. SmE was also able to accommodate O-glycosylation at the P1 position (pie chart, left). Given the higher percentage of smaller O- glycan structures (GalNAc, GalNAc-Gal), it appears that the P1 O-glycosylation tolerance was less permissive at this position. However, these pie charts represent only site-localized glycan structures; site-localization at the C-terminus of the peptide is more challenging due to a lack of positive charge. Thus, the apparent preference for smaller glycan structures at the P1 position is likely due to issues in glycoproteomic analysis as opposed to an inability for SmE to cleave at residues bearing larger glycans.

SmE provides a complementary cleavage profile to StcE which, as mentioned above, cleaves at a T/S*_X_T/S motif. In that work, we also demonstrated that StcE is non-toxic to cells and can be employed to release mucins from the cell surface.[19] Many researchers have since used our mucinase toolkit, especially StcE, to remove mucins from the cell surface and/or degrade mucins in biological samples.[1,33–36] Importantly, the MUC1 repeat sequence HGVTSAPDTRPAPGSTAPPA does not contain StcE's cleavage motif, so limited digestion occurs within this region.[37] Given that SmE has a complementary cleavage motif, a combinatorial treatment strategy could enable further degradation of mucins from various biological samples. Thus, we sought to understand whether SmE is similarly capable of digesting mucins from the cell surface, and

whether the enzyme is likewise non-toxic to cells (Figure 6.1C). Treating HeLa cells with SmE resulted in a reduction in MUC16 staining by Western blot in a manner comparable to that of StcE (Figure 6.1D). We also detected released MUC16 fragments in the supernatant of SmE treated cells (Figure S6.3). Importantly, SmE is not toxic to cells under conditions used previously for StcE, although cell death is observed at higher SmE concentrations over longer treatment durations (Figure 6.1E and Figure S6.4). Taken together, these results indicate that SmE, like StcE, can effectively cleave mucins from the cell surface as a tool to probe mucin biological function.

In summary, SmE is a mucinase with the unique ability to cleave between two glycosylated residues bearing complex O-linked glycans. We anticipate that the characteristics of SmE's cleavage motif and glycan specificities will facilitate not only improved glycoproteomic mapping of mucin-domain glycoproteins, but also clearance of mucins from biological samples.

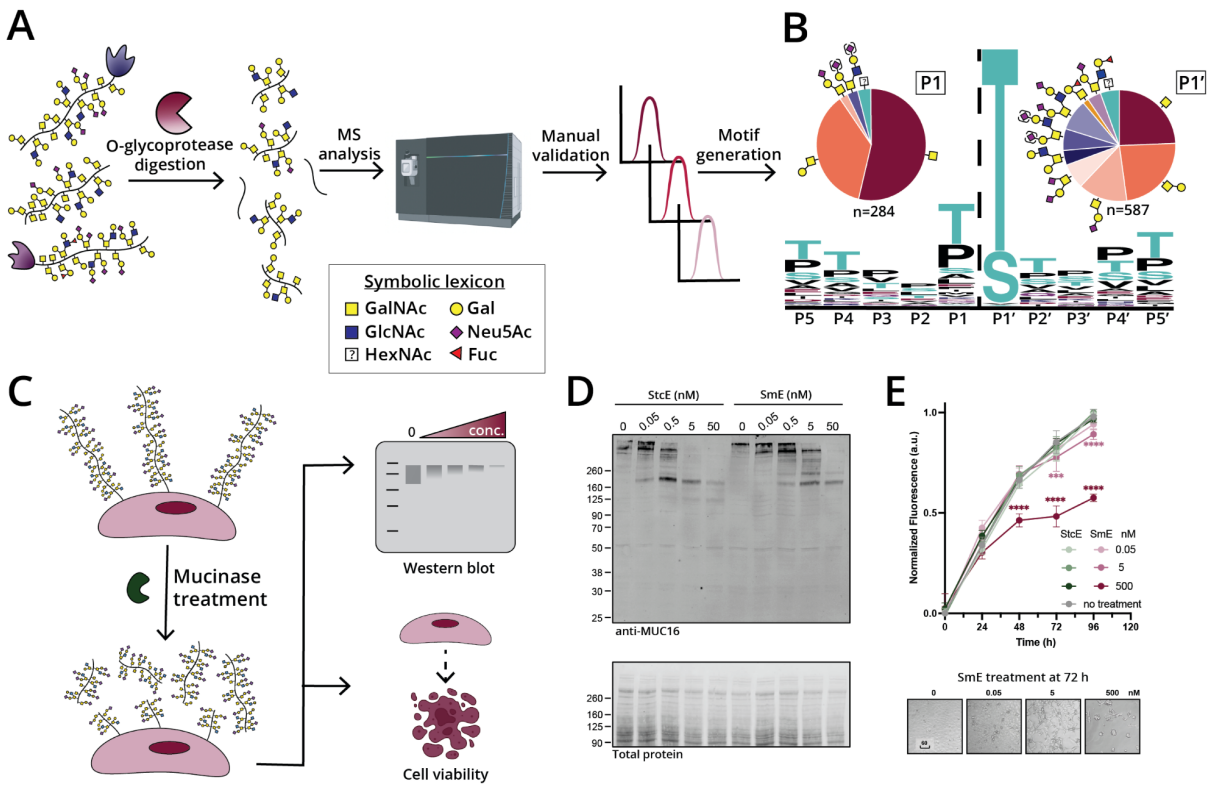


Figure 6.1: Characterization of mucinase SmE for analysis and degradation of mucin-domain glycoproteins

(A) Workflow for generating consensus sequence of SmE. (B) Five recombinant mucin-domain glycoproteins were digested with SmE and subjected to MS analysis. Peptides present in the mucinase-treated samples were used as input for weblogo.berkeley.edu (±5 residues from the site of cleavage). Parentheses around sialic acids (purple diamond) indicate that its linkage site was ambiguous. (C) Workflow to evaluate the toxicity and cell surface activity of SmE. (D) HeLa cells were treated with StcE (left) or SmE (right) at the noted concentrations for 60 min. Following treatment, the cells were lysed in 1X NuPAGE LDS Sample Buffer with 25 mM DTT, subjected to separation by gel electrophoresis, and probed for MUC16 via Western blot. Proteins were transferred to a nitrocellulose membrane using the Trans-Blot Turbo Transfer System (Bio-Rad) at a constant 2.5 A for 15 min. Total protein was quantified using REVERT stain before primary antibody incubation overnight at 4 °C. An IR800 dye-labeled secondary antibody was used according to manufacturer's instructions for visualization on a LiCOR Odyssey instrument. (E) HeLa cells were treated with SmE and StcE at 0, 0.05, 5, and 500 nM. At t = 0, 24, 48, 72, 96 hours post treatment, PrestoBlue was added according to manufacturer's instructions. After 2 hours, the supernatant was transferred to a black 96 well plate and analyzed on a SPECTRAmax GEMINI spectrofluorometer using an excitation wavelength of 544 nm and an emission wavelength of 585 nm. Statistical significance was determined using the two-way ANOVA analysis in Graphpad PRISM software and is reported with respect to the no mucinase control condition. *** p <0.001, **** p <0.0001. Scale bar = 60 µm.

SmE outperforms commercial O-glycoproteases OgpA and ImpA for mucin analysis

To compare the activity of SmE in context with widely used, commercially available O-glycoproteases, we decided to benchmark against both OgpA and ImpA.[21,24,38,39] OgpA was originally identified in Akkermansia muciniphila, a commensal bacterium known to regulate mucin barriers through controlled degradation.[40] As mentioned above, OgpA is reported to cleave N-terminally to O-glycosylated Ser or Thr residues, with highest affinity towards asialylated core 1 species.[41,42] Thus, typical workflows with this enzyme involve removal of sialic acids, which limits its use to site-mapping as opposed to providing information on native glycan structures. ImpA is derived from Pseudomonas aeruginosa, an opportunistic bacterial pathogen that can cause severe infection.[43] Like OgpA, ImpA cleaves N-terminally to an O-glycosylated Ser or Thr residue; however, this enzyme has been reported to accommodate more complex, sialylated glycans, expanding its glycoproteomic potential beyond that of OgpA. Efficiency of cleavage by ImpA has been observed to be influenced by amino acid identity in the P1 position. For example, reduced efficiency was seen when P1 was occupied by Arg or Ile, and cleavage was not observed when occupied by Asp.[24,39]

To directly compare the activities of OgpA, ImpA, and SmE, mucin-domain glycoproteins TIM-1, -3, -4, GP1bα, and C1-Inh were digested in the presence and absence of sialidase, followed by gel electrophoresis (Figure S6.5), MS analysis, and manual glycopeptide validation (see SI Tables S6.1-5 for all annotated glycopeptides). Additionally, we included fetuin to investigate the enzymes' selectivity for mucin glycoproteins (SI Table S6.6). As demonstrated in Figure 6.2A-B, we confirmed the reported cleavage motifs and glycan preferences of both OgpA and ImpA. Notably, after ImpA digestion, we did not detect any glycosites in the P1 position, suggesting that ImpA does not cleave between two glycosylated residues. Given that mucin domains contain many neighboring O-glycosites, this presents a significant limitation in the use of ImpA for mucinomic analysis.

In contrast, we found that SmE activity was not limited by glycan complexity or adjacent glycosylation (Figure 6.1B, 6.2C). Perhaps for this reason, digestion with SmE greatly improved the depth and coverage of the glycoproteomic landscape for each mucin-domain glycoprotein we investigated. Number of cleavage

sites, unique O-glycosites, mapped glycan structures, and total glycoforms were determined (Figure 6.2D, see Figure S6.6 for maps of all cleavage events). Note that for cleavage sites and unique O-glycosites, we included the sialidase treated samples; however, for the total mapped glycan structures and unique glycoforms, we only considered digests without sialidase treatment, as its use inherently limits identification of native glycan complexity. Here, total mapped glycan structures are calculated by counting every O-glycan associated with each O-glycosite. Unique glycoforms refers to the total number of validated glycopeptides that were identified from protein digestions. In our analyses, we found that OgpA allowed for the identification of 113 glycosites, 182 mapped glycan structures, and 358 unique glycoforms. ImpA demonstrated significant improvement over OgpA, enabling localization of 145 glycosites, 300 mapped structures, and 570 glycoforms. Even more impressively, SmE digestion allowed us to identify 205 glycosites, 455 mapped structures, and 887 glycoforms. Notably, use of SmE permitted the identification of 47 unique glycosites, 221 glycan structures, and 498 glycoforms that were not detected using the other enzymes. Previously, glycomic and glycoproteomic analyses of C1-Inh hinted at a total of approximately 25 O-glycosites; however, these were sparingly localized to individual residues.[12,44,45] SmE enabled full glycoproteomic mapping of the C1-Inh mucin domain (Figure S6.7, SI Table S6.1), thus reinforcing the utility of this enzyme. Taken together, SmE greatly outperformed both OgpA and ImpA with regard to glycoproteomic analysis of mucin domains.

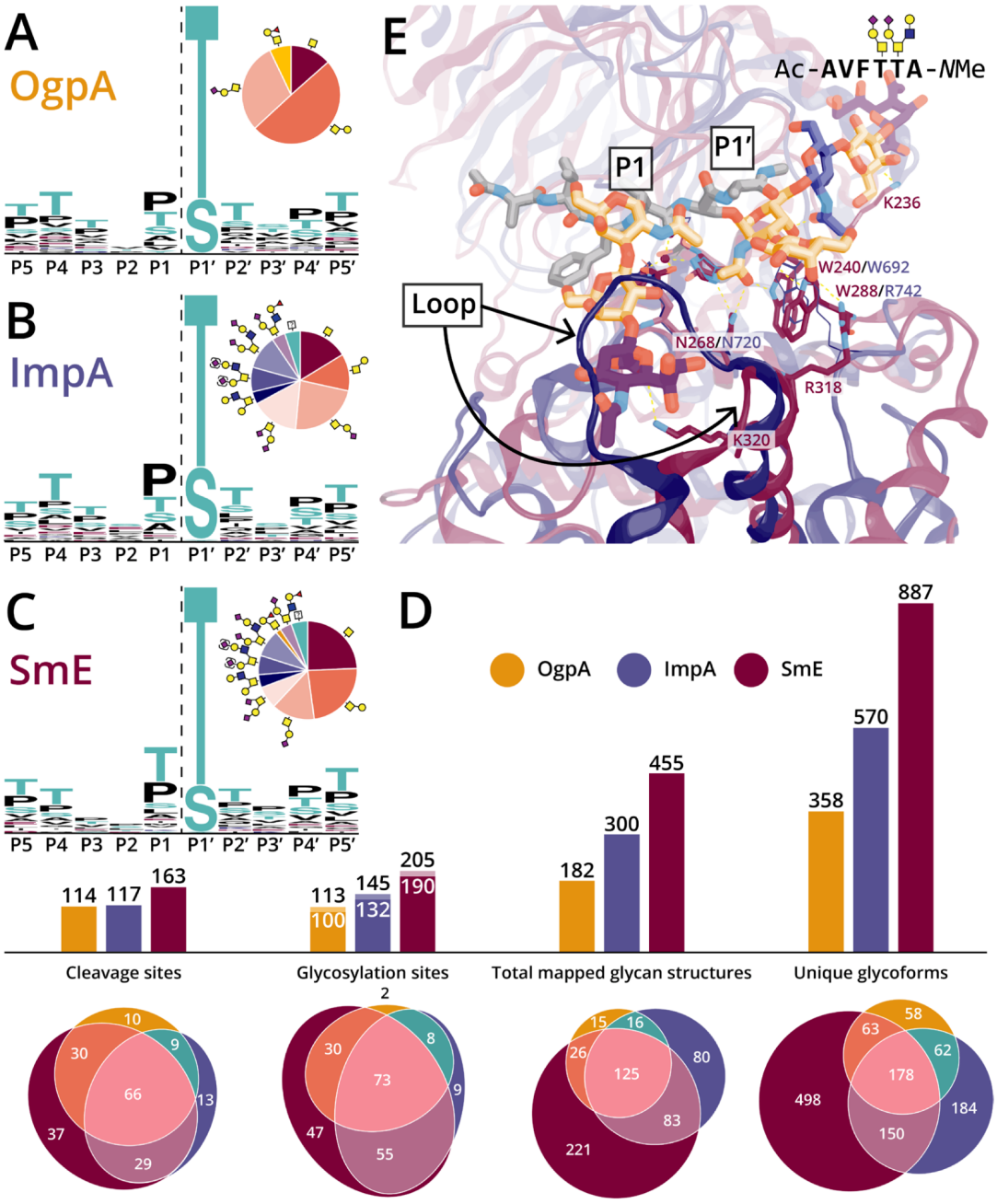


Figure 6.2: SmE outperformed commercial O-glycoproteases due to its structural permissiveness

Cleavage motifs for (A) OgpA, (B) ImpA, and (C) SmE as determined by digestion followed by MS and manual curation of glycopeptides. (D) Bar graphs and Euler plots demonstrating counts and overlap between enzymes regarding the number of observed cleavage sites, localized glycosylation sites, total glycan structures, and unique glycoforms. For the "Glycosylation sites" bar graph, glycosites localized via MS are denoted by white numbers; black numbers above include implied glycosites where cleavage was observed but the glycosite was not localized. (E) A glycopeptide docked in the active site of SmE (maroon) and ImpA (blue), highlighting differences between key loops and residues of the two O-glycoproteases.

Although SmE clearly exhibited benefits for mucin glycoprotein analysis, we observed certain limitations associated with its use. Relative to its performance in mucin domains, the efficiency of SmE was greatly diminished when used on non-mucin glycoproteins, such as fetuin (Figure S6.6; SI Table S6.1). Analysis by MS revealed that SmE was able to cleave non-mucin glycoproteins to a limited extent; however, the abundance of fetuin glycopeptides resulting from SmE digestion was significantly reduced when compared to those of ImpA, and in some cases, OgpA (Figure S6.8). We also observed that SmE digestion was most effective against large (>50 residue) mucin domains. These observations, along with the digestion assay in Figure S6.2, support the notion that SmE is a mucin-selective O-glycoprotease. However, while SmE greatly enhanced sequence coverage and depth, some glycosylation was only localized through the use of OgpA and ImpA (Figure 6.2D). For this reason, we recommend a complementary, multi-enzyme approach to fully elucidate the glycoproteomic landscape of mucin-domain glycoproteins. Additionally, since ImpA was previously reported to have P1 selectivity, we generated "anti-logos" to determine surrounding residues that were unfavorable for cleavage. Here, we considered the total cleavage maps depicted in Figure S6.6, and whenever an enzyme did not cleave at an observed cleavage site, we took the surrounding amino acids and generated a logo. Interestingly, we found that SmE has a lower cleavage efficiency at Ser residues (Figure S6.9). While this could be because Ser is less often decorated with mucin-type O-glycosylation, it might be an important consideration for digestion of mucin domains bearing high levels of Ser residues. Finally, we observed that SmE exhibited reduced proteolytic activity on recombinant proteins expressed in murine-derived cell lines, as exemplified by (a) the digestion of TIM-1 from NS0 and HEK cells over the course of six hours (Figure S6.10) and (b) limited digestion of CD43 derived from NS0 cells (Figure S6.2).

Molecular modeling helps rationalize different substrate selectivity between SmE and ImpA

Previously, we used molecular docking to better understand StcE's substrate selectivity.[19] Given that SmE and ImpA have catalytic domains belonging to the same Pfam family,[46,47] yet have quite different cleavage motifs, we decided to again use molecular modeling to understand the structural basis behind these differences. In addition to its catalytic domain, SmE has two mucin-binding modules (PF03272) while ImpA has unique helical (PF18642) and N-terminal (PF18650) domains. OgpA, by contrast, is a single-domain

enzyme with a catalytic metzincin motif that is not defined by Pfam and is more distantly related to SmE and ImpA, and therefore was excluded from these comparison analyses.

To date, four unique crystal structures of ImpA have been determined,[32,48] including one with a ligand bound at the active site and a second with a ligand bound at an exosite located in the N-terminal domain (PF18650). The structure of SmE, on the other hand, has not yet been elucidated. As such, we aligned structures of all characterized enzymes with a PF13402 catalytic domain,[49–52] including the SmE structure recently predicted by AlphaFold (Figure S6.11).[53,54] We then docked a TIM-4-based bisglycosylated peptide into the predicted SmE structure to provide insight into potential substrate recognition.

In its cocrystal structure with (Gal-GalNAc)Ser,[32] ImpA uses specific residues to recognize glycans branching from P1' (Figure 6.2E, blue). The side chains of the conserved residues Trp692 and Asn720 form polar contacts with the carbonyl of the N-acetyl group of the GalNAc moiety, and the side chain of Arg742 also interacts with the 3-OH and 4-OH of GalNAc; the Gal moiety, on the other hand, does not interact with the enzyme and is projected into solvent. The helix lining the active site of ImpA is also short,[38] indicating that branched glycans – though absent from the crystallized ligand – could be reasonably accommodated by this enzyme in a similar manner to ZmpB and ZmpC.[32] These combined factors likely impart ImpA with activity on substrates bearing mucin-like glycosylation at P1' but little selectivity for particular modifications beyond the initiating GalNAc moiety, consistent with our MS findings.

In our docked structure of SmE with a bisglycosylated TIM-4-based glycopeptide (Figure 6.2E, red), we found analogous interactions between conserved residues Trp240 and Asn268 and the GalNAc initiating from P1'. Unique contacts were found between the 3-OH of GalNAc and the side chain of Trp288, rather than with an Arg residue as seen in crystal structures of ImpA and other PF13402-containing enzymes. While there is an Arg residue (Arg282) nearby in the sequence, it is not predicted to flank the GalNAc moiety in the AlphaFold structure, and we found that this residue is less conserved in PF13402-containing enzymes than previously suggested (Figure S6.11).[49] Interestingly, a different Arg side chain (Arg318) contacted the 4-OH of the Gal residue, likely imparting further specificity for mucin-like glycosylation. Similar to ImpA, SmE is

predicted to have a short active site helix. Here, the branched glycan was well accommodated by the enzyme, and we commonly found orientations that formed contacts between the ligand and the active site helix as well as the preceding loop, similar to what was observed between ZmpB/ZmpC and their branched ligands (Figure S6.12).[32,50] Together, these results suggest that SmE has better recognition of the initiating GalNAc and Gal residues, and that it likely forms additional interactions with branched glycans.

Neither the ImpA crystal structure nor the SmE predicted structure contain a beta hairpin analogous to the one found to recognize P1 glycans by the mucinase AM0627 (Figure S6.11, S6.13), which also allows glycosylated P1 Ser/Thr.[52] Thus, the steric environment in this region is primarily defined by a single loop that is significantly larger in ImpA (Trp770-Leu778) than it is in both SmE (Asn313-Asp316) and AM0627 (Leu384-Asp388). In our docked structure, we observed that the short loop of SmE allowed the enzyme to easily accommodate the P1 glycan, with neither the GalNAc nor the Gal residue forming direct contacts with the enzyme; the sialic acid residue could interact with the enzyme or project toward solvent. This is in contrast to AM0627, which forms direct contacts between its beta hairpin and the GalNAc and Gal residues to impart requirement for P1 glycosylation.[20,49,52] The long loop in ImpA, by contrast, sterically clashes with all three subunits of the P1 glycan, which explains why ImpA is unable to cleave between adjacent residues bearing glycosylation. More broadly, these and prior findings suggest a delicate interplay between the hairpin and loop in determining this enzyme family's tolerance, preference, or requirement for particular glycans at P1 (Figure S6.13). In the case of SmE, the short loop and absence of a hairpin allows the enzyme to tolerate (but not require) larger glycans at the P1 position, which again supports our MS findings.

Molecular modeling identifies potential secondary mucin binding in SmE

Multidomain mucinases are hypothesized to arrange their noncatalytic domains into an architecture that enables specific recognition of secondary sites along the linear bottle-brush of mucins,[50,55] and recombinant StcE lacking one of these noncatalytic domains showed reduced activity on mucin substrates.[56,57] Interestingly, the related metalloprotease MMP-1 (collagenase) required an accessory domain to bind and cleave the linear triple helix of collagen;[58] structural and functional studies revealed the importance of

cooperativity between this enzyme's catalytic and accessory domains as well as specific interactions between each domain and the collagen triple helix.

In one crystal structure, ImpA binds a glycopeptide in an exosite located within its noncatalytic N-terminal domain (PF18650).[48] During our initial docking study, we observed that one of the accessory mucin-binding modules (Asn537-Leu650, PF03272) in SmE is also positioned to recognize additional sites in mucin substrates. At present, there is no experimentally determined structure of a PF03272 domain; thus, the precise details of the domain's structure and ligand recognition remain unknown. As such, we grafted larger segments of TIM-4 (vide infra) onto the docked glycopeptide to determine if this accessory domain can potentially bind the substrate. We observed that several of the larger TIM-4 substrates positioned glycans adjacent to a predicted binding pocket formed from two conserved segments in the mucin-binding module (Figure S6.14, S6.15). While additional work is required to validate this initial finding, the result suggests that SmE could use its mucin-binding module to cooperatively bind mucin substrates and sterically occlude more globular O-glycoproteins. Such a model would explain SmE's observed preference for mucins over non-mucin O-glycoproteins like fetuin.

Glycoproteomic mapping of TIM-1, -3, and -4

With this new tool in-hand, we reasoned that SmE could be used to sequence immune checkpoint mucin-domain glycoproteins at the molecular level. In particular, TIM-1, -3, and -4 are key players in immune cell function and are predicted to be modified by many O-glycosylation sites.[59] Each protein contains an N-terminal variable immunoglobulin (IgV) domain followed by a densely glycosylated mucin domain of varying length, a single transmembrane domain, and a C-terminal intracellular tail.[60] TIM-3 is highly implicated in cancer pathways, thus much of the literature to date has focused on better understanding its molecular interactions. In short, when TIM- 3 is not bound to its extracellular ligands (including phosphatidylserine (PtdSer), high mobility group box 1 protein (HMGB1), and/or galectin-9 (Gal-9)) via its IgV domain, the TIM-3 cytoplasmic tail induces phosphorylation of the T cell receptor (TCR), which promotes T cell proliferation and survival. However, when TIM-3 is bound to one of its ligands, the cytoplasmic tail becomes phosphorylated

and consequently promotes a state of T cell exhaustion that is characteristic of many cancers (Figure 6.3A, left).[61] As such, several antibodies against TIM-3 are currently being investigated as cancer immunotherapies, often in combination with canonical checkpoint inhibitors like PD-1.[62–64]

Compared to TIM-3, little is known about TIM-1 and TIM-4, potentially because these proteins are predicted to bear more O-glycosites than TIM-3, thus complicating their analysis. However, it is known that the combination of TIM-1 blockage and TCR stimulation promotes T cell proliferation and cytokine production;[65] TIM-4 is a PtdSer receptor and binds when PtdSer is exposed on the surface of apoptotic cells (Figure 6.3A, right).[66] While much remains to be discovered about TIM-1 and -4, it is apparent that the entire TIM family plays critical roles in regulating immune responses in normal and dysregulated cellular states. However, only predicted glycosylation sites in the TIM family have been discussed in the literature, leaving their true glycoproteomic landscape a mystery. It follows, then, that we also do not understand how glycosylation contributes to TIM protein ligand binding, structural dynamics, and intracellular signaling. Ultimately, this lack of information hampers our understanding of the TIM family structure and function, which could have strong implications for cancer immunotherapy.

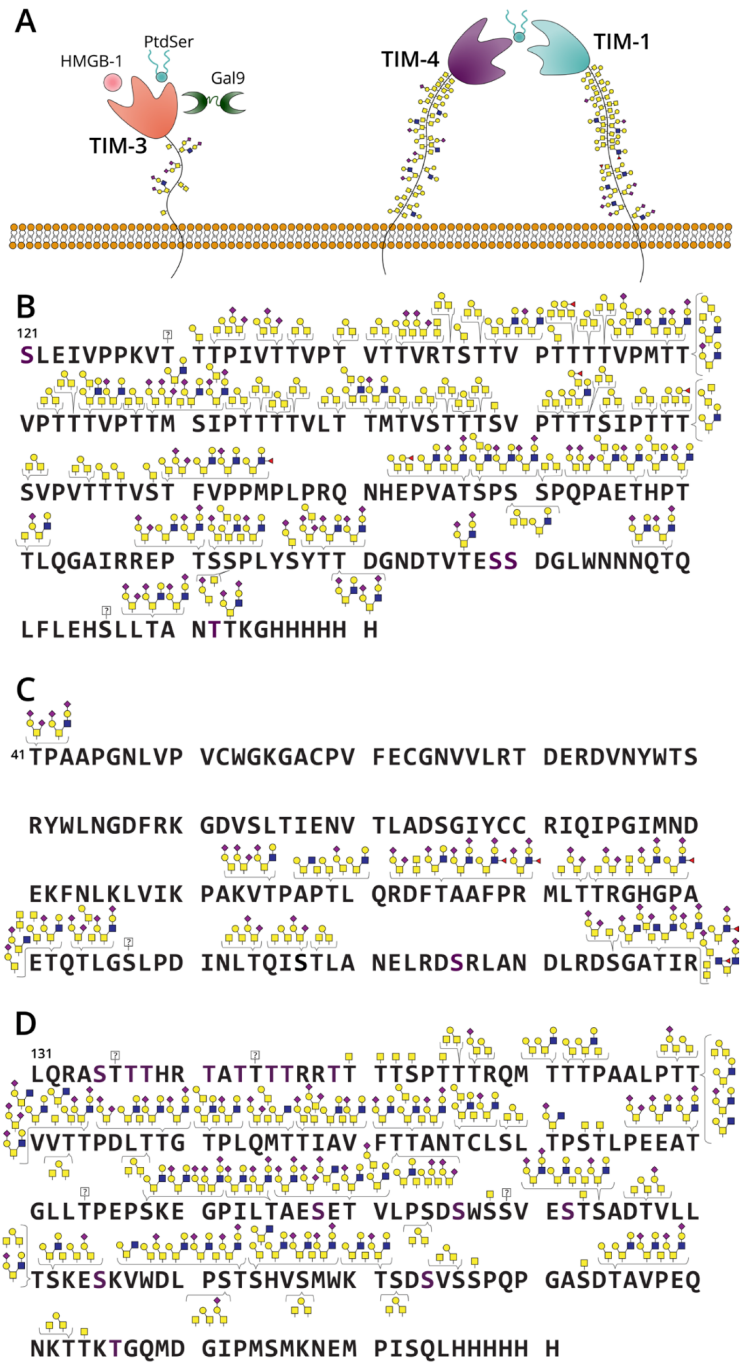


Figure 6.3: Glycoproteomic mapping of TIM family proteins

(A) Cartoon of TIM family structure and ligand interactions. TIM-3 interacts with ligands PtdSer, HMGB1, and/or Gal-9; through intracellular signaling these interactions deactivate T cell function and cytokine release. TIM-1 and TIM-4 purportedly interact through PtdSer to enact effector function. Recombinant TIM-1 (B), TIM-3 (C), and TIM-4 (D) were subjected to digestion with SmE, ImpA, OgpA, and/or trypsin followed by MS analysis and manual data interpretation. Brackets indicate glycans sequenced at each Ser/Thr residue at >5% relative abundance. For full glycoproteomic sequencing data, see SI Tables S6.4, S6.7, and S6.8.

According to NetOGlyc 4.0,[67] TIM-1, -3, and -4 were predicted to bear 67, 8, and 66 O-glycosites, respectively. Additionally, we recently developed a "Mucin Domain Candidacy Algorithm" which takes into account predicted O-glycosites, glycan density, and subcellular location in order to output a "Mucin Score".[10] This value was developed as a method to gauge the likelihood that a protein contains a mucin domain; a protein receiving a score above 2 indicated a high probability. Interestingly, TIM-1 and TIM-4 scored above 6, whereas TIM-3 received a score of 0.[10] Beyond the biological implications of these proteins, we were curious to understand the glycoproteomic landscape of the TIM family given the large disparity in predicted O-glycosites and Mucin Scores. Thus, we digested the three recombinant TIM protein ectodomains with SmE and performed MS analysis followed by manual curation of the glycopeptides. Given our earlier observations regarding SmE's resistance to proteins bearing sparse glycosylation, we digested TIM-3 with ImpA and OgpA to ensure full sequence coverage and O-glycosite identification. As seen in Figure 6.3B with a full list of annotated glycopeptides in SI Tables S6.4, S6.7- 8, we identified all 67 of the predicted O-glycosites on TIM-1; many of these sites were modified by a myriad of O-glycans, thus demonstrating the massive microheterogeneity in mucin domains. TIM-4 was similarly dense in glycosylation and we site-localized a total of 51 O-glycosites (Figure 6.3D). In contrast, only 14 sites of O- glycosylation were detected on TIM-3; the glycosylation was also much less dense, though still quite heterogeneous (Figure 6.3C). Generally, recombinantly expressed proteins are thought to display relatively simple glycosylation (e.g., core 1 or 2 structures). Intriguingly, despite the fact that these proteins were recombinantly expressed in HEK293 cells, we observed not only these glycans, but also highly sialylated and fucosylated structures, and surprisingly glycopeptides with core 4 O-glycans (Figure S6.16). This suggests that when searching MS data, larger glycan databases might be necessary to encompass all of the glycan structures displayed on the complex family of mucin proteins.

MD simulations of TIM-3 and -4 elucidate the structural and dynamical impact of glycosylation

Following the initial glycoproteomic mapping, we asked how these extremely different glycosylation patterns could affect the structures, and potentially functions, of the TIM proteins. While the IgV domain structures have been solved via X-ray crystallography and NMR, the mucin domains were excluded from

analysis, presumably due to the high heterogeneity and density of O-glycosylation.[68,69] After failed attempts to perform cryoEM on the full TIM-3 and -4 ectodomains, we reasoned that molecular modeling and MD simulations could be an alternative method to predict mucin domain structure and better understand how glycosylation contributes to dynamic properties of these proteins. As described above, the microheterogeneity of each glycosite was incredibly high; thus, in order to accurately reconstruct the TIM glycoproteomic landscape, we needed to identify the most abundant glycan at each residue. To do so, we generated extracted ion chromatograms (XICs) of every glycopeptide detected from TIM-1, -3, and -4 to calculate area-under-the-curve relative quantitation using Thermo Xcalibur (SI Tables S6.4, S6.7-8). The relative abundance of every detected glycan, at each O-glycosite, is depicted in Figure 6.4A (right). The density and heterogeneity of glycosylation are apparent in TIM-1 and -4, but less so in TIM-3. Additionally, as others before have suggested, the glycan size and heterogeneity was much lower in areas of dense glycosylation; sparse O-glycosites afforded larger and more diverse glycan structures.[70,71] By obtaining the most abundant O-glycan at each glycosite, we built two all-atom computational models of the fully glycosylated transmembrane glycoproteins TIM-3 and TIM-4 (Figure 6.4A, left) to better understand the contribution of glycan density on the overall flexibility and length of TIMs. Each of these systems contained their respective globular IgV domain, mucin-domain, alpha helical transmembrane domain, and cytoplasmic tail. Approximately 700 and 830 ns of simulation data were generated for TIM-3 and -4, respectively (see SI Methods for full simulation details).

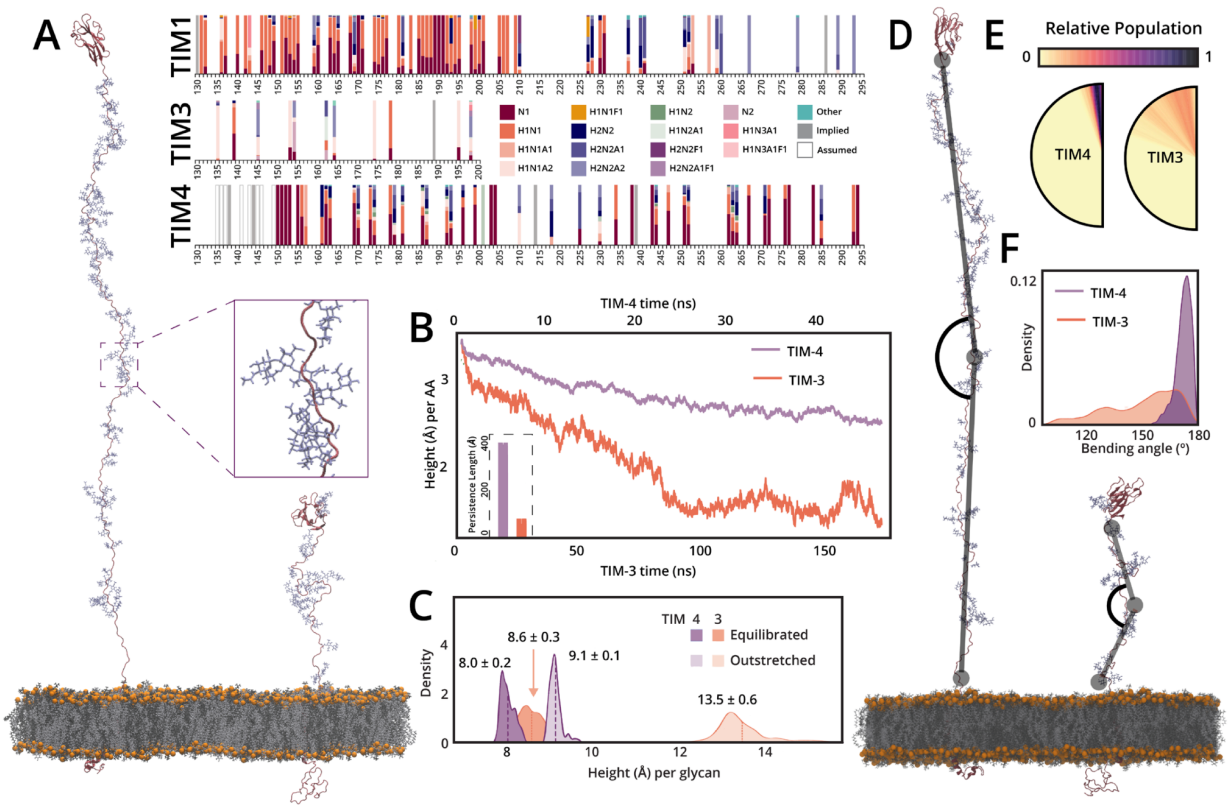


Figure 6.4: MD simulations of TIM-3 and -4 elucidate the structural and dynamical impact of glycosylation

(A, right) XICs were generated for each glycopeptide from TIM-1, -3, and -4 and area-under-the-curve quantitation was performed; glycan composition color legend shown in center. N - HexNAc (GalNAc), H - hexose (galactose), A - NeuAc (sialic acid), F - fucose. (A, left) Image detailing TIM- 4 and TIM-3 models along with an inset view highlighting the dense TIM-4 glycosylation. (B) End-to-end distance of TIM-3 and TIM-4 mucin domains normalized by total length (number of amino acids, AA) within the mucin domains, plotted as a function of simulation length. (inset) Persistence length calculated for TIM-3 and TIM-4 from all simulation replicas. (C) Histograms detailing the end-to-end distance of TIM-3 and TIM-4 mucin domains, normalized by total number of glycans, in the outstretched (starting) conformation (lighter distributions) and equilibrated conformation (darker distributions). (D) Image demonstrating the "bending angle" as calculated in the following panels. (E) Semi-circles graphically detailing the bending angles visited over the complete course of simulations for TIM-3 and -4, angles colored according to relative population. (F) Histograms detailing bending angles sampled by TIM-3 and -4 over the course of all simulations.

To identify the degree to which TIM-3 and TIM-4 mucin domains compress during simulation, we calculated a normalized end-to-end distance of each protein's mucin domain as a function of time (Figure 6.4B). From these results, we see that TIM-3's mucin domain, with only 14 glycans and 71 amino acids, compresses far more significantly than TIM-4's mucin domain, containing 51 glycans and 179 amino acids. We quantified this change by calculating persistence length, which was defined as the distance, in

angstroms, at which the motions of two monomers along a polymeric chain become decorrelated from one another. Strong intramolecular interactions within monomers can lead to highly correlated motions along the polymeric chain, overcoming energetic gains of interactions with solvent or enhanced conformational degrees of freedom, and thus long persistence lengths. Using data from our MD simulations, we calculated the persistence lengths of the TIM-3 and -4 mucin domains to be 81 ± 24 Å and 415 ± 10 Å, respectively. Thus, these two mucin-domains have drastically different degrees of correlation within their protein backbones, likely originating from their varied degrees of glycosylation (Figure 6.4B, inset).

During initial analysis and trajectory visualization, we noticed the TIM-3 mucin domain underwent a significant degree of bending such that the IgV domain tilted toward the membrane. To better quantify this, we calculated the angle between two vectors for both TIM-3 and -4 mucin domains: one drawn from the central residue up to the first residue, and another drawn from the central residue down to the last residue (Figure 6.4D, see SI Methods for complete details). We observed that the TIM-4 mucin domain largely sampled bending angles close to 180°, i.e., the TIM-4 mucin domain was largely linear and bottle-brush like. The TIM-3 mucin domain, however, bent quite significantly and sampled a large range of angles with similar probabilities (Figure 6.4E-F).

The mucin domains of TIM-3 and -4 are variably dense in terms of glycosylation. To investigate this effect on mucin identity and functional dynamics, we aimed to quantify total versus effective glycosylation in a thoroughly glycosylated mucin domain (as in TIM-4) versus in a sparsely glycosylated mucin domain (as in TIM-3). Herein, we define the total glycosylation rate as the ratio of the length of the outstretched, unequilibrated mucin domain protein backbone to the total number of glycans. Similarly, we define effective glycosylation as the ratio of the length on a relaxed, equilibrated mucin domain protein backbone to the total number of glycans. These two values thus illustrate the “height per glycan (Å)” under outstretched and relaxed conditions. As shown in Figure 6.4C, the height per glycan in heavily glycosylated TIM-4 remains nearly the same in both outstretched and equilibrated states: 9.1 ± 0.1 Å and 8.0 ± 0.2 Å, respectively. This indicates that upon relaxation of the mucin domain protein backbone, O-glycans still maintain a similar distribution relative to one another as in the fully outstretched case, i.e., total glycosylation equals effective

glycosylation. However, for TIM-3, the height per glycan distance drops significantly following equilibration, going from 13.5 ± 0.6 Å to 8.6 ± 0.3 Å. In fact, following equilibration, this height per glycan distance seen in TIM-3 becomes similar to those distributions seen in TIM-4. Through trajectory visualization, specific glycan-glycan pairs in TIM-3 were found to be responsible for a large portion of the decrease in height per glycan distance. Specific distant pairs (≥3 glycans away from one another) of glycans are seen to interact via hydrogen bonding, almost as if these glycans are "holding hands," as exemplified by the TIM-3 glycan pair G7 and G4 (glycosylation sites T145 and T162, respectively; Figure S6.17). These results demonstrate the power of MD simulations in characterizing members of the mucinome, as glycoproteomic mapping alone cannot provide atomic-level structural insight, including conformational changes that may allow distant O-glycan pairs to find each other and reach new, functionally significant conformations. These combined methods thus have the potential to classify proteins more rigorously within the mucinome.

6.4 **Discussion**

Historically, numerous challenges have impeded the study of mucin-domain glycoproteins; however, new tools continue to be introduced to unveil mucin glycosylation status, functional roles, and biological impact. Here, we present an addition to this toolkit and use it to better understand mucin-domain glycoprotein structure and dynamics. We first thoroughly characterized the mucinase SmE, which demonstrated a uniquely broad cleavage motif and outperformed commercially available O-glycoproteases, thus enabling unprecedented glycoproteomic mapping of biologically relevant mucin proteins. In particular, we elucidated the glycosylation landscape of clinically relevant immune checkpoint proteins TIM-1, -3, and -4 and used this information to enable glycoproteomic-guided MD simulations for the first time. The data afforded by SmE treatment, in concert with MD simulations, has opened the door to a new realm of atomic-level insight into mucin-domain containing proteins, their structure-function relationships, and their recognition mechanisms by bacterial mucinases. Interplay between glycoproteomic data and molecular modeling offers the potential to expand upon benchmarks for determining mucinome membership, such as mucin-specific persistence

length, effective glycosylation, and flexibility. Ultimately, we developed a powerful workflow to understand detailed molecular structure and guide functional assays for all members of the mucinome.

That said, we have only begun to unlock the potential of this workflow, especially as pertained to the TIM family of proteins. To be sure, aberrant glycosylation is a hallmark of cancer and typical O-glycosylation changes involve truncation of normally elaborated glycan structures.[72] These shortened glycans could strongly impact the "linearity" of the mucin backbone, thus changing TIM protein protrusion from the glycocalyx. As such, transformed O-glycosylation could have implications in how the glycoproteins interact with each other, their ligands, and as a result, intracellular signaling and T cell cytotoxicity. Relatedly, this could greatly influence the efficacy of the 10 anti-TIM-3 antibodies currently being investigated in at least 26 clinical trials (clinicaltrials.gov). Future efforts will be devoted to glycoproteomic mapping of endogenous TIM proteins from primary T cells and patient samples to discover how glycan structures change in health and disease. In concert with biological assays, we will use this information to drive MD simulations that probe how altered glycosylation could affect ligand binding, intracellular interactions, and antibody recognition. Beyond the TIM family of glycoproteins, many other glyco-immune checkpoints have emerged as prominent mechanisms of immune evasion and therapeutic resistance in cancer;[33] we envision that our workflow will also help elucidate structure-function relationships in these proteins.

Aside from MS analysis, SmE encompasses the potential to make a larger impact on the field of glycobiology. In our previous work, we used StcE for clearance of mucins from the cell surface and determined that Siglec-7, but not Siglec-9, selectively bound to mucin-associated sialoglycans.[19] We then upcycled an inactive point mutant of StcE to develop staining reagents for Western blot, immunohistochemistry, and flow cytometry.[20] We also took advantage of the mutant StcE to develop an enrichment procedure that allowed for the selective pulldown of mucin glycoproteins.[10] Finally, and most recently, an engineered version of StcE conjugated to nanobodies was used for targeted degradation of cancer-associated mucins.[56] Given that SmE is similarly active on live cells, has a complementary cleavage motif, two mucin binding domains, and potentially different endogenous targets, future work will be aimed at investigating whether SmE can augment our current mucinase toolkit and therapeutic strategies.

Previously, we developed a "Mucin Domain Candidacy Algorithm" to help identify proteins that have a high probability of bearing a mucin domain.[10] While we recognized at the time that our definition of a mucin domain was novel but rudimentary, the present work has confirmed that our understanding of mucin domains is incomplete. TIM-1 and -4 were predicted by our algorithm to be high confidence mucins, whereas TIM-3 received a score of 0. Here, our glycoproteomic mapping combined with MD simulations demonstrated that absolute glycosylation (i.e., O-glycans per amino acid residue) can be dramatically different than effective glycosylation (i.e., O-glycosylation in relation to total surface area after folding). Thus, while density of O-glycosylation can absolutely be an indication that a mucin domain is present, it does not reveal the entire story, and our definition of a mucin domain continues to develop. While it would be ideal to obtain high-resolution structures of these mucin-domain glycoproteins, that is likely a long-term objective. In the meantime, our workflow is a tangible mechanism for visualizing the enigmatic mucin family to not only better understand the definition of a mucin domain, but to also study the structural dynamics that lie within.

Our ultimate objective is to unravel the complex molecular mechanisms behind glycan structures, patterns, and overall biological functions of mucin domains, but much remains to be accomplished. That said, this work serves as a significant advance toward that overall goal and will find use in furthering our understanding of the mucinome.

6.5 **Materials & Methods**

6.5.1 Materials

Recombinantly expressed TIM-1 and TIM-4 were purchased from R&D Systems (9319-TM, 9407-TM). For structural characterization, TIM-1 was purchased from R&D systems (11157-TM) and TIM-4 was purchased from LifeSpan Biosciences (LS-G139224). TIM-3 was purchased from LifeSpan Biosciences (LS-G97947).

CD43 and TIM-1 recombinantly expressed in NS0 cells were purchased from R&D systems (9680-CD,1750-TM). C1-Inh and fibronectin isolated from human plasma were purchased from Sigma

Aldrich (E0518, F1056). Bovine Fetuin-A was purchased from Promega (V4961). Gp1bα was isolated as described in reference 22.

The plasmid for His-tagged pET28a-SmEnhancin and recombinant StcE protein were kindly provided by the Bertozzi laboratory.

Expression and purification of SmEnhancin

For pET28a-SmEnhancin plasmid extraction, a 50 µL aliquot of chemically competent E. coli DH5α cells (NEB, C2988J) was thawed on ice. Approximately 100 ng of plasmid was added to the cells and incubated on ice for 30 minutes. Cells were then transformed by heat-shock at 42 °C for 30 seconds. Room temperature SOC media (Invitrogen, 15544-034) was added (950 µL) to the cells then incubated at 37 °C with agitation at 250 rpm for one hour. A 150 µL aliquot of the transformed cells were then transferred to a LB-agar (Fisher, BP1425) plate with kanamycin (Sigma Aldrich, K1377) and incubated overnight at 37 °C. Kanamycin was used throughout at a final concentration of 50 µg/mL. A single colony was picked and used to inoculate an overnight culture of 100 mL Luria broth (LB) (Sigma Aldrich, L3022) with kanamycin. The culture was incubated at 37 °C with agitation at 250 rpm. pET28a-SmEnhancin plasmid DNA was extracted using a Qiagen Plasmid Midi Kit (Qiagen, 12143) using the protocol provided by the manufacturer. DNA concentration was determined by NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo-Fisher) then stored at -80 °C. After extraction, the DNA sequence was verified using Plasmidsaurus.

For protein expression of SmEnhancin (SmE), a 20 µL aliquot of competent E. coli BL21(DE3) cells (Millipore Sigma, 69450-4) was thawed on ice. Approximately 10 ng of pET28a-SmEnhancin was added to cells and incubated on ice for 5 minutes. Cells were then transformed by heat-shock at 42 °C for 30 seconds. Room temperature SOC media was added (80 µL) to the cells then incubated at 37 °C with agitation at 250 rpm for one hour. A 100 µL aliquot of the transformed cells were then transferred to a LB-agar plate with kanamycin and incubated overnight at 37 °C for colony growth. A single colony was then picked to inoculate a 10 mL overnight culture of LB with kanamycin and incubated at 37 °C with agitation at 250 rpm. A glycerol stock was made by mixing 4 mL of the overnight culture with 4 mL of 50% glycerol

(Sigma Aldrich, G7893) and stored at -80 °C. The remaining overnight culture was used to inoculate a 1L LB culture with kanamycin. This culture was also incubated at 37 °C with agitation at 250 rpm until it reached an optical density of 0.6-0.8. The maxi culture was then induced with a final concentration of 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) (American Bio, AB00841) and grown overnight at 16 °C with agitation at 250 rpm. The bacterial cells were harvested by centrifugation at 3000 x g for 45 min at 4 °C. The supernatant was decanted and the cell pellet was stored at -80 °C until lysis was performed.

The cell pellet was lysed in a buffer containing 20 mM Tris-HCl at pH 8 (Thermo Scientific, J3636.K2), 200 mM NaCl (Fisher Scientific, S25877), 2 mM magnesium chloride ($MgCl_2$) (American Bio, AB09006), 10% glycerol, and 125 U/mL Benzonase nuclease (Sigma Aldrich E1014). Additionally, 100 µg/mL lysozyme (Thermo Scientific, 89833) and 1% Triton X-100 (Alfa Aesar, J66624) were added to aid cell lysis. For inhibition of protease activity, a cOmplete Mini EDTA-free protease inhibitor cocktail (Roche, 11836170001) was used alongside 1 mM phenylmethylsulfonyl fluoride (PMSF) (American Bio, AB01620). Cells were then resuspended using 1 mL of chilled lysis buffer per gram of cells. The cell suspension was further homogenized by five pulses of probe sonication with 5 seconds of sonication at 35% amplitude followed by 15 second pauses (QSonica Q500). The solution was kept on ice throughout sonication to prevent protein denaturation. Lysate was clarified by spinning at 25,000 x g for 45 min at 4 °C and the supernatant was filtered sequentially through 0.45 µm (Millipore, SLHAR33SS), and 0.2 µm (Cytiva, 10462300) syringe filters.

The protein was purified using an ÄKTA Pure FPLC (Cytiva) with a HisTrap HP column (Cytiva, 17524801). The column was equilibrated for 5 column volumes (CV) at 1 mL/min using buffer A (20 mM Tris-HCl pH8, 500 mM NaCl, 25 mM imidazole (Sigma Aldrich, I202)) prior to loading the sample at a flow rate of 0.5 mL/min. A conditional wash was then performed, rinsing at 1 mL/min with buffer A until the absorbance of the column flowthrough fell below 10 mAU. The protein was then eluted with a 15 CV linear gradient to 100% solvent B (20 mM Tris-HCl pH8, 500 mM NaCl, 500 mM imidazole). During sample load and wash phases, 5 mL fractions of the eluent were collected, while 2 mL fractions were collected during the elution. Fractions containing pure protein were identified by SDS-PAGE gel (BioRad, 3450123). Amicon Ultra 30 kDa MWCO filters (Millipore Sigma, UFC803024) were then used to combine, concentrate, and buffer exchange the

purified protein fractions into 10 mM Tris, pH 7.4 (American Bio, AB14044). Protein concentration was determined by NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo-Fisher) before storage at -80 °C.

Mucinase digestion

All proteins were first digested with either SmE, IMPa (NEB, P0761), or OgpA (Genovis, G2-OP1-020), prior to any further processing. All solutions were prepared using MS grade water (Thermo Scientific, 51140). For the structural characterization and mapping of TIM-4 and TIM-1 as well as the analysis of GP1b, 6-8 µg of protein were used for each digest. All other digests were performed using 2 µg of protein. Each glycoprotein was digested with the O-glycoproteases in a total volume of approximately 15 µL of fresh 50 mM ammonium bicarbonate (AmBic) (Honeywell Fluka, 40867) overnight at 37 °C. Digestions with SmE were conducted at an enzyme to substrate ratio of 1:10, while IMPa and OgpA were digested according to commercial instructions. When sialidase was used, it was added alongside the O-glycoproteases at concentrations in accordance with the manufacturer protocol (NEB, P0720).

SDS-PAGE analysis

Fractions from SmE protein purification were run on a 4-12% Criterion XT BisTris gel (Bio-Rad, 3450123) in MES XT buffer (Bio-Rad, 1610789) at 180 V for 60 min, alongside Precision Plus All Blue Protein Standard (Bio-Rad, 1610373).

Digested proteins were separated on a 4-12% Criterion XT BisTris gel (Bio-Rad, 3450123) in MOPS XT buffer (Bio-Rad, 1610788) at 180 V for 60 min, alongside Blue Easy Protein Ladder (NIPPON Genetics, MWP06).

All protein gels were imaged on a LI-COR Odyssey instrument following a 30 min incubation of Aquastain (Bulldog Bio, AS001000).

Mass spectrometry sample preparation

After mucinase digestion, dithiothreitol (DTT) (Sigma Aldrich, D0632) was added to a concentration of 2mM and allowed to react at 65 °C for 20 min followed by alkylation in 5 mM iodoacetamide (IAA) (Sigma Aldrich, I1149) for 15 minutes in the dark at room temperature.

For samples where presence of N-glycosylation interfered with identification of potential O-glycosylation sites, a PNGase (NEB, P0705) digestion was performed. This included the characterization of C1-Inh, while secondary files for TIM1 and TIM3 were generated after gaps in coverage indicated N-glycosylation. The concentrated enzyme was diluted 1:10 and 1 µL of the diluted stock was used for each 2 µg of protein. After allowing the enzyme to react overnight, the protein was buffer exchanged into 50 mM AmBic using 10 kDa MWCO filters (Merck Millipore, UFC501024).

Proteins with fewer sites of glycosylation (i.e. Fetuin, TIM-3, and C1-Inh) underwent an additional digestion was performed by adding sequencing-grade trypsin (Promega, V5111) in a 1:50 enzyme:substrate ratio for 6 hours at 37 °C. All reactions were quenched by adding 1 µL of formic acid (Thermo Scientific, 85178) and diluted to a volume of 200 µL prior to desalting.

Desalting was performed using 10 mg Strata-X 33 µm polymeric reversed phase SPE columns (Phenomenex, 8B-S100-AAK). Each column was activated using 500 µL acetonitrile (ACN) (Honeywell, LC015) followed by of 500 µL 0.1% formic acid, 500 µL 0.1% formic acid in 40% ACN, and equilibration with two additions of 500 µL 0.1% formic acid. After equilibration, the samples were added to the column and rinsed twice with 200 µL 0.1% formic acid. The columns were transferred to a 1.5 mL tube for elution by two additions of 150 µL 0.1% formic acid in 40% ACN. The eluent was then dried using a vacuum concentrator (LabConco) prior to reconstitution in 10 µL of 0.1% formic acid.

Mass spectrometry data acquisition

Samples were analyzed by online nanoflow liquid chromatography-tandem mass spectrometry using an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to a Dionex UltiMate 3000 HPLC (Thermo Fisher Scientific). For each analysis, 4 µL was injected onto an Acclaim PepMap 100 column packed with 2 cm of 5 µm C18 material (Thermo Fisher, 164564) using 0.1% formic acid in water (solvent A).

Peptides were then separated on a 15 cm PepMap RSLC EASY-Spray C18 column packed with 2 µm C18 material (Thermo Fisher, ES904) using a gradient from 0-35% solvent B (0.1% formic acid with 80% acetonitrile) in 60 minutes.

Full scan MS1 spectra were collected at a resolution of 60,000, an automatic gain control (AGC) target of 3e5, and a mass range from 300 to 1500 m/z. Dynamic exclusion was enabled with a repeat count of 2, repeat duration of 7 s, and exclusion duration of 7 s. Only charge states 2 to 6 were selected for fragmentation. MS2s were generated at top speed for 3 seconds. Higher-energy collisional dissociation (HCD) was performed on all selected precursor masses with the following parameters: isolation window of 2 m/z, 29% normalized collision energy, orbitrap detection (resolution of 7,500), maximum inject time of 50 ms, and a standard AGC target. An additional electron transfer dissociation (ETD) fragmentation of the same precursor was triggered if 1) the precursor mass was between 300 and 1,500 m/z and 2) 3 of 8 HexNAc or NeuAc fingerprint ions (126.055, 138.055, 144.07, 168.065, 186.076, 204.086, 274.092, and 292.103) were present at ± 0.1 m/z and greater than 5% relative intensity. Two files were collected for each sample: the first collected an ETD scan with supplemental energy (EThcD) while the second method collected a scan without supplemental energy. Both used charge-calibrated ETD reaction times, 100 ms maximum injection time, and standard injection targets. EThcD parameters were as follows: Orbitrap detection (resolution 7,500), calibrated charge-dependent ETD times, 15% nCE for HCD, maximum inject time of 150 ms, and a standard precursor injection target. For the second file, dependent scans were only triggered for precursors below 1000 m/z, and data were collected in the ion trap using a normal scan rate.

Mass spectrometry data analysis

Raw files were searched using O-Pair search with MetaMorpheus against directed databases containing the relevant protein sequence. Mass tolerance was set to 10 ppm for MS1's and 20 ppm for MS2's. Methionine oxidation was set as a variable modification and carbamidomethyl cysteine was set as a fixed modification. For samples treated with PNGase F, asparagine deamidation was added as a variable modification. For most samples, we used the default O-glycan database containing 12 common structures.

For analysis of GP1bα, the database was based on the previously-determined glycomic profile. Files initially underwent a nonspecific search in order to determine the cleavage specificity of SmE. After the cleavage motif was determined, files generated using only an O-glycoprotease digestion were searched with semi-specific cleavage N-terminal to serine and threonine and six allowed missed cleavages. Samples treated with trypsin were searched with the same parameters, but also allowed cleavage C-terminal to arginine or lysine. Results were filtered to a q value less than 0.01 and manually validated using Xcalibur software (Thermo Fisher Scientific). Relative abundances were obtained by generating extracted ion chromatograms and determining area under the curve. After abundances were obtained, each file was checked for presence of the identified species. When a peak with matching retention time and mass was present, the peak was validated and the abundance recorded.

Cell culture

HeLa cells (ATCC, CCL-2) were grown in T75 flasks (Falcon, 353136) and maintained at 37 °C in 5% CO_2. The cells were cultured in DMEM (Gibco, 11965-092) supplemented with 10% fetal bovine serum (FBS, Sigma, F0926), 1% sodium pyruvate (Gibco, 11360-070), and 1% penicillin/streptomycin (Cytiva, SV30010).

Western blotting for MUC16 on SmE and StcE treated cells

HeLa cells were seeded in T25 flasks (Falcon, 353109). The following day, the media was removed and dilutions of StcE or SmE (0, 0.05, 0.5, 5, and 50 nM) in Hank's buffered salt solution (HBSS, Gibco, 24020-117) were added for 60 min. After which time, the media (1 mL) was collected into tubes containing 0.75 µL of 0.5 M EDTA (Invitrogen, 15575-038) to quench the enzymatic reaction. The samples were then concentrated in a 3 kDa spin filter (Millipore, UFC500324). The cells remaining in the flask were washed with an enzyme-free dissociation buffer containing EDTA (Millipore, S-004-C), lifted off the plate using the buffer, and transferred to tubes. The cells were pelleted and washed with PBS (Gibco, 14190-144) twice. Then the cells were lysed using 1x NuPAGE LDS Sample Buffer (Invitrogen, NP0008) supplemented with 25 mM DTT and boiling at 95 °C for 5 minutes. Concentrated supernatants were diluted in 4x sample buffer to a final concentration of 1x. Both cell lysates and supernatants were boiled for 5 min at 95 °C and then 30 µL of

sample was loaded onto a 4-12% Criterion XT BisTris gel (Bio-Rad, 3450124) which was run in MOPS XT buffer (Bio-Rad, 1610788) at 180 V for 90 min. Proteins were transferred to a 0.2 µm nitrocellulose membrane (Bio-Rad, 1620112) using the Trans-Blot Turbo Transfer System (Bio-Rad), at a constant 2.5 A for 15 min. Total protein was quantified using Revert 700 stain (LI-COR Biosciences, 926-11011) before primary antibody (Novus Biologicals, NB600-1468) incubation overnight at 4 °C. An IR800 dye-labeled secondary antibody (LI-COR, 926-32210) was used according to manufacturer's instructions for visualization on a LI-COR Odyssey instrument.

Cell viability assay

HeLa cells (CCL-2, ATCC) were seeded in 24-well plates at approximately 20,000 cells per well in 500 µL of media. After 24 hours, SmE and StcE were added at 500, 5, 0.05 and 0 nM. At t = 0, 24, 48, 72, 96 hours post treatment, PrestoBlue (Invitrogen, A13261) was added according to the manufacturer's instructions. After 2 hours, the supernatant was transferred to a black 96 well plate (Thermo Scientific, 237105) for fluorescent readings on a SPECTRAmax GEMINI spectrofluorometer using an excitation wavelength of 544 nm and an emission wavelength of 585 nm. Results were plotted and statistical significance was assessed by two-wayANOVA in GraphPad Prism.

Flow cytometry staining of MUC16 on SmE and StcE treated cells

Cells were fixed in 2% PFA and washed in PBS prior to antibody incubation. Primary (MA5-32121, Invitrogen) and secondary antibody (ab150075, Abcam) incubations were conducted in 0.5% BSA in PBS for 30 min each while rotating at 4 °C. Antibodies were diluted according to manufacturer's protocols. Cells were washed 3 times with 0.5% BSA in PBS between antibody incubations and after staining with the secondary antibody. Cytometry measurements were taken on a CytoFLEX flow cytometer (Beckman Coulter Life Sciences) and the data was analyzed with FlowJo v10 (BD Biosciences).

Computational System Construction

Protein construction: TIM-3 protein model was built from the following component models: X-ray crystal structure of the TIM-3 IgV domain (PDB 7M41), mucin-domain backbone constructed from the BuildPeptide

tool in ROSETTA from the FASTA sequence, and the transmembrane tail and cytoplasmic domains were modeled with AlphaFold (AF Q8TDQ0). TIM-4 protein model was built from the following component models: X-ray crystal structure of the TIM-4 IgV domain (PDB 5F7H), mucin-domain backbone constructed from the BuildPeptide tool in ROSETTA from the FASTA sequence, and the transmembrane tail and cytoplasmic domains were modeled with AlphaFold (AF Q96H15). Each of these domains were then joined together using psfgen in VMDTools. Glycosylation: An FA2 glycan was chosen for each N-linked glycan positions (list out the N-linked glycan positions) as that was consistent with the known glycoprofile at those positions, and fully characterization of N-linked glycan positions is outside the scope of this current work. For all O-linked glycans, the glycan structure at each position with highest population, as determined by MS, was chosen and modeled and constructed onto each site using CHARMM-GUI. Lipid bilayer insertion: Complete TIM-3 and TIM-4 models were then inserted into lipid bilayer patches with compositions similar to that of mammalian cell membranes (56% POPC, 20% CHL, 11% POPI, 9% POPE and 4% PSM). Solvation and Neutralization: Finally, the TIM-3 and TIM-4 systems were embedded into orthorhombic boxes, explicitly solvated with TIP3 water molecules, and neutralized to a concentration of 150mM of NaCl, resulting in systems of 846,793 and 2,122,863 million atoms, respectively. See Table S6.X1 for complete system breakdown:

Table S6.X1: Compositional breakdown of each structure simulated in this work. Lipid bilayer patch size and box size reflect dimensions at t=0, before any simulation was conducted.

	TIM-3	TIM-4
Total #Atoms	846,793	2,122,863
#Protein atoms	4,316	5,336
#Glycan atoms	1,275	3,410
#Water atoms	781,557	2,022,120
#Na/#Cl atoms	830/735	2055/1902
Lipid Bilayer Size (Å x Å)	130 x 130	140 x 140
Box Size (Å x Å x Å)	140.7 x 150.6 x 473.2	177.9 x 170.4 x 818.8

Molecular Dynamics (MD) Simulations

All MD simulations were performed with NAMD2.14 and CHARMM64m all-atom additive force fields on a private supercomputer in the Triton Shared Computing Cluster hosted by the San Diego Supercomputer Center. Lipid tail minimization and melting: All atoms except lipid tails were held fixed according to a Lagrangian constraint (i.e. the "fix" command in NAMD), while lipid tails were subjected to 10,000 steps of Steepest Descent minimization. Then, a heating step was performed wherein, with constraints on all atoms except for lipid tails the system temperature was incrementally increased from 10 K to 310 K for 0.5 ns at 1 fs/step. Total system minimization and equilibration: Following lipid tail melting, the Lagrangian constraints were removed from all atoms, but an energetic restraint (1kcal/mol/Å) was applied to all protein and glycan atoms. The complete system was then subjected to 10,000 steps of Steepest Descent minimization, followed by 0.5 ns of equilibration at 310 K (at a 1 fs timestep). After free total minimization and restrained equilibration, the TIM-3 and TIM-4 structures were branched to perform replicas of the following MD simulation protocols. TIM-3 was branched into 4 replicas and TIM-4 was branched into 3 replicas. Free equilibration: Finally, all restraints were removed (thus no restraints or constraints on the system at all) and all atoms were allowed to equilibrate for 0.5 ns at 310 K (1 fs/step). Production: A total of 700 ns and 830 ns (2 fs/step) were collected for TIM-3 and -4, respectively, see Table S6.X2 for a breakdown of per replica sampling. MDAnalysis was then used to perform all of the resultant analyses from MD simulations.

Table S6.X2: Complete breakdown of total simulation time for TIM-3 and TIM-4.

	TIM-3 (ns)	TIM-4 (ns)
Rep 1	238.8	260.9
Rep 2	152.2	309.9
Rep 3	131.3	261.0
Rep 4 (TIM3 only)	177.1	–
Total	699.7	831.8

Persistence Length Calculations

The polymer module in MDAnalysis was used to calculate persistence length of TIM-3 and TIM-4 mucin domains by defining a mucin polymer as the N, CA, and C, atoms along the backbone for the following residue selections: TIM-3, residues 133 to 198, TIM-4: residues 137 to 310. Per replica and per TIM protein, persistence length was calculated using the last 1000 frames of simulation. We calculated persistence lengths from each replica trajectory, and thus we have reported the average and standard deviation of persistence lengths calculated for each TIM mucin domain over the three replicas.

End-to-end distance calculations

The distances module in MDAnalysis was used to calculate the distance in Angstroms (Å) between the center of mass of the last residue of the globular IgV domain and the center of mass of the first residue of the transmembrane helical domain for every frame in each simulation trajectory. For TIM-3 these first and last residues were selected as 133 and 198, respectively, and for TIM-4 these residues were 137 and 310, respectively. We then normalized the calculated distances by the total number of protein residues within each mucin domain.

Bending Angle Calculations

To calculate the bending angle, we used MDAnalysis to select the center of mass of the first, middle, and final protein residue within each mucin domain. We then used these positions to calculate vectors: one from the middle residue to the first residue of each mucin domain, and one from the middle residue and the last residue of each mucin domain. For TIM-3, the first, middle, and last residues were selected as residues 131, 166, and 202, respectively. For TIM-4, the first, middle, and last residues were selected as residues 135, 225, and 314, respectively. We then calculated the angle between these vectors for each frame for all three replica trajectories. We then plotted these results as normalized density histograms.

Mucinase Structure Overlays, Molecular Docking, & TIM-4 Grafting

Structural comparison and docking were performed using Molecular Operating Environment (MOE) 2020.09. The X-ray crystal structures of IMPa, AM0627, ZmpB, ZmpC, and BT4244 were superimposed with the AlphaFold-predicted structures of SmE, AM0908, and AM1514 using the catalytic histidine and glutamate residues to guide alignment.

Following this, a ligand was generated over several steps for use in docking studies. First, the bisglycosylated ligand cocrystallized with AM0627 was placed into the analogous location of the superimposed SmE model structure. The amino acid residues of the peptide were then mutated to match a TIM-4 sequence (Ala189-Val190-Phe191-Thr192*-Thr193*-Ala194, where the asterisk indicates glycosylation) that was found to be cleaved by SmE but not IMPa. The peptide's N terminus was acetylated and its C terminus was N-methylated to better approximate the steric/electronic environment of a substrate. While holding the SmE model structure fixed, the glycopeptide was allowed to preliminarily minimize in the Amber10:EHT forcefield using restraints to ensure contacts were formed between (1) the backbones of the glycopeptide and the beta strand of the active site as well as (2) the GalNAc moiety and the side chains of the conserved residues Trp240 and Asn268.

During our glycoproteomic mapping, we observed high occupancy of H1N1A1 (or smaller fragments thereof) modifying Thr192 at P1 and H2N2A1 (or smaller fragments thereof) modifying Thr193 at P1' (Supp. F). Similar to previous work,[CITE] the P1 glycan was generated by grafting a sialic acid residue onto the 3-OH of the Gal moiety, using the glycan bound to the GspB siglec domain (PDB 5IUC) as a template.The P1' glycan was similarly generated by grafting the remaining GalNAc, Gal, and Sia residues onto the 6-OH of the GalNAc moiety branching from P1' using the glycan present on PSGL-1 in its cocrystal structure with P-selectin (PDB 1G1S).

Following this, the resulting glycopeptide underwent conformational search, holding all atoms except for the newly-grafted glycans fixed, using LowModeMD to generate a library of over 1300 different conformations of these sugar residues.[CITE] Each conformer underwent virtual screen, holding the mutant

SmED245A enzyme (replacing the catalytic residue to facilitate docking) rigid while allowing the glycopeptide ligand to move freely. The ligand-enzyme complexes with the top 100 docking scores were then used in induced fit docking, keeping both the ligand and the enzyme free. This stepwise search, screen, and docking process allowed us to thoroughly explore conformational space of the ligand while minimizing computational resources.

Following this, the docked ligand was then grafted together with TIM-4 fragments from dynamics simulations to generate larger glycopeptide ligands. First, data from the TIM-4 simulation were analyzed to identify frames containing structures that have Ramachandran angles for Val190 and Phe191 that were within ±15o of those in the docked structure (Val190: ϕ = -140°, ψ = 152°; Phe191: ϕ = -136°, ψ = 102°) (Supp FigX). Corresponding fragments (Pro175-Phe191) of ten different structures were manually superimposed on the docked structure, using Val190 and Phe191 to guide placement. Finally, the docked ligand (Val190-Ala194) was grafted onto each fragment (Pro175-Ala189) to generate ten larger glycopeptides used to identify potential for interaction with the mucin-binding module of SmE.

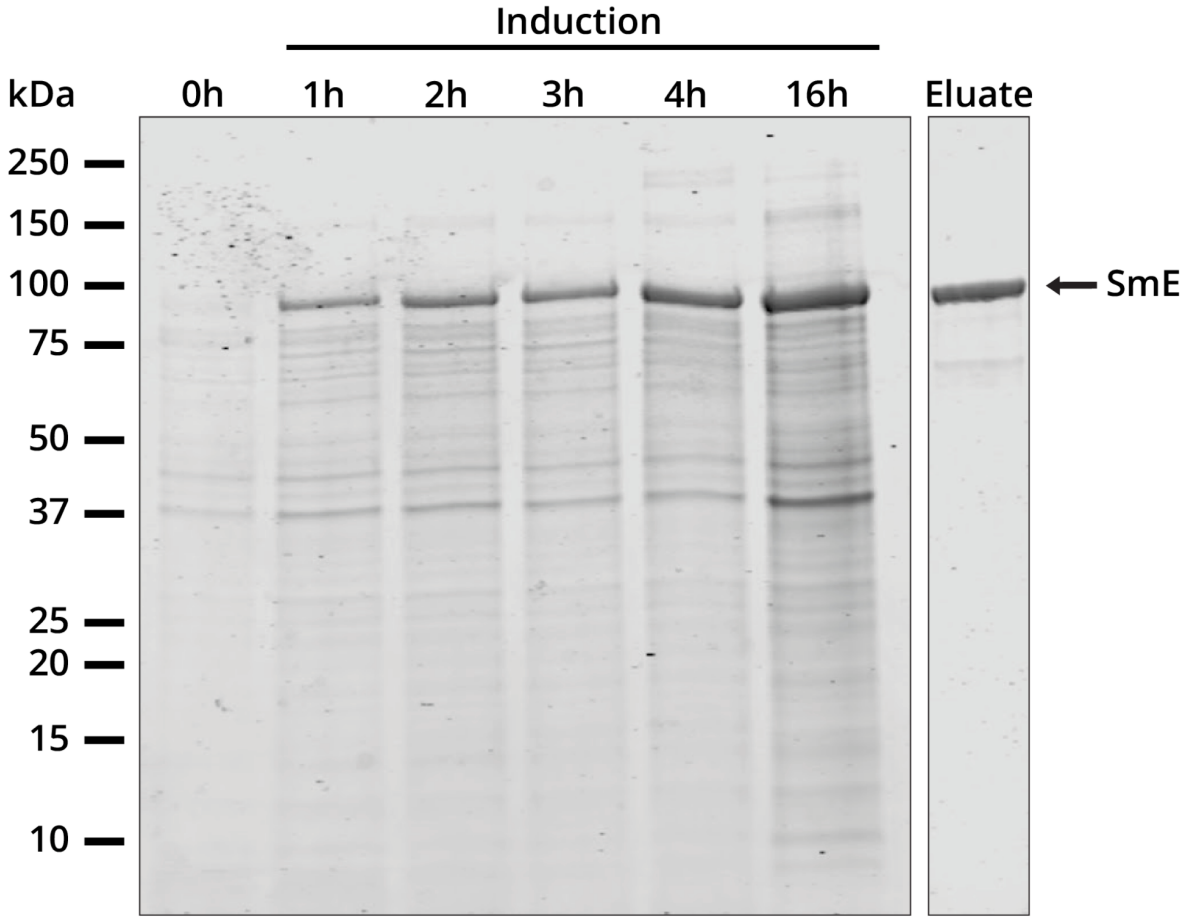


Figure S6.1: Expression and purification of SmE

SmE was expressed O/N at 16 °C. SDS-PAGE gel was stained with Coomassie (Bulldog-Bio) and imaged using an Odyssey CLx Near-Infrared Fluorescence Imaging System (LI-COR Biosciences). SmE ran at the predicted molecular weight of 94 kDa.

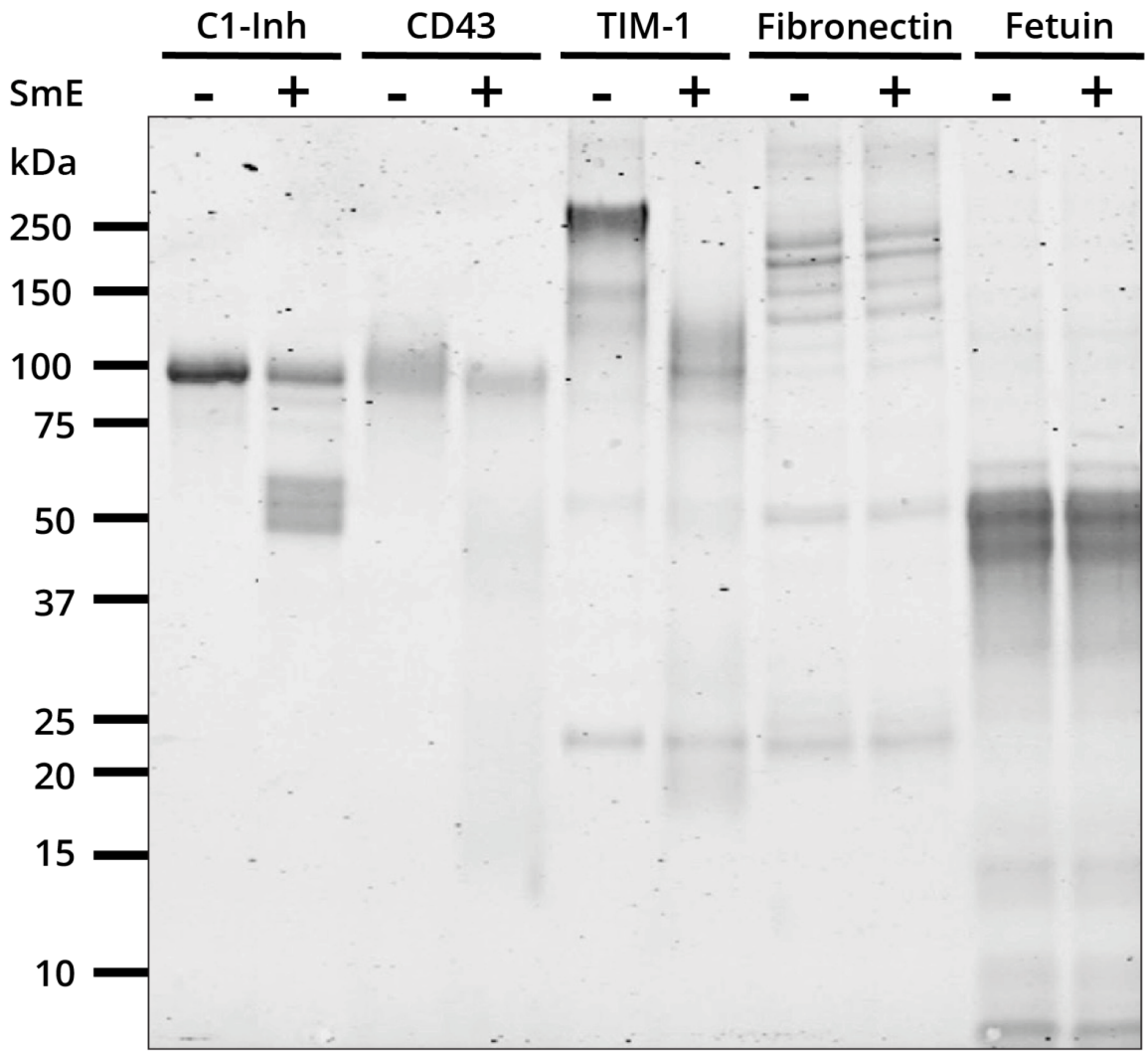


Figure S6.2: SmE selectivity for mucin-domain glycoproteins
The recombinant proteins shown were reacted with SmE at a 1:20 E:S ratio O/N at 37 °C and the digests were separated by SDS-PAGE. SDS-PAGE gel was stained with Coomassie (Bulldog-Bio) and visualized on an Odyssey CLx Near-Infrared Fluorescence Imaging System (LI-COR Biosciences).

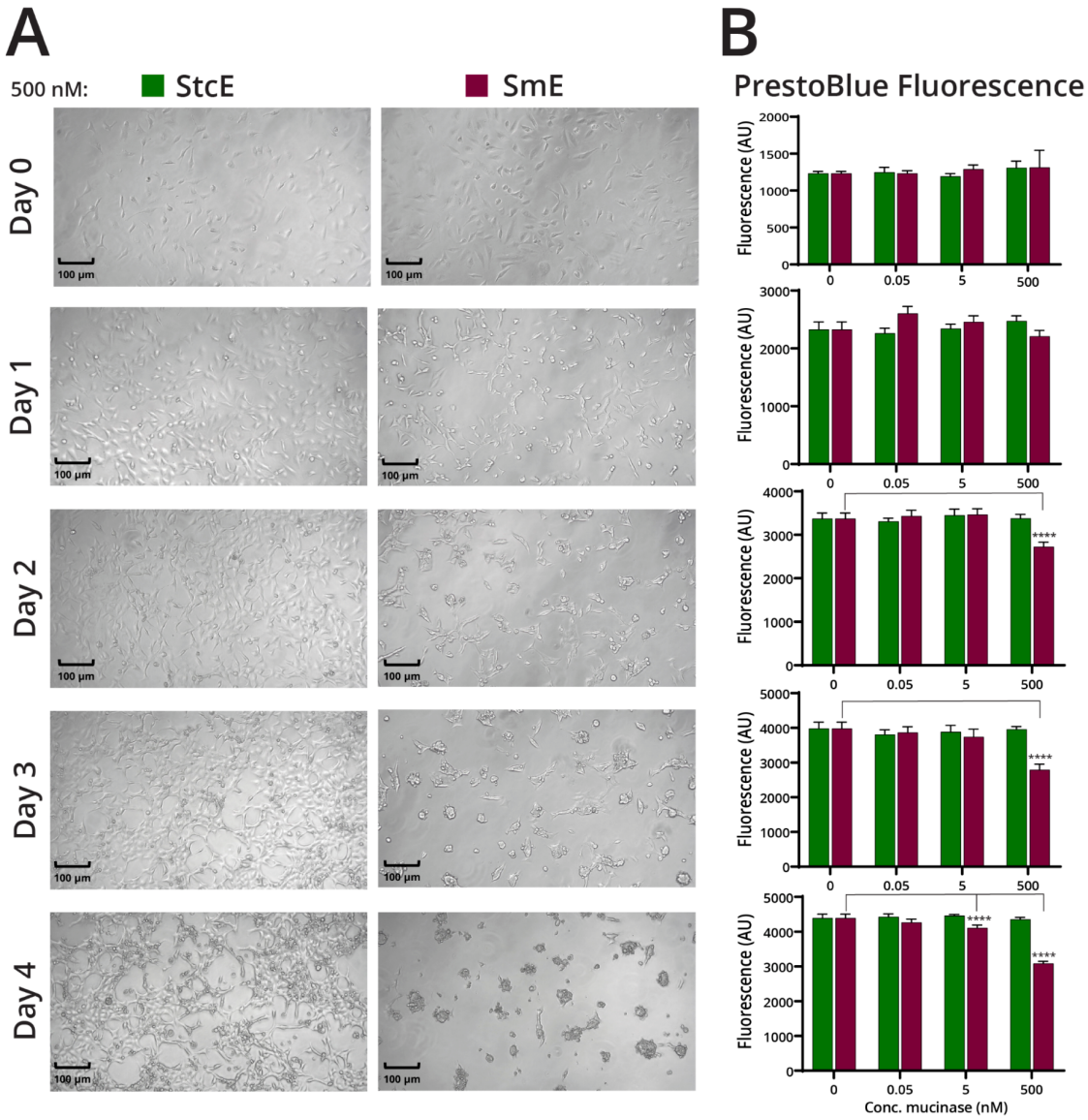


Figure S6.3: SmE treatment was nontoxic to HeLa cells at moderate concentrations and/or durations

(a) Live cell microscopic images of HeLa cells treated with 500 nM StcE or SmE, taken every 24 hours for a total of 4 days. Samples were the same as those used to measure viability in (b). (b) Cellular viability was measured using a resorufin-based dye (PrestoBlue, Thermo Fisher Scientific), at 0, 0.05, 5 and 500 nM StcE or SmE treatment over 4 days. Green: StcE, red: SmE.

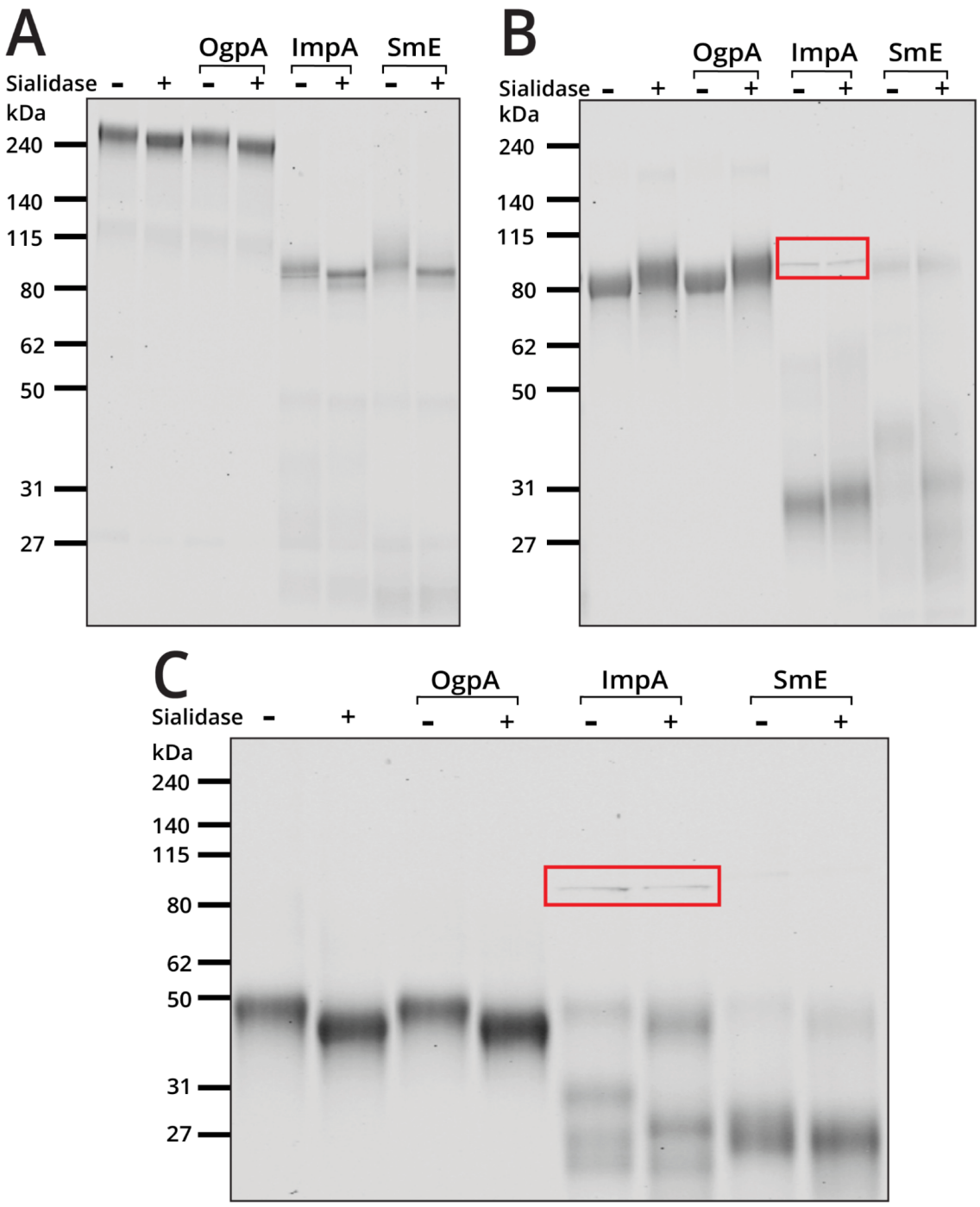


Figure S6.4: Benchmarking OgpA, ImpA and SmE activity on TIM proteins
Recombinant proteins a) TIM-1, b) TIM-4 and c) TIM-3 were reacted with 1:10 E:S ratio of OgpA, ImpA or SmE, with and without sialidase O/N at 37 °C. All digests were separated on an SDS-PAGE gel and stained with Coomassie (Bulldog-Bio). Gels were imaged on an Odyssey CLx Near-Infrared Fluorescence Imaging System (LI-COR Biosciences). ImpA is denoted by the red box.

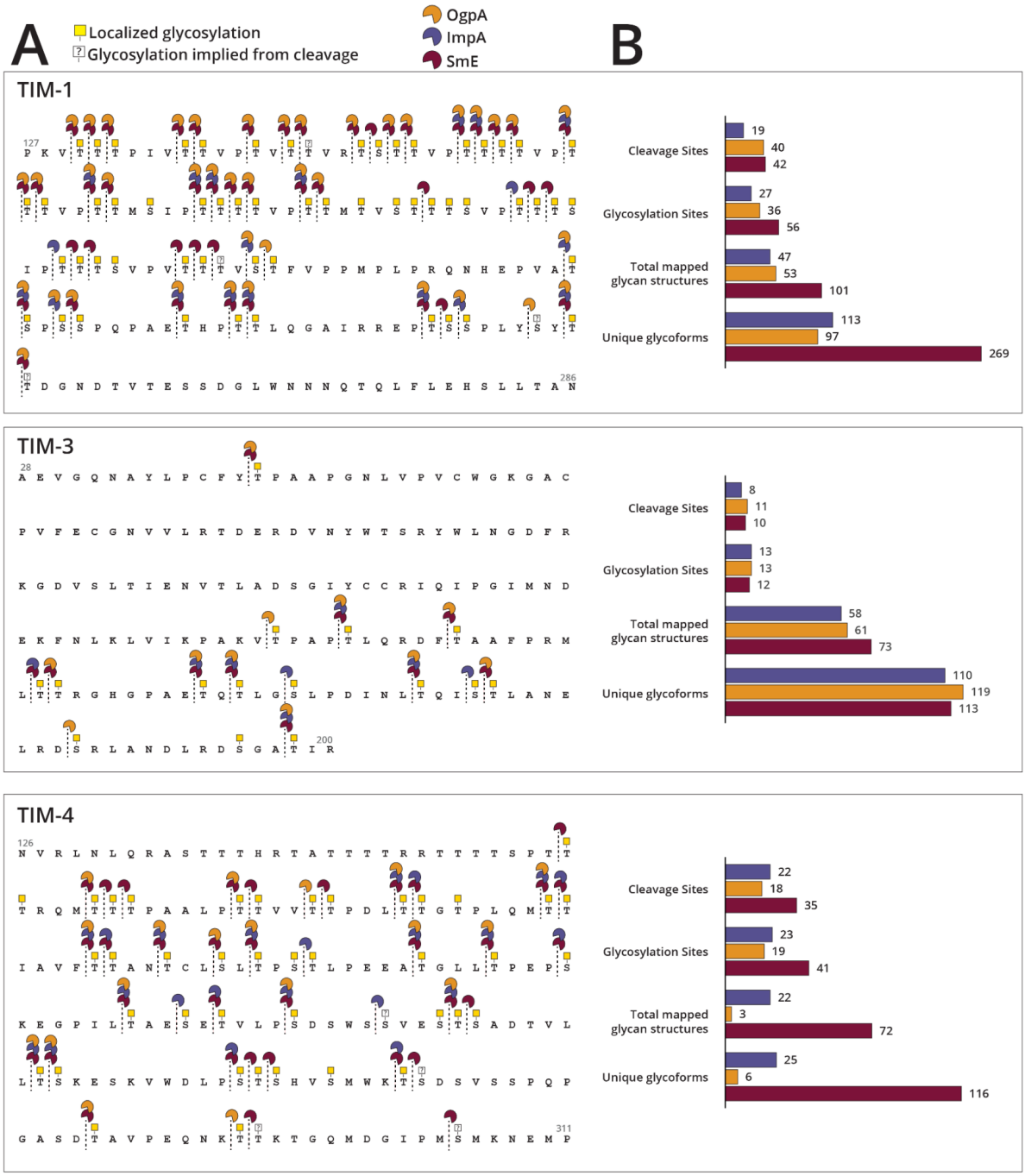


Figure S6.5: Total cleavage events

(a) Cleavage maps depict glycosylation and cleavage sites observed by MS analysis for each enzyme. (b) Graphical interpretation of individual glycoprotein cleavage sites, glycosites, and glycoforms identified by treatment with each enzyme. Red: SmE, blue: ImpA, yellow: OgpA. Asterisk in fetuin cleavage map indicates cleavage attributable to either mucinase or co-enzyme.

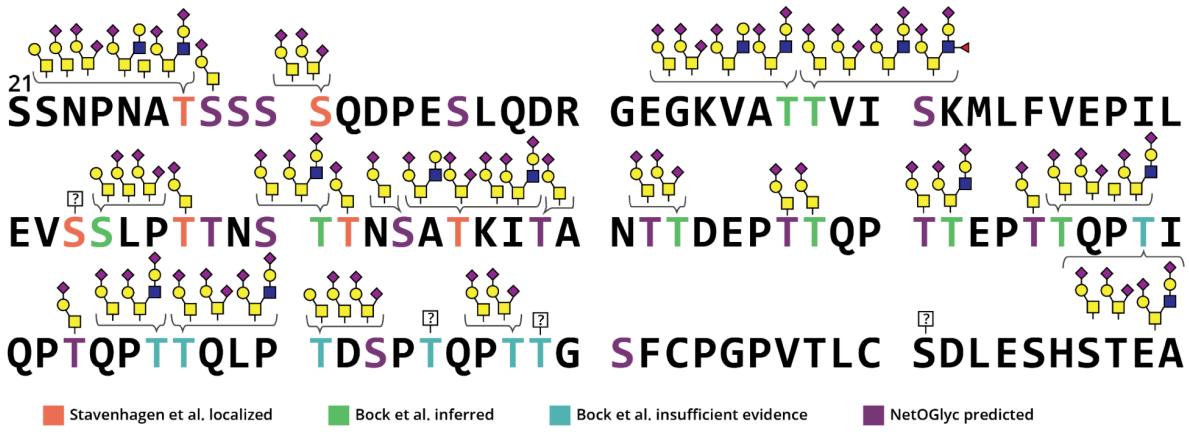


Figure S6.6: C1-INH glycoproteomic landscape

C1-INH isolated from human plasma was subjected to digestion with SmE and trypsin followed by MS analysis and manual data validation. Glycans depicted in brackets were detected on the associated residues. The colored residues were either detected (orange; Stavenhagen *et al.*), inferred (green; Bock *et al.*), hinted at without sufficient evidence (blue; Bock *et al.*), or predicted by NetOGlyc (purple).

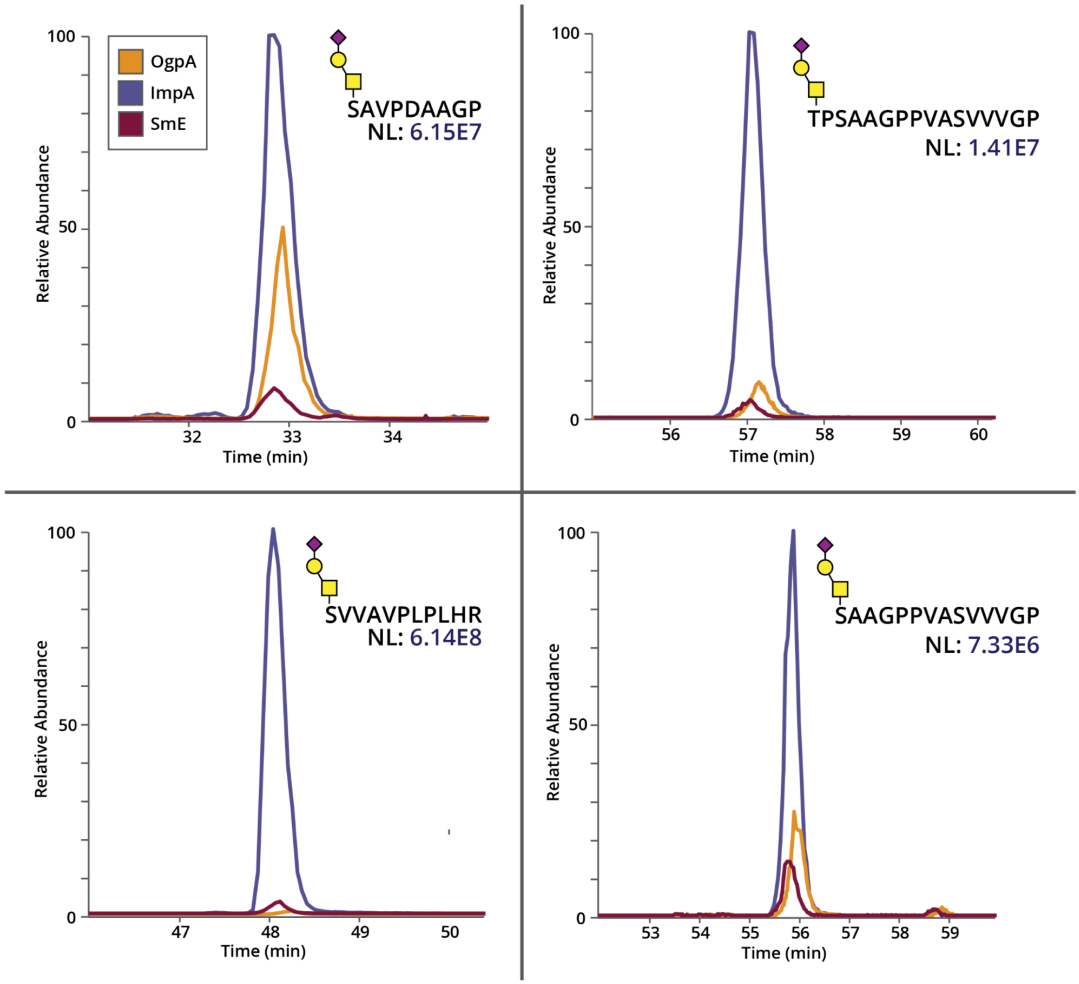


Figure S6.7: Extracted ion chromatograms of cleaved glycopeptides from fetuin
Using Thermo Xcalibur, extracted ion chromatograms (XICs) were generated for four glycopeptides from the OgpA, ImpA, and SmE digest of fetuin. The XICs of specific glycopeptides from the digestion of fetuin were normalized to indicate relative abundance in each analysis. Yellow: OgpA, blue: ImpA, red: SmE.

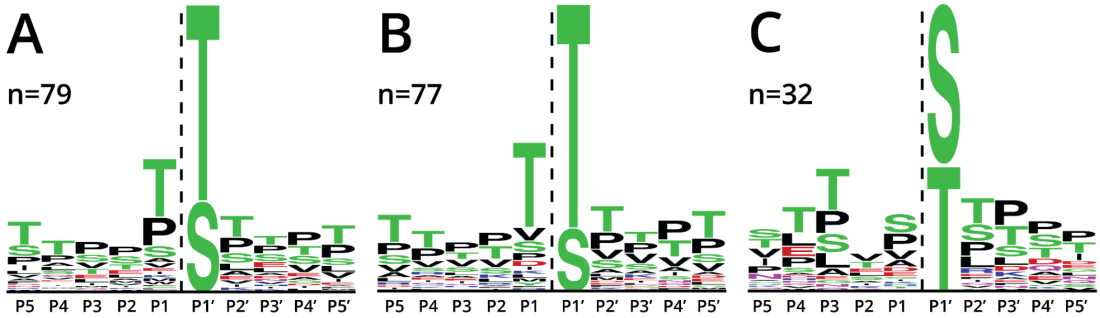


Figure S6.8: Anti-logos demonstrate preference for amino acids at P1 and P1'
Using cleavage maps from Figure S5, all sites without observed cleavage were loaded into weblogo.berkeley.edu. OgpA (A) and ImpA (B) demonstrated less efficiency with Thr P1 position. SmE (C) uniquely showed preference for Thr in the P1' position, and less preference for Ser for either the P1 or P1' position.

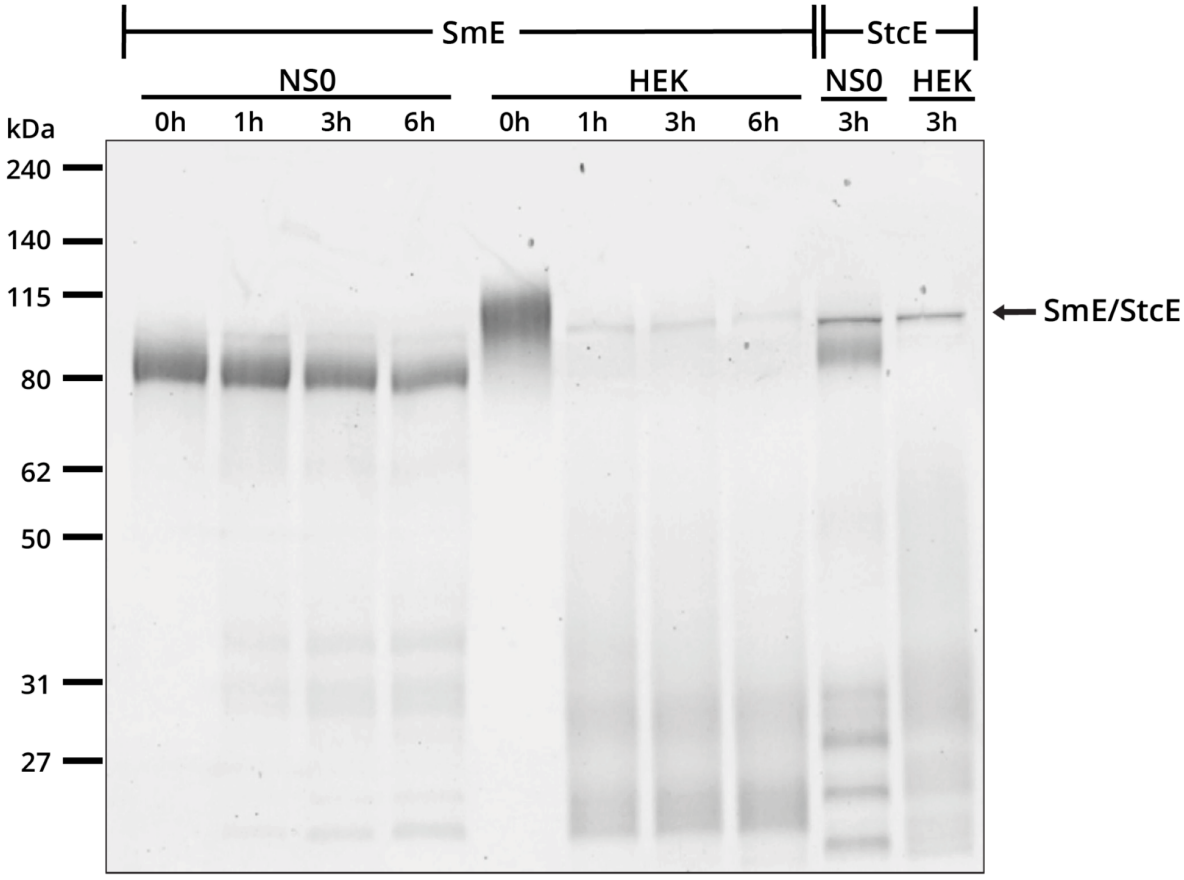


Figure S6.9: Comparison of SmE activity on recombinant mouse (NS0) or human (HEK) TIM-1
Recombinant human TIM-1 expressed in NS0 or HEK293 cells were reacted with SmE at a 1:10 E:S ratio for 1 hour, 3 hours or 6 hours at 37 °C and StcE at a 1:10 E:S ratio for 3 hours at 37 °C. All digests were separated by SDS-PAGE and Coomassie stained (Bulldog-Bio). Gel was visualized on an Odyssey CLx Near-Infrared Fluorescence Imaging System (LI-COR Biosciences).

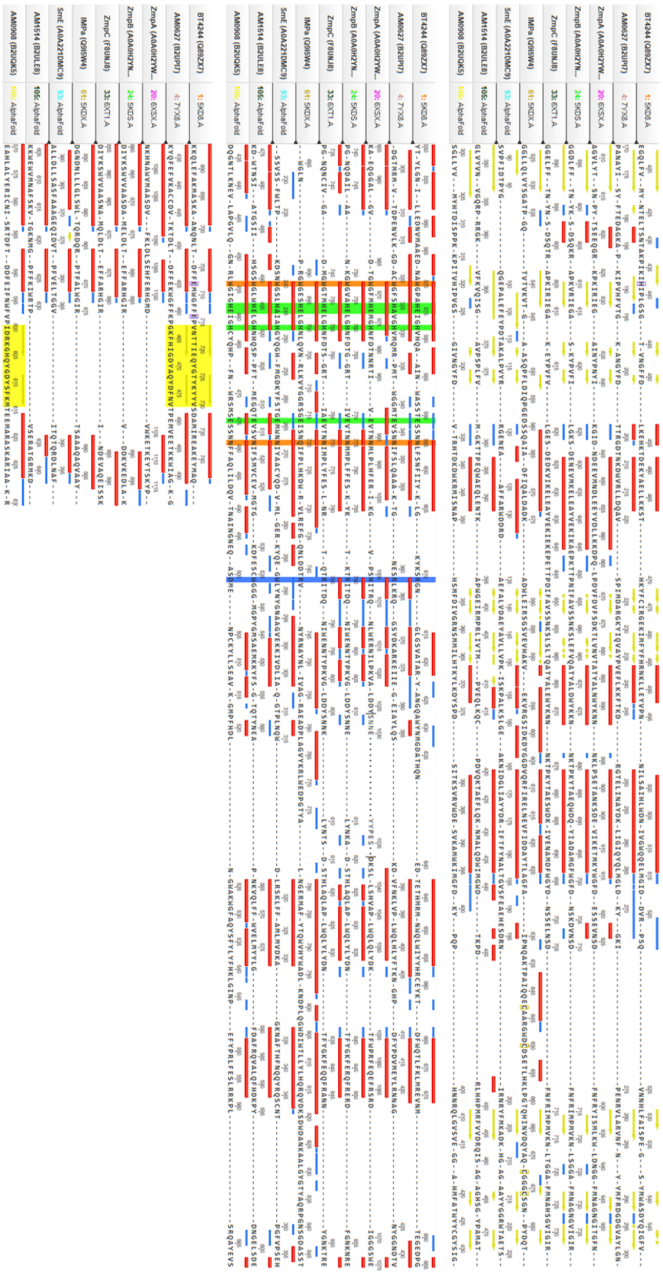


Figure S6.10: Sequence alignment of the catalytic PF13402 domains found in characterized O-glycoproteases

The X-ray crystal structures of BT4244, AM0627, ZmpA, ZmpB, ZmpC, and ImpA as well as the AlphaFold-predicted structures of SmE, AM1514, and AM0908 were structurally overlaid using the conserved zinc-binding and catalytic residues (green). The structural overlays then were used to generate the above sequence alignments, highlighting key secondary structures: alpha helices (red), beta sheets (yellow), and turns (blue). In addition to the catalytic core residues (green), we have highlighted conserved (orange) and semiconserved (blue) residues involved in recognizing P1' glycans as well as residues in a semiconserved beta-hairpin (yellow) that can potentially recognize P1 glycans. Notably, SmE has neither the semiconserved arginine residue nor the semiconserved beta-hairpin.

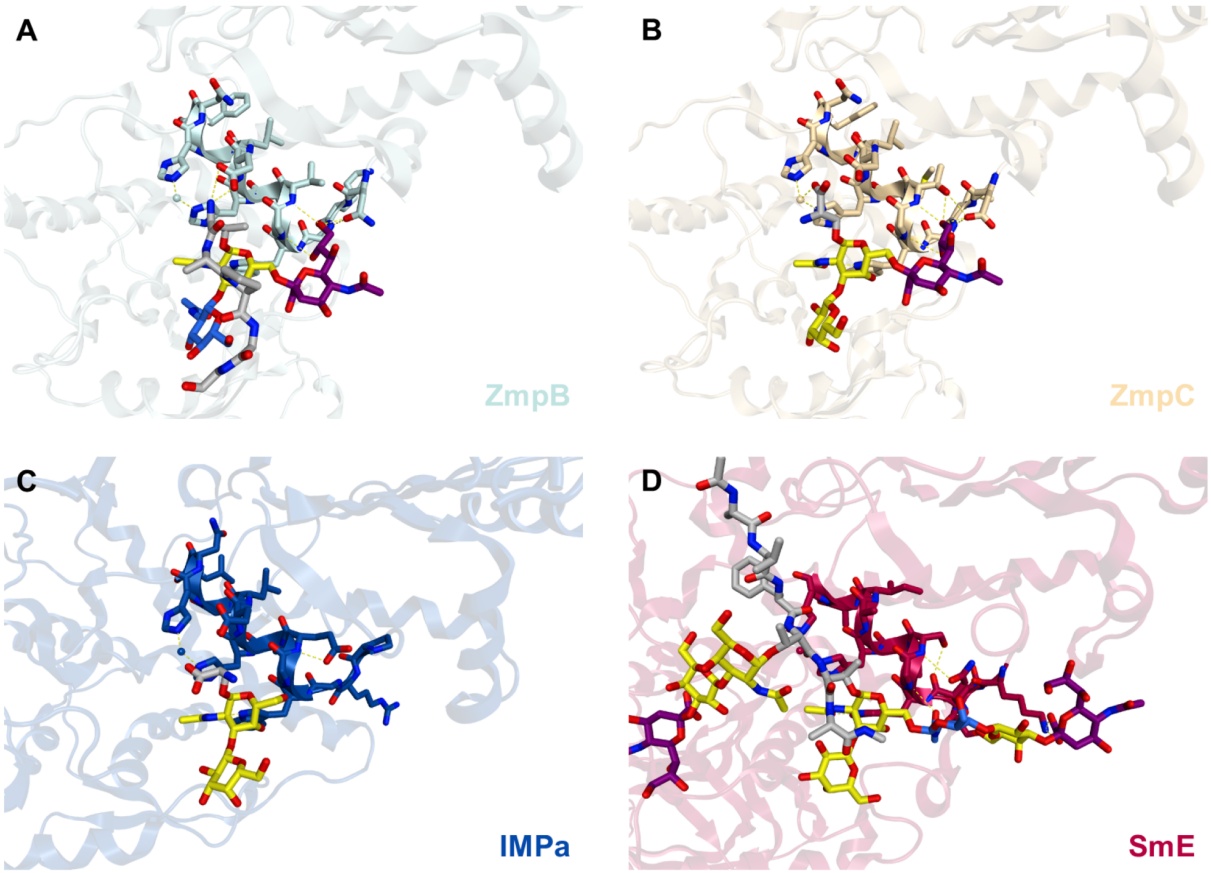


Figure S6.11: Structure of ligands and the catalytic helices in the crystal structures of (A) ZmpB, (B) ZmpC, and (C) ImpA as well as the modeled structure of (D) SmE

In their crystal structures, both ZmpB and ZmpC form specific contacts (yellow dashes) between residues near the terminus of the catalytic helix (colored sticks) and the branching sialic acid residue (purple sticks) found in the ligand at P1'. The ligand in ImpA does not contain sialic acid; similar interactions can likely form given the similar length of its catalytic helix, imparting ImpA with the ability to accommodate branched ligands. In the modeled structure, SmE forms contacts with the GalNAc residue (blue sticks) and additional contacts with the sialic residue (purple sticks) of the glycan, which may explain its ability to accommodate larger branched glycans at P1'.

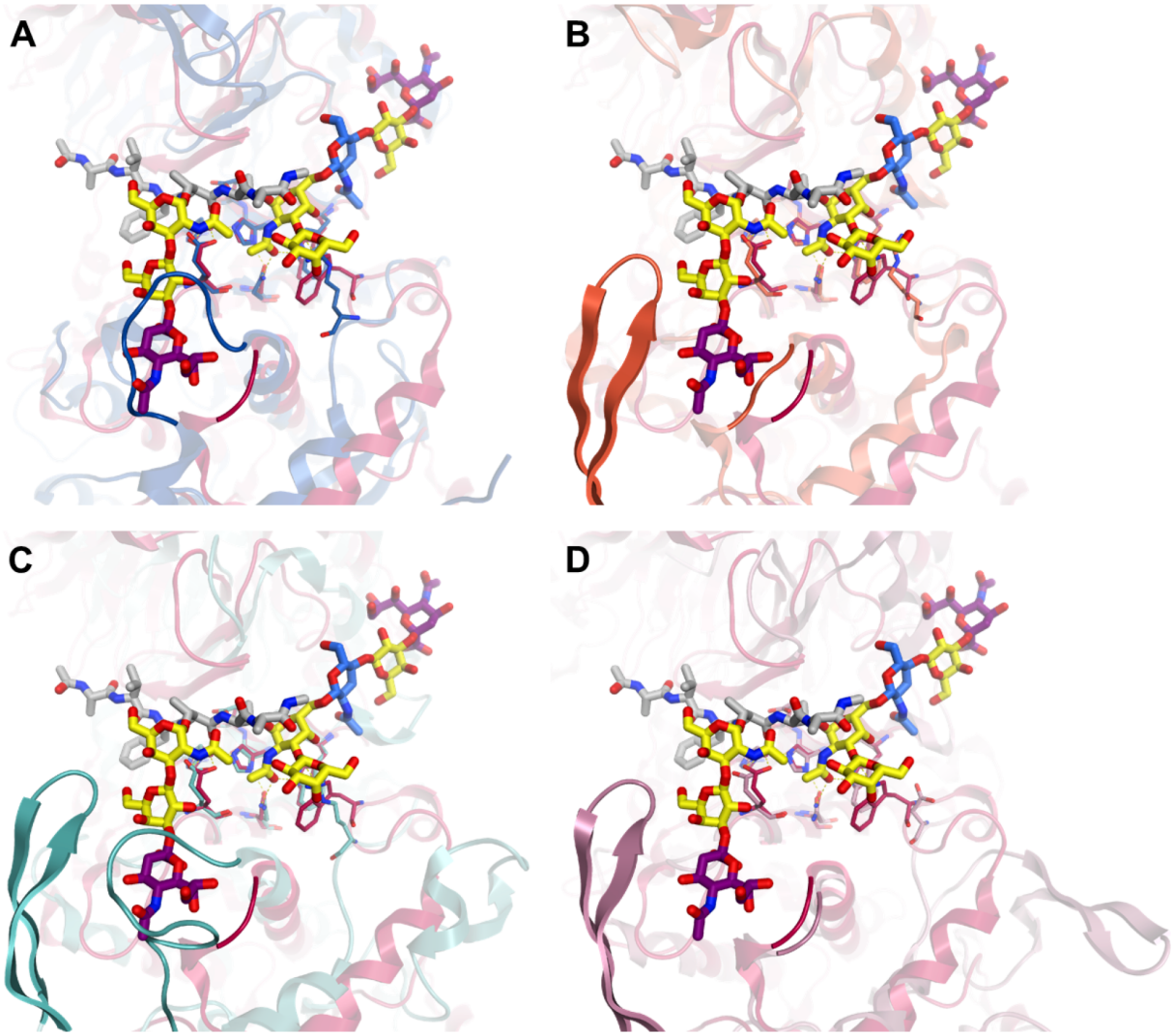


Figure S6.12: Overlays of the docked SmE-glycopeptide complex (maroon) with the crystal structures of (A) ImpA (blue), (B) AM0627 (orange), and (C) BT4244 (mint) as well as the AlphaFold predicted structure of (D) AM0908 (lilac)

The zinc ion (sphere) as well as the residues (thin sticks) of the catalytic core and the conserved residues that bind the P1' glycan are shown with the docked ligand (thick sticks). The loops and beta hairpins that form the steric environment around the P1 glycan are darkened for emphasis. (A) The loop in SmE is predicted to be short, which could explain why SmE can sterically accommodate and act on substrates with glycosylation at P1; the loop of ImpA is long and likely prevents such activity. (B-D) The remaining enzymes have or are predicted to have a hairpin that can recognize glycosylation at P1. (B) AM0627 has a short loop that can accommodate different glycans at P1. (C) The loop of BT4244 is much larger, and this difference could resolve discrepancies reported across the literature and explain why BT4244 showed preference for engineered substrates bearing the smaller Tn-antigen, whereas AM0627 showed preference for analogous substrates modified with the larger T-antigen at P1. (D) AM0908 is predicted to have a short loop similar to AM0627; despite this potential similarity, the two enzymes display different preference for P1 glycosylation, which cannot be explained by this reasoning and may be the result of subtler differences in primary sequence.

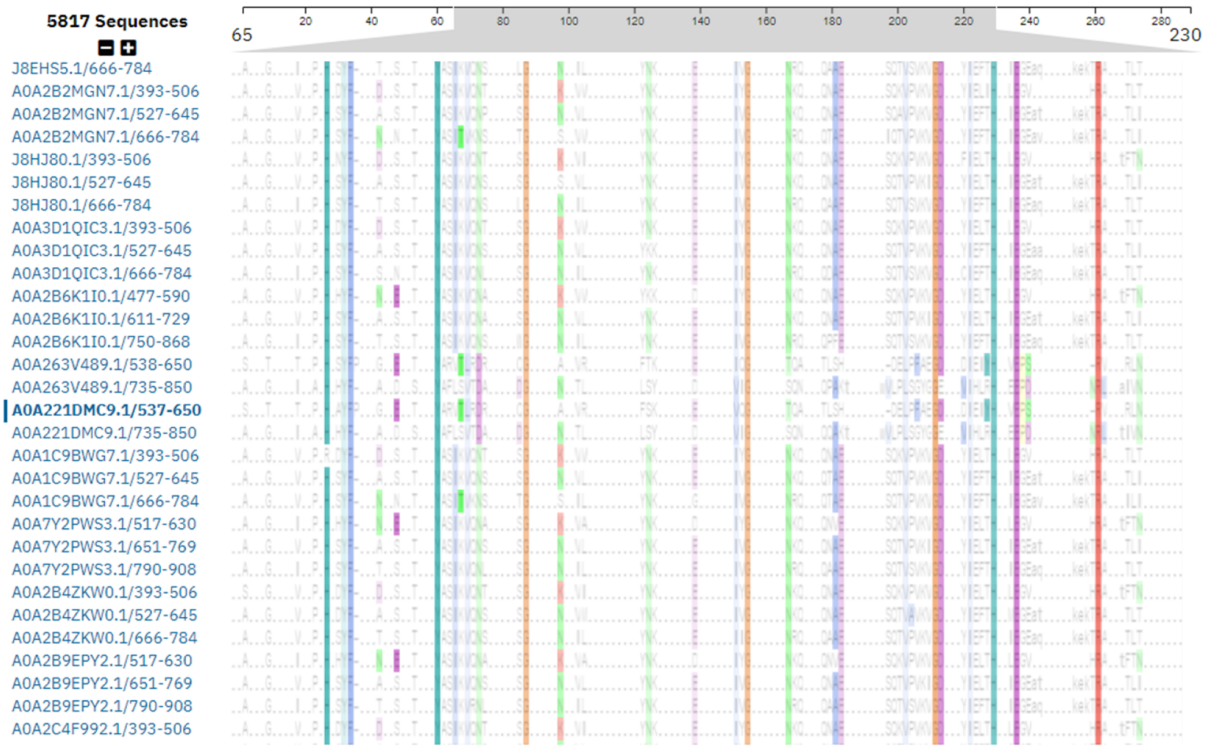


Figure S6.13: Conservation across the 5,817 mucin-binding modules (PF03272) found in proteins, including SmE (A0A221DMC9.1/537-650), listed in UniProt

Degree of conservation is shown using clustal2 coloring, highlighting two conserved motifs (HxxFxxxxY and HxExxR) found in loops clustered together to form a single binding pocket in the AlphaFold structure of SmE. Full results are available at https://www.ebi.ac.uk/interpro/entry/pfam/PF03272/entry_alignments/?type=uniport.

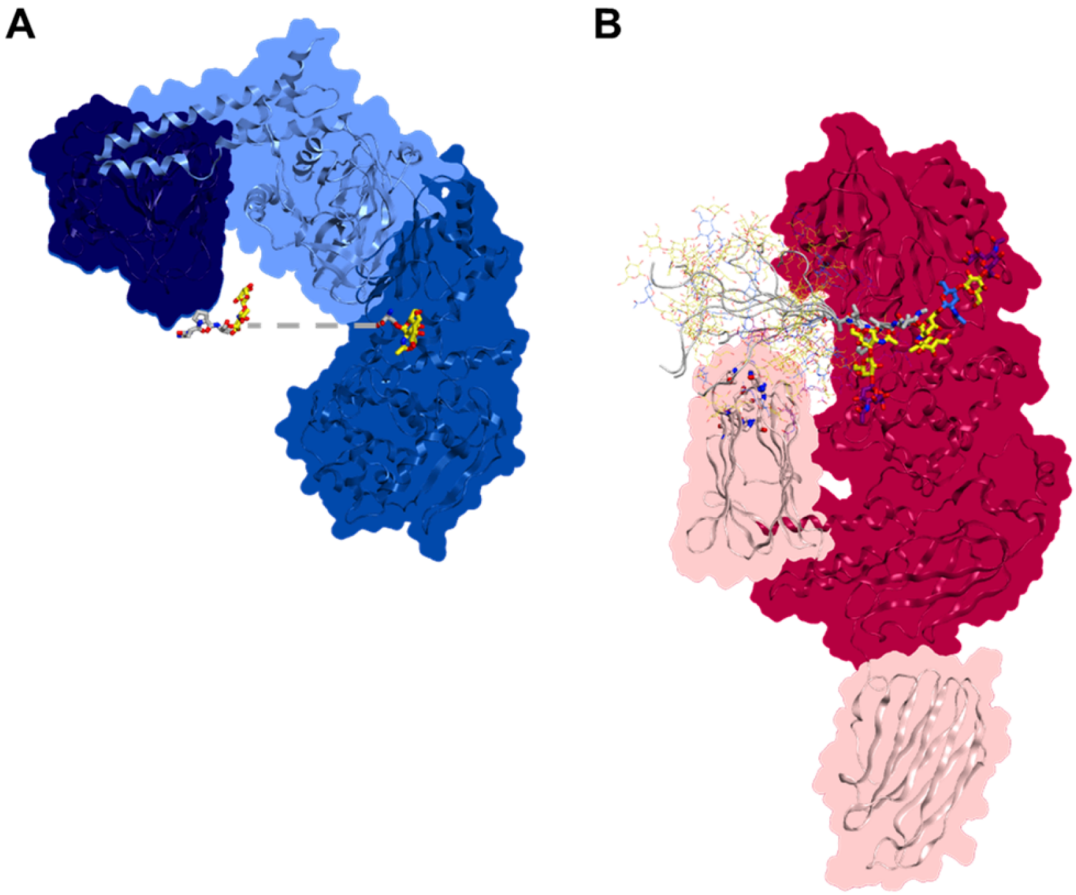


Figure S6.14: Potential cooperativity of ligand binding in ImpA and SmE
(A) Overlay of crystal structures of ImpA with crystallized ligands in the catalytic and accessory domains. The ligand (thick sticks) crystallized in the catalytic domain (blue loops and shaded surface) as well as the ligand (thick sticks) in the accessory N-terminal domain (dark blue loops and shaded surface) are small fragments that likely reflect the pose of larger mucin-like substrates (dashed line), highlighting potential cooperativity between the different domains in ImpA. (B) Grafted TIM-4 fragments overlaid with the modeled structure of SmE. The catalytic domain (maroon loops and shaded surface) and the two mucin-binding modules (rose loops and shaded surface) are shown, including the conserved motifs (HxxFxxxxY and HxExxR, thick sticks) in the loops of the mucin-binding module adjacent to the catalytic site. Also shown are the amino acid and glycan residues of the original docked glycopeptide (thick sticks) that was grafted together with ten different TIM-4 fragments, highlighting only these species' peptide backbone (gray ribbon) and glycans (thin sticks) for clarity. Some glycans were found to flank the conserved residues of the mucin-binding module, suggesting that the catalytic and accessory domains of SmE may cooperatively recognize and cleave mucin-like substrates.

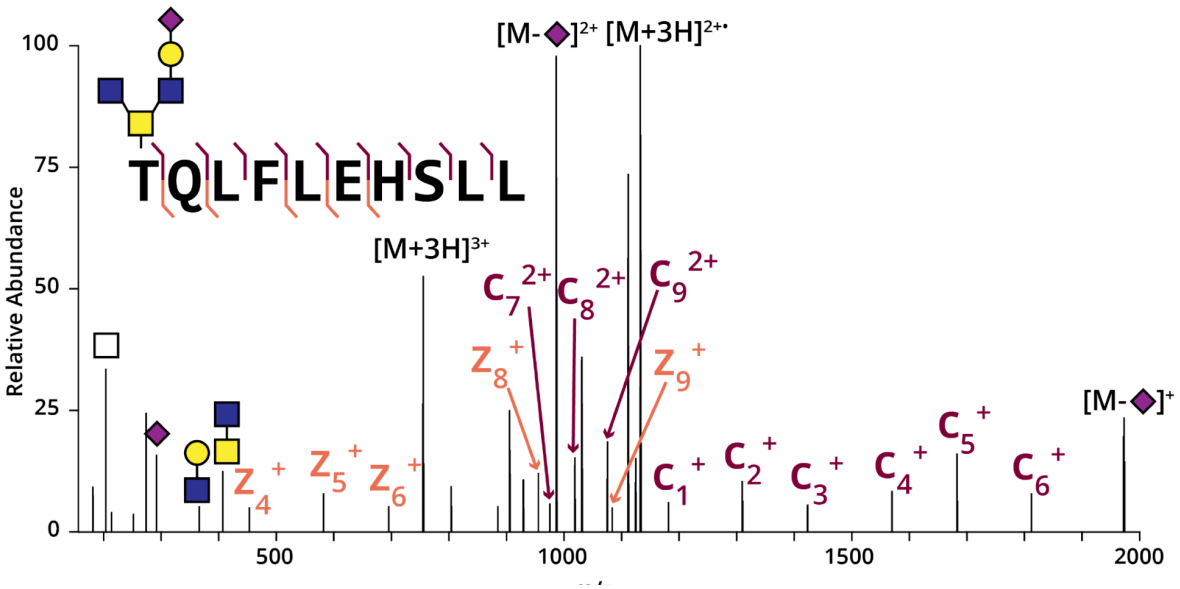


Figure S6.15: Recombinantly expressed TIM proteins contain complex glycan structures
TIM-1 was subjected to digestion with SmE followed by MS analysis and manual data interpretation. An EThcD spectrum from glycopeptide TQLFLEHSLL is depicted above, bearing a core 4 glycan structure. We observed full sequence coverage and unambiguous site-localization.

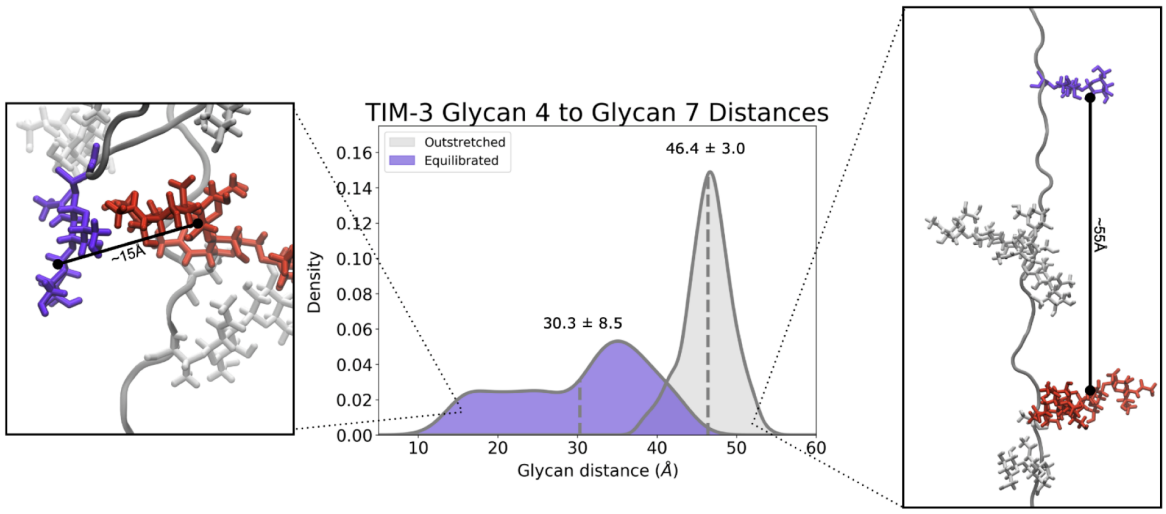


Figure S6.16: Histogram detailing distance between glycans G4 (site T162) and G7 (site T145) in outstretched versus equilibrated conformation, measured throughout molecular dynamics simulation
Left panel is a frame taken from equilibrated conformation with a glycan distance of approximately 15Å between G4 and G7. Right panel is a frame taken from outstretched/not fully equilibrated conformation with glycan distance of approximately 55Å between G4 and G7.

179 TGTPLQMTTIAVFT|TANTCLSLTPSTLP 206

Glycan	Occupancy
H1N1A1	19.6%
H1N1	18.4%
N1	15.7%
n/a	12.9%
H1N1A2	12.0%
H2N2	9.8%
H1N2	3.5%
H2N2A1	3.3%
H1N1F1	2.9%
H1N2A1	1.1%
H2N2A2	0.6%
N2	0.1%

Glycan	Occupancy
H2N2A1	43.0%
N1	26.7%
H2N2	15.3%
H1N1	8.2%
H2N2F1	1.9%
H1N2	1.4%
H1N1A2	1.4%
H1N1A1	1.1%
H1N2A1	0.8%

Figure S6.17: The relative abundance of different glycans branching from T192 and T193 of TIM-4, as revealed through digestion with SmE and subsequent mass spectrometric analysis
The H1N1A1 and H2N2A1 glycans were the most abundant species at these positions, and they were incorporated into the ligand used in docking experiments with SmE.

6.7 **Acknowledgements**

We thank Jeffrey Shabanowitz for his technical expertise and thoughtful conversations during the preparation of this manuscript. We also would like to acknowledge the Keck Biophysical Resource for their assistance in cell viability assays. J.C. is supported by a University Fellowship through Yale University; A.S. is supported by the National Institutes of Health Chemical Biology Training Grant (T32 GM067543); K.E.M. and T.M.L. are supported by Yale Endowed Postdoctoral Fellowships in the Biological Sciences. M.A.R. and F.L.K acknowledge the San Diego Supercomputer at UCSD for providing HPC resources that have contributed to the research results reported within this paper. R.E.A. acknowledges support from NIH GM132826, NSF RAPID MCB-2032054, a UC San Diego Moores Cancer Center 2020 SARS-COV-2 seed grant, and U19-AI171954 from NIAID. M.A.R. is supported by NIH T32 EB009380. This work was also supported by a Sarafan ChEM-H Physician-Scientist fellowship (to M.A.H.), the Stanford Maternal & Child Health Research Institute Instructor K Award Support Program (to M.A.H.), a National Blood Foundation Early Career Scientific Research Grant (to M.A.H.), a NIH NHBLI Pathway to Independence award (1K99HL156029-01 to M.A.H.), and NIH grant R01CA200423 (to C.R.B.). D.J.S. was supported by a US NSF Graduate Research Fellowship

and Stanford Graduate Fellowship. M.J.F. was supported by a University of Redlands Faculty Research Grant. SAM is supported by the Yale Science Development Fund and a NIGMS R35-GM147039.

Chapter 6, in full, is a modified reprint of the materials as it has been submitted for publication in Science as "Joann Chongsaritsinsuk‡, Alexandra D. Steigmeyer‡, Keira E. Mahoney‡, Mia A. Rosenfeld‡, Taryn M. Lucas, Deniz Ince, Fiona L. Kearns, Alexandria S. Battison, Marie A. Hollenhorst, D. Judy Shon, Katherine H. Tiemeyer, Victor Attah, Catherine Kwon, Carolyn R. Bertozzi, Michael J. Ferracane, Rommie E. Amaro, Stacy A. Malaker. Glycoproteomic landscape and structural dynamics of TIM family immune checkpoints enabled by mucinase SmE. 2023." The dissertation author was a primary co-investigator and co-author of this work.

6.8 **References**

1. Shurer, C. R. et al. Physical Principles of Membrane Shape Regulation by the Glycocalyx. Cell 177, 1757- 1770.e21 (2019).

2. Wagner, C. E., Wheeler, K. M. & Ribbeck, K. Mucins and Their Role in Shaping the Functions of Mucus Barriers. Annu. Rev. Cell Dev. Biol. 34, 189–215 (2018).

3. Ince, D., Lucas, T. M. & Malaker, S. A. Current strategies for characterization of mucin-domain glycoproteins. Curr. Opin. Chem. Biol. 69, 102174 (2022).

4. Reily, C., Stewart, T. J., Renfrow, M. B. & Novak, J. Glycosylation in health and disease. Nat. Rev. Nephrol. 15, 346–366 (2019).

5. Rangel-Angarita, V. & Malaker, S. A. Mucinomics as the Next Frontier of Mass Spectrometry. ACS Chem. Biol. acschembio.1c00384 (2021) doi:10.1021/acschembio.1c00384.

6. Kuo, J. C.-H., Gandhi, J. G., Zia, R. N. & Paszek, M. J. Physical biology of the cancer cell glycocalyx. Nat. Phys. 14, 658–669 (2018).

7. Möckl, L. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. Front. Cell Dev. Biol. 8, 253 (2020).

8. Hollingsworth, M. A. & Swanson, B. J. Mucins in cancer: protection and control of the cell surface. Nat. Rev. Cancer 4, 45–60 (2004).

9. Kufe, D. W. Mucins in cancer: function, prognosis and therapy. Nat. Rev. Cancer 9, 874–885 (2009).

10. Malaker, S. A. et al. Revealing the human mucinome. Nat. Commun. 13, 3542 (2022).

11. Hollenhorst, M. A. et al. Comprehensive analysis of platelet glycoprotein Ibα ectodomain glycosylation. J. Thromb. Haemost. S1538783623000375 (2023) doi:10.1016/j.jtha.2023.01.009.

12. Stavenhagen, K. et al. N- and O -glycosylation Analysis of Human C1-inhibitor Reveals Extensive Mucin- type O -Glycosylation. Mol. Cell. Proteomics 17, 1225–1238 (2018).

13. Acharya, N., Sabatos-Peyton, C. & Anderson, A. C. Tim-3 finds its place in the cancer immunotherapy landscape. J. Immunother. Cancer 8, e000911 (2020).

14. Wolf, Y., Anderson, A. C. & Kuchroo, V. K. TIM3 comes of age as an inhibitory receptor. Nat. Rev. Immunol. 20, 173–185 (2020).

15. Oliveira, T., Thaysen-Andersen, M., Packer, N. H. & Kolarich, D. The Hitchhiker's guide to glycoproteomics. Biochem. Soc. Trans. 49, 1643–1662 (2021).

16. Chernykh, A., Kawahara, R. & Thaysen-Andersen, M. Towards structure-focused glycoproteomics. Biochem. Soc. Trans. 49, 161–186 (2021).

17. Bagdonaite, I. et al. Glycoproteomics. Nat. Rev. Methods Primer 2, 48 (2022).

18. Kesimer, M. & Sheehan, J. K. Mass Spectrometric Analysis of Mucin Core Proteins. in Mucins (eds. McGuckin, M. A. & Thornton, D. J.) vol. 842 67–79 (Humana Press, 2012).

19. Malaker, S. A. et al. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. Proc. Natl. Acad. Sci. 116, 7278–7287 (2019).

20. Shon, D. J. et al. An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. Proc. Natl. Acad. Sci. 117, 21299–21307 (2020).

21. Yang, W., Ao, M., Hu, Y., Li, Q. K. & Zhang, H. Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides (EXoO). Mol. Syst. Biol. 14, (2018).

22. Haurat, M. F. et al. The Glycoprotease CpaA Secreted by Medically Relevant Acinetobacter Species Targets Multiple O -Linked Host Glycoproteins. mBio 11, e02033-20, /mbio/11/5/mBio.02033-20.atom (2020).

23. Yang, S. et al. Optimization of O- GIG for O -Glycopeptide Characterization with Sialic Acid Linkage Determination. Anal. Chem. 92, 10946–10951 (2020).

24. Vainauskas, S. et al. A Broad-Specificity O -Glycoprotease That Enables Improved Analysis of Glycoproteins and Glycopeptides Containing Intact Complex O -Glycans. Anal. Chem. 94, 1060–1069 (2022).

25. Carpenter, J. et al. Assembly and organization of the N-terminal region of mucin MUC5AC: Indications for structural and functional distinction from MUC5B. Proc. Natl. Acad. Sci. 118, e2104490118 (2021).

26. Javitt, G. et al. Assembly Mechanism of Mucin and von Willebrand Factor Polymers. Cell 183, 717-729.e16 (2020).

27. Ermund, A. et al. Mucus threads from surface goblet cells clear particles from the airways. Respir. Res. 22, 303 (2021).

28. Fadda, E. Molecular simulations of complex carbohydrates and glycoconjugates. Curr. Opin. Chem. Biol. 69, 102175 (2022).

29. Casalino, L. et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent. Sci. 6, 1722–1734 (2020).

30. Sztain, T. et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. Nat. Chem. 13, 963– 968 (2021).

31. Wu, P. et al. A Gut Commensal Bacterium Promotes Mosquito Permissiveness to Arboviruses. Cell Host Microbe 25, 101-112.e5 (2019).

32. Noach, I. et al. Recognition of protein-linked glycans as a determinant of peptidase activity. Proc. Natl. Acad. Sci. 114, E679–E688 (2017).

33. Biering, S. B. et al. Genome-wide bidirectional CRISPR screens identify mucins as host factors modulating SARS-CoV-2 infection. Nat. Genet. 54, 1078–1089 (2022).

34. Lu, C.-H. et al. Membrane curvature regulates the spatial distribution of bulky glycoproteins. Nat. Commun. 13, 3093 (2022).

35. Imbert, P. R. C. et al. An Acquired and Endogenous Glycocalyx Forms a Bidirectional “Don’t Eat” and “Don’t Eat Me” Barrier to Phagocytosis. Curr. Biol. 31, 77-89.e5 (2021).

36. Yao, Y. et al. Mucus sialylation determines intestinal host-commensal homeostasis. Cell 185, 1172-1188.e28 (2022).

37. Nason, R. et al. Display of the human mucinome with defined O-glycans by gene engineered cells. Nat. Commun. 12, 4070 (2021).

38. Trastoy, B., Naegeli, A., Anso, I., Sjögren, J. & Guerin, M. E. Structural basis of mammalian mucin processing by the human gut O-glycopeptidase OgpA from Akkermansia muciniphila. Nat. Commun. 11, 4844 (2020).

39. Riley, N. M. & Bertozzi, C. R. Deciphering O -glycoprotease substrate preferences with O-Pair Search. Mol. Omics 18, 908–922 (2022).

40. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int. J. Syst. Evol. Microbiol. 54, 1469–1476 (2004).

41. Riley, N. M., Malaker, S. A. & Bertozzi, C. R. Electron-Based Dissociation Is Needed for O-Glycopeptides Derived from OpeRATOR Proteolysis. Anal. Chem. 92, 14878–14884 (2020).

42. Yang, S. et al. Deciphering Protein O -Glycosylation: Solid-Phase Chemoenzymatic Cleavage and Enrichment. Anal. Chem. 90, 8261–8269 (2018).

43. Bardoel, B. W. et al. Identification of an immunomodulating metalloprotease of Pseudomonas aeruginosa (IMPa): PSGL-1 inhibition by P. aeruginosa. Cell. Microbiol. 14, 902–913 (2012).

44. Bock, S. C. et al. Human C.hivin.1 inhibitor: primary structure, cDNA cloning, and chromosomal localization. Biochemistry 25, 4292–4301 (1986).

45. Halim, A., Nilsson, J., Rüetschi, U., Hesse, C. & Larson, G. Human Urinary Glycoproteomics; Attachment Site Specific Analysis of N- and O-Linked Glycosylations by CID and ECD. Mol. Cell. Proteomics 11, M111.013649 (2012).

46. Mistry, J. et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 49, D412–D419 (2021).

47. Cerdà-Costa, N. & Xavier Gomis-Rüth, F. Architecture and function of metallopeptidase catalytic domains: Metallopeptidase Catalytic Domains. Protein Sci. 23, 123–144 (2014).

48. Noach, I. & Boraston, A. B. Structural evidence for a proline-specific glycopeptide recognition domain in an O-glycopeptidase. Glycobiology 31, 385–390 (2021).

49. Shon, D. J., Kuo, A., Ferracane, M. J. & Malaker, S. A. Classification, structural biology, and applications of mucin domain-targeting proteases. Biochem. J. 478, 1585–1603 (2021).

50. Pluvinage, B. et al. Architecturally complex O -glycopeptidases are customized for mucin recognition and hydrolysis. Proc. Natl. Acad. Sci. 118, e2019220118 (2021).

51. Shon, D. J., Fernandez, D., Riley, N. M., Ferracane, M. J. & Bertozzi, C. R. Structure-guided mutagenesis of a mucin-selective metalloprotease from Akkermansia muciniphila alters substrate preferences. J. Biol. Chem. 298, 101917 (2022).

52. Taleb, V. et al. Structural and mechanistic insights into the cleavage of clustered O-glycan patches-containing glycoproteins by mucinases of the human gut. Nat. Commun. 13, 4324 (2022).

53. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

54. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444 (2022).

55. Medley, B. J. et al. A previously uncharacterized O-glycopeptidase from Akkermansia muciniphila requires the Tn-antigen for cleavage of the peptide bond. J. Biol. Chem. 298, 102439 (2022).

56. Pedram, K. et al. Design of a mucin-selective protease for targeted degradation of cancer-associated mucins. http://biorxiv.org/lookup/doi/10.1101/2022.05.20.492748 (2022) doi:10.1101/2022.05.20.492748.

57. Yu, A. C. Y., Worrall, L. J. & Strynadka, N. C. J. Structural Insight into the Bacterial Mucinase StcE Essential to Adhesion and Immune Evasion during Enterohemorrhagic E. coli Infection. Structure 20, 707–717 (2012).

58. Manka, S. W. et al. Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. Proc. Natl. Acad. Sci. 109, 12461–12466 (2012).

59. Kuchroo, V. K., Umetsu, D. T., DeKruyff, R. H. & Freeman, G. J. The TIM gene family: emerging roles in immunity and disease. Nat. Rev. Immunol. 3, 454–462 (2003).

60. Freeman, G. J., Casasnovas, J. M., Umetsu, D. T. & DeKruyff, R. H. TIM genes: a family of cell surface phosphatidylserine receptors that regulate innate and adaptive immunity: TIM genes: a family of PtdSer receptors. Immunol. Rev. 235, 172–189 (2010).

61. He, X. & Xu, C. Immune checkpoint signaling and cancer immunotherapy. Cell Res. 30, 660–669 (2020).

62. Hellmann, M. D. et al. Safety and Immunogenicity of LY3415244, a Bispecific Antibody Against TIM-3 and PD-L1, in Patients With Advanced Solid Tumors. Clin. Cancer Res. 27, 2773–2781 (2021).

63. Harding, J. J. et al. Blocking TIM-3 in Treatment-refractory Advanced Solid Tumors: A Phase Ia/b Study of LY3321367 with or without an Anti-PD-L1 Antibody. Clin. Cancer Res. 27, 2168–2178 (2021).

64. Curigliano, G. et al. Phase I/Ib Clinical Trial of Sabatolimab, an Anti–TIM-3 Antibody, Alone and in Combination with Spartalizumab, an Anti–PD-1 Antibody, in Advanced Solid Tumors. Clin. Cancer Res. 27, 3620–3629 (2021).

65. Du, P., Xiong, R., Li, X. & Jiang, J. Immune Regulation and Antitumor Effect of TIM-1. J. Immunol. Res. 2016, 1–6 (2016).

66. Miyanishi, M. et al. Identification of Tim4 as a phosphatidylserine receptor. Nature 450, 435–439 (2007).

67. Steentoft, C. et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J. 32, 1478–1488 (2013).

68. Cao, E. et al. T Cell Immunoglobulin Mucin-3 Crystal Structure Reveals a Galectin-9-Independent Ligand- Binding Surface. Immunity 26, 311–321 (2007).

69. Gandhi, A. K. et al. High resolution X-ray and NMR structural study of human T-cell immunoglobulin and mucin domain containing protein-3. Sci. Rep. 8, 17512 (2018).

70. Wandall, H. H., Nielsen, M. A. I., King-Smith, S., Haan, N. & Bagdonaite, I. Global functions of O-glycosylation: promises and challenges in O-glycobiology. FEBS J. febs.16148 (2021) doi:10.1111/febs.16148.

71. Bagdonaite, I. et al. Isoform-specific O-glycosylation dictates Ebola virus infectivity. http://biorxiv.org/lookup/doi/10.1101/2022.04.25.489418 (2022) doi:10.1101/2022.04.25.489418.

72. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. Nat. Rev. Cancer 15, 540–555 (2015).