

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Green Cells & Virus Hells: Building Tools for Protein Engineering in Bioproduction and Pandemic-Screening Applications

Permalink

<https://escholarship.org/uc/item/0d65c3n6>

ISBN

9798189233388

Author

Garimella, Shruthi S

Publication Date

2026-06-25

Peer reviewed|Thesis/dissertation

Green Cells & Virus Hells: Building Tools for Protein Engineering in Bioproduction and
Pandemic-Screening Applications

By

SHRUTHI SATYA GARIMELLA
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Priya S. Shah, Chair

Surl-Hee Ahn

Volkmar Heinrich

Committee in Charge

2026

Acknowledgements

My heart is full of love and gratitude to the people who made this possible.

I have the most generous mentors. I am so grateful for my incredible advisor, Dr. Priya Shah. Her infectious energy and creativity outmatch us all. She was always there for a quick laugh or inspiration when I needed someone to push me forward. Especially at times where nothing went to plan, she reminded me of what I am capable of and pushed me to meet what seemed like impossible expectations. Because of her support, I had the chance to develop new skills and explore multiple fields so I could discover what truly excited me along the way. I am forever grateful for your patience and kindness. To my mentor, Dr. Shiaki Minami. He taught me everything I know. I am so appreciative that you took time to answer all of questions, even when I did not have the words to ask them. Thank you for showing me the importance of having backup plans for my backup plans. Because of Priya and Shiaki, I witnessed what great teaching looks like and I can only hope that I can follow in your paths.

I want to thank my dissertation committee members, Dr. Shirley Ahn and Dr. Volkmar Heinrich. Your thoughtful feedback pushed me to think more critically and sharpened my scientific rigor throughout this work. I am also grateful for the rest of my qualifying exam committee members—Dr. Karen McDonald, Dr. Sam Díaz-Muñoz, and Dr. Jiandi Wan—whose guidance helped lay the foundation for these projects and collaborations.

I had the great fortune to mentor some incredible people: Justin Abad Santos, Anusha Khanchandani, and Guadalupe Lauro. Justin, I am so proud of how quickly you learned new ideas and put them into immediate action. Anusha and Guadalupe, thank you for sticking with me when experiments did not work and nothing seemed to make any sense. It made all the moments of

little wins and big successes even sweeter, being able to share them with you both. Thank you for letting me teach you everything I know.

I want to thank the Shah Lab: Dr. Adam Fishburn, Dr. Nitin Sai Beesabathuni, Dr. Neil Alvin Adia, Ritika Gangaraju, Matthew Kenaston, Nick Lopez, Chase Skawinski, Vardhan Pedammallu, Taylor Pullan, Ines Bilkic, Achintya Bharadwaj, Cole Florio, Susannah Schaffer, Nate Burmas, Avery Robertson, Bhargavi Gidugu, Dr. Oanh Pham, Dr. Parvanendhu Pradeep, Dr. Liuba Cherkashchenko, and everyone else who has passed through. The memories we made together would make anyone jealous. I am so grateful for the support, critiques, good food, and the good vibes.

I am so grateful to the friends I made here in Davis: Archan Vyas, Niko Hansen, Meridian Haas, Randi Swanson, Alejandro Gallegos-Tovar, Varun Core, Thomas Nguyen, Hudson Shih, Kevin Gu, Shelly Zhang, Elizabeth McKittrick, and Judith Fishburn. You all are some of my favorite people. I am so glad I found that seat in the middle of the classroom to find you all beside me. This PhD and my time in Davis would not have been the same without your friendship. I want to especially thank the girls for never letting me doubt that I belong here. I also want to thank my oldest and closest friends Sabriya Alam, Rachel MacLeod, Emily Bywater, Emma Lietzke, and Rachel Dyer for their constant support through all the adventures.

Last but certainly not least, thank you to my family. Nanna—my champion—thank you for showing me how far curiosity and hard work can take you. Amma—my superhero—thank you for instilling in me your strength and perseverance. I will never take for granted the opportunities you both have given me. Tammudu—my cheerleader—thank you for being my first friend and stepping up when I could not be there.

Copyright Acknowledgements

Chapter 2 contains material originally published in BMC Biotechnology:

Garimella, S.S., Minami, S.A., Khanchandani, A.N., Abad Santos, J.C., Schaffer, S.R., Shah, P.S.

A simplified two-plasmid system for orthogonal control of mammalian gene expression using light activated CRISPR effector. BMC Biotechnology 25, 58 (2025). <https://doi.org/10.1186/s12896-025-00994-2>

Chapter 3 contains material originally published in Biotechnology Journal:

Minami, S. A., **Garimella, S. S.**, & Shah, P. S. (2024). Computational evaluation of light propagation in cylindrical bioreactors for optogenetic mammalian cell cultures. *Biotechnology Journal*, 19, e2300071. <https://doi.org/10.1002/biot.202300071>

Text and figures have been modified from its original publication to fit the formatting requirements and content of this thesis.

Abstract

Proteins are the primary molecules to facilitate biological processes at the cellular level. Every cell uses a vast array of proteins in a carefully coordinated concert to orchestrate every function to continue their existence. Proteins very rarely work independently and instead rely on cooperative protein interactions for proper control of chemical reactions and larger biological processes. My dissertation focuses on integrating computational and experimental approaches to study these protein interactions and engineer synthetic tools for bioproduction and public health applications. Engineering tools for protein-based applications have become critical for nearly all biological sciences including research, medical, and agricultural endeavors. Chapters 1-3 discuss inducible gene-expression systems, how we optimize a light-inducible system for mammalian gene expression, and how these systems can be applied to industrial bioproduction applications. Chapters 4-6 focus on understanding the viral characteristics of pandemic-potential, developing a computational pipeline screening virus-host protein-protein interactions, and proposing an experimental workflow to validate the interactions.

Artificial control of proteins using light-inducible genetic systems is a relatively new feat of biological engineering. These optogenetic systems use light-responsive proteins to control gene expression, ion channels, protein localization, and signaling with a simple “flip of a switch”. One such tool is the light activated CRISPR effector (LACE) system. Its ability to regulate gene expression in a tunable, reversible, and spatially resolved manner makes it attractive for many applications. We optimized LACE to reduce the number of plasmids needed to deliver all four components from four to two, thus two-plasmid LACE (2pLACE). The behavior of 2pLACE was rigorously tested using optogenetic control of enhanced green fluorescent protein (eGFP) expression as a reporter for gene activation. Using human HEK293T cells, we optimized the ratio of the two plasmids, measured activation as a function of light intensity, and determined the

frequency of the light to activate the maximum fluorescence. Overall, the 2pLACE system showed a similar dynamic range, tunability, and activation kinetics as well as less variability as the original LACE system. This simplified optogenetics system will be more amenable to biotechnology applications where variability and resources need to be minimized in low-resource environments or tissue engineering.

We also considered how different inducible gene-expression systems might be applied in biomanufacturing. Optogenetic tools, like 2pLACE, can offer improved throughput and protein yield in industrial bioproduction, but it is important to consider how these tools can be applied at scale. Previous work simulated how light propagates in a bioreactor at varying geometries to assess how light-inducible mammalian cell cultures utilize light and activate the system. We experimentally tested these simulations with a bench-top photobioreactor model filled with ExpiCHO cells to measure how light scatters through the culture. We observed little change in light intensity within the center of the system, and only some attenuation of light at the system boundaries. Overall, the experiments supported the results of the light propagation experiments which highlight how optogenetics can be applied in industrial bioproduction.

With practical bioproduction in mind we then explored the feasibility of how inducible gene expression systems can be implemented in tissue engineering and cultivated meat applications. However, instead of light, we used glucose was used as the inducer for its availability. A glucose-inducible gene expression system relies on the glucose already present in cell-culture media to activate gene expression using glucose responsive elements in the promoter region. While this promoter design did not strongly activate expression, a similar system could be useful in situations where photobioreactors are not available.

Viruses, like living organisms, rely on protein interactions to facilitate their replication. Most viral processes are done by hijacking the host's proteins through physical interaction, allowing viruses to infect and cause disease. Of particular concern are coronaviruses, which have recently caused three major pandemics, including the COVID-19 pandemic. The size of this virus family and the frequency of coronavirus outbreaks underscore the urgent need to identify emerging strains with the potential for human infection. A critical step in the emergence of a virus is the ability of the viral protein to bind a host receptor and mediate cell entry. For two pandemic-causing coronaviruses (SARS-CoV-1 and SARS-CoV-2), this interaction occurs between the Spike S1 subunit and the human ACE2 receptor. We predicted how 28 emerging coronaviruses may bind to humans using structure predictions, docking programs, and molecular dynamics. After generating the S1 domains for these viruses, we predicted how these proteins might bind ACE2 using pandemic-causing viruses as controls. Contact analysis quantified the predicted binding which highlighted an emerging coronavirus, Zhejiang2013, for which there is no evidence regarding its ability to bind human ACE2. We used molecular dynamics to further validate the docking predictions and model the interaction over time.

Overall, our results indicated that Zhejiang2013 exhibits dynamic interaction patterns consistent with ACE2 binding. Combining structure predictions with protein docking and molecular dynamics enables us to screen uninvestigated proteins in a rapid and consistent manner before experimentation. We then proposed how to express and purify these proteins to test their binding affinity to ACE2 using biolayer interferometry. Given that two ACE2-binding coronaviruses have caused global pandemics within the past two decades, identifying and testing potential ACE2 binders is critical for early warning and pandemic preparedness.

Together, the culmination of this work highlights the power of studying and manipulating protein interactions to both facilitate desired interactions and inhibit detrimental ones.

Table of Contents

ACKNOWLEDGEMENTS.....	II
COPYRIGHT ACKNOWLEDGEMENTS.....	IV
ABSTRACT	V
LIST OF FIGURES AND TABLES.....	XII
CHAPTER 1: INTRODUCTION TO GENE REGULATION AND INDUCIBLE GENE	
EXPRESSION SYSTEMS.....	1
1.1 FOUNDATIONS OF SYNTHETIC BIOLOGY AND GENE REGULATION	1
1.2 CHARACTERISTICS OF CONTROLLED GENE EXPRESSION.....	2
1.3 TYPES OF INDUCIBLE GENE EXPRESSION SYSTEMS	4
1.4 IMPLEMENTATION OF INDUCIBLE GENE EXPRESSION SYSTEMS.....	5
1.5 REFERENCES	6
CHAPTER 2: A SIMPLIFIED TWO-PLASMID SYSTEM FOR ORTHOGONAL CONTROL OF	
MAMMALIAN GENE EXPRESSION USING LIGHT-ACTIVATED CRISPR EFFECTOR.....	10
2.1 BACKGROUND	10
2.2 RESULTS.....	12
2.2.1 Design and plasmid ratio optimization of the 2pLACE system.....	12
2.2.2 Characterization of transient 2pLACE activation dynamics.....	15
2.2.3 Comparison of 2pLACE activation to 4pLACE.....	18
2.3 DISCUSSION.....	19
2.4 CONCLUSIONS.....	22
2.5 METHODS	23
2.5.1 Plasmids	23
2.5.2 Cell Culture	23

2.5.3 Transient transfection	24
2.5.4 Light activation	25
2.5.5 Flow cytometry.....	25
2.5.6 Fluorescent imaging	25
2.5.7 Statistical analysis.....	26
2.6 REFERENCES	27
 CHAPTER 3: APPLICATIONS OF INDUCIBLE GENE EXPRESSION SYSTEMS IN	
BIOPRODUCTION	34
3.1 INTRODUCTION	34
3.2 LIGHT PROPAGATION THROUGH A PHOTOBIOREACTOR	35
3.2.1 Experimental light propagation bench-top setup.....	36
3.2.2 Demonstration of light propagation in the experimental bioreactor	38
3.3 GLUCOSE-INDUCIBLE GENE EXPRESSION SYSTEM FOR MAMMALIAN CELL GROWTH.....	40
3.3.1 Inducible gene expression tools to improve cultivated meat production.....	40
3.3.2 Glucose-inducible promoter design and rationale.....	42
3.3.3 Activation dynamics of the glucose system and future design considerations.....	43
3.4 CONSIDERATIONS IN APPLYING INDUCIBLE GENE EXPRESSION SYSTEMS TO INDUSTRIAL	
BIOPRODUCTION.....	46
3.4 METHODS	48
Glucose inducible system	48
3.5 REFERENCES	50
 CHAPTER 4: INTRODUCTION TO VIRAL SPILLOVER AND PANDEMIC SCREENING	58
4.1 INCREASING RATES OF ZOO NOTIC SPILLOVER AND OUTBREAK PRESSURE	58
4.2 ENVIRONMENTAL AND VIRAL TRAITS DRIVING ZOO NOTIC SPILLOVER.....	60
4.2.1 Conditions for Pandemic Emergence	60

4.2.2 Adaptive Nature of RNA Viruses.....	61
4.3 REFERENCES	64
CHAPTER 5: PREDICTION OF CORONAVIRUS RECEPTOR USAGE USING PROTEIN STRUCTURE PREDICTION AND MOLECULAR DYNAMICS	71
5.1 INTRODUCTION.....	71
5.2 RESULTS.....	74
5.2.1 Selection of Coronaviruses to Study.....	74
5.2.2 Sequence Alignment	77
5.2.3 Protein Structure Prediction	78
5.2.4 Predicting the S1-ACE2 Interaction.....	81
5.2.5 Simulating the S1-ACE2 Interaction Using Molecular Dynamics	86
5.3 DISCUSSION.....	90
5.4 CONCLUSIONS.....	94
5.5 METHODS	95
5.5.1 Protein Structure Predictions.....	95
5.5.2 Protein Interaction Predictions and Restraints.....	95
5.5.3 Sequence Alignments	97
5.5.4 Structure Alignments.....	97
5.5.5 Structure Composition Analysis.....	97
5.5.6 Coarse-Grained MD Simulations.....	97
5.5.7 Steered MD Pulling Simulations.....	98
5.5.8 Statistical Analysis.....	99
5.6 REFERENCES	100
CHAPTER 6: ONGOING EXPERIMENTAL VALIDATION AND FINAL CONCLUSIONS ON PANDEMIC SCREENING	111

6.1 INTRODUCTION	111
6.2 EXPERIMENTALLY TESTING EMERGING CORONAVIRUSES TO A HUMAN RECEPTOR.....	112
6.2.1 Ongoing work: testing protein expression and purification.....	112
6.2.2 Future work: measuring binding affinity	117
6.3 CONCLUSIONS.....	118
6.4 METHODS	120
6.4.1 Plasmid Construction	120
6.4.2 Cell Culture	120
6.4.3 Transfection	120
6.4.5 Protein Harvest and Purification	121
6.4.6 SDS-PAGE and Western blot.....	121
6.5 REFERENCES	122
APPENDIX A: SUPPORTING INFORMATION FOR A SIMPLIFIED TWO-PLASMID SYSTEM FOR ORTHOGONAL CONTROL OF MAMMALIAN GENE EXPRESSION USING LIGHT- ACTIVATED CRISPR EFFECTOR.....	127
APPENDIX B: SUPPLEMENTAL INFORMATION FOR PREDICTION OF CORONAVIRUS RECEPTOR USAGE USING PROTEIN STRUCTURE PREDICTION AND MOLECULAR DYNAMICS.....	133

List of Figures and Tables

CHAPTER 1: INTRODUCTION TO GENE REGULATION AND INDUCIBLE GENE

EXPRESSION SYSTEMS..... 1

FIGURE 1-1. DIAGRAM OF INDUCIBLE GENE EXPRESSION. 3

FIGURE 1-2. METRICS OF INDUCIBLE GENE EXPRESSION SYSTEMS. 4

CHAPTER 2: A SIMPLIFIED TWO-PLASMID SYSTEM FOR ORTHOGONAL CONTROL OF MAMMALIAN GENE EXPRESSION USING LIGHT-ACTIVATED CRISPR EFFECTOR.....10

FIGURE 2-1. SCHEMATIC OF 2PLACE DESIGN.13

FIGURE 2-2. 2PLACE PLASMID RATIO OPTIMIZATION FOR HIGH ABSOLUTE ACTIVATION AND DYNAMIC
RANGE.....14

FIGURE 2-4. COMPARISON OF 2PLACE AND 4PLACE ACTIVATION KINETICS.18

TABLE 2-1 PRIMER SEQUENCES FOR MYCOPLASMA PCR.....23

TABLE 2-2 RELATIVE MASS RATIOS OF THE TRANSIENTLY TRANSFECTED LACE SYSTEM PLASMIDS24

CHAPTER 3: APPLICATIONS OF INDUCIBLE GENE EXPRESSION SYSTEMS IN

BIOPRODUCTION34

FIGURE 3-1. EXPERIMENTAL SETUP FOR LIGHT PROPAGATION MEASUREMENTS.37

FIGURE 3-2. COMPARISONS OF COMPUTATIONAL PREDICTIONS TO EXPERIMENTAL MEASUREMENTS
OF LIGHT INTENSITY IN A CYLINDRICAL SYSTEM.39

FIGURE 3-3. GLUCOSE-INDUCIBLE PROMOTER DESIGN.43

FIGURE 3-4 GLUCOSE-INDUCIBLE PROMOTER ACTIVATION.....45

CHAPTER 4: INTRODUCTION TO VIRAL SPILLOVER AND PANDEMIC SCREENING58

FIGURE 4-1. GENERAL DIAGRAM OF SPILLOVER.....59

FIGURE 4-2. TIMELINE OF EPIDEMICS AND PANDEMIC ACROSS THE 20TH AND 21ST CENTURY.....59

CHAPTER 5: PREDICTION OF CORONAVIRUS RECEPTOR USAGE USING PROTEIN STRUCTURE PREDICTION AND MOLECULAR DYNAMICS	71
FIGURE 5-1. OVERVIEW OF THE STUDY.	75
TABLE 5-1. COMPILATION OF CORONAVIRUSES USED TO TEST HUMAN ACE2 BINDING.	76
FIGURE 5-2. SEQUENCE ALIGNMENT HEATMAP OF S1 (PURPLE) AND RBD (BLUE) USING PAIRWISE ALIGNMENT OF EACH BETACORONAVIRUS.	78
FIGURE 5-3. ALPHAFOLD2 PROTEIN PREDICTIONS OF BETACORONAVIRUS S1 STRUCTURES.	81
FIGURE 5-4. STRUCTURE ANALYSIS OF EMERGING SPIKE S1 AND RBD REGIONS.	82
FIGURE 5-6. COARSE-GRAINING MONTAGE ACROSS 1000 NS SIMULATION OF SARS-CoV-1, MERS- CoV, AND ZHEJIANG2013 S1 BINDING TO HUMAN ACE2.	88
FIGURE 5-7. PULLING SIMULATIONS REVEAL A STRONGER BINDING ENERGY BETWEEN ZHEJIANG2013 AND HUMAN ACE2.	89
CHAPTER 6: ONGOING EXPERIMENTAL VALIDATION AND FINAL CONCLUSIONS ON PANDEMIC SCREENING	111
FIGURE 6-1. PROBING FOR S1 EXPRESSION IN EXPICHO CELL LYSATE AND SUPERNATANT.	114
FIGURE 6-2. PROBING FOR S1 PROTEIN THROUGH THE PROTEIN PURIFICATION PROCESS.	116
APPENDIX A: SUPPORTING INFORMATION FOR A SIMPLIFIED TWO-PLASMID SYSTEM FOR ORTHOGONAL CONTROL OF MAMMALIAN GENE EXPRESSION USING LIGHT- ACTIVATED CRISPR EFFECTOR.....	127
FIGURE A-1. SCHEMATIC OF PLASMID DESIGN FROM THE ORIGINAL FOUR LACE PLASMIDS (4PLACE) TO THE OPTIMIZED TWO PLASMIDS (2PLACE).....	128
FIGURE A-2. EGFP EXPRESSION OF THE OPTIMIZED PLASMID RATIO FOR 2PLACE WITH AND WITHOUT LIGHT ACTIVATION.....	129
FIGURE A-3. CALIBRATION CURVE OF THE OPTOPLATE-96	129

TABLE A-1. P-VALUES FROM ANOVA STATISTICS COMPARING DYNAMIC RANGE OF PLASMID RATIO EGFP EXPRESSION	130
TABLE A-2. P-VALUES FROM ANOVA STATISTICS COMPARING EGFP EXPRESSION OF LED INTENSITY SAMPLES (MW/CM ²)	131
TABLE A-3. P-VALUES FROM ANOVA STATISTICS COMPARING CONDITIONS OF ACTIVATION TIME (HOURS)	132
TABLE A-4. P-VALUES FROM ANOVA STATISTICS COMPARING PULSE NUMBER CONDITIONS.....	132
APPENDIX B: SUPPLEMENTAL INFORMATION FOR PREDICTION OF CORONAVIRUS RECEPTOR USAGE USING PROTEIN STRUCTURE PREDICTION AND MOLECULAR DYNAMICS.....	
FIGURE B-1. SECONDARY STRUCTURE COMPOSITION	133
FIGURE B-2. S1-ACE2 STATIC DOCKING PREDICTIONS OF ALPHAFOLD MULTIMER AND HADDOCK.	134
FIGURE B-3. ROOT MEAN SQUARED DEVIATION (RMSD) OF COARSE-GRAINING SIMULATIONS.....	135
FIGURE B-5. WORK ACTED UPON THE PULLING SIMULATIONS AT A PULLING RATE OF 1 NM/NS.	136
FIGURE B-6. TRACES OF PULL FORCE AS A FUNCTION OF PULL DISTANCE FOR ALL-ATOM STEERED MD SIMULATIONS COMPARING ZHEJIANG2013 AND SARS-CoV-1.	137
TABLE B-1. SEQUENCE IDENTITY MATRIX OF PAIRWISE COMPARISONS OF S1 PROTEINS (%).....	138
TABLE B-2. SEQUENCE IDENTITY MATRIX OF PAIRWISE COMPARISONS OF RBD PROTEINS (%)....	144
TABLE B-3. PAIRWISE COMPARISON OF S1 TM-SCORES	148
TABLE B-4. PAIRWISE COMPARISON OF RBD TM-SCORES	164
TABLE B-5. SECONDARY STRUCTURE COMPOSITION OF S1 AND RBD CALCULATED BY STRIDE .	179
TABLE B-7. RESIDUES USED FOR STATIC DOCKING RESTRAINTS	180
TABLE B-8. NUMBER OF CONTACTS CALCULATED BY PROTEIN DOCKING PLATFORM.....	181

CHAPTER 1: Introduction to gene regulation and inducible gene expression systems

1.1 Foundations of synthetic biology and gene regulation

In 1961, François Jacob and Jacques Monod published a paper that laid out the basic rules of gene expression. This work explained how bacterial genes are expressed and regulated through a timed mechanism that controls protein synthesis in response to the environment ^[1]. This work revolutionized the world of molecular biology by understanding the *lac* operon in bacteria and the universal mechanism of regulatory gene circuits for cell protein synthesis. The key concept is that gene expression is neither static nor linear, but rather dynamically regulated so that genes are activated or repressed in response to environmental stimuli. This framework introduced that cellular behavior is governed by interactions between regulatory elements that could later be modeled and organized as engineered tools.

Advances in molecular biology techniques and genetic engineering have enabled researchers to manipulate gene expression with precision. Biological systems have become programmable in which genetic components can be assembled in a modular fashion to build synthetic tools for specific functions. This boom in synthetic biology has accelerated progress across disciplines such that gene expression is optimized to improve the production of biomolecules, engineer proteins, or model cell functions. Across these fields, the ability to precisely control gene expression has emerged as a central challenge.

1.2 Characteristics of controlled gene expression

Many genes are expressed on constitutive gene expression, under an “always on” design so genes are continuously transcribed regardless of external conditions. This is critical for housekeeping genes to maintain essential cellular functions and enabling strong activation and consistent output. However, in many industrial applications, constitutive gene expression is insufficient as continuous bioproduction can impose metabolic burden, toxicity, or lack of control [2]. Specifically in mammalian systems, reducing the production of recombinant protein improved cell recovery and reduced cell death [3,4]. Systems that allow genes to be controlled in response to a specific signal can optimize performance by enabling dynamic regulation. Inducible gene expression systems give researchers control to tune and reverse gene expression [5–8].

Inducible gene expression systems provide a strategy that achieves the temporal and reversible control of gene expression in an orthogonal manner. These systems generally require an inducer, a corresponding transcription factor, and a regulatory promoter to activate the expression of a gene of interest. Unlike constitutive systems, promoters used in inducible systems are “minimal” or the shortest sequence required to maintain the core promoter function to regulate the specificity of activation [9–11]. Canonically, the presence of the inducer triggers the transcription factor to bind to the promoter to activate expression (Figure 1-1). The opposite can also be true—some systems are designed to repress expression in response to an inducer [5,12–14]. In this situation, an inducer triggers a transcription factor to bind and block gene expression. Additionally, in the absence of an inducer, a repressor can silence the promoter. The addition of the inducer would then neutralize the repressor to activate expression [15–17].

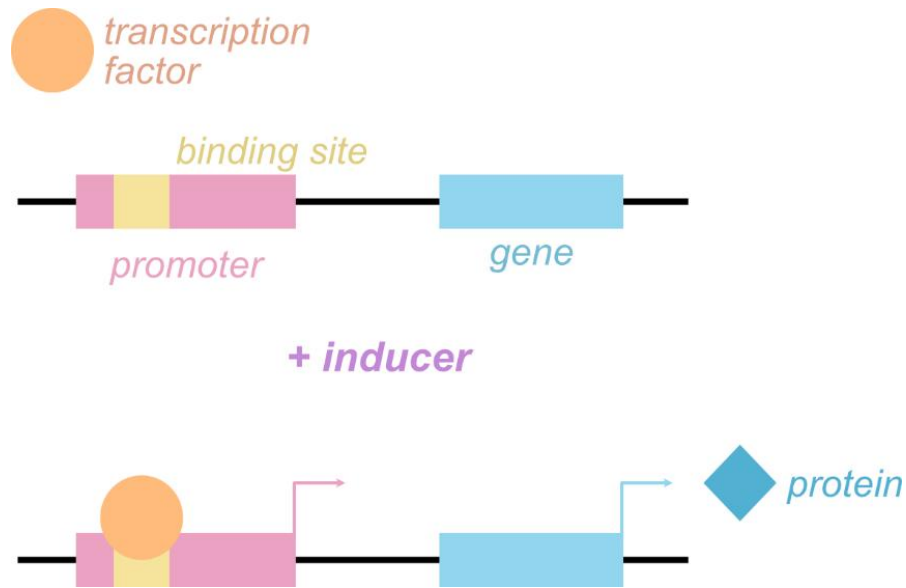


Figure 1-1. Diagram of inducible gene expression. Most inducible expression systems rely on a single inducer and inducer-responsive elements. When the inducer is present, the transcription factor binds to the promoter to activate gene expression to produce protein.

The effectiveness of an inducible system is determined through a few metrics—reversibility, tunability, and dynamic range (Figure 1-2)^[6,18]. Reversibility refers to the ability for the system control expression like an “on/off” switch (Figure 1-2A). Tunability ensures that the system can be proportionally regulated to varying amounts of the inducer (Figure 1-2B). Dynamic range is an important metric, calculated as the ratio between gene expression when the system is “on” versus “off”. High dynamic range indicates strong activation and minimal background expression, or leakiness (Figure 1-2C) ^[6,18]. These properties determine how precisely an inducible system can control expression across different conditions.

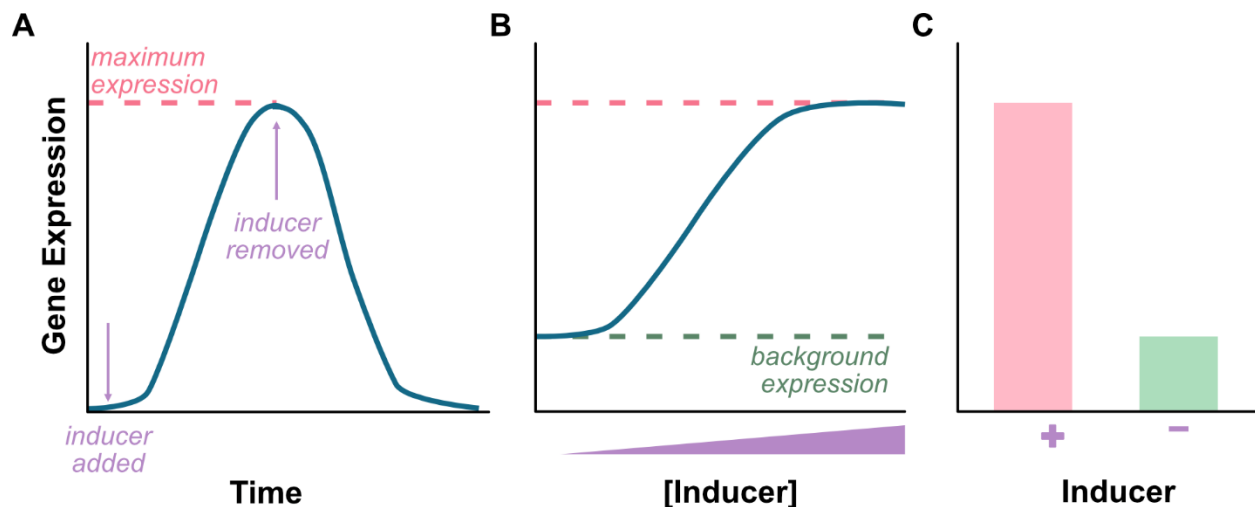


Figure 1-2. Metrics of inducible gene expression systems. A) Reversibility of an inducible system means that the system activates with the addition of an inducer and deactivates when it is removed. B) Gene expression measured in response to changing inducer concentrations. Most inducible systems experience background expression where the system is still activated to a minimal degree, even without the presence of the inducer. The optimal system can balance background expression with the maximum expression, tuned to the inducer. C) Dynamic range is the ratio of “maximum expression” vs “background expression” to understand the magnitude on vs off expression.

1.3 Types of inducible gene expression systems

Inducers are usually external inputs such as a chemical, temperature, or light ^[5]. Chemically induced systems are the most well-established and widely used tools. One of the first inducible systems is Tet-On. The benchmark system uses tetracycline (or its derivative doxycycline) to bind to a tetracycline-responsive promoter to activate gene transcription ^[11,19]. These systems offer robust gene expression with simple implementation ^[10,12]. However, chemicals can have off-target effects and lack spatial resolution ^[18,20]. Additionally, removing the chemical inducer can pose a challenge which may limit its scalability and feasibility in sensitive situations. Temperature-based

inducible systems rely on a sudden temperature difference like a heat or cold shock to stimulate the system. This type of system is a simple design but lacks any specificity as sudden temperature changes can also effect critical cellular processes [5,21]. The unintended effects can complicate comprehension and limit its applications. Light-inducible systems, or optogenetic systems, offer orthogonal gene expression with spatiotemporal control using specific wavelengths of light. Light can be applied and removed rapidly which enables clear reversibility [22]. Light is more accessible and removes the need for additional reagents expanding its chances of scalability. However, optogenetics systems can be sensitive to ambient light which can increase leaky expression if the environment is not well-controlled [7,23,24]. In mammalian systems, challenges such as light penetration and high energy, low wavelength light can complicate its implementation [25,26].

1.4 Implementation of inducible gene expression systems

Synthetic biology and genetic engineering systems give us the means to edit and manipulate genetic information to create novel biological circuits for custom functions. Instead of treating gene expression as a fixed property, inducible systems allow us to ask how timing, dose, and position effect biological processes. This becomes valuable when we are trying to build tools to control cell behavior, tissue patterning, or protein production. It is therefore critical to assess the applicability of these systems and how they might be used outside a research setting.

To bridge the gap between research lab discovery and industrial application, the following two chapters focus on the optimization, scalability, and design of inducible gene expression systems for translation into biotechnology and industrial settings. Chapter 2 discusses how we simplified an optogenetic CRISPR effector into a two-plasmid platform to improve its delivery. By optimizing the lighting conditions and testing for reversibility, tunability, and dynamic range, we showed how this systems can efficiently control gene expression in mammalian cells. Chapter 3 expands on

this work by considering how to scale the systems for bioproduction. We experimentally validated computational simulations on how light propagates through a bioreactor of light-inducible mammalian cells. Additionally, we considered if another type of inducible system could be used in tissue engineering by using a non-antibiotic inducer already found in cell culture media. Taken together, inducible gene expression systems offer new modalities of controlling gene expression, but it is critical to assess how to employ and scale the tools for relevant applications.

1.5 References

1. Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3), 318–356. [https://doi.org/10.1016/S0022-2836\(61\)80072-7](https://doi.org/10.1016/S0022-2836(61)80072-7)
2. Agha-Mohammadi, S., & Lotze, M. T. (2000). Regulatable systems: Applications in gene therapy and replicating viruses. *The Journal of Clinical Investigation*, 105(9), 1177–1183. <https://doi.org/10.1172/JCI10027>
3. Ong, E.-C., Smidt, P., & McGrew, J. T. (2019). Limiting the metabolic burden of recombinant protein expression during selection yields pools with higher expression levels. *Biotechnology Progress*, 35(5), e2839. <https://doi.org/10.1002/btpr.2839>
4. Poulain, A., Mullick, A., Massie, B., & Durocher, Y. (2019). Reducing recombinant protein expression during CHO pool selection enhances frequency of high-producing cells. *Journal of Biotechnology*, 296, 32–41. <https://doi.org/10.1016/j.jbiotec.2019.03.009>
5. Kallunki, T., Barisic, M., Jäättelä, M., & Liu, B. (2019). How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells*, 8(8), 796. <https://doi.org/10.3390/cells8080796>

6. De Carluccio, G., Fusco, V., & di Bernardo, D. (2024). Engineering a synthetic gene circuit for high-performance inducible expression in mammalian systems. *Nature Communications*, 15(1), 3311. <https://doi.org/10.1038/s41467-024-47592-y>
7. Polstein, L. R., & Gersbach, C. A. (2012). Light-Inducible Spatiotemporal Control of Gene Activation by Customizable Zinc Finger Transcription Factors. *Journal of the American Chemical Society*, 134(40), 16480–16483. <https://doi.org/10.1021/ja3065667>
8. Siciliano, V., Menolascina, F., Marucci, L., Fracassi, C., Garzilli, I., Moretti, M. N., & Bernardo, D. di. (2011). Construction and Modelling of an Inducible Positive Feedback Loop Stably Integrated in a Mammalian Cell-Line. *PLOS Computational Biology*, 7(6), e1002074. <https://doi.org/10.1371/journal.pcbi.1002074>
9. Jin, L., Nawab, S., Xia, M., Ma, X., & Huo, Y. (2019). Context-dependency of synthetic minimal promoters in driving gene expression: A case study. *Microbial Biotechnology*, 12(6), 1476–1486. <https://doi.org/10.1111/1751-7915.13489>
10. Mullick, A., Xu, Y., Warren, R., Koutroumanis, M., Guilbault, C., Broussau, S., Malenfant, F., Bourget, L., Lamoureux, L., Lo, R., Caron, A. W., Pilotte, A., & Massie, B. (2006). The cumate gene-switch: A system for regulated expression in mammalian cells. *BMC Biotechnology*, 6(1), 43. <https://doi.org/10.1186/1472-6750-6-43>
11. Gossen, M., & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*, 89(12), 5547–5551. <https://doi.org/10.1073/pnas.89.12.5547>
12. Pedone, E., Postiglione, L., Aulicino, F., Rocca, D. L., Montes-Olivas, S., Khazim, M., di Bernardo, D., Pia Cosma, M., & Marucci, L. (2019). A tunable dual-input system for on-demand dynamic gene expression regulation. *Nature Communications*, 10(1), 4481. <https://doi.org/10.1038/s41467-019-12329-9>

13. Yamada, M., Suzuki, Y., Nagasaki, S. C., Okuno, H., & Imayoshi, I. (2018). Light Control of the Tet Gene Expression System in Mammalian Cells. *Cell Reports*, 25(2), 487-500.e6. <https://doi.org/10.1016/j.celrep.2018.09.026>
14. Peng, Y., Zhang, H., Xu, M., & Tan, M.-W. (2018). A Tet-Off gene expression system for validation of antifungal drug targets in a murine invasive pulmonary aspergillosis model. *Scientific Reports*, 8(1), 443. <https://doi.org/10.1038/s41598-017-18868-9>
15. McShane, E., & Selbach, M. (2014). Gene expression: Degrade to derepress. *The EMBO Journal*, 33(5), 407–408. <https://doi.org/10.1002/embj.201387752>
16. Bellanger, X., Morel, C., Decaris, B., & Guédon, G. (2007). Derepression of Excision of Integrative and Potentially Conjugative Elements from *Streptococcus thermophilus* by DNA Damage Response: Implication of a *ci*-Related Repressor. *Journal of Bacteriology*, 189(4), 1478–1481. <https://doi.org/10.1128/JB.01125-06>
17. Nijland, R., Veening, J.-W., & Kuipers, O. P. (2007). A derepression system based on the *Bacillus subtilis* sporulation pathway offers dynamic control of heterologous gene expression. *Applied and Environmental Microbiology*, 73(7), 2390–2393. <https://doi.org/10.1128/AEM.02266-06>
18. Siddiqui, M., Tous, C., & Wong, W. W. (2022). Small molecule-inducible gene regulatory systems in mammalian cells: Progress and design principles. *Current Opinion in Biotechnology*, 78, 102823. <https://doi.org/10.1016/j.copbio.2022.102823>
19. Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., & Bujard, H. (1995). Transcriptional Activation by Tetracyclines in Mammalian Cells. *Science*, 268(5218), 1766–1769. <https://doi.org/10.1126/science.7792603>
20. Almagor, H., & Paigen, K. (1988). Chemical kinetics of induced gene expression: Activation of transcription by noncooperative binding of multiple regulatory molecules. *Biochemistry*, 27(6), 2094–2102. <https://doi.org/10.1021/bi00406a042>

21. Thaisuchat, H., Baumann, M., Pontiller, J., Hesse, F., & Ernst, W. (2011). Identification of a novel temperature sensitive promoter in CHO cells. *BMC Biotechnology*, 11, 51. <https://doi.org/10.1186/1472-6750-11-51>
22. Motta-Mena, L. B., Reade, A., Mallory, M. J., Glantz, S., Weiner, O. D., Lynch, K. W., & Gardner, K. H. (2014). An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nature Chemical Biology*, 10(3), 196–202. <https://doi.org/10.1038/nchembio.1430>
23. Armbruster, A., Mohamed, A. M., Phan, H. T., & Weber, W. (2024). Lighting the way: Recent developments and applications in molecular optogenetics. *Current Opinion in Biotechnology*, 87, 103126. <https://doi.org/10.1016/j.copbio.2024.103126>
24. Fenno, L., Yizhar, O., & Deisseroth, K. (2011). The Development and Application of Optogenetics. *Annual Review of Neuroscience*, 34(Volume 34, 2011), 389–412. <https://doi.org/10.1146/annurev-neuro-061010-113817>
25. Kuwasaki, Y., Suzuki, K., Yu, G., Yamamoto, S., Otabe, T., Kakiyama, Y., Nishiwaki, M., Miyake, K., Fushimi, K., Bekdash, R., Shimizu, Y., Narikawa, R., Nakajima, T., Yazawa, M., & Sato, M. (2022). A red light-responsive photoswitch for deep tissue optogenetics. *Nature Biotechnology*, 40(11), 1672–1679. <https://doi.org/10.1038/s41587-022-01351-w>
26. Klewer, D. L., & Wu, P. D. Y.-W. (2019). Light-Induced Dimerization Approaches to Control Cellular Processes. *Chemistry (Weinheim an Der Bergstrasse, Germany)*, 25(54), 12452. <https://doi.org/10.1002/chem.201900562>

Chapter 2: A simplified two-plasmid system for orthogonal control of mammalian gene expression using light-activated CRISPR effector

2.1 Background

Inducible mammalian gene expression systems are valuable tools in research, biomedical, and biotechnology applications. They can control the expression of genes involved in processes such as stem cell differentiation, biosynthetic pathway optimization, and improved protein production in bioreactors ^[1–7]. Orthogonal modes to control gene expression can be especially useful depending on the application and can even be multiplexed. Inducible systems that use chemical or temperatures to activate are easy to implement and reversible, but they can have off-target effects and lack precision ^[8–10]. Light-inducible or optogenetic systems function by using a specific wavelength of light to excite a light-activated protein engineered to control biological process ^[11–14]. Optogenetic systems are uniquely suited for reversible and precise spatial control of gene expression because the light can easily be switched off and targeted to specific location via optical setups ^[15–18].

Several optogenetic systems have been developed recently using actuation by different wavelengths, including UV, blue, green, and red spectra ^[17,19–23]. Red light systems may be of interest from an optical standpoint, since the lower energy light is less cytotoxic. In cells, red light also has lower absorption and scattering coefficients compared to blue light. Consequently, red light provides higher spatial control and deeper penetration into thicker tissue samples ^[23,24]. The

iLight and iLight2 systems take advantage of a conformational change in a red light-responsive protein to expose a DNA binding domain and drive gene expression ^[25–27]. The activation of these systems depend on biliverdin (BV) chromophore concentrations in mammalian tissue. Other red/far-red systems removed dependence on BV, such as MagRed and REDLIP, which leverage CRISPR-dCas9 for endogenous mammalian gene expression with minimal background expression ^[24,28]. These systems have shown how red-light inducible gene expression in vivo to further illustrate its therapeutic potential. As red-light inducible systems improve, it becomes important to further develop additional inducible systems that operate in different wavelengths for multiplexing applications.

Blue-light systems are the most widely used optogenetic systems because of their higher chromophore efficiencies and the development of several light-responsive systems engineered to have improved performance ^[29,30]. Light-Activated CRISPR Effector (LACE) is one such system that takes advantage of protein dimerization following excitation by blue light. The light-responsive protein CRY2 is fused to viral protein 64 (VP64), a transcriptional activation domain derived from herpesvirus. CIBN is fused to deactivated-Cas9 (dCas9), which is targeted to a minimal CMV promoter or an endogenous promoter using an appropriate guide RNA (gRNA). CRY2 undergoes a conformational change when stimulated by blue light and dimerizes with CIBN. CRY2-CIBN dimerization brings VP64 to the promoter to activate gene expression. As a reversible system, gene expression turns off when light is removed, when CRY2 dissociates from CIBN ^[20,31].

LACE is a powerful tool that enables modular optogenetics with flexibility to target exogenous and endogenous genes. Endogenous gene expression can be controlled if a suitable gRNA can be designed ^[20]. However, the LACE system requires delivery of four plasmids, which may limit how many cells receive all four components, especially in hard-to-transduce cell types. This limitation can increase variability in the cellular response, and increase the complexity of applying LACE to

cell types relevant for biomedical and biotechnology applications ^[31]. Here, we improve upon the LACE system by decreasing the number of plasmids required and re-optimizing the ratio of the two combined plasmids. We compare two-plasmid LACE (2pLACE) to four-plasmid LACE (4pLACE) in human HEK293T cells and find that 2pLACE has similar dynamic range, tunability, and kinetics to 4pLACE with the added advantage of being more consistent.

2.2 Results

2.2.1 Design and plasmid ratio optimization of the 2pLACE system

We first set out to design a LACE system that would be easier to use and more efficient for various applications. The LACE system requires four different components to be delivered to cells. Originally, these four components were delivered on four separate plasmids by transient transfection^[20,31]. Consequently, the upper limit on LACE activation is dictated by the cells that receive sufficient amounts of all four plasmids. We therefore hypothesized that reducing the number of plasmids used to deliver LACE components would improve LACE behavior. We designed a system to reduce the number of plasmids from four to two by combining the CRY2-VP64 and minCMV-eGFP components and combining the CIBN-dCas9 and gRNA components (Figure 2-1 and A-1). These plasmids were synthesized commercially based on our design using a pCDNA3.1 backbone.

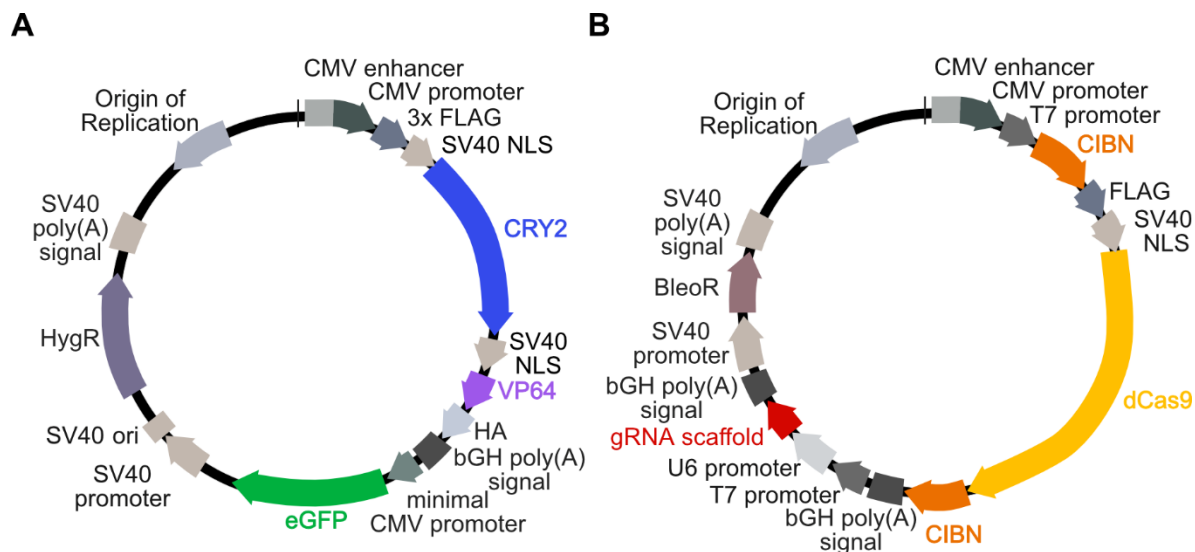


Figure 2-1. Schematic of 2pLACE design. (A) Expression cassettes for CRY2-VP64 and target gene (eGFP) with minimal CMV promoter and gRNA binding sites were combined into a single plasmid (CRY2-eGFP). (B) CIBN-dCas9-CIBN and gRNA expression cassettes were combined into a second plasmid (CIBN-gRNA). Each plasmid was designed to include an SV40 origin of replication (ori) for transient plasmid maintenance in cells expressing Large T-antigen. A distinct selection marker in each plasmid (Hygromycin and Bleocin) was included to enable stable cell line generation.

As an imbalance of the LACE components may influence leaky gene expression, maximal activation, and the dynamic range, we optimized the mass ratios of the two plasmids. Cells were kept in the dark or activated with pulsing blue light for 24 hours, and eGFP reporter fluorescence was analyzed through flow cytometry. As the ratio of CRY2-eGFP to CIBN-gRNA plasmid was increased, the background expression of eGFP in the dark consistently increased. Expression of eGFP following activation also increased as a function of plasmid ratio, but peaks at a ratio of 3:7 and slowly decreased beyond that (Figure 2-2A and A-2). Dynamic range (ratio of light:dark signal) was significantly higher than dynamic range for untransfected cells for each plasmid ratio tested (Figure 2-2B). Interestingly, the dynamic range started high but decreased as the CRY2-eGFP to

CIBN-gRNA ratio increased beyond 6:4. The dynamic range for ratios 6:4 and beyond were significantly lower than dynamic range for the 3:7 ratio (Table A-1). This suggests that there may be several suitable plasmid ratios. To balance high eGFP expression following light activation and high dynamic range, the ratio of 3:7 was used for all subsequent experiments.

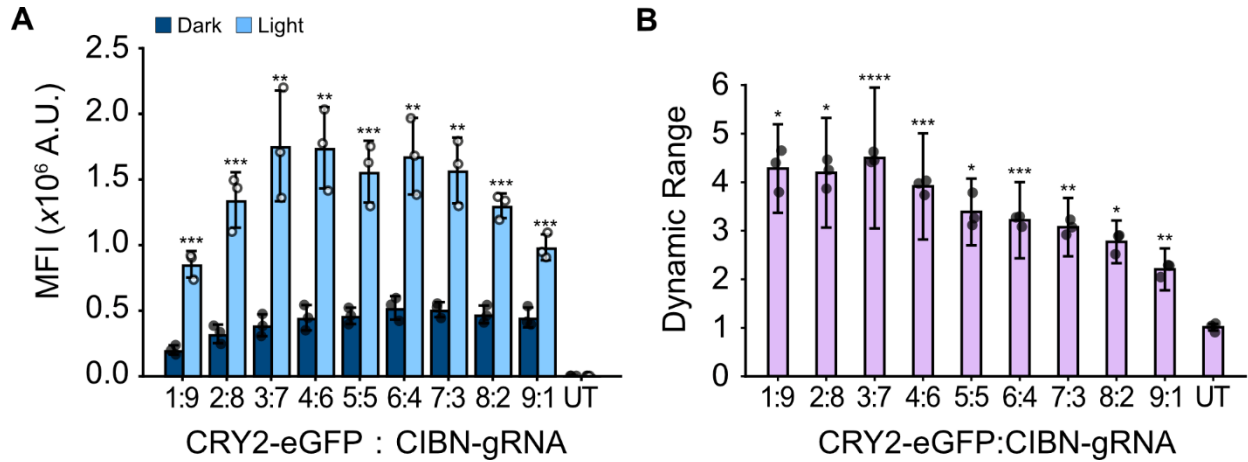


Figure 2-2. 2pLACE plasmid ratio optimization for high absolute activation and dynamic range. (A) Mean fluorescence intensity (MFI) was measured for cells transfected with an increasing ratio of CRY2-eGFP:CIBN-gRNA plasmids followed by 24 hours of incubation in light or dark. (B) Dynamic range for different plasmid ratios from (A) was calculated as MFI in light vs MFI in dark. Plots represent the mean of three independent biological replicates and error bars represent standard deviation. UT (untransfected) is the negative control. MFI significance was calculated with a one-way t-test comparing the light and dark state of a single plasmid ratio. Dynamic range significance was calculated with a one-way ANOVA and Tamhane's T2 post-hoc test to UT. P-values for comparison between other conditions are shown in Table A-1. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, and ns is not significant.

2.2.2 Characterization of transient 2pLACE activation dynamics

Next, we measured the responses of the system to lighting conditions to assess tunability. We first tested the sensitivity of the system to LED intensity, where intensity was varied and eGFP fluorescence was measured 24 hours post-activation. Experiments were performed using the optoPlate to enable high-throughput measurements in a 96-well format ^[32]. We found that the tunable range of eGFP expression occurred at intensities between 0-2 mW/cm², with saturation behavior beyond those intensities (Figure 2-3A). Significant activation occurred with light intensity as low as 0.12 mW/cm², which was the lowest intensity tested (Figure A-3). However, the activation was still quite variable for low intensities of light. Significant activation saturated around 2-3 mW/cm² and maximum system activation occurred 7.26 mW/cm², where comparison between the conditions did not show any significant differences (Table A1). For subsequent experiments, we used an intensity of 9.23 mW/cm². This corresponds to the intensity used to test the optimal mass ratio of 2pLACE (Figure 2-2) as well as the intensity used in previous work ^[31]. Maintaining a similar intensity as the 4pLACE experiments also allows for direct comparison of system response.

We next measured activation kinetics. We activated cells transfected with 2pLACE for increasing amounts of time using the same pulse frequency. eGFP fluorescence was measured by flow cytometry. This time-course experiment showed that the system had an initial lag in expression before reaching significant but minimal expression as early as 4 hours (Figure 2-3B). 2pLACE's activation continued to increase with prolonged activation time for significant expression (Table A-3). The activation kinetics of 2pLACE confirmed that short exposure or activation times turns the system on, but the longer the system is exposed to light, the greater the protein can be expressed.

Our kinetics experiments suggest that prolonged exposure to light maximizes gene expression. However, blue light can be cytotoxic ^[33]. Moreover, some applications may benefit if the system

can be activated with a short exposure to light ^[33–35]. We therefore evaluated the sensitivity of the system to the total number of pulses to determine if there was a critical threshold of pulses for efficient activation while still allowing time after activation for the protein to be expressed. Cells were transfected with 2pLACE and activated with blue light for the specified number of pulses and the same pulse frequency. eGFP fluorescence was measured at 24 hours after pulsing was initiated (Figure 2-3C). Significant activation occurred at 120 pulses, which corresponds to 0.5 hours of activation time. Interestingly, this is lower than the activation time required for significant activation in our kinetics experiment (Figure 2-3B) and suggests that there is a lag between CRY2 association with the promoter and the accumulation of a detectable amount of protein in this system. However, even though significant activation was observed for 0.5 hours of pulsing, the reporter signal only reached substantial values starting at 1440 pulses, corresponding to 6 hours of activation time (Table A-3). Absolute activation and dynamic range continued to increase to 2880 pulses, which corresponds to 12 hours of activation time. Thus, 2pLACE can activate gene expression with as little as 0.5 hours of activation; however there are benefits to prolonged activation.

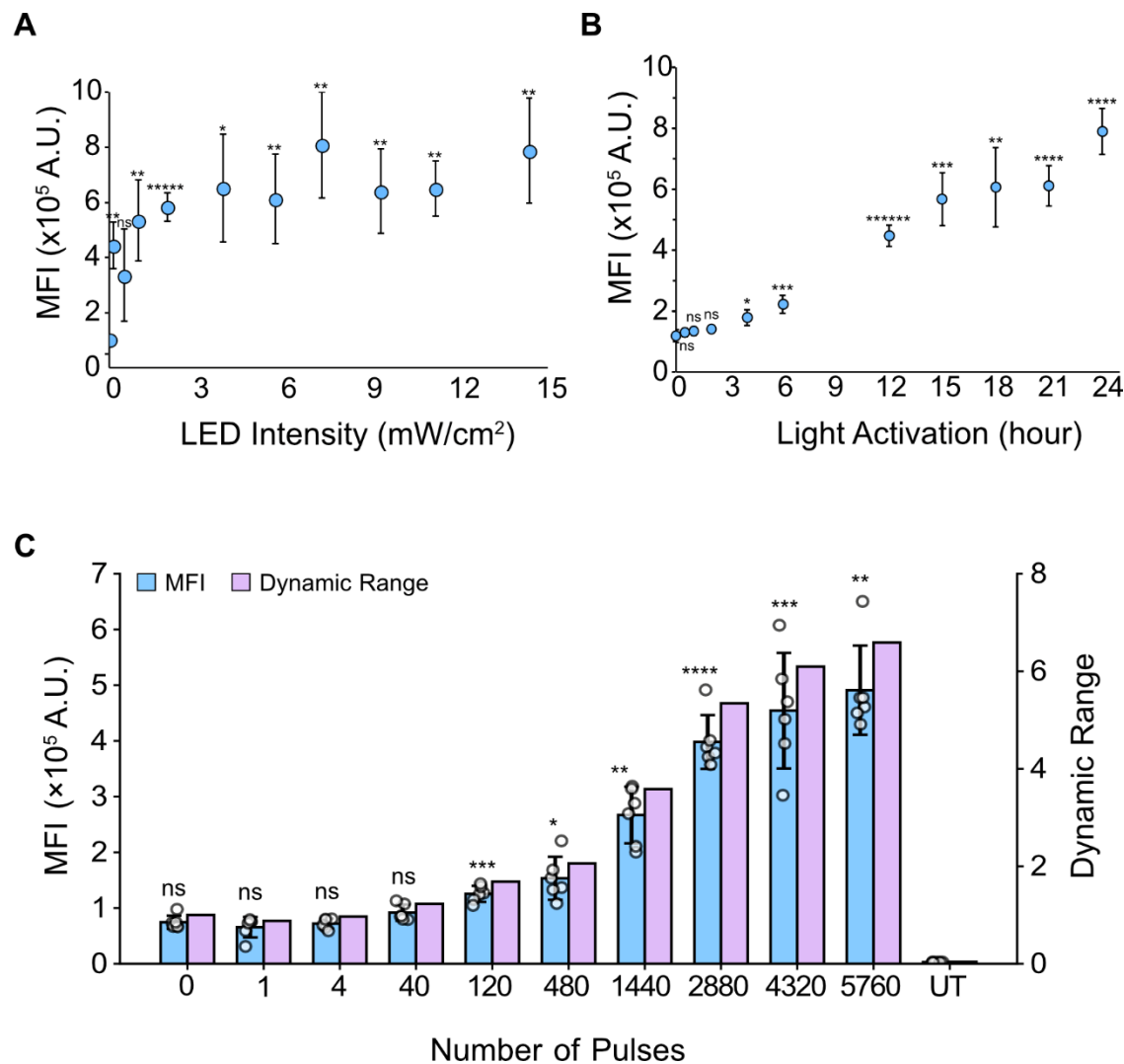


Figure 2-3. Characterization of 2pLACE activation dynamics. (A) eGFP MFI as a function of LED intensity. (B) eGFP MFI as a function of activation time. (C) eGFP MFI and dynamic range as a function of the number of LED pulses applied. UT (untransfected) is the negative control. Plots represent the mean of six technical replicates and error bars represent standard deviation. MFI significance was calculated with a one-way ANOVA and Tamhane's T2 post-hoc test to the dark state condition with 0 hours of light activation or 0 pulses. P-values for comparison between other conditions are shown in Table A-2, Table A-3, and Table A-3, for each panel respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ***** $p < 0.00001$, and ns is not significant.

2.2.3 Comparison of 2pLACE activation to 4pLACE

By reducing the number of plasmids needed, the 2pLACE system offers more ease of use compared to 4pLACE. To this point, we directly compared the light activation characteristics of both systems. We measured eGFP fluorescence by flow cytometry with and without 24 hours of light activation to understand the maximum expression of both systems (Figure 2-4). 2pLACE achieves a similar magnitude of expression as 4pLACE, and notably with less variability (Figure 2-4A). While there is no significant difference between the on states of the two systems, the consistency of 2pLACE is an improvement if the system is to be applied. We also compared activation kinetics of 2pLACE and 4pLACE over a 24 hour period. Both systems activated with similar kinetic behavior (Figure 2-4B). The matching activation profiles indicate the minimal system does not alter the dynamics of the LACE components. Thus, 2pLACE offers a similar activation profile to 4pLACE with improved consistency in activation.

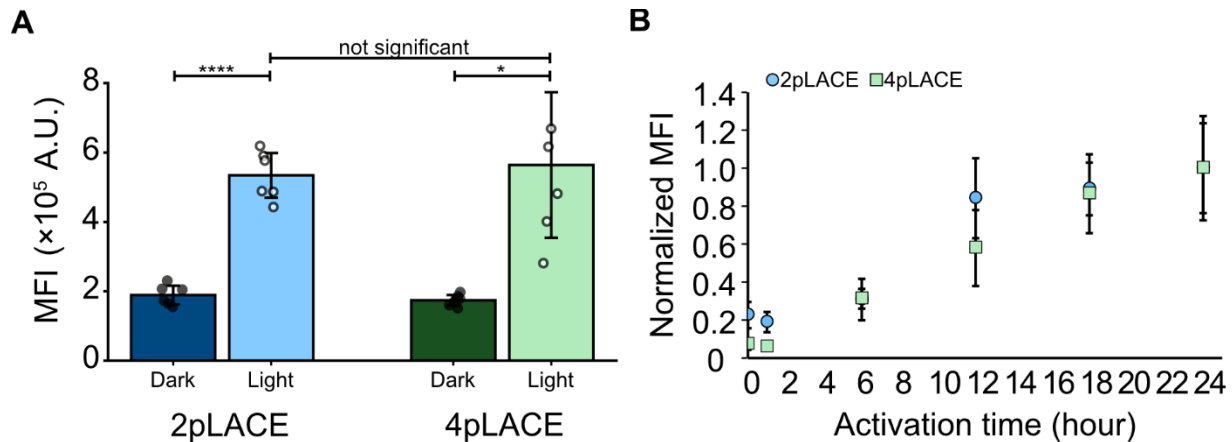


Figure 2-4. Comparison of 2pLACE and 4pLACE activation kinetics. (A) eGFP MFI was measured for cells transfected with 2pLACE and 4pLACE 24 hours post activation with and without blue light. (B) Activation behavior of 2pLACE and 4pLACE over 24 hours of activation. Data is normalized to the maximum activation of cells. Plots represent the mean of six technical replicates and error bars represent standard deviation. Significance was calculated with a one-way ANOVA and Tamhane's T2 post-hoc test. * $p < 0.05$, **** $p < 0.0001$

2.3 Discussion

Optogenetic systems such as LACE are novel systems for inducible gene expression and have the potential to overcome limitations such as off-target effects, reversibility, cost, and lack of spatial control associated with other inducible systems. These advantages make LACE a powerful and flexible tool for a variety of applications, provided all four LACE components can be introduced to the target cells. However, the original LACE system was encoded on four different plasmids, limiting its potential use cases.

Here, we optimized the LACE system by reducing the number of plasmids needed to deliver the key components from four to two plasmids (Figure 2-1). We varied plasmid stoichiometry and found that several plasmid ratios can provide similarly high dynamic range and a 3:7 plasmid ratio provides a balance for strong absolute activation and high dynamic range (Figure 2-2). This reinforces findings by others that plasmid ratio optimization can be an important aspect of system performance for multi-component optogenetic systems ^[36]. The LACE system also relies on the light-induced dimerization of CRY2 and CIBN, however, CRY2 oligomerizes upon blue-light excitation ^[37,38]. Maintaining balance between the LACE components may also minimize the chance of CRY2 oligomerization such that CRY2 and CIBN interact to activate transcription.

The system can also be tuned using different blue light intensities, where significant activation occurs at 0.12 mW/cm² and saturates after 2 mW/cm² (Figure 2-3A). A kinetics time-course showed significant activation by 4 hours but improving magnitude of activation with increasing activation time (Figure 2-3B). Interestingly, when a short period of activation was followed by a prolonged period for gene expression products to accumulate, significant expression was achieved after only 0.5 hours of light activation (Figure 2-3C). This suggests there is a delay

between complex formation at the promoter and translation. These results also reiterate that activation occurs on short timescales, but prolonged activation results in higher gene expression.

Comparing 2pLACE and 4pLACE directly in commonly used HEK293T cells showed that the systems have similar absolute activation and dynamic range after a maximum 24 hours of activation; however the 2pLACE system provided a more consistent signal (Figure 2-4A). We speculate that the improved consistency of 2pLACE comes from reduced variability in cells receiving sufficient amounts of all four LACE components. Reducing the number of plasmids from four to two also does not alter the activation kinetics (Figure 2-4B). We hypothesize that many of the metrics between 2pLACE and 4pLACE are similar because the components have not changed. However, the reduced variability and the ease of use give 2pLACE improved potential for various synthetic biology applications.

Optogenetic systems like 2pLACE are low-cost inducible systems ideal for use in low resource environments. The implementation of optogenetics in industrial applications can ease the high costs of serums and antibiotics that are common to improve yield in bioreactors and activate chemically induced systems ^[35]. For processes that require temporal control of gene expression, optogenetics also avoid costly media changes. 2pLACE's ease-of-use means it can be applied to hard-to-transfect and more biologically relevant cell lines, such as primary cells used in tissue engineering for regenerative medicine and cultivated meat ^[39]. Cultivated meat is a growing area of interest as a sustainable solution for improving food accessibility ^[40,41]. However, cost is critical to the industry's viability. Easier-to-use optogenetic tools like 2pLACE could help bend the cost curve for the cultivated meat industry. For example, muscle stem cells benefit from expression of proliferative proteins for scale up, but these same proteins prevent differentiation into muscle cells ^[40,42]. Temporal control of proliferative gene expression by 2pLACE could be used to strategically promote proliferation during scale up. Differentiation could proceed unimpeded with a flip of the

switch to turn off blue light activation of proliferative gene expression. Stable expression of 2pLACE and efficient optogenetic control of endogenous gene expression will be important to test for these future applications.

By optimizing the number of plasmids needed, 2pLACE shows more consistent behavior that improves its applicability as a tool in some scenarios. Yet, to be broadly applicable, additional improvements are still needed. Applying 2pLACE in industrial biotechnology settings will require stable expression of components to ensure activation over longer timescales. The work presented here characterizes 2pLACE under transient transfection conditions and are not necessarily compatible with long timescale processes. Our previous efforts to create a stable expression system for LACE using lentiviral vectors were hampered due to leaky expression in the dark state, and an inability to efficiently select cells that received all four components ^[31]. To this effect, we designed the 2pLACE system to have antibiotic resistance genes (Hygromycin and Bleocin, Figure 2-1) as distinct selection markers for each plasmid. Again, having fewer plasmids to encode all four components may have some advantages since fewer selection markers are required for stable cell line engineering. Stable expression of 2pLACE with prolonged activation by blue light could also lead to phototoxicity to the cells over long timescale scenarios ^[33]. Using a lower intensity or pulse frequency of blue light could overcome this toxicity.

Additional development and optimization of single component systems could also improve the translation of optogenetics into industrial applications. Recent single component blue light systems in mammalian cells include a pdDronpa-CRISPR system, EL222, and a VP16-fused TULIP system ^[43,44]. The pdDronpa-CRISPR system improves upon the dynamic range of the original LACE system but relies on higher energy violet light for dissociation. EL222 suffered from significantly low maximal reporter expression compared to red and alternative blue light systems in multiple mammalian cell lines. The VP16-fused TULIP system also resulted in less maximal

transient reporter expression in a single component system when compared to a dual component VP16-fused TULIP system. Single component red light inducible systems leverage red light's longer wavelength and minimal scattering to penetrate further in tissue samples and function in vivo. Recently, the red-light system REDLIP demonstrated transcriptional saturation at low intensities ($\sim 1.5 \text{ mW/cm}^2$), suggesting it may be similarly sensitive to LACE and other blue light systems. It also demonstrated minimal expression in UV, blue, green, and far-red light, supporting its potential in multiplexed gene targeting ^[28]. Expanding the optogenetic toolbox with easier to use and more reliable blue light systems, like 2pLACE, could improve the ability to investigate and engineer more complex biological pathways, especially when multiplexed with other optogenetic systems ^[26,45–48].

The simplicity and versatility of optogenetic systems like 2pLACE make them attractive for many applications, but it will still require specialized setups. For large-scale industrial applications, specific photobioreactors would need to be designed or current bioreactors would need to be outfitted with LEDs. We previously showed that blue light can propagate through a bioreactor to activate cells, but scattering is a major limitation that can dictate the upper limit on reactor size ^[35]. Specialized reactor designs with LEDs embedded throughout the reactor could also overcome the limitations associated with light propagation, though that may be costly. Ultimately, specialized smaller-scale applications like just-in-time or personalized medicine biomanufacturing may be more fruitful applications of 2pLACE and similar systems.

2.4 Conclusions

In conclusion, LACE is a reversible and tunable optogenetic tool that can be designed to target a specific gene of interest given an application. By reducing the number of plasmids needed to deliver all necessary components, 2pLACE becomes a flexible tool for synthetic biology and

bioproduction applications. The use of two plasmids ensures consistent signal while maintaining reversibility and tunability of the LACE system in HEK293T cells. Additionally, while the requirement of blue light can limit its applicability in certain applications, 2pLACE broadens the accessibility of inducible gene expression tools for low-resource environments and biologically relevant models.

2.5 Methods

2.5.1 Plasmids

The 2pLACE plasmids, pcDNA3.1-Cry-GFP-Hygro and pcDNA3.1-CibN-gRNA-Zeo, were commercially synthesized (Genscript). Full plasmid maps are included as supplementary files. 4pLACE plasmids, pcDNA3.1-CRY2FL-VP64, pcDNA3.1-CibN-dCas9-CibN, pGL3-Basic-8x-gRNA-eGFP, and gRNA-eGFP-Reporter, were gifted from Charles Gersbach (Addgene plasmids #60554, 60553, 60718, and 60719, respectively).

2.5.2 Cell Culture

HEK293T cells (ATCC-No. CRL-11268) were cultured in DMEM (Gibco) supplemented with 10% FBS (Gibco). Cells were cultured in a humidified incubator at 37°C and 5% CO₂. Cell density and viability were measured prior to cell plating using trypan blue and an automated cell counter (TC20, Bio-Rad). Cells were tested monthly for mycoplasma by MycoStrip (Invivogen) or PCR with the primers targeting 16S rRNA (Table 2-1).

Table 2-1 Primer sequences for mycoplasma PCR

Primer	Sequence
Primer 1	5'- TGCACCATCTGTCACTCTGTTAACCTC -3'

Primer 2	5'- GGAGCAAACAGGATTAGATACCCT-3'
Primer 3	5'- GGCGAATGGGTGAGTAACACG -3'
Primer 4	5'- CGGATAACGCTTGCGACCTATG -3'

2.5.3 Transient transfection

Twenty-four hours before transfection, HEK293T cells were seeded in a 96 well #1.5H glass bottom plate (Cellvis) at a density of 0.035×10^6 cells per well. Each well was transfected with a total of 100 ng of DNA. For experiments in a 12 well plate (VWR), HEK293T cells were seeded at a density of 0.016×10^6 cells per well, with each well transfected with a total of 1000 ng of DNA. HEK293T transfections used PolyJet DNA In Vitro Transfection Reagent (SignaGen) with a PolyJet:DNA volume to mass ratio of 3:1, according to manufacturer instructions.

The plasmid ratios for 2pLACE were relative proportions. Plasmid ratios for 4pLACE were calculated based on mass and molecular weight of plasmids relative to gRNA-eGFP-Reporter (Table 2-2) ^[20].

Table 2-2 Relative mass ratios of the transiently transfected LACE system plasmids

Plasmid Name	Relative Mass Proportions	Mass per 100 ng (ng)	System
pcDNA3.1-Cry-GFP-Hygro	3	30	2pLACE
pcDNA3.1-CibN-gRNA-Zeo	7	70	2pLACE
pcDNA3.1-CRY2FL-VP64	2.06	20.6	4pLACE
pcDNA3.1-CibN-dCas9-CibN	2.94	29.4	4pLACE
pGL3-Basic-8x-gRNA-eGFP	4	40	4pLACE
gRNA-eGFP-Reporter	1	10	4pLACE

Transfections used PolyJet DNA In Vitro Transfection Reagent (SignaGen) with a PolyJet:DNA volume to mass ratio of 3:1 according to manufacturer instructions.

2.5.4 Light activation

Cells were activated with LEDs at 465 nm twenty-four hours post transfection. For the optimization of plasmid ratios, a breadboard with 8 mm LEDs at a pulse frequency of 0.067 Hz was used to light activate the cells in a 12-well plate. Dark state samples were wrapped in aluminum foil to ensure isolation from light. For all other experiments, light activation was performed using an optoPlate-96 ^[32]. LED intensities and pulse frequencies were regulated with an Arduino Micro microcontroller. For experiments performed in 96-well plates, samples in the dark were wrapped in aluminum foil to ensure isolation from light. LED intensities were measured using the 1931-C Newport optical power meter.

2.5.5 Flow cytometry

Under red light, cells were trypsinized (0.05% Trypsin-EDTA, Gibco) and resuspended in cold DPBS (Gibco) with 1% FBS (FACS buffer). After centrifugation at 4°C for 10 minutes at 1000 RPM, 80% of the supernatant volume was removed and pellets were resuspended in 250 µL cold FACS buffer. HEK293T samples were wrapped in aluminum foil to ensure isolation from light following preparation for flow cytometry. A 488 nm laser in a CytoFLEX S Flow Cytometer was used for all analytical flow cytometry experiments. A target value of 10,000 total events was reached for every sample before gating for live, single, fluorescing cells. Cells were gated at an intensity of 5×10^3 on the FITC-GFP channel. This corresponded to ~0.5% of cells falling into this gate for an untransfected negative control.

2.5.6 Fluorescent imaging

Fluorescent images were taken using a ZOE Fluorescent Cell Imager (BioRad) after 24 hours of illumination, unless otherwise specified, using the green and brightfield channels.

2.5.7 Statistical analysis

Statistical analysis was done using a one-way ANOVA test with Tamhane's T2 post-hoc test for its conservative analysis because it assumes unequal variances where each condition was compared to all other conditions. Python packages `scipy.stats`, `f_oneway`, and `scikit_posthocs` were used for the ANOVA and post-hoc test. For statistical analysis using a one-tailed t-test, each light state was compared to its dark state. Python package `scipy.stats`' `ttest_ind` function was used for the one-tailed t-test. The one-tailed t-test was done assuming the mean of the dark state was less than the mean of the light state making it a right-tailed t-test. Data were considered statistically significant if $p < 0.05$.

For the biological replicates, standard deviation was calculated with error propagated:

$$\sigma_f = |f| \sqrt{\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 - \frac{2\sigma_{ab}}{AB}} \quad (1)$$

σ_f is the standard deviation of the dynamic range, f is the dynamic range value, A is the MFI value for a light activated sample, B is the MFI value for the dark state sample, σ_A and σ_B are their standard deviations, respectively. σ_{AB} is the covariance of the light and dark conditions. We assumed that the two conditions are not correlated, therefore the covariance is 0.

2.6 References

1. Randolph, L. N., Bao, X., Zhou, C., & Lian, X. (2017). An all-in-one, Tet-On 3G inducible PiggyBac system for human pluripotent stem cells and derivatives. *Scientific Reports*, 7(1), 1549. <https://doi.org/10.1038/s41598-017-01684-6>
2. Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R., & Tabor, J. J. (2014). Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nature Methods*, 11(4), 449–455. <https://doi.org/10.1038/nmeth.2884>
3. Toettcher, J. E., Gong, D., Lim, W. A., & Weiner, O. D. (2011). Light-based feedback for controlling intracellular signaling dynamics. *Nature Methods*, 8(10), 837–839. <https://doi.org/10.1038/nmeth.1700>
4. Michaels, Y. S., Barnkob, M. B., Barbosa, H., Baeumler, T. A., Thompson, M. K., Andre, V., Colin-York, H., Fritzsche, M., Gileadi, U., Sheppard, H. M., Knapp, D. J. H. F., Milne, T. A., Cerundolo, V., & Fulga, T. A. (2019). Precise tuning of gene expression levels in mammalian cells. *Nature Communications*, 10(1), 818. <https://doi.org/10.1038/s41467-019-08777-y>
5. De Carluccio, G., Fusco, V., & di Bernardo, D. (2024). Engineering a synthetic gene circuit for high-performance inducible expression in mammalian systems. *Nature Communications*, 15(1), 3311. <https://doi.org/10.1038/s41467-024-47592-y>
6. Majors, B. S., Betenbaugh, M. J., Pederson, N. E., & Chiang, G. G. (2008). Enhancement of transient gene expression and culture viability using Chinese hamster ovary cells overexpressing Bcl-x(L). *Biotechnology and Bioengineering*, 101(3), 567–578. <https://doi.org/10.1002/bit.21917>
7. Mellahi, K., Cambay, F., Brochu, D., Gilbert, M., Perrier, M., Ansorge, S., Durocher, Y., & Henry, O. (2019). Process development for an inducible rituximab-expressing Chinese

- hamster ovary cell line. *Biotechnology Progress*, 35(1), e2742.
<https://doi.org/10.1002/btpr.2742>
8. Kallunki, T., Barisic, M., Jäättelä, M., & Liu, B. (2019). How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells*, 8(8), 796.
<https://doi.org/10.3390/cells8080796>
 9. Pedone, E., Postiglione, L., Aulicino, F., Rocca, D. L., Montes-Olivas, S., Khazim, M., di Bernardo, D., Pia Cosma, M., & Marucci, L. (2019). A tunable dual-input system for on-demand dynamic gene expression regulation. *Nature Communications*, 10(1), 4481.
<https://doi.org/10.1038/s41467-019-12329-9>
 10. Thaisuchat, H., Baumann, M., Pontiller, J., Hesse, F., & Ernst, W. (2011). Identification of a novel temperature sensitive promoter in cho cells. *BMC Biotechnology*, 11, 51.
<https://doi.org/10.1186/1472-6750-11-51>
 11. Huang, Z. D., & Bugaj, L. J. (2024). Optogenetic Control of Condensates: Principles and Applications. *Journal of Molecular Biology*, 436(23), 168835.
<https://doi.org/10.1016/j.jmb.2024.168835>
 12. Klewer, D. L., & Wu, P. D. Y.-W. (2019). Light-Induced Dimerization Approaches to Control Cellular Processes. *Chemistry (Weinheim an Der Bergstrasse, Germany)*, 25(54), 12452.
<https://doi.org/10.1002/chem.201900562>
 13. Emiliani, V., Entcheva, E., Hedrich, R., Hegemann, P., Konrad, K. R., Lüscher, C., Mahn, M., Pan, Z.-H., Sims, R. R., Vierock, J., & Yizhar, O. (2022). Optogenetics for light control of biological systems. *Nature Reviews Methods Primers*, 2(1), 1–25.
<https://doi.org/10.1038/s43586-022-00136-4>
 14. Armbruster, A., Mohamed, A. M., Phan, H. T., & Weber, W. (2024). Lighting the way: Recent developments and applications in molecular optogenetics. *Current Opinion in Biotechnology*, 87, 103126. <https://doi.org/10.1016/j.copbio.2024.103126>

15. Shimizu-Sato, S., Huq, E., Tepperman, J. M., & Quail, P. H. (2002). A light-switchable gene promoter system. *Nature Biotechnology*, 20(10), 1041–1044. <https://doi.org/10.1038/nbt734>
16. Wang, X., Chen, X., & Yang, Y. (2012). Spatiotemporal control of gene expression by a light-switchable transgene system. *Nature Methods*, 9(3), 266–269. <https://doi.org/10.1038/nmeth.1892>
17. Legnini, I., Emmenegger, L., Zappulo, A., Rybak-Wolf, A., Wurmus, R., Martinez, A. O., Jara, C. C., Boltengagen, A., Hessler, T., Mastrobuoni, G., Kempa, S., Zinzen, R., Woehler, A., & Rajewsky, N. (2023). Spatiotemporal, optogenetic control of gene expression in organoids. *Nature Methods*, 20(10), 1544–1552. <https://doi.org/10.1038/s41592-023-01986-w>
18. Yazawa, M., Sadaghiani, A. M., Hsueh, B., & Dolmetsch, R. E. (2009). Induction of protein-protein interactions in live cells using light. *Nature Biotechnology*, 27(10), 941–945. <https://doi.org/10.1038/nbt.1569>
19. Crefcoeur, R. P., Yin, R., Ulm, R., & Halazonetis, T. D. (2013). Ultraviolet-B-mediated induction of protein–protein interactions in mammalian cells. *Nature Communications*, 4(1), 1779. <https://doi.org/10.1038/ncomms2800>
20. Polstein, L. R., & Gersbach, C. A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology*, 11(3), 198–200. <https://doi.org/10.1038/nchembio.1753>
21. Chatelle, C., Ochoa-Fernandez, R., Engesser, R., Schneider, N., Beyer, H. M., Jones, A. R., Timmer, J., Zurbriggen, M. D., & Weber, W. (2018). A Green-Light-Responsive System for the Control of Transgene Expression in Mammalian and Plant Cells. *ACS Synthetic Biology*, 7(5), 1349–1358. <https://doi.org/10.1021/acssynbio.7b00450>
22. Guntas, G., Hallett, R. A., Zimmerman, S. P., Williams, T., Yumerefendi, H., Bear, J. E., & Kuhlman, B. (2015). Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proceedings of the National Academy of Sciences*, 112(1), 112–117. <https://doi.org/10.1073/pnas.1417910112>

23. Müller, K., Engesser, R., Metzger, S., Schulz, S., Kämpf, M. M., Busacker, M., Steinberg, T., Tomakidi, P., Ehrbar, M., Nagy, F., Timmer, J., Zubriggen, M. D., & Weber, W. (2013). A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. *Nucleic Acids Research*, 41(7), e77. <https://doi.org/10.1093/nar/gkt002>
24. Kuwasaki, Y., Suzuki, K., Yu, G., Yamamoto, S., Otabe, T., Kakihara, Y., Nishiwaki, M., Miyake, K., Fushimi, K., Bekdash, R., Shimizu, Y., Narikawa, R., Nakajima, T., Yazawa, M., & Sato, M. (2022). A red light-responsive photoswitch for deep tissue optogenetics. *Nature Biotechnology*, 40(11), 1672–1679. <https://doi.org/10.1038/s41587-022-01351-w>
25. Bhoo, S.-H., Davis, S. J., Walker, J., Karniol, B., & Vierstra, R. D. (2001). Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. *Nature*, 414(6865), 776–779. <https://doi.org/10.1038/414776a>
26. Kaberniuk, A. A., Baloban, M., Monakhov, M. V., Shcherbakova, D. M., & Verkhusha, V. V. (2021). Single-component near-infrared optogenetic systems for gene transcription regulation. *Nature Communications*, 12(1), 3859. <https://doi.org/10.1038/s41467-021-24212-7>
27. Baloban, M., Kasatkina, L. A., & Verkhusha, V. V. (2024). iLight2: A near-infrared optogenetic tool for gene transcription with low background activation. *Protein Science*, 33(5), e4993. <https://doi.org/10.1002/pro.4993>
28. Qiao, L., Niu, L., Wang, M., Wang, Z., Kong, D., Yu, G., & Ye, H. (2024). A sensitive red/far-red photoswitch for controllable gene therapy in mouse models of metabolic diseases. *Nature Communications*, 15, 10310. <https://doi.org/10.1038/s41467-024-54781-2>
29. Park, H., Kim, N. Y., Lee, S., Kim, N., Kim, J., & Heo, W. D. (2017). Optogenetic protein clustering through fluorescent protein tagging and extension of CRY2. *Nature Communications*, 8(1), 30. <https://doi.org/10.1038/s41467-017-00060-2>

30. Kawano, F., Suzuki, H., Furuya, A., & Sato, M. (2015). Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nature Communications*, 6(1), 6256. <https://doi.org/10.1038/ncomms7256>
31. Minami, S., & Shah, P. (2021). Transient light-activated gene expression in Chinese hamster ovary cells. *BMC Biotechnology*, 21(13). [https://doi.org/https://doi.org/10.1186/s12896-021-00670-1](https://doi.org/10.1186/s12896-021-00670-1)
32. Bugaj, L. J., & Lim, W. A. (2019). High-throughput multicolor optogenetics in microwell plates. *Nature Protocols*, 14(7), 2205–2228. <https://doi.org/10.1038/s41596-019-0178-y>
33. Icha, J., Weber, M., Waters, J. C., & Norden, C. (2017). Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 39(8). <https://doi.org/10.1002/bies.201700003>
34. Kitajima, K., Kawahira, N., Lee, S.-W., Tamura, K., Morishita, Y., & Ohtsuka, D. (2021). Light-induced local gene expression in primary chick cell culture system. *Development, Growth & Differentiation*, 63(3), 189–198. <https://doi.org/10.1111/dgd.12721>
35. Minami, S. A., Garimella, S. S., & Shah, P. S. (2024). Computational evaluation of light propagation in cylindrical bioreactors for optogenetic mammalian cell cultures. *Biotechnology Journal*, 19(1), 2300071. <https://doi.org/10.1002/biot.202300071>
36. An-adirekkun, J. (My), Stewart, C. J., Geller, S. H., Patel, M. T., Melendez, J., Oakes, B. L., Noyes, M. B., & McClean, M. N. (2020). A yeast optogenetic toolkit (yOTK) for gene expression control in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, 117(3), 886–893. <https://doi.org/10.1002/bit.27234>
37. Bugaj, L. J., Choksi, A. T., Mesuda, C. K., Kane, R. S., & Schaffer, D. V. (2013). Optogenetic protein clustering and signaling activation in mammalian cells. *Nature Methods*, 10(3), 249–252. <https://doi.org/10.1038/nmeth.2360>
38. Más, P., Devlin, P. F., Panda, S., & Kay, S. A. (2000). Functional interaction of phytochrome B and cryptochrome 2. *Nature*, 408(6809), 207–211. <https://doi.org/10.1038/35041583>

39. Stout, A. J., Arnett, M. J., Chai, K., Guo, T., Liao, L., Mirliani, A. B., Rittenberg, M. L., Shub, M., White, E. C., Yuen, J. S. K. Jr., Zhang, X., & Kaplan, D. L. (2023). Immortalized Bovine Satellite Cells for Cultured Meat Applications. *ACS Synthetic Biology*, 12(5), 1567–1573. <https://doi.org/10.1021/acssynbio.3c00216>
40. Jara, T. C., Park, K., Vahmani, P., Van Eenennaam, A. L., Smith, L. R., & Denicol, A. C. (2023). Stem cell-based strategies and challenges for production of cultivated meat. *Nature Food*, 4(10), 841–853. <https://doi.org/10.1038/s43016-023-00857-z>
41. Negulescu, P. G., Risner, D., Spang, E. S., Sumner, D., Block, D., Nandi, S., & McDonald, K. A. (2023). Techno-economic modeling and assessment of cultivated meat: Impact of production bioreactor scale. *Biotechnology and Bioengineering*, 120(4), 1055–1067. <https://doi.org/10.1002/bit.28324>
42. Pan, G., & Thomson, J. A. (2007). Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Research*, 17(1), 42–49. <https://doi.org/10.1038/sj.cr.7310125>
43. Zhou, X. X., Zou, X., Chung, H. K., Gao, Y., Liu, Y., Qi, L. S., & Lin, M. Z. (2018). A Single-Chain Photoswitchable CRISPR-Cas9 Architecture for Light-Inducible Gene Editing and Transcription. *ACS Chemical Biology*, 13(2), 443–448. <https://doi.org/10.1021/acscchembio.7b00603>
44. Beyer, H. M., Kumar, S., Nieke, M., Diehl, C. M. C., Tang, K., Shumka, S., Koh, C. S., Fleck, C., Davies, J. A., Khammash, M., & Zurbriggen, M. D. (2024). Genetically-stable engineered optogenetic gene switches modulate spatial cell morphogenesis in two- and three-dimensional tissue cultures. *Nature Communications*, 15(1), 10470. <https://doi.org/10.1038/s41467-024-54350-7>
45. Dwijayanti, A., Zhang, C., Poh, C. L., & Lautier, T. (2022). Toward Multiplexed Optogenetic Circuits. *Frontiers in Bioengineering and Biotechnology*, 9, 804563. <https://doi.org/10.3389/fbioe.2021.804563>

46. Redchuk, T. A., Omelina, E. S., Chernov, K. G., & Verkhusha, V. V. (2017). Near-infrared optogenetic pair for protein regulation and spectral multiplexing. *Nature Chemical Biology*, 13(6), 633–639. <https://doi.org/10.1038/nchembio.2343>
47. Mao, D., Sun, F., Driscoll, B., Li, Z., & Xu, G. (2023). Close-packed dual-color micro-LEDs enable cortical-layer-specific bidirectional *in vivo* optogenetic electrophysiology. *Cell Reports Physical Science*, 4(12), 101702. <https://doi.org/10.1016/j.xcrp.2023.101702>
48. Mermet-Joret, N., Moreno, A., Zbela, A., Nazari, M., Ellendersen, B. E., Comaposada-Baro, R., Krauth, N., Philipsborn, A. von, Sørensen, A. T., Piriz, J., Lin, J. Y., & Nabavi, S. (2025). Dual-color optical activation and suppression of neurons with high temporal precision. *eLife*, 12. <https://doi.org/10.7554/eLife.90327.2>

CHAPTER 3: Applications of inducible gene expression systems in bioproduction

3.1 Introduction

Chapter 2 discussed the optimization of 2pLACE, a simplified optogenetic tool at the lab-scale that could be used in low-resource environments to reduce the costly serums and reagents required for bioprocessing. Expanding on how such a tool can be used in bioproduction, this chapter focuses on how to apply these inducible systems tools and the respective design considerations for industrial processes.

For optogenetic gene expression tools, like 2pLACE, to be applied in industrial conditions, we need to consider its prerequisites—how will the light source be introduced to the culture? Will the light penetrate the culture to induce sufficient expression? Photobioreactors are more commonly used outdoors to harness sunlight for the growth of fermenters or photosynthetic microbes ^[1–4]. As optogenetic systems gain popularity, there is a greater need to scale up bioproduction and design photobioreactors with artificial light ^[2,5,6].

Companies like Fermeate, PCI Biotech, and Prolific Machines are working to develop new optogenetic technology to increase yield and measure growth dynamics for precise modulation of bioproduction^[5,7,7]. Fermeate is developing integrative technology to repurpose existing bioreactors for optogenetic control of microbial fermentation. PCI Biotech is developing light-inducible drug delivery mechanisms at industry-scale. Within the last year, Prolific Machines has been able to successfully demonstrate a 200L optogenetic production run, and set a new

benchmark of antibody production at 21 g/L in Chinese hamster ovary (CHO) cells^[8,9]. As optogenetics becomes a more popular industrial technology, there is a greater need for understanding how to better design photobioreactors to balance costs and energy consumption while maintaining optimal growth kinetics and mixing ^[3,10–12]. The first section of this chapter discusses a project where we experimentally demonstrated how light propagates through a mammalian cell culture in a representative benchtop model to support computational simulations of light propagation in a photobioreactor.

In situations where photobioreactors are not feasible to engineer, we considered a glucose-inducible expression tool for mammalian cell culture. A limiting factor for cell growth in bioreactors is the need for costly antibiotics and serums to maintain cell viability and protein production ^[13–15]. The second section of this chapter focuses on a project where we designed a glucose-inducible gene expression system. Glucose is an inexpensive and accessible component already used in cell culture media so that toxic off-target effects are avoided.

This chapter concludes with future directions of these inducible gene expression projects. One proposed idea is printing inducible mammalian cells in an agarose-based matrix to build muscle scaffolds for cultivated meat and tissue engineering applications. By taking advantage of the spatiotemporal and reversible control of light, we can control the growth and expression of muscle cells into a specific structure at scale.

3.2 Light propagation through a photobioreactor

In Minami *et al.*, we addressed the issue of whether light-inducible mammalian cell culture can be scaled up to be effective at production-scale in industry ^[16]. Using computational modeling, we evaluated how light propagates through cylindrical bioreactors, accounting for absorption,

scattering, geometry, radiation, and cell density. We found that reduction of light intensity as it propagates through the bioreactor is primarily caused by light absorption rather than scattering. We simulated varying reactor sizes and showed that the average light intensity decreases as the reactor size and cell density increases. The computational predictions showed that optogenetic activation can be maintained in bioreactors up to about 80,000 L at high cell densities as long as sufficient light activation is provided at the boundaries of the reactor and thorough mixing within the reactor is ensured. These results suggest that optogenetic activation is feasible in bioreactors because even intermittent light exposure is sufficient to penetrate the cell culture and continuous illumination is not essential.

3.2.1 Experimental light propagation bench-top setup

Minami *et al.* previously simulated how bioreactor size could limit the application of optogenetics by varying height-to-diameter ratios and modeling how light propagates through the culture considering changes in absorption and scattering. The simulations predicted sufficient light penetration to activate light-inducible cells in bioreactors as big as 80,000 L^[16]. To validate the computational models, we performed a set of small-scale experiments to determine whether our predictions were realistic. Light propagation through a bioreactor was tested experimentally using a 2.28 L cylindrical acrylic container (GLOTRENDS USA). The top of the cylinder contained a slit along the diameter to measure the diffusion of light and an additional opening to add and remove suspension cultures (Figure 3-1A). The entire surface of the cylinder was wrapped in remote-controlled LEDs evenly dispersed by 2.5 cm (RGB Color Changing Flexible LED Sheet 52W, Solid Apollo LED) (Figure 3-1B). The light intensity was measured using LI-250A Light Meter (LI-COR Biosciences) and the US-SQS/L Spherical Micro Quantum Sensor (Heinz Walz GmbH, S/N SQSA1208) (Figure 3-1C). The sensor was calibrated with Dulbecco's phosphate buffered saline (DPBS; Thermo Fisher Scientific, Cat. No. 21600069). For calibration, the intensity of the blue light was adjusted such that the incident radiation into the cylinder was 100 W/m², converted from

460 $\mu\text{mol}\cdot\text{m}^2/\text{s}$ on the light sensor and using a calibration coefficient of -465.8 as specified by the quality control documentation. Measurements were taken along the radius in triplicate, taking measurements over 15 second averages every 2 cm from the center.

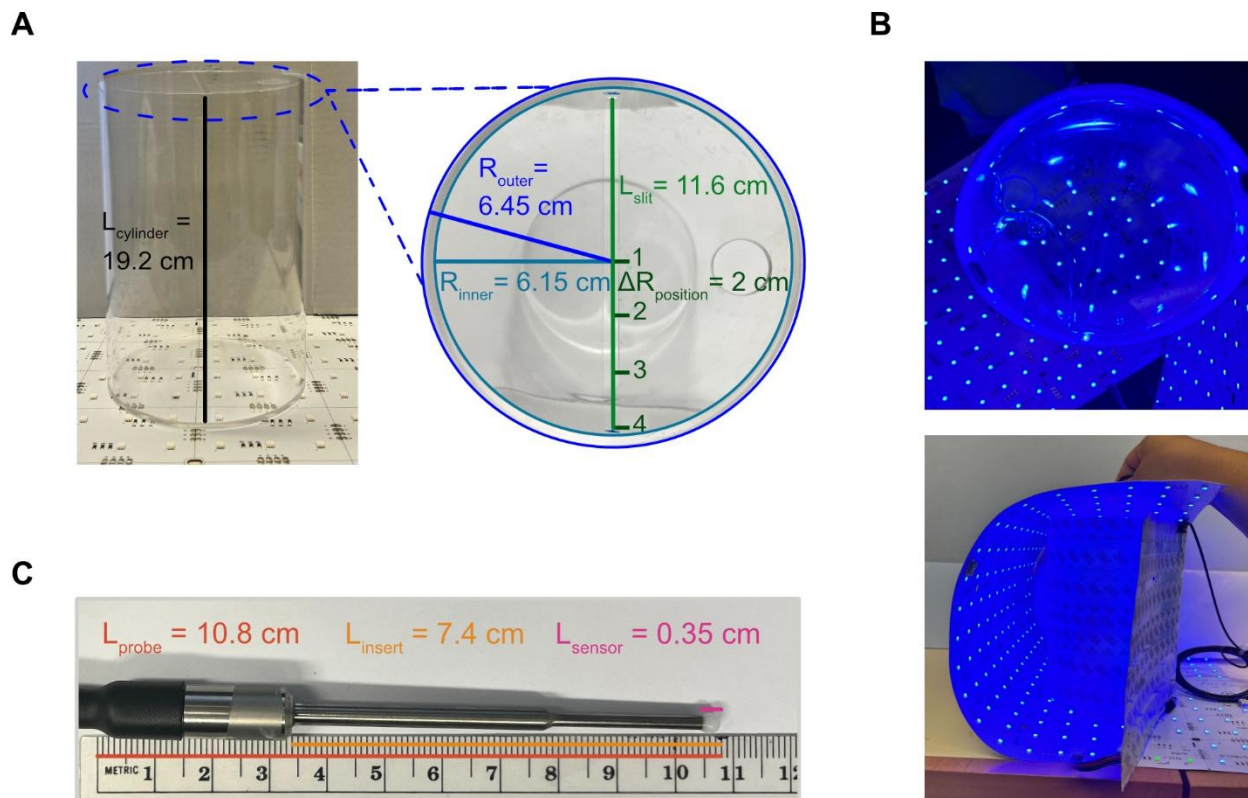


Figure 3-1. Experimental setup for light propagation measurements. (A) Cylindrical acrylic vessel used as a model bioreactor. (B) LED sheets used to illuminate the cylinder. (C) Quantum light probe used for light intensity measurements.

ExpiCHO-S cells (Thermo Fisher Scientific, Cat. No. A29127) were used to fill the reactor to measure light diffusion. They were grown in 1 L vented culture flasks (Thermo Fisher Scientific, Cat. No. 4115-1000) with 280 mL of culture in ExpiCHO Expression Medium (Thermo Fisher Scientific, Cat. No. A2910001). A volume of 1.12 L of cells was cultured to a density of

approximately 106 cells/mL at 37°C in 8% CO₂. DPBS was used for dilutions. The solution was manually mixed between measurements at different positions to capture a well-mixed volume.

3.2.2 Demonstration of light propagation in the experimental bioreactor

We performed experiments on a small scale using a suspension of CHO cells to assess the dependence of light propagation on cell density and radial position in a model bioreactor (Figure 3-1A). Similar to the trends observed in the computational simulations, measured light intensity decreased as cell density increased (Figure 3-2A) ^[16]. We observed little variation in intensity in the radial direction (~10% decrease) through the cell medium, consistent with high scattering anisotropy for a system of this size and with the lower cell densities used in the experiments. Given that the highest cell density for our experimental conditions was 15% of the maximum cell density simulated in the computations, our experimental measurements were consistent with the simulations that predicted little radial variation. We note, however, that even at the outer radius ($r = 6$ cm), before any reduction due to propagation through the cell medium has occurred, the measured light intensities in the cell suspensions were significantly lower than the incident radiation outside the beaker (Figure 3-2A).

We would expect the internal boundary of the beaker to show an intensity that matches the incident radiation. Since this was not observed and to better understand the discrepancy, we performed a simulation replicating the experimental conditions at various values of g , the scattering anisotropy factor (Figure 3-2B). This factor quantifies the amount of forward direction retained after a scattering event; $g = 1$ implies complete forward scattering, whereas $g = 0$ implies isotropic scattering. These simulations also showed little radial variation in intensity at high g , with a maximum decrease of 20% at lower values of g . However, the radiation intensities from the computations were higher than those observed experimentally, suggesting the reason for the

discrepancy being dissipation of light intensity at the reactor wall (likely due to reflection) in the experiments not having been accounted for in the model.

Future work could include a more sophisticated experimental setup that can delineate light attenuation from reflection, absorption, and scattering, such as by using an integrating sphere or measuring sufficiently thin samples such that measurements are taken within the single-scattering limit [17–20]. Additional experimental validation includes measuring light propagation over a timecourse to also check cell viability over time since dead cells and cell secretions can scatter light differently than live cells [10,20].

The relatively uniform intensity distribution within the reactor shown here suggests that optical properties of the cell suspension are like those used in the simulations. For the parameter scan performed for g , the experimentally measured distribution most closely matches higher values of g (Figure 3-2B). This is also consistent with values reported in the literature for mammalian cells.

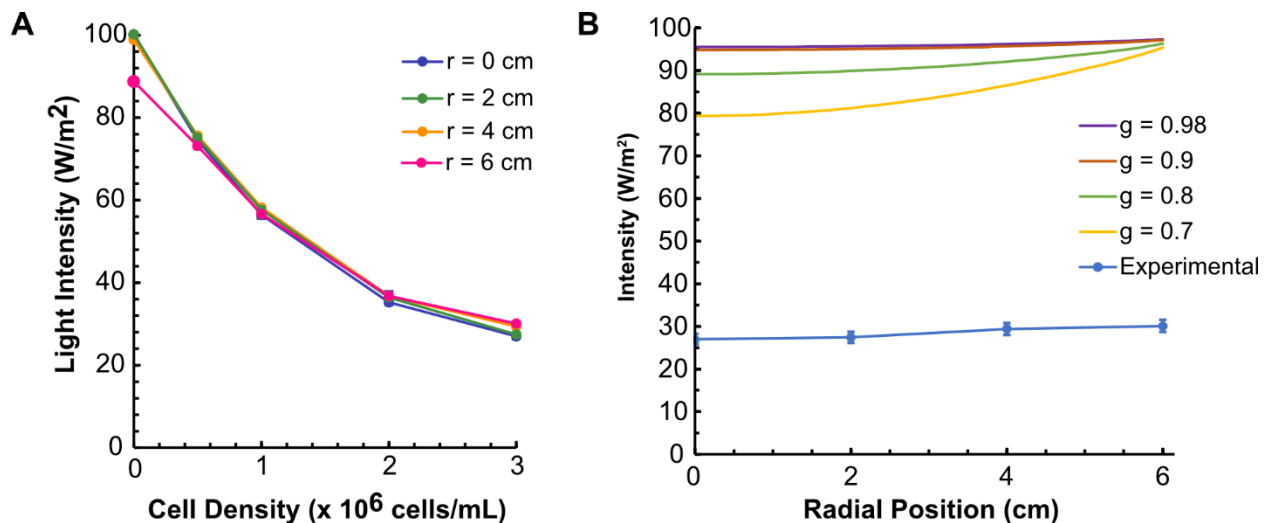


Figure 3-2. Comparisons of computational predictions to experimental measurements of light intensity in a cylindrical system. (A) Measurements of light intensity as a function of cell

density at $r = 0, 2, 4,$ and 6 cm. Error bars representing the standard deviation are too small to be visualized. (B) Light intensity as a function of radial position for the highest cell concentration tested experimentally (3×10^6 cells/mL) and for simulations with identical physical dimensions and different values of g . Error bars on the experimental data represent one standard deviation from the mean for three technical replicates.

In summary, our work suggests that the simulations realistically capture the behavior of light propagation within the reactor. Manufacturing a photobioreactor is not a trivial venture, but this work indicates that implementing optogenetics is a possibility for industrial bioproduction.

3.3 Glucose-inducible gene expression system for mammalian cell growth

In applications where photobioreactors with optogenetic gene expression may not be an appropriate solution, we now consider a glucose-inducible expression approach for mammalian cell culture. A limiting factor for cell growth in bioreactors is the need for costly antibiotics and serums to maintain cell viability and protein production. This section of the chapter focuses on a project where we designed an inducible gene expression system to be activated with glucose. As stated earlier, glucose is an inexpensive and accessible component already used in cell culture media so that toxic off-target effects are avoided by its use.

3.3.1 Inducible gene expression tools to improve cultivated meat production

As the global population surpasses 8 billion people and effects of climate change become more noticeable each day, it is imperative to find more sustainable food production systems ^[21]. The meat industry is a major contributor to greenhouse gas emissions and consumer of natural

resources. Current practices are unlikely to provide a long-term solution to minimize costs and consumption while meeting the food demand of a 60% population increase by 2050.

Cultivated meat has emerged as a promising solution by reducing environmental impact while improving food accessibility, potentially replacing 40% of consumer meat products [21–23]. Cultivated meat enables overall product optimization—improving nutritional benefits while minimizing additional chemical usage and energy consumption [22]. However, scalability remains a challenge especially when balancing long-term cell growth with material costs [22–25].

A key limitation when scaling up production of cultivated meat, is the lack of dynamic control over cellular processes such as proliferation and differentiation for muscle growth [24,26,27]. While advances in immortalized muscle cell lines are being explored, there is still a need for tools that can regulate gene expression in a tunable and reversible manner in response to environmental cues [28–30]. Such control systems could ultimately be applied to optimize cell growth and production in cultivated meat systems and other biomanufacturing platforms to maintain viability of cultures while keeping costs and materials low. To address this need, we designed a synthetic glucose-responsive promoter system that enables tunable and reversible control of gene expression in mammalian cells.

Chemical-inducible gene expression systems offer a reversible method of activating and tuning gene expression to a specific inducer. However, a common limitation of chemical-based system is the material costs and off-target effects, especially when trying to scale the system up for industrial applications [31,32]. For cultivated meat and other tissue engineering applications, finding an inducer that is already commonly present in the culture can simplify requirements. Glucose is a low-cost, accessible, non-antibiotic chemical inducer that can be used for controllable gene expression [28]. Glucose is a primary source of energy for mammalian cells so adverse or off-target effects are well-known or negligible. Cells are already equipped with the necessary enzymes and

processes to break down and use glucose for energy in glycolysis. By using glucose as the inducer, we can regulate the process by varying the concentration of glucose in the cell media where higher concentrations of glucose would lead to greater activation in gene expression.

3.3.2 Glucose-inducible promoter design and rationale

Some glucose-inducible promoters have already been well-studied in diabetes and metabolism research, such as liver-type pyruvate kinase (PKL) ^[33–37]. We also looked at muscle-type pyruvate kinase (PKM) since we are interested in tissue engineering and cultivated meat applications ^[38]. Pyruvate kinase activation is induced by glucose through two promoter elements—an upstream enhancer and a downstream core along with other glucose-responsive elements and consensus sequences to enable transcription. Both pyruvate kinase genes have been characterized and identified through the USCS Genome Browser (Table 3-1).

To build the promoter region, we started with an upstream enhancer and a downstream core promoter from PKL and PKM—pklCore and pklBasal for the liver-type pyruvate kinase and pkmELS and pkmPLS for the muscle-type pyruvate kinase. To improve responsiveness to glucose, consensus sites and linker sequences from carbohydrate or glucose responsive elements were used in between the promoter sequences. These responsive elements are commonly paired with PKL for diabetes research, but their usage with PKM could improve muscle growth with the glucose inducer ^[39–41]. Therefore, we paired different enhancer box DNA response element (eBox) sites from different glucose responsive genes (mouseLPK, mouse S14, ratACCPI, and ratS14) to increase transcription sites for the transcription factor to bind ^[41–44]. The transcription factor, max-like protein (MLX), is a glucose-responsive protein that promotes myogenesis and is used in mammalian cells ^[44,45]. We only used the eBox sites to the PKM enhancer region because it did not contain as many MLX binding sites and already coded for a non-canonical eBox for transcription factor binding ^[44,46]. With these eight glucose-inducible

promoter elements, we were able to build twelve plasmids that all expressed the same fluorescent reporter, eGFP, to measure promoter activation (Figure 3-3).

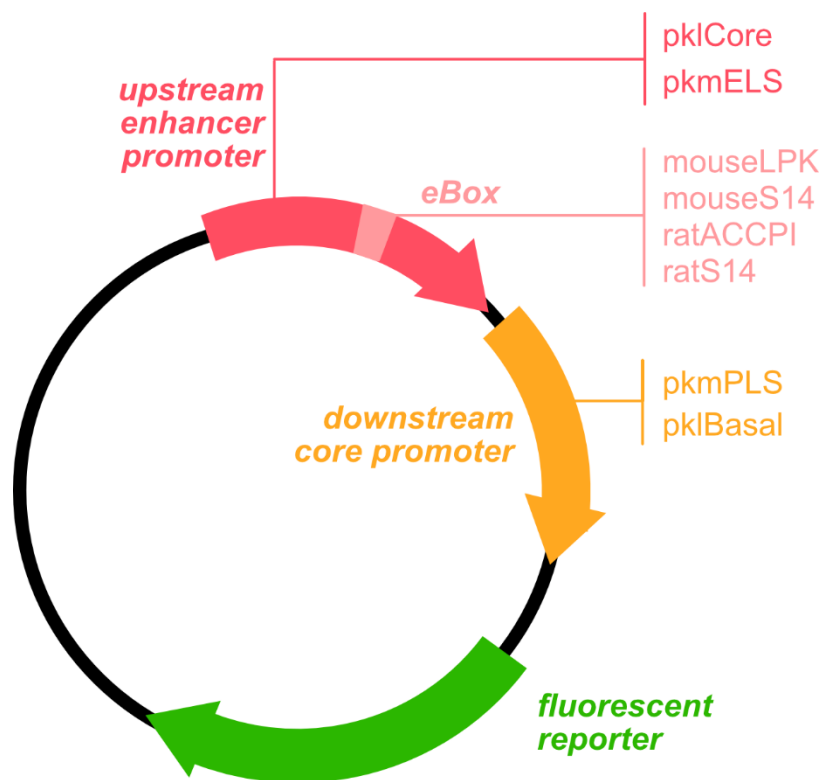


Figure 3-3. Glucose-inducible promoter design. Plasmid design included three promoter elements designed to test different combinations of an upstream enhancer, with or without an eBox, and a downstream core promoter element. Promoter activation was measured with the activation of a fluorescent reporter.

3.3.3 Activation dynamics of the glucose system and future design considerations

We tested the activation of the twelve constructs in high-glucose and zero-glucose cell culture media using flow cytometry to measure eGFP expression as a proxy for promoter activation. The high-glucose cell culture medium has a concentration of 4.5 mg/mL of cell culture-grade glucose. We cultured HEK293T cells in a regular medium before switching the medium to the respective concentrations for transfection and an additional 24 hours for sufficient activation. pklCore-

pkIBasal served as a positive control for glucose inducibility since this promoter has already been well studied.

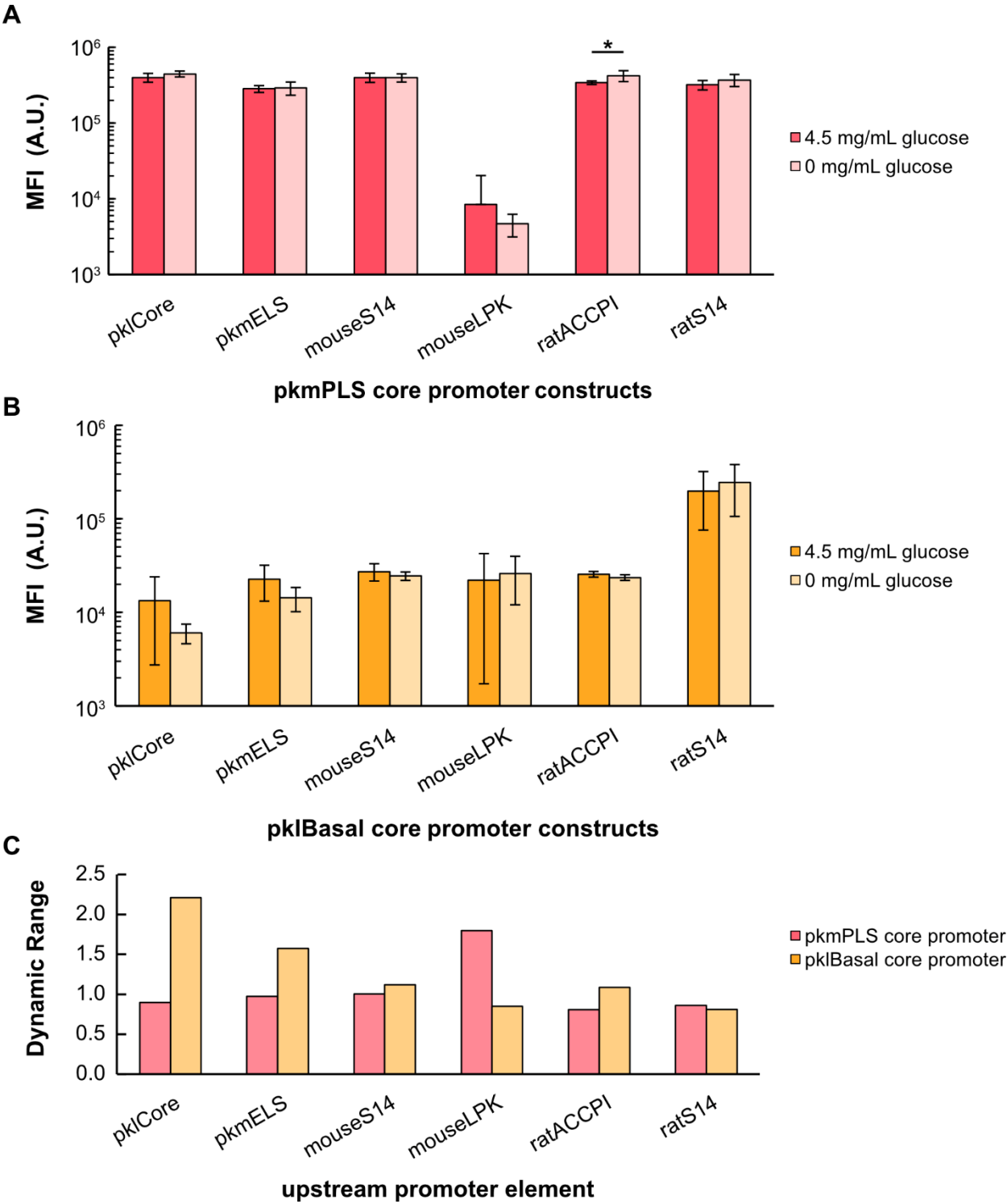


Figure 3-4 Glucose-inducible promoter activation. A-B) Mean fluorescence intensity (MFI) of eGFP was measured in HEK293T cells transfected with different combinations of the promoter elements in high glucose (4.5 mg/mL) media or media with 0 mg/mL of glucose. A) shows promoter combinations with pkmPLS as the downstream core promoter and B) shows combinations with pklBasal as the downstream core promoter. C) Dynamic range of each construct from (A) and (B) calculated as MFI in high glucose vs no glucose media. Plots represent the mean of six technical replicates and error bars represent standard deviation. MFI significance was calculated with a one-way ANOVA with Tukey's multiple comparisons test. * $p < 0.05$.

Overall, we found no construct sufficiently activated between the high-glucose and no glucose media, the respective on and off states (Figure 3-4). We measured eGFP reporter fluorescence with flow cytometry to quantify promoter activation. For the constructs with the pkmPLS core promoter, all the constructs except for mouseLPK eBox activated to a similar magnitude in high glucose media. The mouseLPK eBox construct achieved a magnitude lower expression. The constructs showed similar behavior when in the media with no glucose (Figure 3-4A). The construct with the ratACCPI eBox was the only construct that showed significant difference between its active and inactive states. Similar trends were also observed for the constructs with the pklBasal core promoter (Figure 3-4B). The ratS14 eBox construct expressed one magnitude greater than the other constructs. None of the pklBasal constructs had a significant difference in activation between their active and inactive states. The dynamic range (ratio of high to zero glucose) across all the constructs was about 1-2 (Figure 3-3C). The ratACCPI eBox with the pkmPLS core promoter was the only construct that had significant difference in expression and had a dynamic range of 0.8. This means its inactive state expressed more eGFP than its active state so the promoter did not activate well. Take together, the lack of difference in eGFP expression indicates that these promoter combinations are not responding to the differences in glucose levels in the cell culture media.

Glucose-inducible genes have been widely tested and utilized, but a glucose-inducible promoter system to eventually control muscle cell proliferation and differentiation has not been well studied. We tested these promoter constructs in HEK293T cells but designed the regions using mouse genes for muscle development so testing these constructs in mouse myoblasts, like C2C12s, may give more accurate results about promoter activation. If the system we tested had responded to varying inducer levels, our next step would have been to test NANOG, a gene known to control cell proliferation and differentiation signals [29,47,48]. There may be other combinations or promoter elements that are better suited to this type of chemical-inducible system. Similarly, using glucose as the inducer may introduce too many complications given its critical role in glycolysis and metabolism which may indicate that a glucose-inducible gene expression system is not practical for cultivated meat. Regulating muscle cell growth with an inducible system may thus be better suited to an optogenetic system, as discussed in Chapter 2.

3.4 Considerations in applying inducible gene expression systems to industrial bioproduction

Large, industrial scale processes are costly operations that require antibiotics, media replacements, and chemical additives to maintain the viability, sterility, and overall protein yields. Inducible gene expression systems can replace some of the media requirements while maintaining the production scale. The simplification of materials also means that inducible gene expression systems, especially optogenetic systems, can be helpful in low-resource environments like space missions where payloads are limited.

In Chapter 2, we established that LACE regulates transcriptional activation for a gene of interest with reversible and tunable control. Using optogenetics in bioreactors can be challenging since

light has to penetrate through the culture for effective gene activation. Our work simulating reactor geometry and light scattering experimentally and through computations shows that higher light intensities with pulsing can overcome the limitations. The experimental model revealed that reflection of light at the wall could reduce the intensity of light entering the cell medium. However, once in the medium, light intensity is relatively consistent throughout the reactor, assuming a well-mixed culture. Future directions should include higher cell densities and mixed cell viability to measure different light absorption and scattering patterns. Overall, optogenetic tools like 2pLACE can be a strategy to improve production conditions in more complex applications like cultivated meat, tissue engineering, and recombinant protein production.

The glucose-inducible promoter project approached the same goal from a different angle by asking whether a common component of cell culture media can be a sufficient inducer of gene expression. Literature on pyruvate kinase promoters also highlights the difficulty of separating glucose-responsive regulation from endogenous expression [33,38,49,50]. Our preliminary work indicates that our plasmid design scheme is not highly responsive to the presence or lack of glucose in the cell culture media. Future work to optimize the system includes trying greater glucose concentrations and combining other responsive elements for similar carbohydrates like sucrose. However, this promoter system may not be specific enough to control gene expression.

The ideas of large-scale optogenetics production to stimulate cell proliferation and differentiation scaffolded for tissue engineering applications could be combined to print inducible mammalian cells into a matrix scaffold [51,52]. One of the main advantages of optogenetics is that gene expression can be spatially and temporally controlled, which makes 2pLACE the optimal system to apply in a 3D environment [53–56]. To date, there is no single example of light-inducible, myogenic cells in a bioprinted hydrogel for industrial applications, but literature already shows that

optogenetics can control differentiation and signaling in mammalian cells in a matrix system [51,52,54,57].

Our work here suggests that inducible gene expression in mammalian cells can evolve from an experimental tool to a control strategy for bioproduction. Taken together, these studies reinforce the idea that light-inducible systems remain attractive because they combine reversibility, external control, and spatial precision in ways that chemical and metabolic inputs generally do not.

3.4 Methods

Glucose inducible system

Plasmid Construction

Sequences for the upstream and downstream promoter elements were sourced from the USCS Genome Browser in the *Mus musculus* genome: L-PK core, L-PK Basal, M-PK Enhancer, and M-PK Core (Table 3-1). These elements were determined based on their ENCODE scores as either a “promoter-like signal” or “enhancer-like signal”. Sequences were not codon optimized to maintain any intrinsic binding site or transcription start sites already present in the sequences. Promoter fragments were synthesized with overhangs for a pcDNA backbone (TWIST Biosciences). The fragments were cloned into a pcDNA-eGFP-2xStrep backbone using Gibson Assembly after being digested with MluI-HF and BsmBI (New England Biolabs).

Table 3-1. Locations of promoter elements as denoted by UCSC Genome Browser.

Promoter Element Name	Sequence Location
L-PK Core	chr3:89,135,966-89,136,264
L-PK Basal	chr3:89,136,285-89,136,586

M-PK Enhancer	chr9:59,656,041-59,656,216
M-PK Core	chr9:59,656,251-59,656,600

Cell Culture

HEK293T cells were cultured in DMEM (Gibco) with 10% FBS (Gibco) in a humidified incubator at 37°C and 5% CO₂. Cells from passage 3 to 15 were used for experiments. Cells were tested for mycoplasma every month by MycoStrip (Invivogen).

Transfection

HEK293T cells were seeded in a 96 well plate twenty-four hours before transfection at a density of 0.035×10^6 cells per well. Transfections used PolyJet DNA In Vitro Transfection Reagent (SignaGen) with a PolyJet/DNA ratio of 3:1.

Flow Cytometry

Cells were trypsinized and resuspended in cold DPBS with 1% FBS (FACS buffer). After centrifugation at 4°C for 10 minutes at 1000 RPM, the supernatant was partially removed, and pellets were resuspended in cold FACS buffer. A 488 nm laser in a CytoFLEX S Flow Cytometer was used for all analytical flow cytometry experiments. A target value of 10000 total events was reached for every sample before gating for live, single, fluorescing cells. Cells were gated at an intensity of 5×10^3 on the FITC-GFP channel. This corresponded to ~0.5% of cells falling into this gate for an untransfected negative control.

Statistical analysis

Statistical analysis was done using a one-way ANOVA test with Tukey's multiple comparisons test. Python packages `scipy.stats`, `f_oneway`, and `scikit_posthocs` were used. Data were considered statistically significant if $p < 0.05$.

3.5 References

1. *Bioreactor concepts, types, and modeling*. (n.d.). <https://doi.org/10.1016/B978-0-323-85690-4.00004-X>
2. Chanquia, S. N., Vernet, G., & Kara, S. (2021). Photobioreactors for cultivation and synthesis: Specifications, challenges, and perspectives. *Engineering in Life Sciences*, 22(12), 712–724. <https://doi.org/10.1002/elsc.202100070>
3. Posten, C. (2009). Design principles of photo-bioreactors for cultivation of microalgae. *Engineering in Life Sciences*, 9(3), 165–177. <https://doi.org/10.1002/elsc.200900003>
4. Xu, L., Weathers, P. J., Xiong, X.-R., & Liu, C.-Z. (2009). Microalgal bioreactors: Challenges and opportunities. *Engineering in Life Sciences*, 9(3), 178–189. <https://doi.org/10.1002/elsc.200800111>
5. *Software & Illumination | Prolific Machines*. (n.d.). Retrieved May 21, 2026, from <https://www.prolific-machines.com/software-and-illumination>
6. Wang, H., & Yang, Y.-T. (2017). Mini Photobioreactors for in Vivo Real-Time Characterization and Evolutionary Tuning of Bacterial Optogenetic Circuit. *ACS Synthetic Biology*, 6(9), 1793–1796. <https://doi.org/10.1021/acssynbio.7b00091>
7. *TECHNOLOGY*. (n.d.). PCI Biotech. Retrieved May 23, 2026, from <https://www.pcibiotech.no/technology>
8. *Prolific Machines Sets New mAb Productivity Benchmark at 21 g/L*. (n.d.). The Pharma Navigator - Navigating The Key News Stories In Pharma. Retrieved May 23, 2026, from

<https://www.thepharmanavigator.com/news/prolific-machines-sets-new-mab-productivity-benchmark-at-21-gl>

9. Prolific Machines Announces the Completion of Optogenetic Bioproduction at 200L, Marking the World's Largest Demonstration of Optogenetics to Date. (n.d.). Retrieved May 23, 2026, from <https://www.businesswire.com/news/home/20250417177385/en/Prolific-Machines-Announces-the-Completion-of-Optogenetic-Bioproduction-at-200L-Marking-the-Worlds-Largest-Demonstration-of-Optogenetics-to-Date>
10. Kiviat, K., Block, D. E., & Manikantan, H. (2025). Simulation of bioreactor flow environments for large-scale animal cell culture conditions relevant to cultivated meat production. *Biochemical Engineering Journal*, 220, 109753. <https://doi.org/10.1016/j.bej.2025.109753>
11. Ranganathan, P., Pandey, A. K., Sirohi, R., Tuan Hoang, A., & Kim, S.-H. (2022). Recent advances in computational fluid dynamics (CFD) modelling of photobioreactors: Design and applications. *Bioresource Technology*, 350, 126920. <https://doi.org/10.1016/j.biortech.2022.126920>
12. Gernigon, V., Chekroun, M. A., Cockx, A., Guiraud, P., & Morchain, J. (2019). How Mixing and Light Heterogeneity Impact the Overall Growth Rate in Photobioreactors. *Chemical Engineering & Technology*, 42(8), 1663–1669. <https://doi.org/10.1002/ceat.201900102>
13. Crater, J. S., & Lieverse, J. C. (2018). Scale-up of industrial microbial processes. *FEMS Microbiology Letters*, 365(13), fny138. <https://doi.org/10.1093/femsle/fny138>
14. Pouzet, S., Banderas, A., Le Bec, M., Lautier, T., Truan, G., & Hersen, P. (2020). The Promise of Optogenetics for Bioproduction: Dynamic Control Strategies and Scale-Up Instruments. *Bioengineering*, 7(4), 151. <https://doi.org/10.3390/bioengineering7040151>
15. Jiang, X., Ji, J., Yang, Q., Dou, Y., Li, Y., Yang, X., Liu, C., Dou, S., & Dong, L. (2025). Key Technologies of Synthetic Biology in Industrial Microbiology. *Microorganisms*, 13(10), 2343. <https://doi.org/10.3390/microorganisms13102343>

16. Minami, S. A., Garimella, S. S., & Shah, P. S. (2024). Computational evaluation of light propagation in cylindrical bioreactors for optogenetic mammalian cell cultures. *Biotechnology Journal*, 19(1), 2300071. <https://doi.org/10.1002/biot.202300071>
17. Cone, M. T., Mason, J. D., Figueroa, E., Hokr, B. H., Bixler, J. N., Castellanos, C. C., Noojin, G. D., Wigle, J. C., Rockwell, B. A., Yakovlev, V. V., & Fry, E. S. (2015). Measuring the absorption coefficient of biological materials using integrating cavity ring-down spectroscopy. *Optica*, 2(2), 162–168. <https://doi.org/10.1364/OPTICA.2.000162>
18. Röttgers, R., & Gehnke, S. (2012). Measurement of light absorption by aquatic particles: Improvement of the quantitative filter technique by use of an integrating sphere approach. *Applied Optics*, 51(9), 1336–1351. <https://doi.org/10.1364/AO.51.001336>
19. Jacques, S. L. (2013). Optical properties of biological tissues: A review. *Physics in Medicine & Biology*, 58(11), R37. <https://doi.org/10.1088/0031-9155/58/11/R37>
20. Hohmann, M., Lengenfelder, B., Muhr, D., Späth, M., Hauptkorn, M., Klämpfl, F., & Schmidt, M. (2021). Direct measurement of the scattering coefficient. *Biomedical Optics Express*, 12(1), 320–335. <https://doi.org/10.1364/BOE.410248>
21. Rischer, H., Szilvay, G. R., & Oksman-Caldentey, K.-M. (2020). Cellular agriculture—Industrial biotechnology for food and materials. *Current Opinion in Biotechnology*, 61, 128–134. <https://doi.org/10.1016/j.copbio.2019.12.003>
22. Post, M. J., Levenberg, S., Kaplan, D. L., Genovese, N., Fu, J., Bryant, C. J., Negowetti, N., Verzijden, K., & Moutsatsou, P. (2020). Scientific, sustainability and regulatory challenges of cultured meat. *Nature Food*, 1(7), 403–415. <https://doi.org/10.1038/s43016-020-0112-z>
23. Martins, B., Bister, A., Dohmen, R. G. J., Gouveia, M. A., Hueber, R., Melzener, L., Messmer, T., Papadopoulos, J., Pimenta, J., Raina, D., Schaeken, L., Shirley, S., Bouchet, B. P., & Flack, J. E. (2024). Advances and Challenges in Cell Biology for Cultured Meat. *Annual Review of Animal Biosciences*, 12(Volume 12, 2024), 345–368. <https://doi.org/10.1146/annurev-animal-021022-055132>

24. Jara, T. C., Park, K., Vahmani, P., Van Eenennaam, A. L., Smith, L. R., & Denicol, A. C. (2023). Stem cell-based strategies and challenges for production of cultivated meat. *Nature Food*, 4(10), 841–853. <https://doi.org/10.1038/s43016-023-00857-z>
25. Negulescu, P. G., Risner, D., Spang, E. S., Sumner, D., Block, D., Nandi, S., & McDonald, K. A. (2023). Techno-economic modeling and assessment of cultivated meat: Impact of production bioreactor scale. *Biotechnology and Bioengineering*, 120(4), 1055–1067. <https://doi.org/10.1002/bit.28324>
26. Xie, Y., Ding, S., Wang, J., Li, X., Guo, R., Tang, C., Wang, C., Zhou, G., & Li, C. (2026). *Technological Advances and the Challenges for Large-Scale Cultured Meat Production*. <https://doi.org/10.1146/annurev-food-053124-085815>
27. Ravikumar, M., Powell, D., & Huling, R. (2024). Cultivated meat: Research opportunities to advance cell line development. *Trends in Cell Biology*, 34(7), 523–526. <https://doi.org/10.1016/j.tcb.2024.04.005>
28. O'Neill, E. N., Ansel, J. C., Kwong, G. A., Plastino, M. E., Nelson, J., Baar, K., & Block, D. E. (2022). Spent media analysis suggests cultivated meat media will require species and cell type optimization. *Npj Science of Food*, 6(1), 46. <https://doi.org/10.1038/s41538-022-00157-z>
29. Marucci, L. (2017). Nanog Dynamics in Mouse Embryonic Stem Cells: Results from Systems Biology Approaches. *Stem Cells International*, 2017(1), 7160419. <https://doi.org/10.1155/2017/7160419>
30. Stout, A. J., Arnett, M. J., Chai, K., Guo, T., Liao, L., Mirliani, A. B., Rittenberg, M. L., Shub, M., White, E. C., Yuen, J. S. K. Jr., Zhang, X., & Kaplan, D. L. (2023). Immortalized Bovine Satellite Cells for Cultured Meat Applications. *ACS Synthetic Biology*, 12(5), 1567–1573. <https://doi.org/10.1021/acssynbio.3c00216>

31. Kallunki, T., Barisic, M., Jäätelä, M., & Liu, B. (2019). How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells*, 8(8), 796. <https://doi.org/10.3390/cells8080796>
32. De Carluccio, G., Fusco, V., & di Bernardo, D. (2024). Engineering a synthetic gene circuit for high-performance inducible expression in mammalian systems. *Nature Communications*, 15(1), 3311. <https://doi.org/10.1038/s41467-024-47592-y>
33. Yamada, K., & Noguchi, T. (1999). Nutrient and hormonal regulation of pyruvate kinase gene expression. *Biochemical Journal*, 337(Pt 1), 1–11.
34. Liu, Z., Thompson, K. S., & Towle, H. C. (1993). Carbohydrate regulation of the rat L-type pyruvate kinase gene requires two nuclear factors: LF-A1 and a member of the c-myc family. *Journal of Biological Chemistry*, 268(17), 12787–12795. [https://doi.org/10.1016/S0021-9258\(18\)31457-1](https://doi.org/10.1016/S0021-9258(18)31457-1)
35. Han, J., McLane, B., Kim, E.-H., Yoon, J.-W., & Jun, H.-S. (2011). Remission of Diabetes by Insulin Gene Therapy Using a Hepatocyte-specific and Glucose-responsive Synthetic Promoter. *Molecular Therapy*, 19(3), 470–478. <https://doi.org/10.1038/mt.2010.255>
36. Xu, J., Christian, B., & Jump, D. B. (2006). Regulation of Rat Hepatic L-Pyruvate Kinase Promoter Composition and Activity by Glucose, n-3 Polyunsaturated Fatty Acids, and Peroxisome Proliferator-activated Receptor- α Agonist. *The Journal of Biological Chemistry*, 281(27), 18351–18362. <https://doi.org/10.1074/jbc.M601277200>
37. Lan, M. S., Wang, H.-W., Chong, J., & Breslin, M. B. (2007). Coupling of glucose response element from L-type pyruvate kinase and G6Pase promoter enhances glucose responsive activity in hepatoma cells. *Molecular and Cellular Biochemistry*, 300(1), 191–196. <https://doi.org/10.1007/s11010-006-9383-4>
38. Yamada, K., & Noguchi, T. (1999). Regulation of Pyruvate Kinase M Gene Expression. *Biochemical and Biophysical Research Communications*, 256(2), 257–262. <https://doi.org/10.1006/bbrc.1999.0228>

39. Vaulont, S., & Kahn, A. (1994). Transcriptional control of metabolic regulation genes by carbohydrates. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 8(1), 28–35. <https://doi.org/10.1096/fasebj.8.1.8299888>
40. Kawaguchi, T., Takenoshita, M., Kabashima, T., & Uyeda, K. (2001). Glucose and cAMP regulate the L-type pyruvate kinase gene by phosphorylation/dephosphorylation of the carbohydrate response element binding protein. *Proceedings of the National Academy of Sciences*, 98(24), 13710–13715. <https://doi.org/10.1073/pnas.231370798>
41. Nagaoka, M., Shiraishi, Y., & Sugiura, Y. (2001). Selected base sequence outside the target binding site of zinc finger protein Sp1. *Nucleic Acids Research*, 29(24), 4920–4929. <https://doi.org/10.1093/nar/29.24.4920>
42. Chaudhary, J., & Skinner, M. K. (1999). Basic helix-loop-helix proteins can act at the E-box within the serum response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli cells. *Molecular Endocrinology*, 13(5), 774–786. <https://doi.org/10.1210/mend.13.5.0271>
43. O'Callaghan, B. L., Koo, S. H., Wu, Y., Freake, H. C., & Towle, H. C. (2001). Glucose regulation of the acetyl-CoA carboxylase promoter PI in rat hepatocytes. *The Journal of Biological Chemistry*, 276(19), 16033–16039. <https://doi.org/10.1074/jbc.M101557200>
44. Stoeckman, A. K., Ma, L., & Towle, H. C. (2004). Mlx is the functional heteromeric partner of the carbohydrate response element-binding protein in glucose regulation of lipogenic enzyme genes. *The Journal of Biological Chemistry*, 279(15), 15662–15669. <https://doi.org/10.1074/jbc.M311301200>
45. Hunt, L. C., Xu, B., Finkelstein, D., Fan, Y., Carroll, P. A., Cheng, P.-F., Eisenman, R. N., & Demontis, F. (2015). The glucose-sensing transcription factor MLX promotes myogenesis via myokine signaling. *Genes & Development*, 29(23), 2475–2489. <https://doi.org/10.1101/gad.267419.115>

46. Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenko, V. V., & Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 128(6), 1231–1245. <https://doi.org/10.1016/j.cell.2006.12.048>
47. Pan, G., & Thomson, J. A. (2007). Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Research*, 17(1), 42–49. <https://doi.org/10.1038/sj.cr.7310125>
48. Shahini, A., Choudhury, D., Asmani, M., Zhao, R., Lei, P., & Andreadis, S. T. (2018). NANOG restores the impaired myogenic differentiation potential of skeletal myoblasts after multiple population doublings. *Stem Cell Research*, 26, 55–66. <https://doi.org/10.1016/j.scr.2017.11.018>
49. Thompson, K. S., & Towle, H. C. (1991). Localization of the carbohydrate response element of the rat L-type pyruvate kinase gene. *Journal of Biological Chemistry*, 266(14), 8679–8682. [https://doi.org/10.1016/S0021-9258\(18\)31498-4](https://doi.org/10.1016/S0021-9258(18)31498-4)
50. Shih, H.-M., Liu, Z., & Towle, H. C. (1995). Two CACGTG Motifs with Proper Spacing Dictate the Carbohydrate Regulation of Hepatic Gene Transcription (*). *Journal of Biological Chemistry*, 270(37), 21991–21997. <https://doi.org/10.1074/jbc.270.37.21991>
51. K Handral, H., Hua Tay, S., Wan Chan, W., & Choudhury, D. (2022). 3D Printing of cultured meat products. *Critical Reviews in Food Science and Nutrition*, 62(1), 272–281. <https://doi.org/10.1080/10408398.2020.1815172>
52. Kawecki, N. S., Chen, K. K., Smith, C. S., Xie, Q., Cohen, J. M., & Rowat, A. C. (2024). Scalable Processes for Culturing Meat Using Edible Scaffolds. *Annual Review of Food Science and Technology*, 15(Volume 15, 2024), 241–264. <https://doi.org/10.1146/annurev-food-072023-034451>
53. Mironov, V., Kasyanov, V., & Markwald, R. R. (2011). Organ printing: From bioprinter to organ biofabrication line. *Current Opinion in Biotechnology, Tissue, Cell and Pathway Engineering*, 22(5), 667–673. <https://doi.org/10.1016/j.copbio.2011.02.006>

54. Fan, R., Piou, M., Darling, E., Cormier, D., Sun, J., & Wan, J. (2016). Bio-printing cell-laden Matrigel-agarose constructs. *Journal of Biomaterials Applications*, 31(5), 684–692. <https://doi.org/10.1177/0885328216669238>
55. Polstein, L. R., & Gersbach, C. A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology*, 11(3), 198–200. <https://doi.org/10.1038/nchembio.1753>
56. Beyer, H. M., Kumar, S., Nieke, M., Diehl, C. M. C., Tang, K., Shumka, S., Koh, C. S., Fleck, C., Davies, J. A., Khammash, M., & Zurbriggen, M. D. (2024). Genetically-stable engineered optogenetic gene switches modulate spatial cell morphogenesis in two- and three-dimensional tissue cultures. *Nature Communications*, 15(1), 10470. <https://doi.org/10.1038/s41467-024-54350-7>
57. Sakar, M. S., Neal, D., Boudou, T., Borochin, M. A., Li, Y., Weiss, R., Kamm, R. D., Chen, C. S., & Asada, H. H. (2012). Formation and optogenetic control of engineered 3D skeletal muscle bioactuators. *Lab on a Chip*, 12(23), 4976–4985. <https://doi.org/10.1039/C2LC40338B>

CHAPTER 4: Introduction to viral spillover and pandemic screening

RNA viruses have caused almost every epidemic or pandemic over the last century. Emerging infectious diseases have increased over the last century, and zoonotic viruses account for a majority of these^[1–3]. Zoonotic spillover rates are increasing with greater urbanization, rapid deforestation, and climate change which in turn increases the chance of pathogens spilling over from zoonotic reservoirs to cause outbreaks. This chapter explains how zoonotic spillover increases the possibility of pandemic emergence, specifically with respect to RNA viruses.

4.1 Increasing rates of zoonotic spillover and outbreak pressure

About 60% of all emerging pathogens have spilled over from zoonotic reservoirs, the natural host population where a pathogen circulates without harming the host^[3,4]. Disease spreads through repeated spillover events usually from an animal population to the human population, known as zoonotic spillover. Disease appears from an initial spillover event that is amplified through infection with the contagion and eventual spread of the disease across population and location (Figure 4-1). As a pathogen continues to grow and amplify, the outbreak grows to become an epidemic or pandemic. Once the exponential spread of the virus transitions to becoming a more constant presence, it becomes endemic^[5,6]. Over recent decades, the rate of emerging viruses spilling over from zoonotic reservoirs has increased as global warming and overconsumption have grown alongside global population growth (Figure 4-2)^[1,4]. As the world continues to urbanize and human beings push into historically undisturbed areas, our increased interactions with wild animal species increase the chances of being infected with new diseases ^[2,5,7–10].

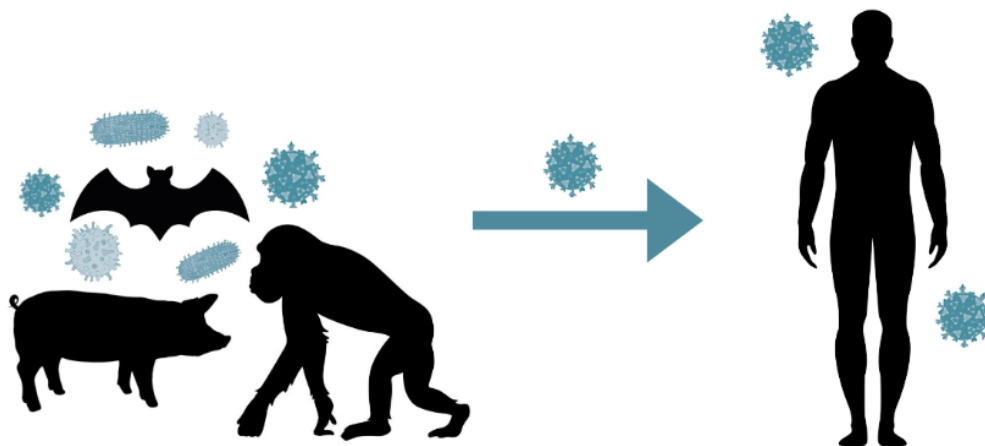


Figure 4-1. General diagram of spillover. Initially, a virus circulates within a wild animal population, the zoonotic reservoir. The virus can accrue different mutations that allow it to be able to infect humans in addition to other epidemiological and ecological factors aligning. This is called a spillover event when the virus jumps from the zoonotic reservoir to the human population.

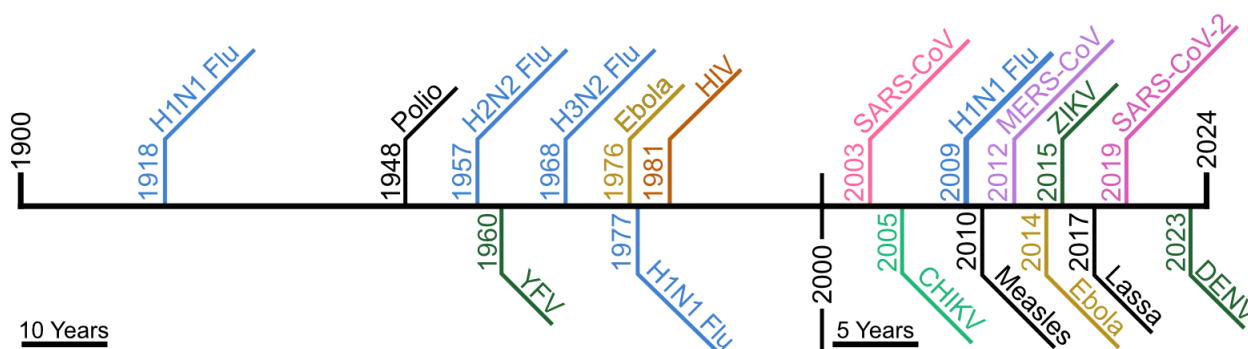


Figure 4-2. Timeline of epidemics and pandemics across the 20th and 21st century. The rate of these viral events has increased in the last 20 years. The 20th century saw eight major -demic events (scaled for every ten years). Since 2000, there have been ten events (scaled for every five years). The timeline has been scaled to highlight the rapid increase in epidemic and pandemic events.

Increased spillover attempts, or prolonged human exposures to a zoonotic virus, increase the chances for a virus to select for desirable traits to support sustainable human transmission. While

only a few spillover events truly succeed at human transmission and spread, rising outbreak pressure increases their chances of succeeding^[5]. Outbreak pressure includes frequent and persistent wildlife contact, ecological disruption, and the ability of a virus to adapt and select for traits for continued animal-to-human and then human-to-human transmission^[11,12].

4.2 Environmental and viral traits driving zoonotic spillover

The positive feedback loop between frequent spillover events and subsequent outbreaks demonstrates that a virus's sustained introduction from a reservoir can effectively overcome the barrier to human transmission by providing the virus with opportunities to adapt to a human host^[1,5]. But what is driving this increased outbreak pressure?

4.2.1 Conditions for Pandemic Emergence

Human expansion, exponential population growth, and increasing industrialization have led to the overconsumption of natural resources, encroachment into wild habitats, and rising global temperatures. A growing population inevitably increases contact between humans and wildlife, primarily driven by deforestation and agricultural expansion^[4,5,13–15]. This severe change in land-use has induced ecological disruption and habitat loss, increasing the probability of human-pathogen interfacing. In the last twenty years, about 1 million square kilometers of global forests were lost due to agricultural expansion^[16]. For example, the Amazon rainforest has lost about 17% of its original landscape due to deforestation, rapidly displacing local wild species from their habitats^[17]. The extreme land-use changes have been directly correlated to which has been linked to about 25% of emerging infectious disease events since the cleared land increases the direct contact between zoonotic reservoirs and humans^[18,19].

Simultaneously, rising temperatures allow zoonotic vectors to range far beyond their native regions. Rising global temperatures have increased by approximately 1°C and may reach 2°C in the next couple of decades. This shift in global temperatures allows zoonotic vectors, such as mosquitoes and bats, to shift and expand their ranges across broader stretches beyond their historic limits. For example, *Aedes aegypti*, the mosquito vector of dengue and Zika, have expanded their coverage from the global south and equatorial regions to reach parts of Europe and North America^[15,20–22]. This is in direct response to global warming where zoonotic reservoirs are encroaching into previously temperate zones and introducing novel disease risks^[21,23,24].

A virus achieves pandemic-potential when exposure, host compatibility, high virulence, and transmission efficiency align to support sustained spread in human populations^[11,25]. However, crossing the species barrier is the first step. For a virus to amplify to epidemic or pandemic scale, it often requires specific biological adaptation. The virus must possess the machinery to bind human receptors and efficiently replicate while evading immune defenses^[5,6,20]. Additionally, to achieve sustained human-to-human transmission, the virus needs to be released in large amounts and spread through specific transmission routes like the air or through vectors. When these ecological pressures converge with highly desired viral traits, a localized outbreak can escalate to a global public health threat.

4.2.2 Adaptive Nature of RNA Viruses

RNA viruses pose the biggest risk for increased pandemic probability, accounting for about 45% of all emerging pathogens^[1,2]. They pose a unique threat to public health because of their tendency to combine zoonotic transmission, shorter replication cycles, and high mutation rates^[2,26]. Unlike most DNA viruses, RNA viruses generally encode their own replication machinery which is optimized to fit their needs and usually does not include any proofreading capabilities^[27]. Consequently, they are prone to high mutation rates with an average range

between 10^{-6} to 10^{-4} mutations per nucleotide per replication cycle^[27,28]. This rapid diversification allows RNA viruses to exist as dynamic “quasispecies” (the result of mutations creating a diverse population of variants) granting them evolutionary flexibility to quickly adapt^[29,30]. To evaluate pandemic risk across the landscape of RNA viruses, screening efforts must account for diverse genetic mechanisms and ecological factors.

Overcoming this barrier for that initial handshake between a virus and a host requires rapid molecular adaptation. For a virus to successfully infect its host, its surface protein must bind to a human cell host receptor^[5,11,31,32]. The constant pressure to adapt makes RNA viruses prime candidates for pandemic emergence, driven by error-prone replication resulting in high mutation rates. The standard RNA-dependent RNA polymerase (RdRP) lacks an intrinsic exoribonuclease domain, that is normally responsible for proofreading during replication^[33]. Furthermore, the structural flexibility of an RdRP means that it can prioritize speed over strand alignment, resulting in a high rate of template switching during replication. The promiscuity of the polymerase is optimized for faster replication rates at the expense of genomic fidelity^[34,35]. This rapid evolution allows RNA viruses to produce genetic variants which increases the probability of acquiring the necessary traits to breach the species barrier and cause spillover^[2,3,27,36]. Because these mutations create a pool of variants, proactive screening pipelines are essential to monitor reservoirs to flag potential pathogens before infecting humans.

However, the mechanisms that allow viruses to adapt also create blind spots for viral surveillance. The gradual accumulation of mutations drives genetic drift. This can cause issues with screening efforts if a pool of viruses has a large-scale genetic shift. Mechanisms such as genome segmentation and recombination can instantly change a virus's host-receptor tropism. This means that surveillance frameworks can overlook structural changes if they only rely on phylogeny, sequence similarity, or genome alignments.

Influenza A is one of the strongest models for pandemic emergence, where strains arise through constant host-switching, reassortment, and adaptation^[37,38]. As an orthomyxovirus, it has a segmented negative-sense RNA genome. When two influenza strains co-infect a single host cell, these genomic segments can be swapped during viral assembly. As a recent example, the 2009 H1N1 virus was a result of gene segments from different host backgrounds combining into a pandemic strain^[37,39]. For screening pipelines, reassortment can complicate phylogenetic tracking. A screening system cannot rely on a virus's phylogeny historical lineage to predict its risk if it can acquire the ability to bind a host protein^[37,40]. However, because the primary bottleneck for zoonotic spillover remains to be the protein-protein interaction between the viral surface protein and host cell receptor, such as influenza's hemagglutinin binding to sialic acid. This binding interface can dictate whether a virus can cross the species barrier, screening pipelines can bypass this genomic reshuffling and directly evaluate spillover risk.

Coronaviruses are positive-sense, single-stranded RNA viruses with the largest viral RNA genome of around 30 kilobasepairs. Coronaviruses uniquely code for an exoribonuclease to ensure replication accuracy to maintain its large genome. However, the activity of the enzyme seems to change across hosts which may impair its ability to reconcile genome mutations as it spills over^[41]. The increased chance of recombination events during coronavirus replication increases the chances of a new virus emerging^[23]. This, coupled with their ability to spill over from bats, has warranted research on key interactions that are responsible for virus spillover and infection ^[5,23,31,42,43].

There are only seven known coronaviruses that can infect humans, three of which have caused pandemics in the last two decades, underscoring the accelerated nature of their emergence as a public health threat (Figure 4-2)^[32,44,45]. This rapid succession of outbreaks demonstrates that

coronavirus spillover is a continuing threat rather than a series of historical anomalies. Because these viruses utilize diverse cell surface receptors to cross species barriers, they represent a shifting evolutionary target. Waiting for a novel virus to emerge before characterizing it is no longer a viable strategy. Therefore, developing predictive tools to proactively screen emerging coronaviruses is essential for intercepting the next imminent emergence event.

The next two chapters discuss a workflow to screen the initial virus-host protein-protein interaction of emerging coronaviruses to a human receptor. Chapter 5 focuses on the development of a prediction pipeline for screening emerging coronaviruses for pandemic-potential in comparison to SARS-CoV-1, SARS-CoV-2, and MERS-COV—known pandemic-causing coronaviruses. Chapter 6 proposes experimental validation to determine if these interactions are real and whether the emerging coronaviruses could be considered a threat.

4.3 References

1. Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990–993. <https://doi.org/10.1038/nature06536>
2. Carrasco-Hernandez, R., Jácome, R., López Vidal, Y., & Ponce de León, S. (2017). Are RNA Viruses Candidate Agents for the Next Global Pandemic? A Review. *ILAR Journal*, 58(3), 343–358. <https://doi.org/10.1093/ilar/ilx026>
3. Weiss, R. A., & Sankaran, N. (2022). Emergence of epidemic diseases: Zoonoses and other origins. *Faculty Reviews*, 11, 2. <https://doi.org/10.12703/r/11-2>
4. Cutler, S. J., Fooks, A. R., & van der Poel, W. H. M. (2010). Public Health Threat of New, Reemerging, and Neglected Zoonoses in the Industrialized World. *Emerging Infectious Diseases*, 16(1), 1–7. <https://doi.org/10.3201/eid1601.081467>

5. Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, 15(8), Article 8. <https://doi.org/10.1038/nrmicro.2017.45>
6. Pandit, P. S., Anthony, S. J., Goldstein, T., Olival, K. J., Doyle, M. M., Gardner, N. R., Bird, B., Smith, W., Wolking, D., Gilardi, K., Monagin, C., Kelly, T., Uhart, M. M., Epstein, J. H., Machalaba, C., Rostal, M. K., Dawson, P., Hagan, E., Sullivan, A., ... Johnson, C. K. (2022). Predicting the potential for zoonotic transmission and host associations for novel viruses. *Communications Biology*, 5(1), 844. <https://doi.org/10.1038/s42003-022-03797-9>
7. Duque, M., De Coninck, D., Montero-Zamora, P., Sahbaz, S., Bautista, T., Ertanir, B., Garcia, M. F., Alpysbekova, A., Scaramutti, C., Laboy, M., Vo, D., Acaf, Y., Vignoles, V., & Schwartz, S. J. (2024). The spillover effect of implicit immigrant stereotypes and incidental pandemic threat. *International Journal of Intercultural Relations*, 99, 101939. <https://doi.org/10.1016/j.ijintrel.2024.101939>
8. Grange, Z. L., Goldstein, T., Johnson, C. K., Anthony, S., Gilardi, K., Daszak, P., Olival, K. J., O'Rourke, T., Murray, S., Olson, S. H., Togami, E., Vidal, G., Expert Panel, PREDICT Consortium, & Mazet, J. A. K. (2021). Ranking the risk of animal-to-human spillover for newly discovered viruses. *Proceedings of the National Academy of Sciences*, 118(15), e2002324118. <https://doi.org/10.1073/pnas.2002324118>
9. Spervovasilis, N., Markaki, I., Papadakis, M., Tsioutis, C., & Markaki, L. (2021). Epidemics and pandemics: Is human overpopulation the elephant in the room? *Ethics, Medicine, and Public Health*, 19, 100728. <https://doi.org/10.1016/j.jemep.2021.100728>
10. Gippet, J. M. W., Carlson, C. J., Klaftenberger, T., Schweizer, M., Eskew, E. A., Gore, M. L., & Bertelsmeier, C. (2026). Wildlife trade drives animal-to-human pathogen transmission over 40 years. *Science*, 392(6794), 178–182. <https://doi.org/10.1126/science.adw5518>
11. Morse, S. S., Mazet, J. A., Woolhouse, M., Parrish, C. R., Carroll, D., Karesh, W. B., Zambrana-Torrel, C., Lipkin, W. I., & Daszak, P. (2012). Prediction and prevention of the

- next pandemic zoonosis. *The Lancet*, 380(9857), 1956–1965. [https://doi.org/10.1016/S0140-6736\(12\)61684-5](https://doi.org/10.1016/S0140-6736(12)61684-5)
12. Wolfe, N. D., Dunavan, C. P., & Diamond, J. (2007). Origins of major human infectious diseases. *Nature*, 447(7142), 279–283. <https://doi.org/10.1038/nature05775>
 13. Morand, S., & Lajaunie, C. (2021). Outbreaks of Vector-Borne and Zoonotic Diseases Are Associated With Changes in Forest Cover and Oil Palm Expansion at Global Scale. *Frontiers in Veterinary Science*, 8. <https://doi.org/10.3389/fvets.2021.661063>
 14. Farnham, A., Cossa, H., Dietler, D., Engebretsen, R., Leuenberger, A., Lyatuu, I., Nimako, B., Zabre, H. R., Brugger, F., & Winkler, M. S. (2020). Investigating Health Impacts of Natural Resource Extraction Projects in Burkina Faso, Ghana, Mozambique, and Tanzania: Protocol for a Mixed Methods Study. *JMIR Research Protocols*, 9(4), e17138. <https://doi.org/10.2196/17138>
 15. Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., Hudson, P., Jolles, A., Jones, K. E., Mitchell, C. E., Myers, S. S., Bogich, T., & Ostfeld, R. S. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468(7324), 647–652. <https://doi.org/10.1038/nature09575>
 16. Rohr, J. R., Barrett, C. B., Civitello, D. J., Craft, M. E., Delius, B., DeLeo, G. A., Hudson, P. J., Jouanard, N., Nguyen, K. H., Ostfeld, R. S., Remais, J. V., Riveau, G., Sokolow, S. H., & Tilman, D. (2019). Emerging human infectious diseases and the links to global food production. *Nature Sustainability*, 2(6), 445–456. <https://doi.org/10.1038/s41893-019-0293-3>
 17. Albert, J. S., Carnaval, A. C., Flantua, S. G. A., Lohmann, L. G., Ribas, C. C., Riff, D., Carrillo, J. D., Fan, Y., Figueiredo, J. J. P., Guayasamin, J. M., Hoorn, C., de Melo, G. H., Nascimento, N., Quesada, C. A., Ulloa Ulloa, C., Val, P., Arieira, J., Encalada, A. C., & Nobre, C. A. (2023). Human impacts outpace natural processes in the Amazon. *Science*, 379(6630), eabo5003. <https://doi.org/10.1126/science.abo5003>

18. Plowright, R. K., Reaser, J. K., Locke, H., Woodley, S. J., Patz, J. A., Becker, D. J., Oppler, G., Hudson, P. J., & Tabor, G. M. (2021). Land use-induced spillover: A call to action to safeguard environmental, animal, and human health. *The Lancet Planetary Health*, 5(4), e237–e245. [https://doi.org/10.1016/S2542-5196\(21\)00031-0](https://doi.org/10.1016/S2542-5196(21)00031-0)
19. Pfenning-Butterworth, A., Buckley, L. B., Drake, J. M., Farner, J. E., Farrell, M. J., Gehman, A.-L. M., Mordecai, E. A., Stephens, P. R., Gittleman, J. L., & Davies, T. J. (2024). Interconnecting global threats: Climate change, biodiversity loss, and infectious diseases. *The Lancet Planetary Health*, 8(4), e270–e283. [https://doi.org/10.1016/S2542-5196\(24\)00021-4](https://doi.org/10.1016/S2542-5196(24)00021-4)
20. Fishburn, A. T., Pham, O. H., Kenaston, M. W., Beesabathuni, N. S., & Shah, P. S. (2022). Let's Get Physical: Flavivirus-Host Protein-Protein Interactions in Replication and Pathogenesis. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.847588>
21. Zitzmann, C., Adia, N. A. B., Shah, P. S., & Manore, C. (2025). Opportunities in multiscale modeling of mosquito-borne flaviviruses. *BioSystems*, 257, 105593. <https://doi.org/10.1016/j.biosystems.2025.105593>
22. Ryan, S. J., Carlson, C. J., Mordecai, E. A., & Johnson, L. R. (2019). Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLOS Neglected Tropical Diseases*, 13(3), e0007213. <https://doi.org/10.1371/journal.pntd.0007213>
23. Ruiz-Aravena, M., McKee, C., Gamble, A., Lunn, T., Morris, A., Snedden, C. E., Yinda, C. K., Port, J. R., Buchholz, D. W., Yeo, Y. Y., Faust, C., Jax, E., Dee, L., Jones, D. N., Kessler, M. K., Falvo, C., Crowley, D., Bharti, N., Brook, C. E., ... Plowright, R. K. (2022). Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology*, 20(5), Article 5. <https://doi.org/10.1038/s41579-021-00652-2>
24. Luis, A. D., Hayman, D. T. S., O'Shea, T. J., Cryan, P. M., Gilbert, A. T., Pulliam, J. R. C., Mills, J. N., Timonin, M. E., Willis, C. K. R., Cunningham, A. A., Fooks, A. R., Rupprecht, C. E., Wood, J. L. N., & Webb, C. T. (2013). A comparison of bats and rodents as reservoirs of

- zoonotic viruses: Are bats special? *Proceedings of the Royal Society B: Biological Sciences*, 280(1756), 20122753. <https://doi.org/10.1098/rspb.2012.2753>
25. Neumann, G., & Kawaoka, Y. (2023). Which Virus Will Cause the Next Pandemic? *Viruses*, 15(1), 199. <https://doi.org/10.3390/v15010199>
 26. Woolhouse, M. E. J., & Gowtage-Sequeria, S. (2005). Host Range and Emerging and Reemerging Pathogens. *Emerging Infectious Diseases*, 11(12), 1842–1847. <https://doi.org/10.3201/eid1112.050997>
 27. Duffy, S. (2018). Why are RNA virus mutation rates so damn high? *PLOS Biology*, 16(8), e3000003. <https://doi.org/10.1371/journal.pbio.3000003>
 28. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral Mutation Rates. *Journal of Virology*, 84(19), 9733–9748. <https://doi.org/10.1128/jvi.00694-10>
 29. Singh, K., Mehta, D., Dumka, S., Chauhan, A. S., & Kumar, S. (2023). Quasispecies Nature of RNA Viruses: Lessons from the Past. *Vaccines*, 11(2), 308. <https://doi.org/10.3390/vaccines11020308>
 30. Sardanyés, J., Perales, C., Domingo, E., & Elena, S. F. (2024). Quasispecies theory and emerging viruses: Challenges and applications. *Npj Viruses*, 2(1), 54. <https://doi.org/10.1038/s44298-024-00066-w>
 31. Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K., & Munster, V. J. (2020). Bat-borne virus diversity, spillover and emergence. *Nature Reviews Microbiology*, 18(8), 461–471. <https://doi.org/10.1038/s41579-020-0394-z>
 32. Menachery, V. D., Graham, R. L., & Baric, R. S. (2017). Jumping species—A mechanism for coronavirus persistence and survival. *Current Opinion in Virology*, 23, 1–7. <https://doi.org/10.1016/j.coviro.2017.01.002>
 33. Moeller, N. H., Shi, K., Demir, Ö., Belica, C., Banerjee, S., Yin, L., Durfee, C., Amaro, R. E., & Aihara, H. (2022). Structure and dynamics of SARS-CoV-2 proofreading exoribonuclease

- ExoN. *Proceedings of the National Academy of Sciences*, 119(9), e2106379119. <https://doi.org/10.1073/pnas.2106379119>
34. Cheng, C.-P., & Nagy, P. D. (2003). Mechanism of RNA Recombination in Carno- and Tombusviruses: Evidence for Template Switching by the RNA-Dependent RNA Polymerase In Vitro. *Journal of Virology*, 77(22), 12033–12047. <https://doi.org/10.1128/JVI.77.22.12033-12047.2003>
 35. Moustafa, I. M., Korboukh, V. K., Arnold, J. J., Smidansky, E. D., Marcotte, L. L., Gohara, D. W., Yang, X., Sánchez-Farrán, M. A., Filman, D., Maranas, J. K., Boehr, D. D., Hogle, J. M., Colina, C. M., & Cameron, C. E. (2014). Structural Dynamics as a Contributor to Error-prone Replication by an RNA-dependent RNA Polymerase. *The Journal of Biological Chemistry*, 289(52), 36229–36248. <https://doi.org/10.1074/jbc.M114.616193>
 36. Olival, K. J., Hosseini, P. R., Zambrana-Torrel, C., Ross, N., Bogich, T. L., & Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature*, 546(7660), 646–650. <https://doi.org/10.1038/nature22975>
 37. Saunders-Hastings, P. R., & Krewski, D. (2016). Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission. *Pathogens*, 5(4), 66. <https://doi.org/10.3390/pathogens5040066>
 38. Russell, C. J. (2016). Orthomyxoviruses: Structure of Antigens. *Reference Module in Biomedical Sciences*, B978-0-12-801238-3.95721-0. <https://doi.org/10.1016/B978-0-12-801238-3.95721-0>
 39. Neumann, G., Noda, T., & Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature*, 459(7249), 931–939. <https://doi.org/10.1038/nature08157>
 40. Wells, H. L., Letko, M., Lasso, G., Ssebide, B., Nziza, J., Byarugaba, D. K., Navarrete-Macias, I., Liang, E., Cranfield, M., Han, B. A., Tingley, M. W., Diuk-Wasser, M., Goldstein, T., Johnson, C. K., Mazet, J. a. K., Chandran, K., Munster, V. J., Gilardi, K., & Anthony, S. J. (2021). The

- evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. *Virus Evolution*, 7(1), veab007. <https://doi.org/10.1093/ve/veab007>
41. Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., & Ng, W.-L. (2020). Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Molecular Cell*, 79(5), 710–727. <https://doi.org/10.1016/j.molcel.2020.07.027>
 42. Anthony, S. J., Johnson, C. K., Greig, D. J., Kramer, S., Che, X., Wells, H., Hicks, A. L., Joly, D. O., Wolfe, N. D., Daszak, P., Karesh, W., Lipkin, W. I., Morse, S. S., PREDICT Consortium, Mazet, J. A. K., & Goldstein, T. (2017). Global patterns in coronavirus diversity. *Virus Evolution*, 3(1), vex012. <https://doi.org/10.1093/ve/vex012>
 43. da Silva, P. G., Mesquita, J. R., de São José Nascimento, M., & Ferreira, V. A. M. (2021). Viral, host and environmental factors that favor anthroponotic spillover of coronaviruses: An opinionated review, focusing on SARS-CoV, MERS-CoV and SARS-CoV-2. *Science of The Total Environment*, 750, 141483. <https://doi.org/10.1016/j.scitotenv.2020.141483>
 44. Chen, Z., Boon, S. S., Wang, M. H., Chan, R. W. Y., & Chan, P. K. S. (2021). Genomic and evolutionary comparison between SARS-CoV-2 and other human coronaviruses. *Journal of Virological Methods*, 289, 114032. <https://doi.org/10.1016/j.jviromet.2020.114032>
 45. Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), Article 3. <https://doi.org/10.1038/s41579-018-0118-9>

Chapter 5: Prediction of coronavirus receptor usage using protein structure prediction and molecular dynamics

5.1 Introduction

Nearly every major pandemic of the past century has been caused by RNA viruses. A pandemic begins when a disease escapes its point of origin and spreads from a single geographical region to whole countries and continents simultaneously, disrupting lives and livelihoods. Often, these outbreaks start with a zoonotic spillover followed by amplification through human transmission and eventual widespread dissemination ^[1]. As the world continues to urbanize and human populations encroach further into historically undisturbed areas, our increased interactions with wildlife increase the chances of new spillover events. Public health responses have focused on reacting to outbreaks through containment, vaccination, and therapeutic interventions. But what if we could anticipate spillover threats before they emerge?

Coronaviruses are a clear example of why this question matters. In the last 20 years, three global outbreaks have been caused by coronaviruses^[2]. Severe acute respiratory syndrome coronavirus (SARS-CoV and now referred to as SARS-CoV-1) was first identified in southern China in November 2002, marking the emergence of the first pandemic of the 21^[3-5]. Middle East respiratory syndrome coronavirus (MERS-CoV) emerged in 2012 ^[6]. Most recently, in December 2019, SARS-CoV-2 emerged to cause the COVID-19 pandemic, one of the deadliest in recent

history ^[7–10]. These events have firmly established coronaviruses as a persistent and pressing threat, underscoring the importance of assessing their potential for future emergence.

Coronaviruses are positive-sense, single-stranded RNA viruses with the largest viral RNA genome of around 30 kilobases. Such a large genome is replicated by an RNA-dependent RNA polymerase with an estimated error rate of around 10^4 mutations per nucleotide. An exoribonuclease is also encoded to maintain genome fidelity ^[11–13]. The frequent template switching that occurs during replication increases the chance of recombination events^[14]. This drives rapid genetic variety but also increases the likelihood of emerging viruses. For instance, there is some speculation that recombination of different Spike sequences leads to the emergence of bat sarbecoviruses ^[11]. This, coupled with the pandemic potential of coronavirus spillover, has warranted research into key interactions that are responsible for virus spillover and infection ^[1,3,15,16].

There are only seven known coronaviruses that can infect humans, spanning either the alpha or betacoronavirus genus. Human coronaviruses (HCoVs) 229E and NL63 from the alphacoronavirus genus are also associated with the “common cold” ^[17–20]. The betacoronavirus genus has five subgenera, three of which are responsible for human-infecting coronaviruses: embecovirus, merbecovirus, and sarbecovirus ^[14]. Merbecoviruses include MERS-CoV, and sarbecoviruses include SARS-CoV-1 and SARS-CoV-2, which have caused respiratory virus outbreaks and pandemics ^[2,21,22]. Embecoviruses include HCoV-OC43 and HCoV-HKU1, which bind to N-acetyl-9-O-acetylneuraminic acid and are associated with the “common cold” ^[17,23–25].

The interaction between Spike and a host receptor, such as ACE2, is essential for coronavirus infection, though receptor usage can vary among different clades and strains of coronaviruses ^[3,11,26,27]. Evidence suggests that ACE2 usage is common and an ancestral trait among

sarbecoviruses, though there are several lineages that have lost ACE2 usage^[28,29]. Some sarbecoviruses can enter cells via receptor independent -mechanisms^[27,30]. Some sarbecovirus RBDs only need a few mutations to enable binding to human ACE2^[28,31–33]. Binding to ACE2 has also been studied in detail with respect to mutants of SARS-CoV-2 and has elucidated key residues and structural characteristics that are critical for binding ^[32,33]. Studying emerging coronaviruses under a similar lens can determine whether a virus has greater potential for human ACE2 binding and thus emergence.

Computationally predicting protein binding between coronaviruses and potential receptors can expedite the screening and, consequently, prioritize viral proteins for downstream experimental testing. Indeed, recent developments in computational biology have made it much easier to study protein interactions, especially when experimental data are limited. In recent years, there have been advances in deep-learning based protein prediction tools and more accessible, web-based servers. AlphaFold Multimer can predict structures and protein complexes from amino acid sequences^[34–36]. ClusPro and HADDOCK both model how proteins interact with binding energy scoring metrics and amino acid restraints^[37–45] to compare binding predictions to build a more comprehensive understanding of whether emerging antigens may interact with human receptors. These tools can be applied to predict protein interactions between different coronaviruses and potential receptors.

While these prediction tools offer an initial view of potential binding orientations, they fail to capture the intrinsic flexibility, intermolecular forces, and dynamics of proteins on any time scale. Docking treats structures as rigid, overlooking conformational changes, solvent effects, and dynamic interactions that influence binding stability. Molecular dynamics (MD) simulations can validate docking predictions by allowing complexes to evolve over time under physiological conditions.

MD helps assess the stability and plausibility of docked complexes, offering a more realistic understanding of protein-protein interactions and their intermolecular forces.

In this study, we aim to expand on current pandemic risk assessment efforts by testing how various betacoronavirus S1 proteins may bind human ACE2 using computational protein prediction methods. By focusing on the S1 region, we could bypass the difficulties associated with looking at full-length Spike across different lineages while still incorporating nuances that may not be captured with RBD alone, since its location may vary across viruses^[46,47]. We achieved this by compiling a representative set of betacoronaviruses across three subgenera, with some predicted and known to bind ACE2, and some that do not, and applying a computational pipeline to these representative sequences. We compared static protein-docking and protein folding platforms and found that ClusPro best recapitulated known ACE2 binding predictions. We also identified a hibecovirus, Zhejiang2013, as a candidate ACE2 binder that was previously unreported. We then prioritized this candidate for molecular dynamics (MD) simulations to validate the physical feasibility of the binding predictions. Through this systematic approach to binding predictions, we determined that Zhejiang2013 behaves similarly to known pandemic-causing coronaviruses. Overall, this work demonstrates that accessible computational tools can screen and prioritize emerging antigens for downstream testing as we shift from reactive to proactive surveillance of viral threats.

5.2 Results

5.2.1 Selection of Coronaviruses to Study

Our overall goal was to predict which emerging betacoronaviruses bind to human ACE2 using protein structure prediction, static protein docking, and sophisticated MD simulations (Figure 5-1A). Previous work identified a set of sarbecoviruses, merbecovirus, and a hibecovirus as

emerging coronaviruses with varying receptor usage predictions ^[29]. Using this previous work as a reference, we compiled the S1 and RBD sequences for 28 coronaviruses from NCBI (Table 5-1). Our selections were made to test known ACE2 binders (positive controls), known ACE2 non-binders (negative controls), and a sampling of coronaviruses for which ACE2 binding/non-binding had been predicted based on phylogeny but not tested experimentally (Figure 5-1B-C).

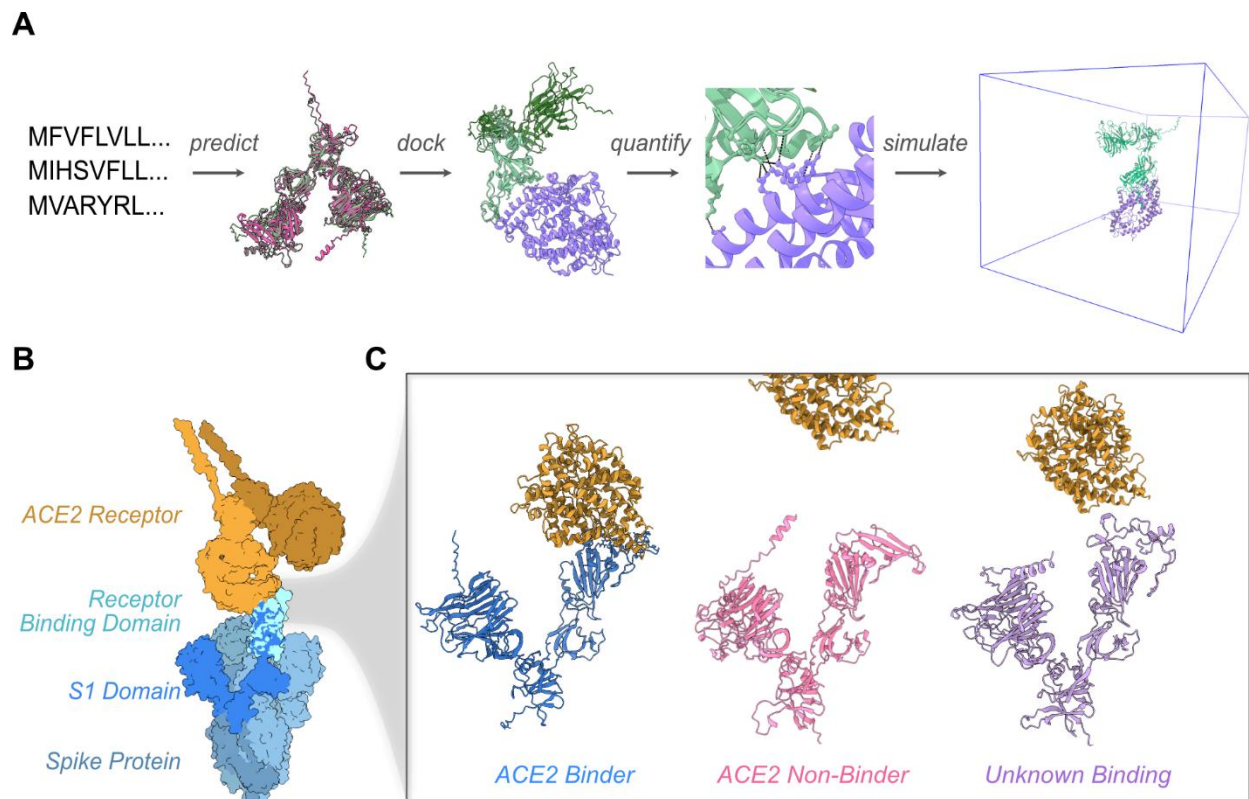


Figure 5-1. Overview of the study. A) Workflow of protein and interaction prediction to quantify binding. B) Example of Spike-ACE2 binding using SARS-CoV-2 Spike trimer and human ACE2 dimer using PDB structures 6VXX, 61MD, 6M0J, and 6VSB. C) Example of coronavirus sequences as ACE2 binders, non-binders, and unknowns tested in this pipeline. AlphaFold predictions, from left to right, of S1 domains from sarbecovirus binder (SARS-CoV-2), merbecovirus non-binder (MERS-CoV), and hibecovirus unknown binding (Zhejiang2013) subgenera.

Table 5-1. Compilation of coronaviruses used to test human ACE2 binding. We tested 24 sarbecoviruses, 3 merbecoviruses, and 1 hibecovirus. The NCBI accession number is listed for each Spike protein, along with whether it binds human ACE2. Pandemic-causing coronaviruses that bind to ACE2 are marked with * and non-binders noted with **.

Subgenus	Name	NCBI Accession	ACE2 Binder [28,29,31,32,48–50]
Sarbecovirus	SARS-CoV-1	YP_009825051.1	Yes*
	SARS-CoV-2	QHD43416.1	Yes*
	LYRa11	KF569996	Yes
	LYRa3	KF569997.1	Yes
	RaTG13	QHR63300.2	Yes
	P2V	QIQ54048.1	Yes
	Rs3367	AGZ48818.1	Yes
	Rs7327	ATO98231.1	Yes
	YN2018B	QDF43825.1	Yes
	Rm1	ABD75332.1	Yes
	Rs672	ACU31032.1	Yes
	HKU3	ADE34812.1	Yes
	SX2013C	AIA62300.1	No
	YNLF31C	AKZ19076.1	No
	YNLF34C	AKZ19087.1	No
	JTMC15	ANA96027.1	Yes
	BtKY72	APO40579.1	Yes
	Rs4255	ATO98193.1	Yes

	SC2018	QDF43815.1	No
	YN2018A	QDF43820.1	Yes
	YN2018C	QDF43830.1	Yes
	YN2018D	QDF43835.1	Yes
	Pang17	QIA48632.1	Yes
	W1V1	AGZ48831.1	Yes
Merbecovirus	MERS-CoV	NC_019843.3	No**
	HKU4	YP_001039953.1	No
	BtCoV133	ABG47052.1	No
Hibecovirus	Zhejiang2013	AIL94216.1	Unknown

5.2.2 Sequence Alignment

We first compared the primary protein sequences of these S1 and RBD sequences using multiple sequence alignment and pairwise sequence identity calculations. We annotated which sequences were predicted/known to be ACE2 binders or non-binders (Figure 5-2). ACE2 non-binding sarbecoviruses formed one sub-group. Sarbecovirus ACE2 binders formed another sub-group. The ACE2 non-binding hibecovirus and merbecoviruses formed an outgroup.

The RBD lies near the middle of the S1 domain (Figure 5-1B). This position remains relatively consistent across the set of viruses. Overall, we observe at least 60% sequence similarity in the S1 across all sarbecoviruses (Figure 5-2). Merbecoviruses and the hibecovirus show much lower similarity to sarbecoviruses and to each other. The merbecoviruses only have a 20% similarity to either subgenus and serve as negative controls for binding to ACE2 since MERS-CoV uses DPP4 as its receptor. The hibecovirus shares about 27% sequence similarity to sarbecoviruses that bind ACE2 (Table B-1). We also calculated pairwise sequence identities using the entire S1 sequence

across all the proteins. Trends observed with the S1 region were largely recapitulated by RBD sequence alignments (Figure 5-2 and Table B-2).

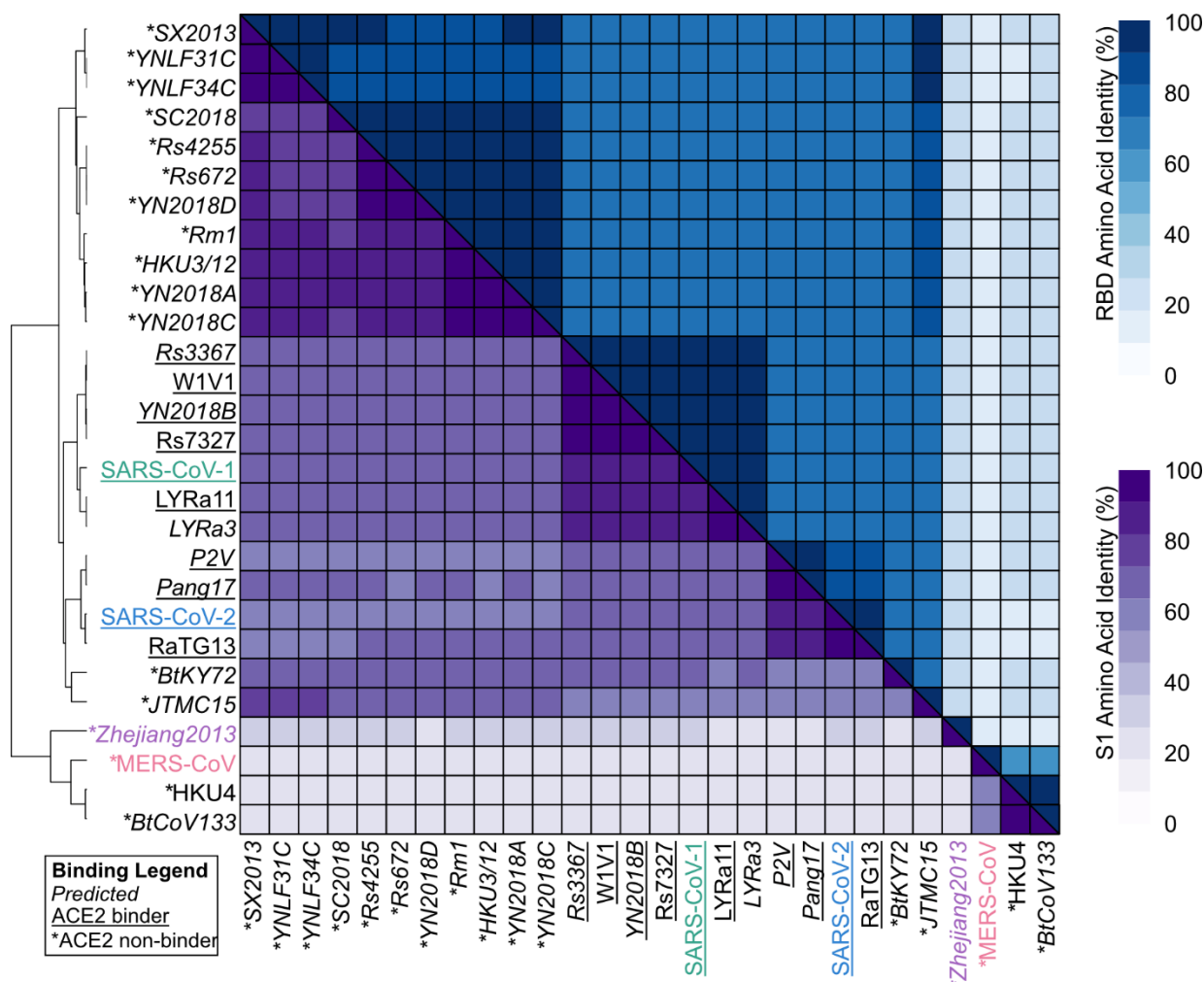


Figure 5-2. Sequence alignment heatmap of S1 (purple) and RBD (blue) using pairwise alignment of each betacoronavirus.

5.2.3 Protein Structure Prediction

We next predicted protein structures using AlphaFold2 (Figure 5-1A). Visual inspection of the predicted proteins showed that the S1 structures had a relatively conserved shape (Figure 5-3). Merbecoviruses generally have an additional beta sheet in their RBD compared to the

sarbecoviruses and the hibecovirus. Based on these predictions, the hibecovirus has more intrinsically disordered regions in the RBD.

Using these structures, we took a closer look at the structural similarity between the proteins. We used US-Align to obtain a template modeling (TM)-score to assess whether the trends observed with the sequence alignment are recapitulated with the protein structures. The TM-score is a length-independent metric that focuses on global alignment than local structural differences. A score of 1 indicates that two proteins are a perfect match where anything above 0.5 can indicate similarly folded proteins ^[51,52]. The trends observed for sequence similarity (Figure 5-2) are also evident with structure similarity (Figure 5-4). Sarbecoviruses share more similarity with each other than with hibecovirus and merbecoviruses.

Interestingly, structural alignments revealed interesting relationships between Zhejiang2013 and other viruses. Zhejiang2013 is a hibecovirus that is phylogenetically more similar to ACE2 non-binders and its predicted binding to ACE2 is unknown ^[29]. Zhejiang2013 S1 and RBD sequence alignments showed it was distinct from ACE2-binding sarbecoviruses (Figure 5-2). The mean sequence identity to sarbecovirus S1 and RBD was 22% and 27%, respectively. However, Zhejiang2013 structural alignments showed more similarity to ACE2-binding sarbecoviruses than expected compared to other ACE2 non-binding viruses with an average TM-score of 0.8 (Figure 5-4 and Table B-3). This structural similarity was more pronounced when looking at the RBD region specifically (Figure 5-4 and Table B-4).

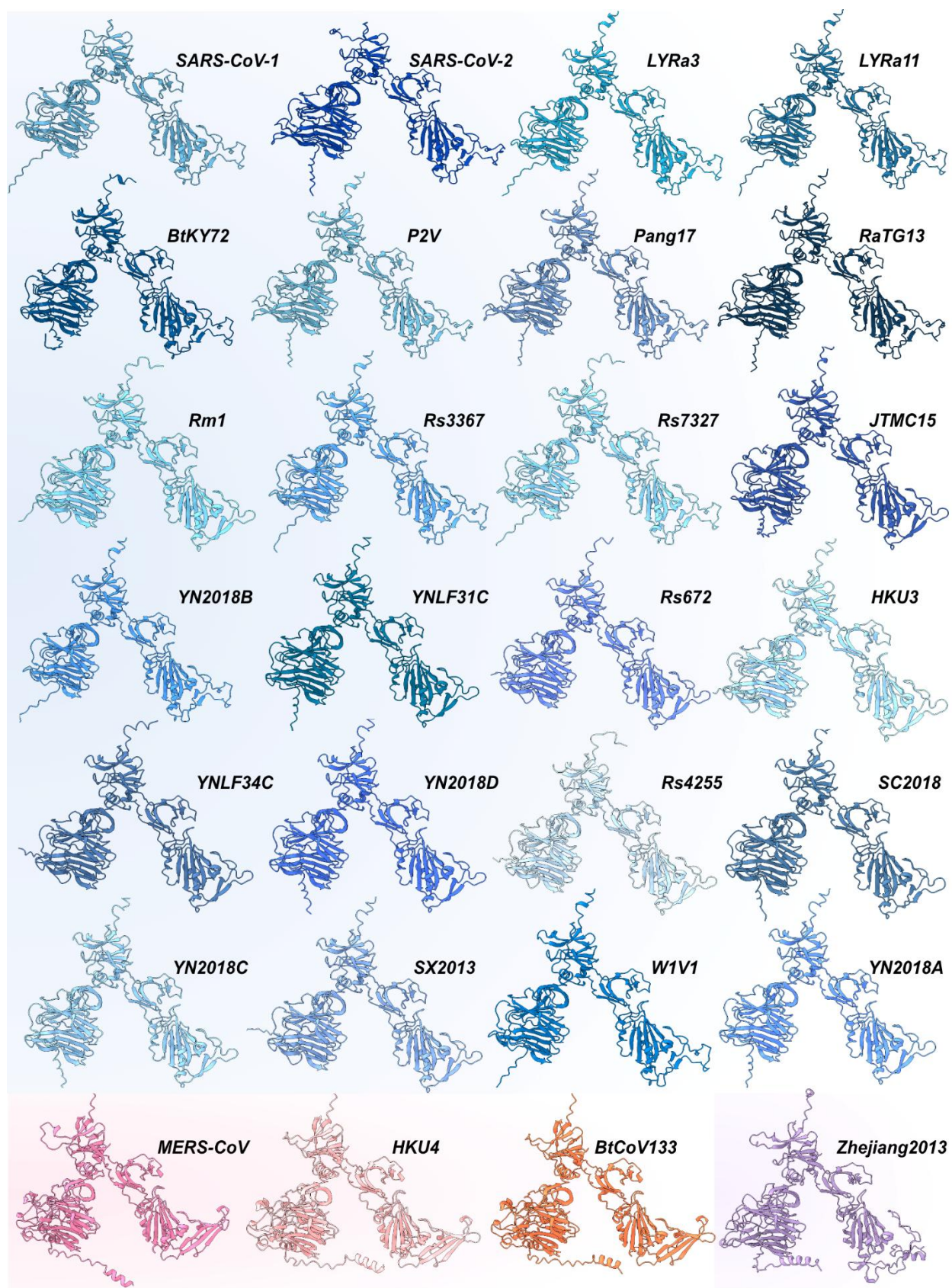


Figure 5-3. AlphaFold2 protein predictions of betacoronavirus S1 structures. Structures are oriented so that the RBD is towards the bottom right of the S1 domain. From left to right, the first 24 structures in blue are sarbecoviruses, the next three structures in pink are merbecoviruses, and the last structure in purple is a hibeovirus.

To achieve greater resolution of structural similarities and differences. We utilized the STRIDE web server, which implements a DSSP algorithm, to quantify alpha helices, beta strands, turns, and “other” (coils, 310-helices, and bridges). S1s were mostly comprised of beta strands (about 44%) (Figure B-1A). RBDs exhibited a markedly higher proportion of intrinsically disordered regions, falling in the “other” category with an average of about 39% (S1s averages around 23%) (Figure B-1B and Table B-5). Nevertheless, there was no obvious correlation between S1 or RBD secondary structure content and known ACE2 binding.

5.2.4 Predicting the S1-ACE2 Interaction

We then predicted how each S1 might bind to human ACE2. Spike binds to its host receptor at the RBD located in the S1 domain, so we assumed that this function is maintained across other S1s. However, we reasoned that using S1 to model this protein interaction may affect the total interaction energy of the interaction and how the algorithm calculates the prediction.

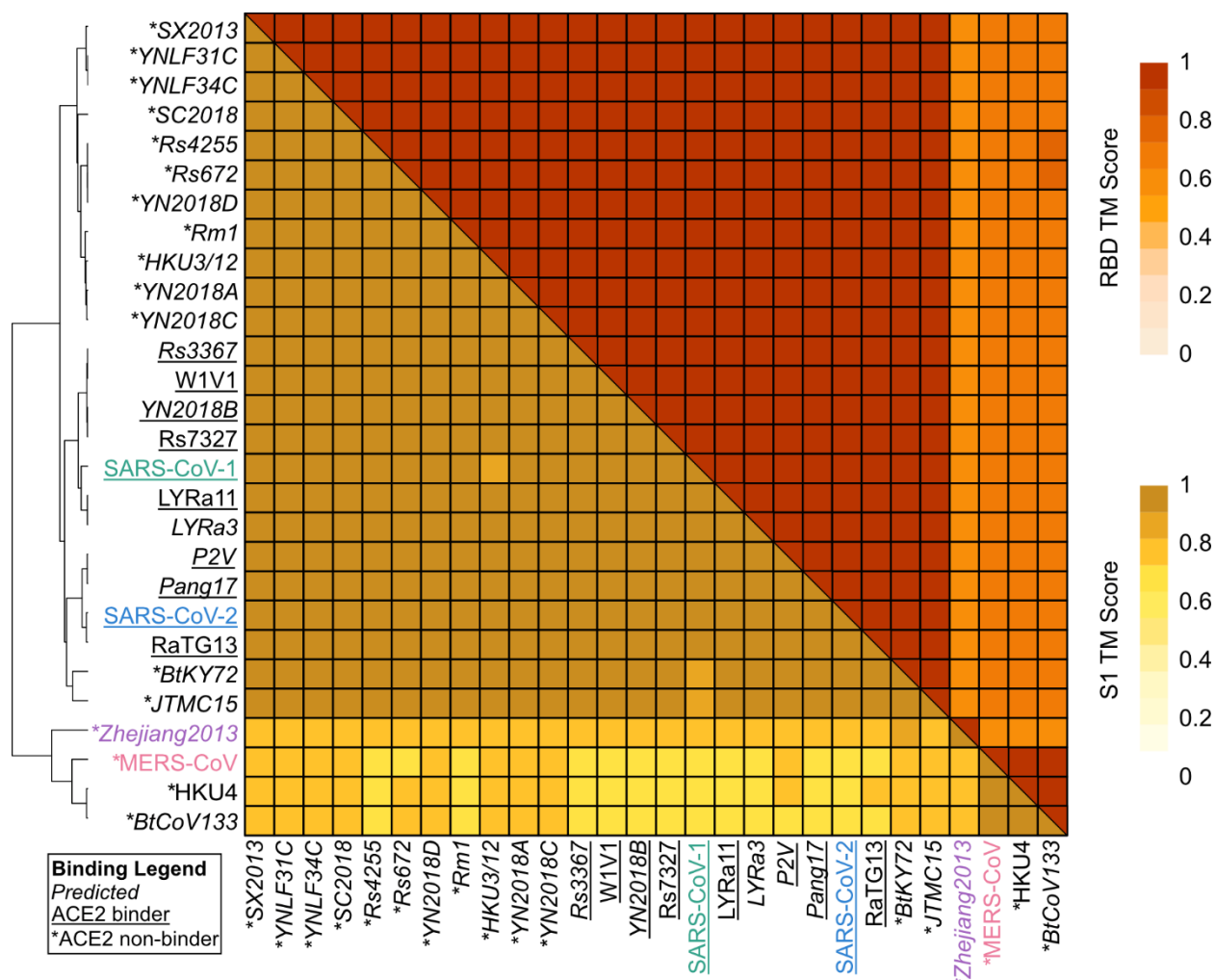


Figure 5-4. Structure analysis of emerging Spike S1 and RBD regions. Structure alignment heatmap of S1 and RBD shows the subgenera grouped together.

While there are many programs that can model protein-protein interactions, we used ClusPro, HADDOCK, and AlphaFold Multimer to predict the interactions due to their ease of use, requisite requirements, and web-based interface. Multimer simply requires the amino acid sequences, which can be helpful for predicting proteins and complexes when little structural information is available. ClusPro applies rigid-body docking using Fast Fourier Transform to evaluate thousands of orientations of the two proteins by scoring each orientation using van der Waals, electrostatics, and hydrophobicity [37–41]. It filters out the lowest energy complexes until it determines the most

physically plausible orientation. HADDOCK further integrates known information through restraints or binding interfaces to guide more biologically informed docking, serving as a semi-flexible docking platform^[44,45]. By focusing on free, web-based protein prediction and docking servers to compare binding predictions to build a more comprehensive understanding of whether emerging antigens may interact with human receptors.

The results of each program gave varying degrees of interaction clarity. Predictions without any guidance towards the binding interface often resulted in physically impossible complexes. To guide the orientation of proteins in docking algorithms, we used restraints, or single residue pairings, to orient the RBD in S1 to the binding interface of ACE2^[43] (Table B-6). As a structure prediction algorithm, AlphaFold Multimer does not support restraints. ClusPro makes restraints optional, and HADDOCK requires the information to simulate a protein interaction. For ClusPro and HADDOCK, we used a combination of three amino acids in the receptor binding motif of each S1 paired to two amino acids at the top of ACE2 (H34 and G354) to orient the proteins into a physically feasible complex. Each docking prediction yielded multiple complexes which were ranked by static energy calculations.

We analyzed the top-ranked model of each prediction, ranked based on minimal free energy. The lack of binding interface input for Multimer predictions meant that many of the predicted complexes were physically impossible either by the orientation of ACE2 or how the S1 was folded with ACE2 (Figure B-2). Like Multimer, unrestrained ClusPro predictions result in impossible complexes. The ClusPro predictions with minimal restraints left more positional flexibility and kept the proteins in more plausible configurations. HADDOCK requires a known binding interface for its docking, which can force how two proteins are predicted to interact.

To quantify protein binding, we calculated the total number of contacts between S1 and ACE2 for each platform (Figure 5-5). Contacts predict which amino acids may be interacting between molecules. Pandemic-causing viruses were used as controls to establish a binding affinity threshold. Known ACE2 binders served as positive controls, whereas known non-binders served as negative controls. Contacts were calculated in ChimeraX, defining contacts as interatomic distances of less than 0.4 Å. We categorized the viruses based on their predicted ability to bind ACE2 as a way to understand if we can infer binding based on the number of contacts between the S1 and ACE2^[29]. Given the structural similarity Zhejiang2013 has with known ACE2 binding sarbecoviruses, we categorized the hibeovirus as “N/A” to compare.

The different protein interaction prediction tools produced different ranges of contacts for ACE2 binders and non-binders (Table B-7). For ClusPro, the median number of contacts of predicted ACE2-binders was significantly greater than the median of predicted non-binders (Figure 5-5A). With HADDOCK, the median number of contacts was not significantly different, which may be due to the greater number of restraints placed on docking, including non-binders (Figure 5-5B). The number of contacts calculated for the Multimer predictions showed greater variability and spread (Figure 5-5C), likely driven by physically unrealistic predictions for many viruses (Figure B-2).

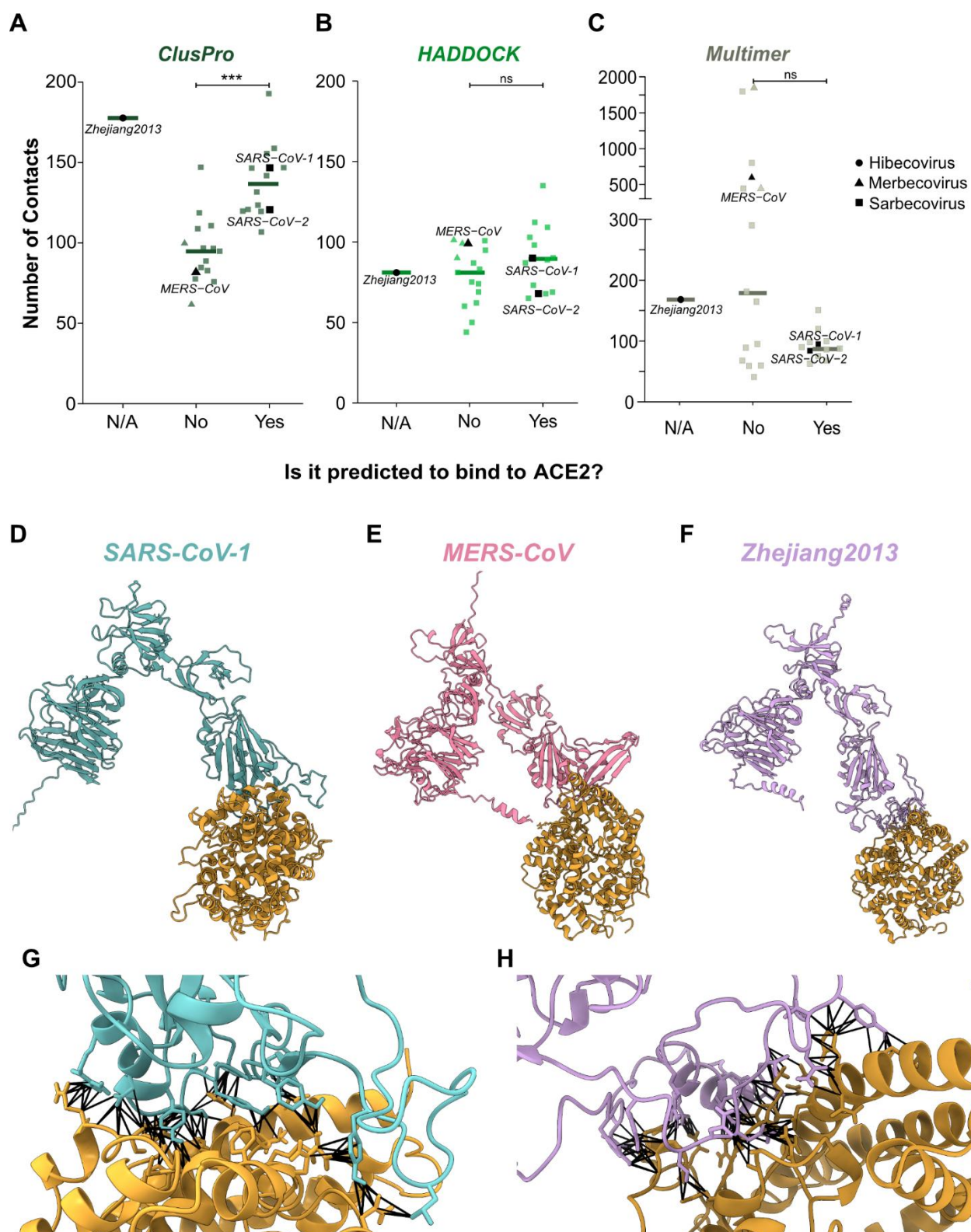


Figure 5-5. S1-ACE2 Docking Predictions. A-C) Number of contacts calculated between each S1 to ACE2 for A) ClusPro, B) HADDOCK, and C) AlphaFold Multimer. The bar indicates the

median number of contacts for each subset. Significance was calculated using a Mann-Whitney U test, *** $p < 0.001$. D-F) Docking predictions from ClusPro for D) SARS-CoV-1, E) MERS-CoV, and F) Zhejiang2013 S1 binding to ACE2. G-H) Close up look at the binding interface and the predicted contacts, highlighted in black, between each S1 to ACE2 as identified in ChimeraX for G) SARS-CoV-1 and H) Zhejiang2013 to ACE2.

We next analyzed protein interaction prediction for Zhejiang2013 more closely. This virus has not been studied so its ability to bind ACE2 is unknown, however its structure and sequence indicated it might be an ACE2 binder. To more closely analyze the human ACE2-binding potential of Zhejiang2013, we focused on the ClusPro predictions, which showed statistically significant difference between binders and the non-binders. The minimal restraints allowed for some flexibility while maintaining a physically feasible orientation (Figure 5-5D-F). Compared with controls and overall contact behavior, the number of contacts between Zhejiang2013 and human ACE2 is much greater than any other interaction (Figure 5-5G-H). In addition to the similarity between Zhejiang2013 and the sarbecoviruses at the RBD, this may indicate its potential to bind to human ACE2.

5.2.5 Simulating the S1-ACE2 Interaction Using Molecular Dynamics

To further test whether Zhejiang2013 S1 can bind human ACE2, we performed molecular dynamics (MD) simulations using coarse-grained methods. In contrast to static docking or protein folding approaches, MD simulations are computational algorithms that use Newton's laws of motion to inform system dynamics over time, providing detailed insights into binding affinity and stability. As a dynamic approach, MD provides an atomistic understanding of how binding occurs in a biological system^[53]. Coarse-graining provides a simplified representation of molecular systems in which four atoms are represented as one bead, preserving essential properties while significantly reducing the degree of freedom. This reduction enables the use of larger integration

timesteps and extended timescales^[54,55]. On average, the S1 proteins can have about 12,000 atoms. Coarse-graining reduces the model to about 3000 beads, which reduces the computational cost and time of the simulation. To keep the model simplified, we did not include the glycans that are known to normally decorate the Spike protein^[56,57].

The coarse-grained simulations were run for 1000 ns in triplicate. By approximately 350 ns of the simulation, the MERS-CoV S1 subunit began to separate from ACE2 and remained detached thereafter (Figure 5-6). In contrast, the SARS-CoV-1 S1 subunit maintained continuous contact with ACE2 throughout the entire simulation. Interestingly, the Zhejiang2013 S1 subunit also remained bound to ACE2 for the full 1000 ns, supporting our hypothesis based on the number of S1–ACE2 contacts observed. The root mean squared deviation (RMSD), which measures how much the complexes deviate from their initial conformations over time, shows that both SARS-CoV-1 and Zhejiang2013 had lower RMSDs, resulting from remaining in a complex like their original structure. In contrast, MERS-CoV and ACE2 had a higher RMSD due to unbinding throughout the simulation (Figure B-3).

While coarse-graining provided an initial indication of whether the S1 domains bind to human ACE2, we lose information about the electrostatic forces by reducing the atoms to beads. We used all-atom steered MD (pulling) simulations to better understand the binding forces between Zhejiang2013 S1 and human ACE2 in comparison to known ACE2 binder SARS-CoV-1. Pulling simulations model the force required to pull two proteins apart in a complex. We can use this simulation to calculate the energy required for the two proteins to bind (Figure 5-7).

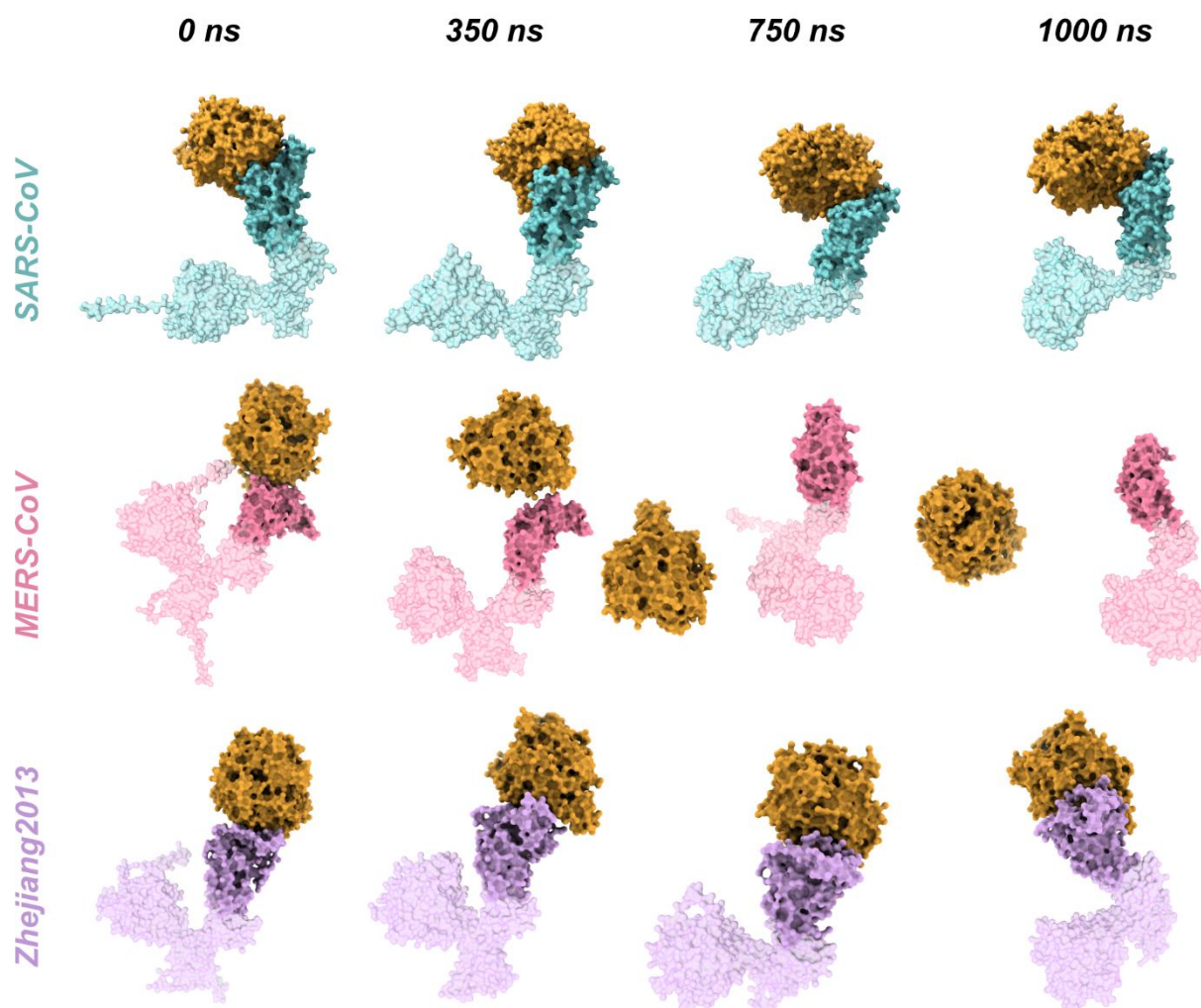


Figure 5-6. Coarse-graining montage across 1000 ns simulation of SARS-CoV-1, MERS-CoV, and Zhejiang2013 S1 binding to human ACE2. By the end of the simulation, SARS-CoV-1 (teal) and Zhejiang2013 (purple) remain attached to ACE2 (gold) at the RBD, whereas MERS-CoV (pink) is detached from ACE2. The RBD is emphasized as more opaque than the lower transparency S1.

For these pulling simulations, we also included glycans on the proteins to get a more accurate prediction of the binding mechanism. Glycan positioning is well-known and experimentally resolved for pandemic-causing coronaviruses, but unknown for the emerging coronaviruses

[49,53,56–59]. Because Zhejiang2013 is an unstudied virus, there is no experimental evidence available about its glycan positioning. We therefore predicted the glycans for both proteins using GLYCAM^[59–62]. We compared the GLYCAM-predicted glycosylation pattern for SARS-CoV-1 to the known glycosylated structure, which showed agreement in glycosylation sites and continued to predict glycans on Zhejiang2013 S1 (Figure B-4). We did not simulate MERS-CoV again because it does not normally bind ACE2 and the two proteins did not stay bound in the coarse-graining simulations (Figure 5-6). We used the pulling simulations to get better resolution of the binding energy between S1 and ACE2.

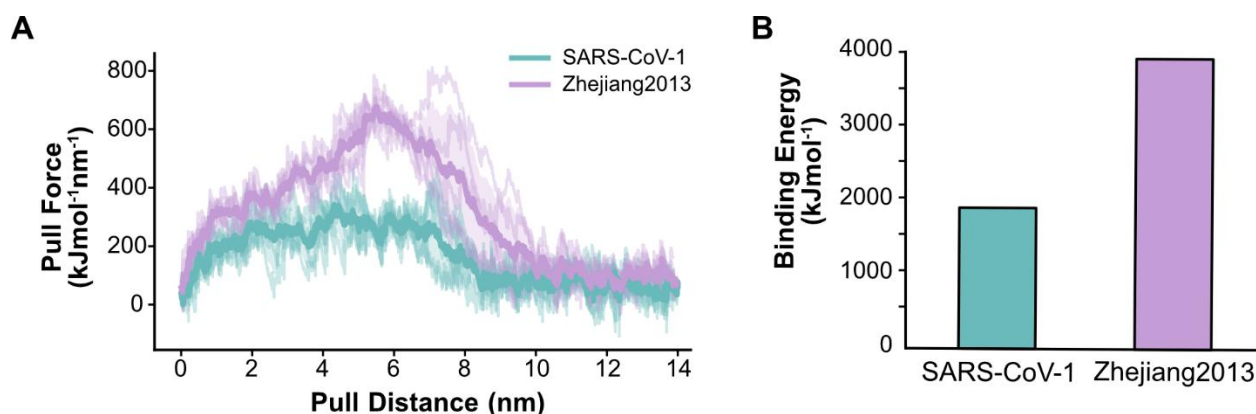


Figure 5-7. Pulling simulations reveal a stronger binding energy between Zhejiang2013 and human ACE2. A) Traces of pull force as a function of pull distance at a pulling rate of 1 nm/ns. We show five replicates per system and highlight the mean of the replicates and shade the standard deviation region. B) Jarzynski inequality calculation of pulling energy estimates (kJ/mol) for each system over five replicates.

By plotting the pull force as a function of pull distance between SARS-CoV-1 and Zhejiang2013 S1 versus ACE2, we observed that both complexes separated at around 8 nm and 10 nm, respectively, at a rate of 1 nm/ns (Figure 5-7A). The distances were calculated using the center of mass of both proteins and then normalized to highlight overall behaviors. Interestingly,

Zhejinag2013 achieved a higher peak force than SARS-CoV-1 meaning the complex requires more work to break the interaction (Figure B-5). This behavior persisted even when we pulled the proteins at higher pulling rates (Figure B-6).

We next calculated the binding energies from these pulling force trajectories. Pulling simulations are faster than traditional MD simulations because they do not wait for the system to reach equilibrium at every point along the reaction coordinate. This can result in an overestimation of the true binding free energy ^[63]. To improve accuracy, we applied the Jarzynski equality to multiple independent pulling simulations, combined the trajectories using an exponential average, and estimated the equilibrium free energy change. This exponential reduces the influence of high-work trajectories and places greater weight on lower-work trajectories that are closer to equilibrium behavior ^[64]. When we calculated the energy required for the interactions, Zhejiang2013 exhibited a greater binding energy than SARS-CoV-1 (Figure 5-7B).

5.3 Discussion

Here, we screened various emerging betacoronaviruses to assess their potential to bind ACE2 using protein docking platforms and molecular dynamics simulations. We started by aligning the sequences to see how they might cluster and identify similar binding behaviors (Figure 5-2). We used pandemic-causing viruses as controls to set a threshold in binding affinity—SARS-CoV-1 and SARS-CoV-2 serving as positive controls of ACE2 binding and MERS-CoV as a negative control. We then predicted 28 Spike S1 domains and aligned their structures to assess whether any conserved regions might mediate ACE2 binding (Figures 5-3 and 5-4). Using the predicted structures, we docked each S1 to ACE2 to predict potential interactions and quantified binding by counting the number of contacts between the proteins (Figure 5-5). The number of contacts between each S1 and ACE2 indicated that Zhejiang2013, a hibeovirus of unknown ACE2 binding

ability, has more interactions with ACE2 than controls. To validate the docking predictions, we used MD simulations to examine the intermolecular forces at play and calculated the binding energy between S1 and ACE2 (Figure 7). Taken together, our results suggest that Zhejiang2013 is indeed an ACE2 binder, despite its evolutionary relatedness to ACE2 non-binders^[29].

Focusing on the S1 domain offers a practical solution to overcome the difficulties of examining full-length Spike across different coronaviruses while maintaining a broader view that includes the RBD, especially for uninvestigated viruses. Studying the S1 region of Spike can open the possibility of identifying new binding modalities in cases where the RBD may be positioned slightly differently within the S1 structure^[46,47]. If both the RBD location and overall S1 structures are relatively conserved across a representative set of betacoronaviruses, then concentrating on the RBD could streamline and alleviate computational costs by simplifying modeling efforts. By studying the S1 domain and the RBD together, we can better understand how these structures shape interaction with the host receptor.

Evolutionary analyses suggest that Zhejiang2013 does not bind ACE2, though its ability to bind ACE2 has not been tested^[29]. Our computational modeling revealed unexpected similarities to SARS-CoV, a known ACE2 binder. The predicted S1 domain of Zhejiang2013 more closely resembled that of known ACE2 binders than non-binders. Both Zhejiang2013 and SARS-CoV share a disordered region, and an α -helix at the binding interface, which contrasts the β -sheets in MERS-CoV at the same interface. These similarities indicated that Zhejiang2013 may retain structural elements necessary for ACE2 recognition, even if its lineage is not usually associated with ACE2 binding. These results emphasized that small structural differences within the S1 domain can significantly influence receptor engagement.

This work also highlights the utility of easy-to-use docking platforms. Docking platforms are valuable for screening emerging antigens and generating a reproducible screening pipeline. They are fast, computationally efficient, and require minimal prior information—an advantage for poorly characterized proteins with limited experimental information. AlphaFold Multimer only requires the protein sequence to predict a protein complex, but the lack of energy calculations to guide the prediction resulted in biologically improbable orientations. Conversely, HADDOCK provides a flexible docking approach for predicted structures but requires the positions of a known binding interface. Its predictions are more guided and thus more accurate for true interactions, but the requirement of a known binding motif makes it difficult to use for largely uninvestigated proteins. This could also “force” an interaction through the required restraints when none should exist. ClusPro provided a Goldilocks solution by requiring only the predicted structure with some options for limited restraints. The capability of unrestrained docking, optimized by energy minimization, becomes a useful alternative for unknown proteins. ClusPro also provides many options to guide the docking if more information is available about the two proteins of interest. We found that using residue-based restraints afforded the predictions more flexibility and clustered a variety of possible orientations (Figure 5-5 and Table B-6). This meant that we could discern statistically significant differences between predictions across the emerging antigens. Overall, protein docking platforms afford researchers quick methods to test protein binding. However, docking platforms generally treat proteins as rigid structures and ignore the temporal dynamics of protein-protein interactions, which oversimplifies the model.

The initial protein docking simulations should be interpreted as a preliminary attempt at understanding a protein-protein interaction. The simplified structures and the lack of temporal binding dynamics mean that docking captures only a snapshot of the interaction. For well-known proteins like SARS-CoV-1 S1, we acknowledge that a glycan-free, static prediction oversimplifies the interaction, and we lose nuance on an already truncated model using S1 instead of the Spike

trimer. However, in the case of an emerging antigen, there is very little experimental or structural data to draw on to inform a predictive simulation. Static docking evens the playing field by not requiring additional information and quickly calculating the most thermodynamically feasible complex between two proteins.

Our pulling simulations revealed that the binding energy between Zhejiang2013 S1 and ACE2 was nearly twice that between SARS-CoV-1 S1 and ACE2, thereby requiring significantly more work to disrupt the interaction (Figure 5-7B and Figure B-4). This suggests that Zhejiang2013 may bind ACE2 strongly. In the context of pandemic screening, the ability to bind a known host cell receptor indicates that it possesses a critical prerequisite of zoonotic spillover. Zhejiang2013 exhibits a binding profile much greater than SARS-CoV-1 suggesting that its RBD may contain the machinery for human receptor usage. This mirrors similar trends already measured with SARS-CoV-2 having a greater binding affinity to ACE2 than SARS-CoV-1^[16,46,65–67].

Post-translational modifications (PTMs) play critical roles in protein folding, binding and functions. Viral fusion proteins (i.e. class 1 viral fusion proteins like coronavirus Spike, HIV Env, and influenza HA) rely on PTMs like glycans for shielding important antigen epitopes from host immune defenses, influencing target binding, increasing binding affinity, and stabilizing conformations^[57,61,68–72]. More specifically for pandemic-causing coronaviruses, glycosylation at the RBD directly affects binding strength to the human receptor^[49,56,57,59,61,73]. However, structural data and residue positioning of PTMs can be inconsistent, especially for less-studied proteins and viruses. To initially streamline our structure and docking predictions, we did not consider any glycans to the S1 domain or RBD. While this omission was necessary for protein docking and computational efficiency, we acknowledge that the glycan-free models do not accurately reflect binding dynamics. However, as we continued MD simulations, we predicted the positions of

glycans across the protein using GLYCAM to more accurately measure the intermolecular binding forces (Figure B-4). While these MD results provide a more accurate representation of binding forces, PTM positioning can be inconsistent. Additionally, the lack of experimental data to support the glycan positioning on Zhejiang2013 means that these results should not be taken as fact. Further work is necessary to experimentally resolve the Zhejiang2013 binding interface in the presence of relevant PTMs and to validate its binding to human ACE2. Furthermore, the simulation data strongly suggest that Zhejiang2013 can bind ACE2 and attach to the host cell, but high binding energy does not guarantee successful viral entry or replication. This is simply the first step that must be satisfied for viral emergence and highlights the potential risk posed by Zhejiang2013.

Ultimately, this work focuses solely on computational methods for predicting binding between an emerging antigen and a human receptor. This pipeline demonstrates that computational screening of emerging coronaviruses can help to prioritize high-risk candidates for further experimental investigation. It combines different levels of computational rigor to screen and calculate potential binding to a human receptor. While this workflow does simplify the protein interaction dynamics, it does allow for screening uninvestigated proteins and emerging viruses for which there is no prior information available. By flagging Zhejiang2013 as a potential ACE2-binder, we can see how structural predictions can bridge the gap between screening a library of antigens for experimental validation and accelerating pandemic preparedness efforts with minimal resources.

5.4 Conclusions

We predicted the structure of 28 emerging coronaviruses S1 proteins to dock human ACE2 to screen for pandemic-potential. We quantified the predicted protein interactions using contact analysis that highlighted the potential binding of the hibeovirus, Zhejiang2013, not known to bind

ACE2. When modeled with MD simulations, Zhejiang2013 has greater binding energy compared to SARS-CoV-1, suggesting its potential to bind ACE2. Further work is needed to experimentally measure the binding of the antigen to the human receptor.

5.5 Methods

5.5.1 Protein Structure Predictions

Coronavirus Spike S1 domain structures were predicted using AlphaFold2 on the UC Davis College of Biological Sciences High-Performance Computer (HPC) to increase the throughput of predictions. This computational tool requires the amino acid sequence as input and generates the corresponding protein structure ^[34]. To prevent biased decisions and ensure reproducibility, we continued our analysis with the best-ranked model, a choice we consistently maintained throughout the computational analysis.

The ACE2 structure was based on the crystal structure of ACE2 bound to the RBD of SARS-CoV-2 (RBD 6M0J). The RBD structure was removed, and sequences of the ACE2 structure were used to predict a new structure using AlphaFold to apply the same treatment across all the structure files for the docking predictions.

5.5.2 Protein Interaction Predictions and Restraints

5.5.2.1 AlphaFold3-Multimer

To predict an S1-ACE2 interaction, we input the amino acid sequence of the S1 and ACE2's sequence into the AlphaFold3 web server as separate molecules.

5.5.2.2 ClusPro Docking

ClusPro docking requires a .pdb protein structure for the two interacting molecules. S1 was the designated receptor with chain A, and ACE2 was the designated ligand with chain L. With just the structures, this is an unrestrained prediction. Designing restraints for a complex can guide the orientation of the two interacting proteins. The restraint file was produced using the restraint generator within ClusPro, which makes a .json file. We used the highest-ranked complex for further analysis.

5.5.2.3 HADDOCK

HADDOCK requires both protein structures files and interacting residues to guide the prediction of the protein interaction. For each prediction, we uploaded each .pdb file and specified each molecule as a protein and specified the chain label—A for the S1 file and L for ACE2. We specified that the active residues of S1 were the protein's RBD and the residues 26-48 and 318-358 to guide the correct orientation of the proteins.

5.5.2.4 Docking Restraints

These restraints were based on structural and sequence alignments between each S1 and SARS-CoV S1, using known key binding residues. Overall, the sarbecovirus S1 maintained the same residues between the structure and sequence alignment. However, the merbecoviruses and the hibeovirus had amino acid residues when aligned to SARS-CoV S1 by its secondary structure or its protein sequence. For every program we used, we selected the highest-ranked model for subsequent analysis.

5.5.2.5 Contact Analysis

Contacts were calculated using UCSF ChimeraX. Contacts are defined as interatomic distances of less than 4 Å between chains.

5.5.3 Sequence Alignments

The amino acid sequences of each S1 or RBD were aligned using ClustalOmega.

5.5.4 Structure Alignments

The PDB files from the AlphaFold structure predictions were uploaded to the USAlign tool on the on UC Davis College of Biological Science HPC to increase throughput of comparisons.

5.5.5 Structure Composition Analysis

To analyze differences in secondary structure composition among emerging coronaviruses, we used STRIDE, a web server for secondary structure assignments. STRIDE is based on the dictionary of secondary structure of proteins (DSSP) algorithm, which determines secondary structure by identifying hydrogen-bonding patterns and backbone geometry from atomic coordinates. This tool takes as input a PDB file containing the predicted protein structure and outputs the secondary structure type assigned to each residue. The program generated an Excel file listing the proportions of α -helices, β -strands, turns, coils, 3_{10} -helices, and bridges, which we manually consolidated into the into four main classes: α -helices, β -strands, turns, and other. The same procedure was applied to the RBDs.

5.5.6 Coarse-Grained MD Simulations

We constructed four coarse-grained (CG) complexes of the S1 subunit ACE2 for SARS-CoV-1, MERS-CoV, and Zhejiang2013 on the Expanse at San Diego Supercomputer Center. The transmembrane domain of ACE2 was excluded from all systems. We adjusted the pH of each complex to 7.4 and used the MARTINI3 force field to coarse-grain the proteins^[74,75]. An elastic network was implemented to maintain the secondary and tertiary structures, with parameters set to -ef 700.0, -el 0.5, and -eu 0.9. The coarse-grained systems were then solvated in a cubic water

box with at least 5 nm between the protein and the box edges, using 0.15 M NaCl. Energy minimizations were performed until the convergence criteria were met.

Energy minimization was then performed to remove steric clashes and relax the system. The minimization was carried out for 50,000 steps with an energy step size of 0.001 ps and a convergence criterion of 100.0 kJ/mol/nm (emtol), meaning that the maximum force on any atom was kept below 100.0 kJ/mol/nm. To verify the success of minimization, potential energy was plotted using Python, showing a steady decrease followed by a plateau.

Following minimization, equilibration at constant pressure was performed using a molecular dynamics integrator with a timestep (dt) of 0.02 ps and $nsteps = 2,500,000$. During this stage, the system was gradually brought to the target temperature of 310 K using the V-rescale thermostat. The equilibration process was evaluated by monitoring the time evolution of density, kinetic energy, potential energy, pressure, and temperature

For each S1–ACE2 complex, the production simulation was performed using a timestep (dt) of 0.02 ps and $nsteps = 250,000,000$, corresponding to an aggregate simulation time of 5 μ s. For each complex, three independent replicates were conducted to ensure reproducibility. Trajectory data were recorded for subsequent analysis of the system's structural and dynamical behavior. RMSD and RMSF were plotted for each simulation run.

5.5.7 Steered MD Pulling Simulations

Man5 glycans were added to the S1–ACE2 complex using the GLYCAM web interface. The ClusPro-generated docked S1–ACE2 complex was used as input for glycan attachment. The GLYCAM output for SARS-CoV-1 was compared to the crystalized structure for validation. ACPYPE was used to convert the GLYCAM output files into GROMACS-compatible formats.

The S1–ACE2 complex was placed in a simulation box and solvated using the TIP3P water model and 0.15 M NaCl, while maintaining a minimum distance of 3 nm between the solute and the box edges, resulting in a box dimension of 20 x 20 x 40 nm. After the box was generated, the complexes were rotated so that the pulling direction was along the x-axis. Then we performed a steep minimization step of 50,000 steps with an energy step size of 0.01 ps and a convergence criterion of 1000.0 kJ/mol/nm (emtol). This was followed by an NPT equilibration step using a molecular dynamics integrator with a timestep (dt) of 2 fs and nsteps = 50,000.

Pulling simulations were performed under NPT conditions using 2 fs timesteps, a Nose-Hoover thermostat at 310K, and a Parrinello-Rahman barostat at 1 atm. In all runs, ACE2 was restrained while S1 was pulled with a pulling rate of 1000 kJ/mol/nm². A total of 45 pulling simulations were performed at three different pulling rates (1, 5, and 10 nm/ns) over 14 nm. Simulations were repeated five times and averaged.

5.5.8 Statistical Analysis

Statistical analysis was done using a two-way Mann-Whitney U test.

5.6 References

1. Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, 15(8), Article 8. <https://doi.org/10.1038/nrmicro.2017.45>
2. Zhu, Z., Lian, X., Su, X., Wu, W., Marraro, G. A., & Zeng, Y. (2020). From SARS and MERS to COVID-19: A brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respiratory Research*, 21(1), 224. <https://doi.org/10.1186/s12931-020-01479-w>
3. Anthony, S. J., Johnson, C. K., Greig, D. J., Kramer, S., Che, X., Wells, H., Hicks, A. L., Joly, D. O., Wolfe, N. D., Daszak, P., Karesh, W., Lipkin, W. I., Morse, S. S., PREDICT Consortium, Mazet, J. A. K., & Goldstein, T. (2017). Global patterns in coronavirus diversity. *Virus Evolution*, 3(1), vex012. <https://doi.org/10.1093/ve/vex012>
4. Cherry, James. D. (2004). The chronology of the 2002–2003 SARS mini pandemic. *Paediatric Respiratory Reviews*, 5(4), 262–269. <https://doi.org/10.1016/j.prrv.2004.07.009>
5. Cherry, J. D., & Krogstad, P. (2004). SARS: The First Pandemic of the 21st Century. *Pediatric Research*, 56(1), 1–5. <https://doi.org/10.1203/01.PDR.0000129184.87042.FC>
6. de Groot, R. J., Baker, S. C., Baric, R. S., Brown, C. S., Drosten, C., Enjuanes, L., Fouchier, R. A. M., Galiano, M., Gorbalenya, A. E., Memish, Z. A., Perlman, S., Poon, L. L. M., Snijder, E. J., Stephens, G. M., Woo, P. C. Y., Zaki, A. M., Zambon, M., & Ziebuhr, J. (2013). Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. *Journal of Virology*, 87(14), 7790–7792. <https://doi.org/10.1128/JVI.01244-13>
7. Meganck, R. M., & Baric, R. S. (2021). Developing therapeutic approaches for twenty-first-century emerging infectious viral diseases. *Nature Medicine*, 27(3), Article 3. <https://doi.org/10.1038/s41591-021-01282-0>

8. *COVID-19 and Your Health*. (2020, February 11). Centers for Disease Control and Prevention.
<https://www.cdc.gov/coronavirus/2019-ncov/your-health/end-of-phe.html>
9. Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, 19(3), 141–154. <https://doi.org/10.1038/s41579-020-00459-7>
10. Ludwig, S., & Zarbock, A. (2020). Coronaviruses and SARS-CoV-2: A Brief Overview. *Anesthesia and Analgesia*, 10.1213/ANE.0000000000004845. <https://doi.org/10.1213/ANE.0000000000004845>
11. Graham, R. L., & Baric, R. S. (2010). Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. *Journal of Virology*, 84(7), 3134–3146. <https://doi.org/10.1128/jvi.01394-09>
12. Bradley, C. C., Wang, C., Gordon, A. J. E., Wen, A. X., Luna, P. N., Cooke, M. B., Kohn, B. F., Kennedy, S. R., Avadhanula, V., Piedra, P. A., Lichtarge, O., Shaw, C. A., Ronca, S. E., & Herman, C. (2024). Targeted Accurate RNA Consensus sequencing (tARC-seq) reveals mechanisms of replication error affecting SARS-CoV-2 divergence. *Nature Microbiology*, 9(5), 1382–1392. <https://doi.org/10.1038/s41564-024-01655-4>
13. Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., & Ng, W.-L. (2020). Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Molecular Cell*, 79(5), 710–727. <https://doi.org/10.1016/j.molcel.2020.07.027>
14. Ruiz-Aravena, M., McKee, C., Gamble, A., Lunn, T., Morris, A., Snedden, C. E., Yinda, C. K., Port, J. R., Buchholz, D. W., Yeo, Y. Y., Faust, C., Jax, E., Dee, L., Jones, D. N., Kessler, M. K., Falvo, C., Crowley, D., Bharti, N., Brook, C. E., ... Plowright, R. K. (2022). Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology*, 20(5), Article 5. <https://doi.org/10.1038/s41579-021-00652-2>
15. Menachery, V. D., Graham, R. L., & Baric, R. S. (2017). Jumping species—A mechanism for coronavirus persistence and survival. *Current Opinion in Virology*, 23, 1–7. <https://doi.org/10.1016/j.coviro.2017.01.002>

16. Shirbhate, E., Pandey, J., Patel, V. K., Kamal, M., Jawaid, T., Gorain, B., Kesharwani, P., & Rajak, H. (2021). Understanding the role of ACE-2 receptor in pathogenesis of COVID-19 disease: A potential approach for therapeutic intervention. *Pharmacological Reports: PR*, 73(6), 1539–1550. <https://doi.org/10.1007/s43440-021-00303-6>
17. Mesel-Lemoine, M., Millet, J., Vidalain, P.-O., Law, H., Vabret, A., Lorin, V., Escriou, N., Albert, M. L., Nal, B., & Tangy, F. (2012). A Human Coronavirus Responsible for the Common Cold Massively Kills Dendritic Cells but Not Monocytes. *Journal of Virology*, 86(14), 7577–7587. <https://doi.org/10.1128/JVI.00269-12>
18. Li, Z., Tomlinson, A. C., Wong, A. H., Zhou, D., Desforges, M., Talbot, P. J., Benlekbi, S., Rubinstein, J. L., & Rini, J. M. (n.d.). The human coronavirus HCoV-229E S-protein structure and receptor binding. *eLife*, 8, e51230. <https://doi.org/10.7554/eLife.51230>
19. Hofmann, H., Pyrc, K., van der Hoek, L., Geier, M., Berkhout, B., & Pöhlmann, S. (2005). Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proceedings of the National Academy of Sciences*, 102(22), 7988–7993. <https://doi.org/10.1073/pnas.0409465102>
20. Wu, K., Li, W., Peng, G., & Li, F. (2009). Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proceedings of the National Academy of Sciences*, 106(47), 19970–19974. <https://doi.org/10.1073/pnas.0908837106>
21. Carrasco-Hernandez, R., Jácome, R., López Vidal, Y., & Ponce de León, S. (2017). Are RNA Viruses Candidate Agents for the Next Global Pandemic? A Review. *ILAR Journal*, 58(3), 343–358. <https://doi.org/10.1093/ilar/ilx026>
22. Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), Article 3. <https://doi.org/10.1038/s41579-018-0118-9>
23. Woo, P. C. Y., Lau, S. K. P., Yip, C. C. Y., Huang, Y., & Yuen, K.-Y. (2009). More and More Coronaviruses: Human Coronavirus HKU1. *Viruses*, 1(1), Article 1. <https://doi.org/10.3390/v1010057>

24. Hulswit, R. J. G., Lang, Y., Bakkers, M. J. G., Li, W., Li, Z., Schouten, A., Ophorst, B., van Kuppeveld, F. J. M., Boons, G.-J., Bosch, B.-J., Huizinga, E. G., & de Groot, R. J. (2019). Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proceedings of the National Academy of Sciences*, 116(7), 2681–2690. <https://doi.org/10.1073/pnas.1809667116>
25. Lang, Y., Li, W., Li, Z., Koerhuis, D., van den Burg, A. C. S., Rozemuller, E., Bosch, B.-J., van Kuppeveld, F. J. M., Boons, G.-J., Huizinga, E. G., van der Schaar, H. M., & de Groot, R. J. (2020). Coronavirus hemagglutinin-esterase and spike proteins coevolve for functional balance and optimal virion avidity. *Proceedings of the National Academy of Sciences*, 117(41), 25759–25770. <https://doi.org/10.1073/pnas.2006299117>
26. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2), 271-280.e8. <https://doi.org/10.1016/j.cell.2020.02.052>
27. Guo, H., Li, A., Dong, T.-Y., Su, J., Yao, Y.-L., Zhu, Y., Shi, Z.-L., & Letko, M. (2022). ACE2-Independent Bat Sarbecovirus Entry and Replication in Human and Bat Cells. *mBio*, 13(6), e02566-22. <https://doi.org/10.1128/mbio.02566-22>
28. Starr, T. N., Zepeda, S. K., Walls, A. C., Greaney, A. J., Alkhovsky, S., Veisler, D., & Bloom, J. D. (2022). ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature*, 603(7903), Article 7903. <https://doi.org/10.1038/s41586-022-04464-z>
29. Wells, H. L., Letko, M., Lasso, G., Ssebide, B., Nziza, J., Byarugaba, D. K., Navarrete-Macias, I., Liang, E., Cranfield, M., Han, B. A., Tingley, M. W., Diuk-Wasser, M., Goldstein, T., Johnson, C. K., Mazet, J. a. K., Chandran, K., Munster, V. J., Gilardi, K., & Anthony, S. J. (2021). The evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. *Virus Evolution*, 7(1), veab007. <https://doi.org/10.1093/ve/veab007>

30. Guo, H., Li, A., Dong, T.-Y., Si, H.-R., Hu, B., Li, B., Zhu, Y., Shi, Z.-L., & Letko, M. (2023). Isolation of ACE2-dependent and -independent sarbecoviruses from Chinese horseshoe bats. *Journal of Virology*, 97(9), e00395-23. <https://doi.org/10.1128/jvi.00395-23>
31. Letko, M., Marzi, A., & Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*, 5(4), Article 4. <https://doi.org/10.1038/s41564-020-0688-y>
32. Su, C., He, J., Wang, L., Hu, Y., Cao, J., Bai, B., Qi, J., Gao, G. F., Yang, M., & Wang, Q. (2024). Structural characteristics of BtKY72 RBD bound to bat ACE2 reveal multiple key residues affecting ACE2 usage of sarbecoviruses. *mBio*, 15(9), e01404-24. <https://doi.org/10.1128/mbio.01404-24>
33. Starr, T. N., Greaney, A. J., Stewart, C. M., Walls, A. C., Hannon, W. W., Veessler, D., & Bloom, J. D. (2022). Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. *PLOS Pathogens*, 18(11), e1010951. <https://doi.org/10.1371/journal.ppat.1010951>
34. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), Article 7873. <https://doi.org/10.1038/s41586-021-03819-2>
35. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493–500. <https://doi.org/10.1038/s41586-024-07487-w>
36. Omid, A., Møller, M. H., Malhis, N., Bui, J. M., & Gsponer, J. (2024). AlphaFold-Multimer accurately captures interactions and dynamics of intrinsically disordered protein regions.

- Proceedings of the National Academy of Sciences*, 121(44), e2406407121.
<https://doi.org/10.1073/pnas.2406407121>
37. Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Vajda, S. (2013). How Good is Automated Protein Docking? *Proteins*, 81(12), 2159–2166.
<https://doi.org/10.1002/prot.24403>
38. Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols*, 12(2), 255–278. <https://doi.org/10.1038/nprot.2016.169>
39. Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Kozakov, D. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins*, 85(3), 435–444. <https://doi.org/10.1002/prot.25219>
40. Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., & Vajda, S. (2020). Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure*, 28(9), 1071-1081.e3. <https://doi.org/10.1016/j.str.2020.06.006>
41. Jones, G., Jindal, A., Ghani, U., Kotelnikov, S., Egbert, M., Hashemi, N., Vajda, S., Padhorny, D., & Kozakov, D. (2022). Elucidation of protein function using computational docking and hotspot analysis by ClusPro and FTMap. *Acta Crystallographica. Section D, Structural Biology*, 78(Pt 6), 690–697. <https://doi.org/10.1107/S2059798322002741>
42. Ashizawa, R., Kotelnikov, S., Khan, O., Li, S. X., Glukhov, E., Cao, X., Lazou, M., Bekar-Cesaretli, A., Hailegeorgis, D., Averkava, V., Zhu, Y., Jones, G., Yu, H., Kalitin, D., Stepanenko, D., Koirala, K., Patsahan, T., Beglov, D., Lukin, M., ... Kozakov, D. (2026). Modeling Protein-Protein and Protein-Ligand Interactions by the ClusPro Team in CASP16. *Proteins*, 94(1), 183–191. <https://doi.org/10.1002/prot.70066>
43. Xia, B., Vajda, S., & Kozakov, D. (2016). Accounting for pairwise distance restraints in FFT-based protein–protein docking. *Bioinformatics*, 32(21), 3342–3344. <https://doi.org/10.1093/bioinformatics/btw306>

44. Honorato, R. V., Koukos, P. I., Jiménez-García, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., Rosato, A., & Bonvin, A. M. J. J. (2021). Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Frontiers in Molecular Biosciences*, 8. <https://doi.org/10.3389/fmolb.2021.729513>
45. Honorato, R. V., Trellet, M. E., Jiménez-García, B., Schaarschmidt, J. J., Giulini, M., Reys, V., Koukos, P. I., Rodrigues, J. P. G. L. M., Karaca, E., van Zundert, G. C. P., Roel-Touris, J., van Noort, C. W., Jandová, Z., Melquiond, A. S. J., & Bonvin, A. M. J. J. (2024). The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nature Protocols*, 19(11), 3219–3241. <https://doi.org/10.1038/s41596-024-01011-0>
46. Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veersler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2), 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
47. Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Boons, G.-J., Bosch, B.-J., Rey, F. A., de Groot, R. J., & Veersler, D. (2019). Structural basis for human coronavirus attachment to sialic acid receptors. *Nature Structural & Molecular Biology*, 26(6), 481–489. <https://doi.org/10.1038/s41594-019-0233-y>
48. Gonzalez-Isunza, G., Jawaid, M. Z., Liu, P., Cox, D. L., Vazquez, M., & Arsuaga, J. (2023). Using machine learning to detect coronaviruses potentially infectious to humans. *Scientific Reports*, 13(1), Article 1. <https://doi.org/10.1038/s41598-023-35861-7>
49. Allen, J. D., Ivory, D. P., Song, S. G., He, W., Capozzola, T., Yong, P., Burton, D. R., Andrabi, R., & Crispin, M. (2023). The diversity of the glycan shield of sarbecoviruses related to SARS-CoV-2. *Cell Reports*, 42(4), 112307. <https://doi.org/10.1016/j.celrep.2023.112307>
50. Demogines, A., Farzan, M., & Sawyer, S. L. (2012). Evidence for ACE2-Utilizing Coronaviruses (CoVs) Related to Severe Acute Respiratory Syndrome CoV in Bats. *Journal of Virology*, 86(11), 6350–6353. <https://doi.org/10.1128/JVI.00311-12>

51. Zhang, C., Freddolino, L., & Zhang, Y. (2026). A graphic and command line protocol for quick and accurate comparisons of protein and nucleic acid structures with US-align. *Nature Protocols*, 21(2), 517–541. <https://doi.org/10.1038/s41596-025-01189-x>
52. Zhang, C., Shine, M., Pyle, A. M., & Zhang, Y. (2022). US-align: Universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods*, 19(9), 1109–1115. <https://doi.org/10.1038/s41592-022-01585-1>
53. Vankadari, N., & Wilce, J. A. (2020). Emerging WuHan (COVID-19) coronavirus: Glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. *Emerging Microbes & Infections*, 9(1), 601–604. <https://doi.org/10.1080/22221751.2020.1739565>
54. Guenza, M. G. (2025). Everything You Want to Know About Coarse-Graining and Never Dared to Ask: Macromolecules as a Key Example. *WIREs Computational Molecular Science*, 15(2), e70022. <https://doi.org/10.1002/wcms.70022>
55. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., & de Vries, A. H. (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*, 111(27), 7812–7824. <https://doi.org/10.1021/jp071097f>
56. Hsu, Y.-P., Frank, M., Mukherjee, D., Shchurik, V., Makarov, A., & Mann, B. F. (2022). Structural remodeling of SARS-CoV-2 spike protein glycans reveals the regulatory roles in receptor-binding affinity. *Glycobiology*, 33(2), 126–137. <https://doi.org/10.1093/glycob/cwac077>
57. Cho, B. G., Gautam, S., Peng, W., Huang, Y., Goli, M., & Mechref, Y. (2021). Direct Comparison of N-Glycans and Their Isomers Derived from Spike Glycoprotein 1 of MERS-CoV, SARS-CoV-1, and SARS-CoV-2. *Journal of Proteome Research*, 20(9), 4357–4365. <https://doi.org/10.1021/acs.jproteome.1c00323>
58. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, 369(6501), 330–333. <https://doi.org/10.1126/science.abb9983>

59. Watanabe, Y., Berndsen, Z. T., Raghwani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., McLellan, J. S., Wilson, I. A., Bowden, T. A., Ward, A. B., & Crispin, M. (2020). Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nature Communications*, 11(1), 2688. <https://doi.org/10.1038/s41467-020-16567-0>
60. Grant, O. C., Wentworth, D., Holmes, S. G., Kandel, R., Sehnal, D., Wang, X., Xiao, Y., Sheppard, P., Grelsson, T., Coulter, A., Miller, G., Foley, B. L., & Woods, R. J. (2025). *Generating 3D Models of Carbohydrates with GLYCAM-Web* (p. 2025.05.08.652828). bioRxiv. <https://doi.org/10.1101/2025.05.08.652828>
61. Huang, Y., Harris, B. S., Minami, S. A., Jung, S., Shah, P. S., Nandi, S., McDonald, K. A., & Faller, R. (2022). SARS-CoV-2 spike binding to ACE2 is stronger and longer ranged due to glycan interaction. *Biophysical Journal*, 121(1), 79–90. <https://doi.org/10.1016/j.bpj.2021.12.002>
62. Shajahan, A., Pepi, L. E., Kumar, B., Murray, N. B., & Azadi, P. (2023). Site specific N- and O-glycosylation mapping of the spike proteins of SARS-CoV-2 variants of concern. *Scientific Reports*, 13(1), 10053. <https://doi.org/10.1038/s41598-023-33088-0>
63. *Non-equilibrium pulling—GROMACS 2026.2 documentation*. (n.d.). Retrieved May 22, 2026, from <https://manual.gromacs.org/current/reference-manual/special/pulling.html>
64. Park, S., Khalili-Araghi, F., Tajkhorshid, E., & Schulten, K. (2003). Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics*, 119(6), 3559–3566. <https://doi.org/10.1063/1.1590311>
65. Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., Qin, H., Wang, M., Lu, Q., Li, X., Sun, Q., Liu, J., Zhang, L., Li, X., Huang, W., & Wang, Y. (2020). The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell*, 182(5), 1284-1294.e9. <https://doi.org/10.1016/j.cell.2020.07.012>

66. Nguyen, H. L., Lan, P. D., Thai, N. Q., Nissley, D. A., O'Brien, E. P., & Li, M. S. (2020). Does SARS-CoV-2 Bind to Human ACE2 More Strongly Than Does SARS-CoV? *The Journal of Physical Chemistry. B*, 124(34), 7336–7347. <https://doi.org/10.1021/acs.jpcb.0c04511>
67. Xie, Y., Karki, C. B., Du, D., Li, H., Wang, J., Sobitan, A., Teng, S., Tang, Q., & Li, L. (2020). Spike Proteins of SARS-CoV and SARS-CoV-2 Utilize Different Mechanisms to Bind With Human ACE2. *Frontiers in Molecular Biosciences*, 7. <https://doi.org/10.3389/fmolb.2020.591873>
68. Feng, T., Zhang, J., Chen, Z., Pan, W., Chen, Z., Yan, Y., & Dai, J. (2022). Glycosylation of viral proteins: Implication in virus-host interaction and virulence. *Virulence*, 13(1), 670–683. <https://doi.org/10.1080/21505594.2022.2060464>
69. Watanabe, Y., Bowden, T. A., Wilson, I. A., & Crispin, M. (2019). Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1863(10), 1480–1497. <https://doi.org/10.1016/j.bbagen.2019.05.012>
70. Newby, M. L., Allen, J. D., & Crispin, M. (2024). Influence of glycosylation on the immunogenicity and antigenicity of viral immunogens. *Biotechnology Advances*, 70, 108283. <https://doi.org/10.1016/j.biotechadv.2023.108283>
71. Prabakaran, P., Dimitrov, A. S., Fouts, T. R., & Dimitrov, D. S. (2007). Structure and Function of the HIV Envelope Glycoprotein as Entry Mediator, Vaccine Immunogen, and Target for Inhibitors. *Advances in Pharmacology (San Diego, Calif.)*, 55, 33–97. [https://doi.org/10.1016/S1054-3589\(07\)55002-7](https://doi.org/10.1016/S1054-3589(07)55002-7)
72. Negi, G., Sharma, A., Dey, M., Dhanawat, G., & Parveen, N. (2022). Membrane attachment and fusion of HIV-1, influenza A, and SARS-CoV-2: Resolving the mechanisms with biophysical methods. *Biophysical Reviews*, 14(5), 1109–1140. <https://doi.org/10.1007/s12551-022-00999-7>

73. Yang, Q., Kelkar, A., Manicassamy, B., & Neelamegham, S. (2025). Conserved role of spike S2 domain N-glycosylation across betacoronaviruses. *Npj Viruses*, 3(1), 4. <https://doi.org/10.1038/s44298-024-00085-7>
74. Souza, P. C. T., Limongelli, V., Wu, S., Marrink, S. J., & Monticelli, L. (2021). Perspectives on High-Throughput Ligand/Protein Docking With Martini MD Simulations. *Frontiers in Molecular Biosciences*, 8, 657222. <https://doi.org/10.3389/fmolb.2021.657222>
75. Souza, P. C. T., Thallmair, S., Conflitti, P., Ramírez-Palacios, C., Alessandri, R., Raniolo, S., Limongelli, V., & Marrink, S. J. (2020). Protein-ligand binding with the coarse-grained Martini model. *Nature Communications*, 11(1), 3714. <https://doi.org/10.1038/s41467-020-17437-5>

CHAPTER 6: Ongoing experimental validation and final conclusions on pandemic screening

6.1 Introduction

Chapter 5 discussed how we screened emerging betacoronaviruses for pandemic-potential using structure prediction, protein docking tools, and molecular dynamics simulations. From this work, we identified Zhejiang2013—a hibeovirus, previously not thought to bind to ACE2. However, docking and simulations still need to be validated because these models only replicate protein interactions using limited physical or thermodynamic parameters. Simulations cannot fully capture the complexity of a cell. Computational predictions suggest possibilities and cannot definitively prove that these interactions occur in biological systems.

We need to experimentally validate the predictions to determine if they accurately reflect the S1-ACE2 protein interaction. This chapter will consider how to validate these computational predictions and the preliminary work done to date on these validation efforts. We first need to express Zhejiang2013 S1 in a mammalian system to accurately capture post-translation modifications and folding. We can then purify the protein to use in binding assays to measure its interaction with ACE2 in comparison to the pandemic-causing controls, SARS-CoV-1 and MERS-CoV S1. The chapter ends with a discussion of screening pipelines and predicting pandemic potential in emerging viruses.

6.2 Experimentally testing emerging coronaviruses to a human receptor

The previous chapter discussed a prediction pipeline to assess binding between viral proteins to a human cell receptor. Our predictions used both static and dynamic simulations to model the S1-ACE2 protein interaction across emerging betacoronaviruses and predicted that Zhejiang2013 binds ACE2 to a comparable degree to pandemic-causing coronaviruses. However, computational screening is still a prediction of a biological system. To test this hypothesis, must express and purify the protein and measure its binding affinity to ACE2.

Several strategies are available. Large-scale mammalian culture (125-250 mL) provides high protein quantities with native post-translation modifications (PTMs) but can be slow and costly to express multiple proteins^[1]. Using a higher throughput format like a 96-deep well plate (1-2 mL) enables parallel expression of multiple proteins though at a lower yield. Alternatively, yeast surface display or phage display can rapidly screen protein libraries against a single human receptor using fluorescently activated cell sorting (FACS), ultimately testing protein interactions^[2]. While these display methods would optimize throughput for screening viral proteins and their interactions, they would not capture relevant PTMs or maintain cellular environments and therefore lose critical biophysical context. For our experimental pipeline, we decided to express the S1 proteins in a mammalian system to then test the interaction with ACE2 using biolayer interferometry.

6.2.1 Ongoing work: testing protein expression and purification

We selected a small panel of coronavirus S1s along with controls to test protein expression (Table 5-1). We designed expression plasmids containing an N-terminus secretion peptide and a C-terminus histidine (His) affinity tag. The secretion peptide directs the protein into the cell culture

supernatant where it can be more easily collected and purified away from intracellular proteins. This affinity tag enables standardized purification across different S1 proteins using the same affinity chromatography protocol ^[1,3].

To increase throughput and protein yield, we also transiently transfected the S1 domains into Chinese Hamster Ovary (CHO) cells using the ExpiCHO expression systems ^[1]. CHO cells are frequently used in industrial applications for high-yield production of recombinant proteins for therapeutics^[4]. The cells were cultured in a 96-deep well plate so that multiple coronavirus S1 proteins could be simultaneously expressed and processed in parallel to increase experimental throughput. The S1 proteins have a mass around 75 kDa, but with glycosylation, we expected bands around 100 to 130 kDa. We used untransfected samples, cells that are not transiently expressing S1, as a negative control to compare background bands from the His-tag antibody.

Our expression data from a 96-deep well plate showed some S1 expression in the cell supernatant at around 100-135 kDa (Figure 6-1). There are bands at the expected size for the S1 except for Rs7327 and YN2018B in the supernatant. We see that the untransfected sample has a band around 65 kDa which would suggest some endogenous protein expression with histidines that is not our protein of interest. This band is found across all the S1 samples in the supernatant. We do not see any bands for the untransfected sample in the lysate which is expected since S1 was not being expressed. We did not observe strong expression in the lysate which suggests that the protein secreted into the cell culture media.

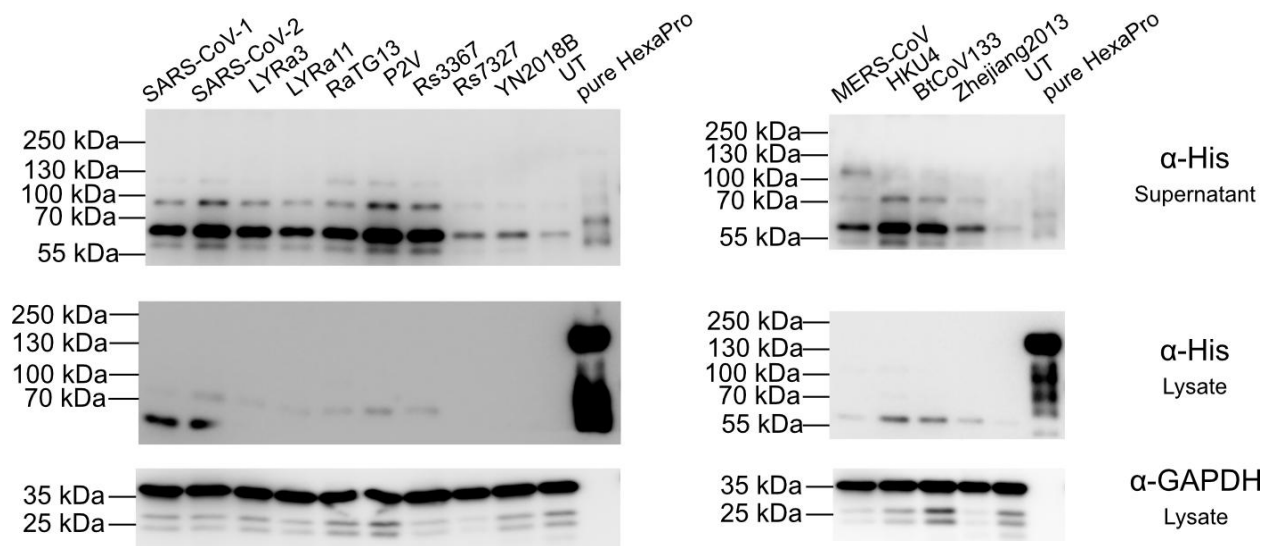


Figure 6-1. Probing for S1 expression in ExpiCHO cell lysate and supernatant. We expected S1 to run between 100 and 135 kDa due to glycosylation in the expression system. HexaPro Spike is a Spike protein stabilized with six proline substitutions that was used as a positive expression control (~180 kDa) based on previous work^[1,5,6]. UT is untransfected sample which served as a negative control and GAPDH was the loading control. We probed for S1 and Spike using His affinity tag.

We next purified the S1 proteins in the same 96-well format to maintain throughput using Ni-NTA affinity chromatography (Figure 6-2). We used Western blot to probe for our S1s of interest—Zhejiang2013 and the controls, SARS-CoV-1 and MERS-CoV—to test whether the purification worked before proceeding with the other proteins. We also purified HexaPro Spike because of our previous work optimizing its expression and purification in larger ExpiCHO cultures^[6]. HexaPro Spike is the SARS-CoV-2 full-length Spike protein that is stabilized with six proline substitutions that improve stability and expression^[5].

Western blot analysis showed bands only in the flowthrough (FT) fraction with nothing detectable in the wash (W) or elution (E) fractions (Figure 6-2). Again, we expected to see bands in the range

of 100 to 135 kDa for the S1 protein, and around 180 kDa for HexaPro Spike. For the HexaPro Spike sample, there is a faint band around 180 kDa in the flowthrough fraction and then multiple smaller bands. No other bands are visible across the other fractions for the S1 samples. For MERS-CoV S1, there seems to be a faint band around 100 kDa but it is not clear. Across the samples, there is a band is around 65 kDa in the FT fraction, which we also saw in our expression blots (Figure 6-1). However, for SARS-CoV-1 and Zhejiang2013 S1, we did not see the band we expected for S1, not even in the FT fraction.

The lack of bands across the S1 samples and fractions, especially the lack of bands at the expected size, indicated that there was probably no protein loaded in the column. Interestingly, HexaPro Spike also failed to elute. We have previously expressed and purified HexaPro Spike in larger ExpiCHO cultures and expected it to express and purify easily to be used as a positive control^[6]. This suggested that protein expression did not work in the 96-well format. The lack of an S1 band in the FT fraction indicated that there was not sufficient protein expressed in the culture or loaded in the column. Therefore, there was no protein to purify. If protein was present in the loaded sample but failed to bind to the resin, we would have seen stronger bands expressed at least in the FT fraction as well as an earlier wash fraction (W1) ^[3,7].

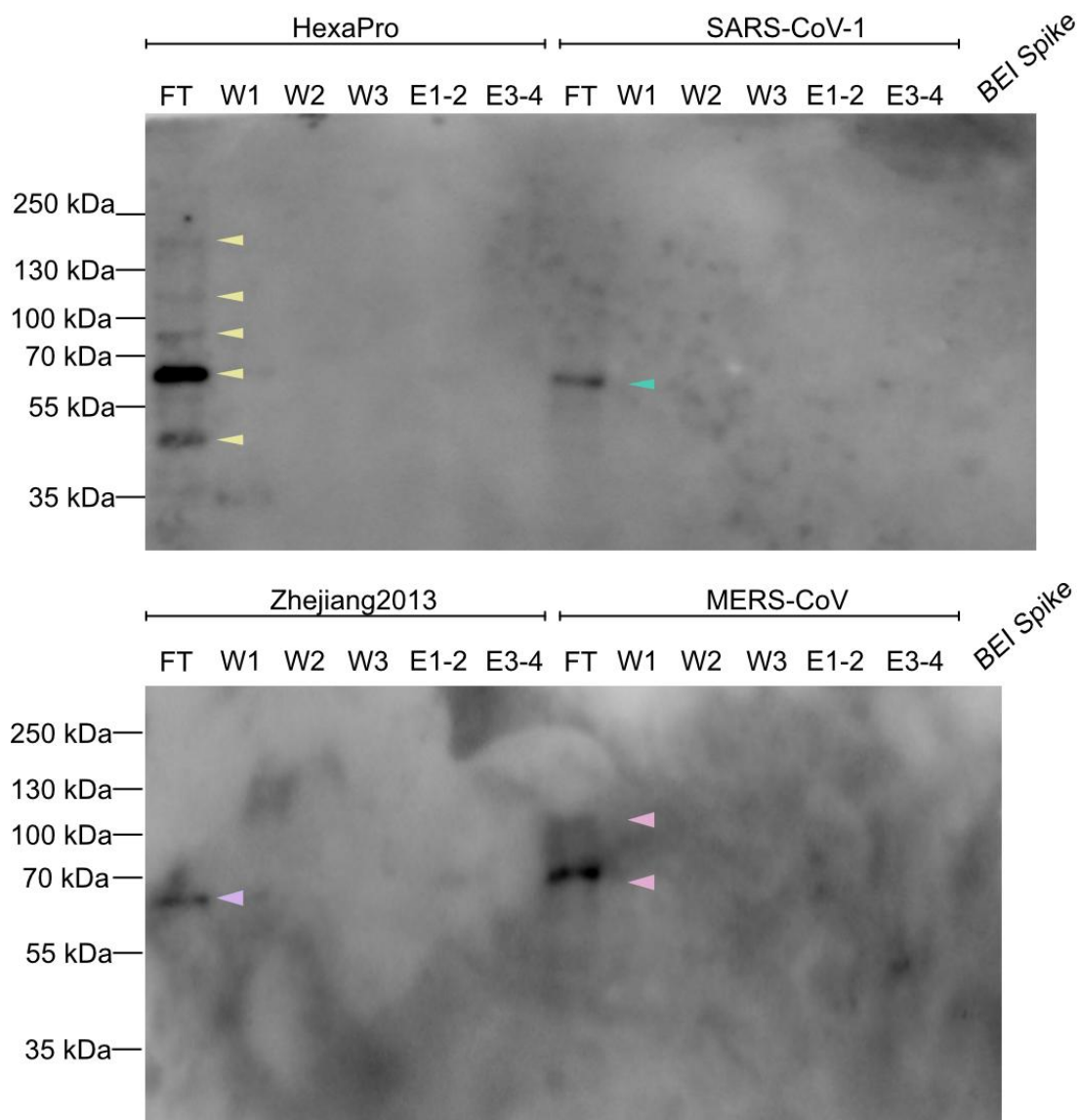


Figure 6-2. Probing for S1 protein through the protein purification process. We purified S1 of SARS-CoV-1, Zhejiang2013, and MERS-CoV and purified HexaPro Spike as a purification control. BEI Spike was a positive control for the Western blot. We expected bands between 100-135 kDa for S1 and around 180 kDa for HexaPro Spike and BEI Spike. FT is flow-through, W is wash, and E is elution. Colored arrows indicate the presence of bands due to the low contrast of the blot. We probed S1 and Spike using the His affinity tag. All the samples have a band at ~65 kDa in the FT fraction and no bands in the other fractions.

Future work is still needed to optimize the expression and purification of these proteins for downstream assays. First, we could optimize the 96-well format by addressing buffer conditions to improve His-tag accessibility and nickel binding or switching to 24-well format to increase culture volume. Second, if protein expression remains problematic in ExpiCHO cells, it may be worth using HEK293T cells. High-yield protein secretion in ExpiCHO cells occurs on a time-scale of 7-10 days whereas protein secreted from HEK293T cells takes 1-2 days which can allow for more rapid optimization^[1,8]. Third, we can simplify the expression construct by testing the RBD instead of the full S1 domain. Our computational work demonstrated that the trends observed in S1 were maintained with the RBD so its binding affinity to ACE2 would recapitulate the interaction. While testing the S1-ACE2 interaction would be a more accurate measurement, testing RBD with ACE2 would preserve the core binding interface while simplifying protein production^[1,2].

6.2.2 Future work: measuring binding affinity

Measuring the binding affinity between S1 and ACE2 can determine how feasible it is for a virus like Zhejiang2013 to attach to human cells. This is essential to experimentally validate the binding of the S1-ACE2 interaction to corroborate the trends observed in the computational predictions. Specifically understanding the binding kinetics of Zhejiang2013 will establish if it is a credible pandemic threat based on how it engages a human receptor. This work is currently in early stages, but it is still important to consider the implications.

We can measure the binding affinity of the interaction with biolayer interferometry (BLI). BLI is an optical biosensor “dip and read” technique that uses the light interference patterns to measure the association and dissociation rates between proteins^[9]. From the shift in interference patterns following the binding and unbinding of proteins, we can calculate an equilibrium dissociation constant, K_D . K_D is inversely related to binding affinity, so a lower K_D indicates stronger binding.

From literature, we know that ACE2 binds SARS-CoV-1 and SARS-CoV-2 RBD with an equilibrium dissociation constant of 5.0 nM and 1.2 nM, respectively, measured with BLI^[10]. While not tested directly with BLI, MERS-CoV does not bind ACE2 or enter cells expressing ACE2, and shows robust entry with cells expressing the human receptor dipeptidyl peptidase-4 (DPP4)^[11–14]. These pandemic-causing viruses with known ACE2 binding behaviors provide binding affinity limits. Based on the binding energy calculated from the pulling simulations (Figure 5-7), we predict that Zhejiang2013 would bind ACE2 with an affinity similar or stronger than SARS-CoV-1.

6.3 Conclusions

For a virus to successfully infect its host, it first has to attach to the receptor, enter the cells, replicate itself and evade host defenses, and then exit the cell. Understanding whether an emerging zoonotic virus can successfully infect humans means that it includes the required machinery to satisfy the aforementioned steps. We can use these steps to screen emerging viruses to assess if they pose a genuine threat to human health^[15–17]. As a first assessment, we built a pipeline that screens the initial attachment interaction. Chapter 5 discusses how we could predict binding of emerging coronaviruses binding ACE2. This chapter expands on this workflow by proposing experiments to validate the binding predictions.

Although protein expression and purification challenges remain to be resolved, this chapter outlines ongoing work to validate the S1-ACE2 predictions described in Chapter 5. The MD simulations predicted that Zhejiang2013 would bind ACE2, in a manner similar to or exceeding that of SARS-CoV-1. Testing this hypothesis requires experimentation that bridges the gap with computational predictions before testing further for replication. These experiments can provide key results to assess Zhejiang2013 infecting human cells via ACE2. If BLI measurements confirm that Zhejiang2013 binds ACE2, this would validate the computational pipeline and confirm that

Zhejiang2013 satisfies the first step to human emergence. Further studies are needed to confirm if Zhejiang2013 can continue to enter and replicate in humans using pseudovirus to display the Spike protein for attachment and entry^[18–20]. This approach optimizes resources by filtering out low-risk viruses early, only high-risk candidates advance to expensive experimental assays.

If BLI measurements show weak or absent binding, the computational predictions would be refuted and we would need to refine our modeling approach. We currently quantify predicted binding using a distance-based contact analysis. This method calculates residues pairs within 4 angstroms of each other across the proteins to identify which interact. However, this approach includes transient interactions that may not contribute to binding. A more conservative approach would be to calculate hydrogen bonds^[12]. Hydrogen bonds are directional interactions and could be more indicative of true biomolecular binding forces. By comparing hydrogen bonding to a human receptor across viruses, we could more accurately distinguish binders and refine our quantification metric for screening. Additionally, our static docking and structure predictions do not include glycans because their positioning is unknown for uninvestigated viruses though glycosylation is critical for host receptor binding^[21,22]. The lack of glycans on the protein would effect which residues would be interacting between S1 and ACE2. There have been recent developments to allow PTM modeling in docking simulations, such as with Alphafold3^[23].

While this work focuses on emerging betacoronaviruses and their potential interaction with ACE2, the pipeline is adaptable to other viral families with pandemic potential. For example, paramyxoviruses like Nipah virus require efficient binding to ephrin-B receptors^[24,25]. Filoviruses like Ebola depend on Niemann-Pick C1 receptor interactions^[26,27]. Influenza hemagglutinin must bind sialic acid receptors^[28–30]. Currently, the computational pipeline is best suited to screen viral surface proteins that engage with only one or two human receptors, but future experimental work can explore testing the interaction of multiple receptors.

For zoonotic pathogens, the underlying question remains the same: does the viral surface protein bind human receptors sufficiently to establish infection? We can substitute the relevant viral protein and human receptor into this framework. Computational screening can inform which proteins to measure binding which leads to entry assays and then formal animal transmissibility studies. This this workflow we can systematically evaluate emerging threats across viral families and transforms the pipeline from a coronavirus-specific tool into a generalizable framework for pandemic risk assessment across diverse emerging pathogens.

6.4 Methods

6.4.1 Plasmid Construction

HexaPro Spike was gifted from Ilya Finkelstein at UT Austin^[5]. S1 fragments were synthesized with overhangs for a pCAGGS backbone (TWIST Biosciences) and cloned using Gibson Assembly into a backbone with a N-terminus secretion tag and a C-terminus 8x histidine tag.

6.4.2 Cell Culture

ExpiCHO cells were cultured in a humidified incubator at 37°C and 8% CO₂. ExpiCHOs were maintained in a 125mL vented shaker flask with 30 mL of culture in ExpiCHO Expression Medium (Thermo Fisher Scientific) on an orbital shaker at 120 rpm on a 19 mm shaking diameter. Cells were tested for mycoplasma every month by MycoStrip (Invivogen).

6.4.3 Transfection

For transfection in 96-deep well plates (Genesee Scientific), 0.8 mL of ExpiCHO cells were plated at 6×10^6 cells/mL on the day of transfection. Cells were cultured on an orbital shaker (Benchmark Scientific) at 900 rpm at a shaking diameter of 3 mm. Cells were transfected according to

manufacturer instructions using the Expifectamine CHO Transfection Kit (Thermo Fisher Scientific).

6.4.5 Protein Harvest and Purification

ExpiCHO samples were harvested 4 days post-transfection by centrifuging the cultures at 300 rcf for 5 mins and collecting the supernatant and lysing the cells with RIPA buffer (150 mM NaCL, 50 mM Tris Base, 1% TritonX-100, 0.5% sodium deoxycholate). Collected samples were incubated on ice for 30 mins and then centrifuged at 10,000 g for 10 mins at 4°C. ExpiCHO cells protein concentration was normalized by cell count (1 uL of RIPA/1000 cells).

ExpiCHO protein samples were purified with the HisPur Ni-NTA spin plate (Thermo Fisher Scientific) using imidazole to elute the proteins according to manufacturer's instructions.

6.4.6 SDS-PAGE and Western blot

Samples were prepared by adding NuPAGE LDS sample buffer supplemented with TCEP and boiled at 95°C for 10 mins. Samples were run on a 4-20% gradient polyacrylamide hand-cast gels for 15 mins at 115 V then 50 mins at 150 V. Proteins were transferred onto polyvinylidene difluoride membranes in a wet sandwich and blocked using 5% non-fat milk for 1 hour at room temperature. Membranes were then incubated overnight at 4°C with primary antibodies (BioRad MCA1396, Fisher Scientific PIMA515738) at a 1:1000 dilution and then for 1 hour at room temperature with a secondary antibody at a 1:4000 dilution (SouthernBiotech 6170-05). Membranes were developed using Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) and imaged using an Amersham Imager 600 (Cytiva).

6.5 References

1. Minami, S. A., Jung, S., Huang, Y., Harris, B. S., Kenaston, M. W., Faller, R., Nandi, S., McDonald, K. A., & Shah, P. S. (2022). Production of novel SARS-CoC-2 Spike truncation in Chinese hamster ovary cells leads to high expression and binding to antibodies. *Biotechnology Journal*, (17). <https://doi.org/10.1002/biot.202100678>
2. Starr, T. N., Greaney, A. J., Stewart, C. M., Walls, A. C., Hannon, W. W., Veelsler, D., & Bloom, J. D. (2022). Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. *PLOS Pathogens*, 18(11), e1010951. <https://doi.org/10.1371/journal.ppat.1010951>
3. [16] Purification of proteins using polyhistidine affinity tags. (2000). In *Methods in Enzymology* (Vol. 326, pp. 245–254). Academic Press. [https://doi.org/10.1016/S0076-6879\(00\)26058-8](https://doi.org/10.1016/S0076-6879(00)26058-8)
4. Kim, J. Y., Kim, Y.-G., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Applied Microbiology and Biotechnology*, 93(3), 917–930. <https://doi.org/10.1007/s00253-011-3758-5>
5. Hsieh, C.-L., Goldsmith, J. A., Schaub, J. M., DiVenere, A. M., Kuo, H.-C., Javanmardi, K., Le, K. C., Wrapp, D., Lee, A. G., Liu, Y., Chou, C.-W., Byrne, P. O., Hjorth, C. K., Johnson, N. V., Ludes-Meyers, J., Nguyen, A. W., Park, J., Wang, N., Amengor, D., ... McLellan, J. S. (2020). Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science (New York, N.Y.)*, 369(6510), 1501–1505. <https://doi.org/10.1126/science.abd0826>
6. Ophir, Y., Wong, J. H., Haddad, K. R., Huuskonen, A., Karmaker, A., Gore, V., Jung, S., Oloumi, A., Liu, Y., Fu, J., Zhang, L., Huang, P., Minami, S. A., Garimella, S. S., Thyagatur, A., Zaini, P. A., Vitikainen, M., Tchelet, R., Valbuena, N., ... Nandi, S. (2025). Expression and characterization of SARS-CoV-2 spike protein in *Thermothelomyces heterothallica* C1. *Vaccine*, 65, 127784. <https://doi.org/10.1016/j.vaccine.2025.127784>

7. Janson, J.-C. (2012). Protein Purification: Principles, High Resolution Methods, and Applications. John Wiley & Sons.
8. Kenaston, M. W., Cherkashchenko, L., Skawinski, C. L. S., Fishburn, A. T., Peddamallu, V., Florio, C. J., Robertson, A. E., Bhattacharya, T., Young, J. M., Malik, H. S., & Shah, P. S. (2025). *Yellow Fever Virus Interactomes Reveal Common and Divergent Strategies of Replication and Evolution for Mosquito-borne Flaviviruses* (p. 2025.06.14.659623). bioRxiv. <https://doi.org/10.1101/2025.06.14.659623>
9. Sultana, A., & Lee, J. E. (2015). Measuring Protein-Protein and Protein-Nucleic Acid Interactions by Biolayer Interferometry. *Current Protocols in Protein Science*, 79(1), 19.25.1-19.25.26. <https://doi.org/10.1002/0471140864.ps1925s79>
10. Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Velesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2), 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
11. Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X., Arledge, K. C., Chen, Y.-H., Zhang, L., & Wang, X. (2013). Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Research*, 23(8), 986–993. <https://doi.org/10.1038/cr.2013.92>
12. Li, Y., Zhang, Z., Yang, L., Lian, X., Xie, Y., Li, S., Xin, S., Cao, P., & Lu, J. (2020). The MERS-CoV Receptor DPP4 as a Candidate Binding Target of the SARS-CoV-2 Spike. *iScience*, 23(6), 101160. <https://doi.org/10.1016/j.isci.2020.101160>
13. Xiong, Q., Cao, L., Ma, C., Tortorici, M. A., Liu, C., Si, J., Liu, P., Gu, M., Walls, A. C., Wang, C., Shi, L., Tong, F., Huang, M., Li, J., Zhao, C., Shen, C., Chen, Y., Zhao, H., Lan, K., ... Yan, H. (2022). Close relatives of MERS-CoV in bats use ACE2 as their functional receptors. *Nature*, 612(7941), 748–757. <https://doi.org/10.1038/s41586-022-05513-3>
14. Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H. W., Müller, M. A., Dijkman, R., Muth, D., Demmers, J. A. A., Zaki, A., Fouchier, R. A. M., Thiel, V., Drosten, C., Rottier, P. J. M.,

- Osterhaus, A. D. M. E., Bosch, B. J., & Haagmans, B. L. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*, 495(7440), 251–254. <https://doi.org/10.1038/nature12005>
15. Mathez, G., & Cagno, V. (2021). Viruses Like Sugars: How to Assess Glycan Involvement in Viral Attachment. *Microorganisms*, 9(6), 1238. <https://doi.org/10.3390/microorganisms9061238>
 16. Kumar, P. K. R. (2017). Systematic screening of viral entry inhibitors using surface plasmon resonance. *Reviews in Medical Virology*, 27(6), e1952. <https://doi.org/10.1002/rmv.1952>
 17. Peng, T. (2010). Strategies for antiviral screening targeting early steps of virus infection. *Virologica Sinica*, 25(4), 281–293. <https://doi.org/10.1007/s12250-010-3135-z>
 18. Starr, T. N., Zepeda, S. K., Walls, A. C., Greaney, A. J., Alkhovsky, S., Veessler, D., & Bloom, J. D. (2022). ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature*, 603(7903), Article 7903. <https://doi.org/10.1038/s41586-022-04464-z>
 19. Menachery, V. D., Yount, B. L., Sims, A. C., Debbink, K., Agnihothram, S. S., Gralinski, L. E., Graham, R. L., Scobey, T., Plante, J. A., Royal, S. R., Swanstrom, J., Sheahan, T. P., Pickles, R. J., Corti, D., Randell, S. H., Lanzavecchia, A., Marasco, W. A., & Baric, R. S. (2016). SARS-like WIV1-CoV poised for human emergence. *Proceedings of the National Academy of Sciences*, 113(11), 3048–3053. <https://doi.org/10.1073/pnas.1517719113>
 20. Menachery, V. D., Dinnon, K. H., Yount, B. L., McAnarney, E. T., Gralinski, L. E., Hale, A., Graham, R. L., Scobey, T., Anthony, S. J., Wang, L., Graham, B., Randell, S. H., Lipkin, W. I., & Baric, R. S. (2020). Trypsin Treatment Unlocks Barrier for Zoonotic Bat Coronavirus Infection. *Journal of Virology*, 94(5), e01774-19. <https://doi.org/10.1128/JVI.01774-19>
 21. Hsu, Y.-P., Frank, M., Mukherjee, D., Shchurik, V., Makarov, A., & Mann, B. F. (2022). Structural remodeling of SARS-CoV-2 spike protein glycans reveals the regulatory roles in receptor-binding affinity. *Glycobiology*, 33(2), 126–137. <https://doi.org/10.1093/glycob/cwac077>

22. Huang, Y., Harris, B. S., Minami, S. A., Jung, S., Shah, P. S., Nandi, S., McDonald, K. A., & Faller, R. (2022). SARS-CoV-2 spike binding to ACE2 is stronger and longer ranged due to glycan interaction. *Biophysical Journal*, 121(1), 79–90. <https://doi.org/10.1016/j.bpj.2021.12.002>
23. Huang, C., Kannan, N., & Moremen, K. W. (2025). Modeling glycans with AlphaFold 3: Capabilities, caveats, and limitations. *Glycobiology*, 35(10), cwaf048. <https://doi.org/10.1093/glycob/cwaf048>
24. Salleh, M. Z. (2025). Structural biology of Nipah virus G and F glycoproteins: Insights into therapeutic and vaccine development. *European Journal of Microbiology & Immunology*, 15(2), 83–93. <https://doi.org/10.1556/1886.2025.00017>
25. Bowden, T. A., Crispin, M., Harvey, D. J., Aricescu, A. R., Grimes, J. M., Jones, E. Y., & Stuart, D. I. (2008). Crystal Structure and Carbohydrate Analysis of Nipah Virus Attachment Glycoprotein: A Template for Antiviral and Vaccine Design. *Journal of Virology*, 82(23), 11628–11636. <https://doi.org/10.1128/jvi.01344-08>
26. Peng, W., Rayaprolu, V., Parvate, A. D., Pronker, M. F., Hui, S., Parekh, D., Shaffer, K., Yu, X., Saphire, E. O., & Snijder, J. (2022). Glycan shield of the ebolavirus envelope glycoprotein GP. *Communications Biology*, 5(1), 785. <https://doi.org/10.1038/s42003-022-03767-1>
27. Lee, J. E., & Saphire, E. O. (2009). Ebolavirus glycoprotein structure and mechanism of entry. *Future Virology*, 4(6), 621–635. <https://doi.org/10.2217/fvl.09.56>
28. Dales, S., & Choppin, P. W. (1962). Attachment and penetration of influenza virus. *Virology*, 18(3), 489–493. [https://doi.org/10.1016/0042-6822\(62\)90041-7](https://doi.org/10.1016/0042-6822(62)90041-7)
29. Skehel, J. J., & Wiley, D. C. (2000). Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. *Annual Review of Biochemistry*, 69(Volume 69, 2000), 531–569. <https://doi.org/10.1146/annurev.biochem.69.1.531>
30. Negi, G., Sharma, A., Dey, M., Dhanawat, G., & Parveen, N. (2022). Membrane attachment and fusion of HIV-1, influenza A, and SARS-CoV-2: Resolving the mechanisms with

biophysical methods. *Biophysical Reviews*, 14(5), 1109–1140.
<https://doi.org/10.1007/s12551-022-00999-7>

Appendix A: Supporting Information for A simplified two-plasmid system for orthogonal control of mammalian gene expression using light-activated CRISPR effector

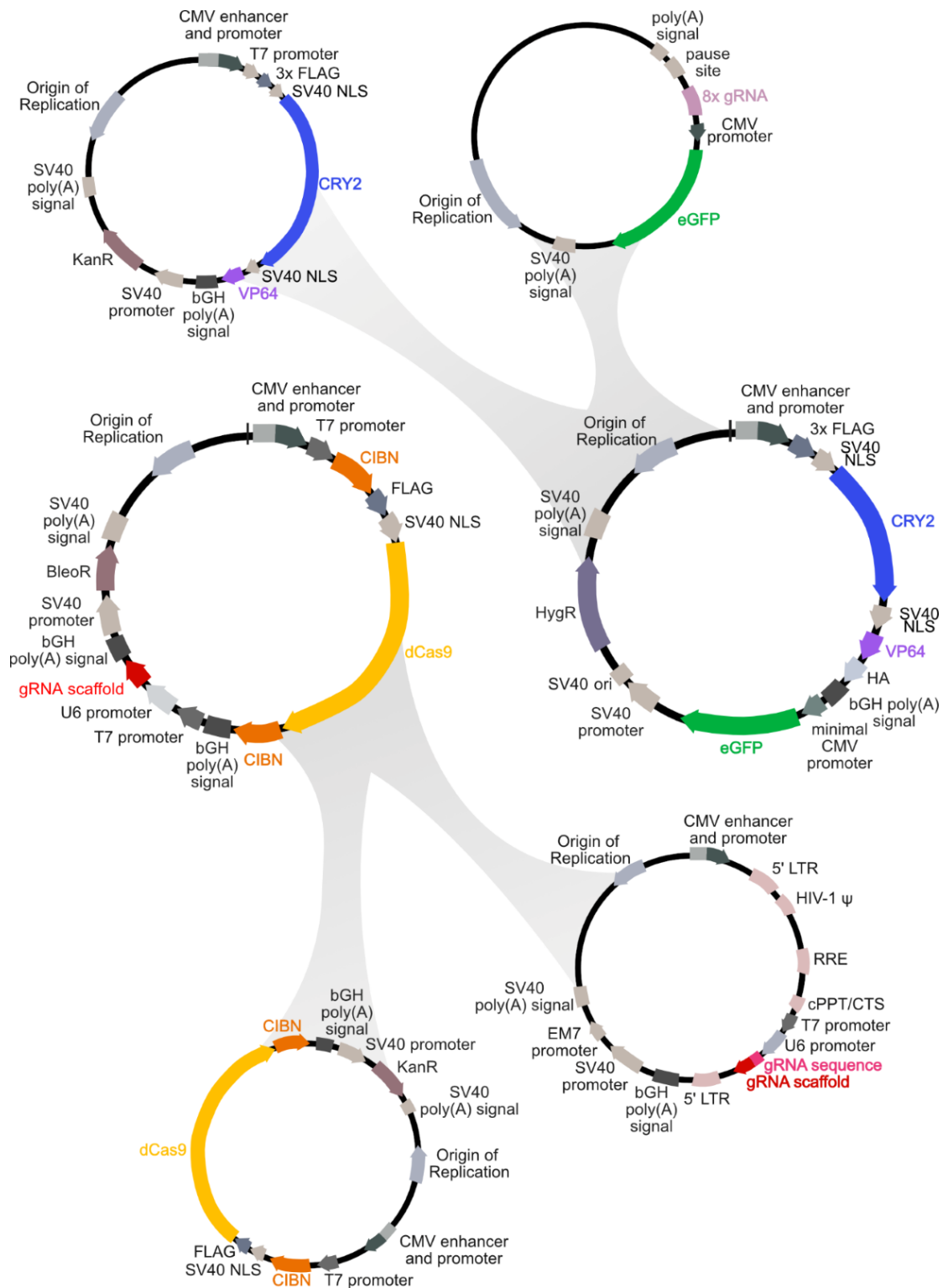


Figure A-1. Schematic of plasmid design from the original four LACE plasmids (4pLACE) to the optimized two plasmids (2pLACE).

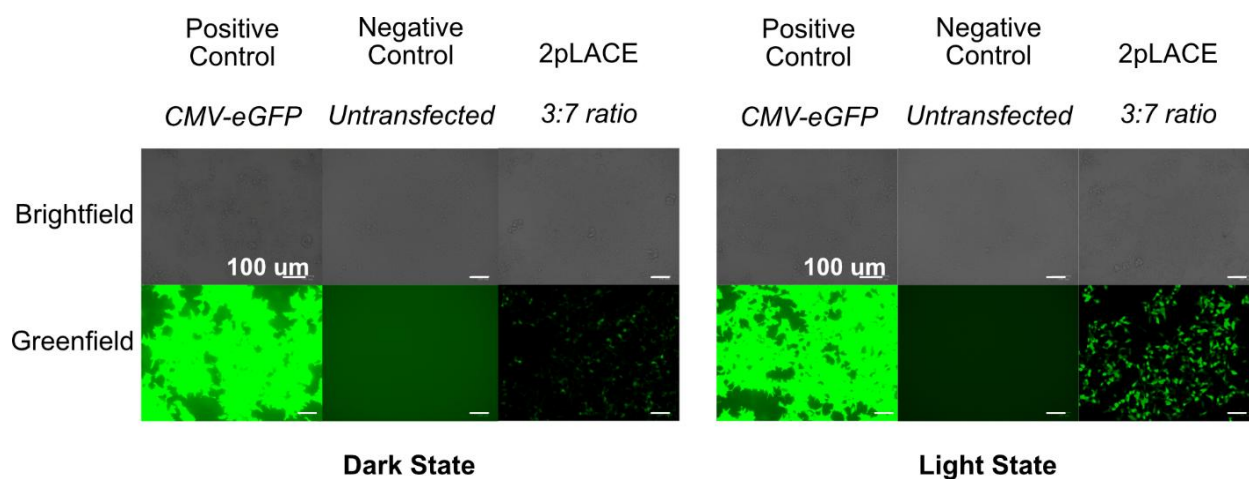


Figure A-2. eGFP expression of the optimized plasmid ratio for 2pLACE with and without light activation. Images were taken 24 hours post activation.

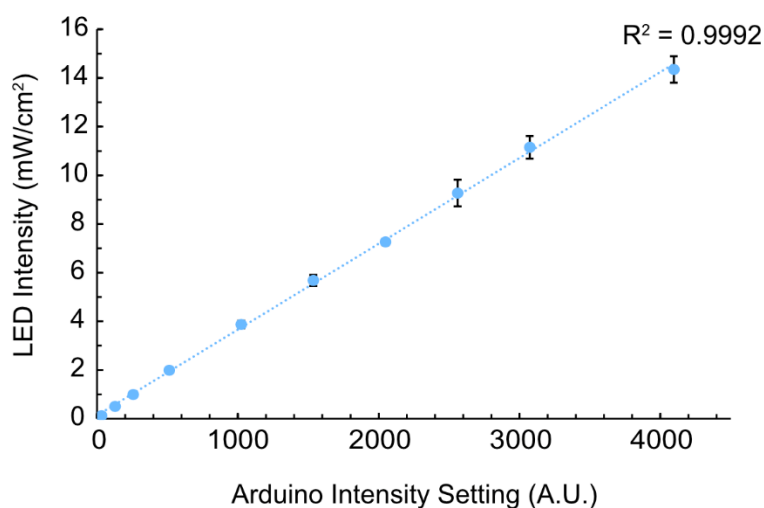


Figure A-3. Calibration curve of the optoPlate-96. Four LEDs were sampled from the optoPlate-96. Error bars represent standard deviation. Data were fitted with a line and the intercept was fixed to be 0.

Table A-1. P-values from ANOVA statistics comparing dynamic range of plasmid ratio eGFP expression

	UT	1:9	2:8	3:7	4:6	5:5	6:4	7:3	8:2	9:1
UT	1.00	4.39E-2	1.44E-2	3.05E-5	5.46E-4	4.63E-2	2.24E-4	1.57E-3	1.72E-2	5.23E-3
1:9	4.39E-2	1.00	1.00	9.99E-1	9.64E-1	4.26E-1	3.66E-1	2.66E-1	1.29E-1	8.98E-2
2:8	1.44E-2	1.00	1.00	9.14E-1	9.42E-1	3.37E-1	1.78E-1	1.03E-1	3.67E-2	2.50E-2
3:7	3.05E-5	9.99E-1	9.14E-1	1.00	7.81E-2	2.08E-1	1.74E-3	3.47E-3	1.22E-2	3.40E-4
4:6	5.46E-4	9.64E-1	9.42E-1	7.81E-2	1.00	6.72E-1	4.33E-2	2.65E-2	2.68E-2	1.56E-3
5:5	4.63E-2	4.26E-1	3.37E-1	2.08E-1	6.72E-1	1.00	9.99E-1	9.51E-1	5.33E-1	1.62E-1
6:4	2.24E-4	3.66E-1	1.78E-1	1.74E-3	4.33E-2	9.99E-1	1.00	9.59E-1	4.12E-1	7.42E-3
7:3	1.57E-3	2.66E-1	1.03E-1	3.47E-3	2.65E-2	9.51E-1	9.59E-1	1.00	7.55E-1	2.01E-2
8:2	1.72E-2	1.29E-1	3.67E-2	1.22E-2	2.68E-2	5.33E-1	4.12E-1	7.55E-1	1.00	2.37E-1
9:1	5.23E-3	8.98E-2	2.50E-2	3.40E-4	1.56E-3	1.62E-1	7.42E-3	2.01E-2	2.37E-1	1.00

Table A-2. P-values from ANOVA statistics comparing eGFP expression of LED intensity samples (mW/cm²)

	0	0.12	0.50	0.99	1.99	3.87	5.68	7.26	9.27	11.15	14.35
0	1.00	1.36E-3	1.89E-1	8.15E-3	6.75E-6	1.06E-2	6.14E-3	3.04E-3	3.61E-3	3.52E-4	3.28E-3
0.12	1.36E-3	1.00	9.07E-1	9.38E-1	8.52E-2	4.29E-1	4.69E-1	4.46E-2	2.50E-1	3.72E-2	5.59E-2
0.50	1.89E-1	9.07E-1	1.00	4.48E-1	1.37E-1	1.39E-1	1.56E-1	1.25E-2	8.53E-2	4.28E-2	1.61E-2
0.99	8.15E-3	9.38E-1	4.48E-1	1.00	9.99E-1	9.70E-1	9.97E-1	2.15E-1	9.58E-1	8.32E-1	2.85E-1
1.99	6.75E-6	8.52E-2	1.37E-1	9.99E-1	1.00	9.98E-1	1.00	3.25E-1	9.97E-1	9.01E-1	4.13E-1
3.87	1.06E-2	4.29E-1	1.39E-1	9.70E-1	9.98E-1	1.00	1.00	9.09E-1	1.00	1.00	9.62E-1
5.68	6.14E-3	4.69E-1	1.56E-1	9.97E-1	1.00	1.00	1.00	6.43E-1	1.00	1.00	7.62E-1
7.26	3.04E-3	4.46E-2	1.25E-2	2.15E-1	3.25E-1	9.09E-1	6.43E-1	1.00	7.87E-1	7.46E-1	1.00
9.27	3.61E-3	2.50E-1	8.53E-2	9.58E-1	9.97E-1	1.00	1.00	7.87E-1	1.00	1.00	8.86E-1
11.15	3.52E-4	3.72E-2	4.28E-2	8.32E-1	9.01E-1	1.00	1.00	7.46E-1	1.00	1.00	8.59E-1
14.35	3.28E-3	5.59E-2	1.61E-2	2.85E-1	4.13E-1	9.62E-1	7.62E-1	1.00	8.86E-1	8.59E-1	1.00

Table A-3. P-values from ANOVA statistics comparing conditions of activation time (hours)

	0	0.5	1	2	4	6	12	15	18	21	24
0	1.00	9.83E-1	8.50E-1	4.35E-1	1.57E-2	6.17E-4	3.55E-7	3.15E-4	2.29E-3	2.52E-5	1.20E-5
0.5	9.83E-1	1.00	1.00	7.47E-1	4.08E-2	2.14E-3	4.10E-6	5.17E-4	2.90E-3	6.31E-5	2.67E-5
1	8.50E-1	1.00	1.00	9.91E-1	6.97E-2	2.95E-3	4.36E-6	5.44E-4	3.04E-3	6.59E-5	2.75E-5
2	4.35E-1	7.47E-1	9.91E-1	1.00	1.45E-1	8.02E-3	2.26E-5	7.04E-4	3.44E-3	1.07E-4	4.18E-5
4	1.57E-2	4.08E-2	6.97E-2	1.45E-1	1.00	2.05E-1	9.18E-7	5.39E-4	4.04E-3	3.25E-5	1.24E-5
6	6.17E-4	2.14E-3	2.95E-3	8.02E-3	2.05E-1	1.00	4.09E-6	8.84E-4	6.51E-3	4.40E-5	1.32E-5
12	3.55E-7	4.10E-6	4.36E-6	2.26E-5	9.18E-7	4.09E-6	1.00	1.76E-1	2.78E-1	8.89E-3	2.06E-4
15	3.15E-4	5.17E-4	5.44E-4	7.04E-4	5.39E-4	8.84E-4	1.76E-1	1.00	1.00	9.92E-1	9.15E-3
18	2.29E-3	2.90E-3	3.04E-3	3.44E-3	4.04E-3	6.51E-3	2.78E-1	1.00	1.00	1.00	1.76E-1
21	2.52E-5	6.31E-5	6.59E-5	1.07E-4	3.25E-5	4.40E-5	8.89E-3	9.92E-01	1.00	1.00	1.61E-2
24	1.20E-5	2.67E-5	2.75E-5	4.18E-5	1.24E-5	1.32E-5	2.06E-4	9.15E-3	1.76E-1	1.61E-2	1.00

Table A-4. P-values from ANOVA statistics comparing pulse number conditions

	0	1	4	40	120	480	1440	2880	4320	5760
0	1.00	9.85E-1	1.00	4.01E-1	6.39E-4	3.08E-2	1.60E-3	6.82E-5	2.58E-3	4.16E-4
1	9.85E-1	1.00	9.98E-1	1.96E-1	1.10E-3	1.41E-2	7.55E-4	2.26E-5	2.01E-3	2.72E-4
4	1.00	9.98E-1	1.00	2.01E-1	4.34E-4	2.97E-2	1.80E-3	9.18E-5	2.62E-3	4.48E-4
40	4.01E-1	1.96E-1	2.01E-1	1.00	2.36E-2	9.04E-2	2.20E-3	6.51E-5	3.09E-3	4.60E-4
120	6.39E-4	1.10E-3	4.34E-4	2.36E-2	1.00	7.89E-1	6.91E-3	1.31E-4	4.98E-3	7.32E-4
480	3.08E-2	1.41E-2	2.97E-2	9.04E-2	7.89E-1	1.00	1.68E-2	2.96E-5	4.34E-3	2.87E-4
1440	1.60E-3	7.55E-4	1.80E-3	2.20E-3	6.91E-3	1.68E-2	1.00	1.00E-2	4.86E-2	3.32E-3
2880	6.82E-5	2.26E-5	9.18E-5	6.51E-5	1.31E-4	2.96E-5	1.00E-2	1.00	9.55E-1	3.36E-1
4320	2.58E-3	2.01E-3	2.62E-3	3.09E-3	4.98E-3	4.34E-3	4.86E-2	9.55E-1	1.00	9.99E-1
5760	4.16E-4	2.72E-4	4.48E-4	4.60E-4	7.32E-4	2.87E-4	3.32E-3	3.36E-1	9.99E-1	1.00

Appendix B: Supplemental Information for Prediction of coronavirus receptor usage using protein structure prediction and molecular dynamics

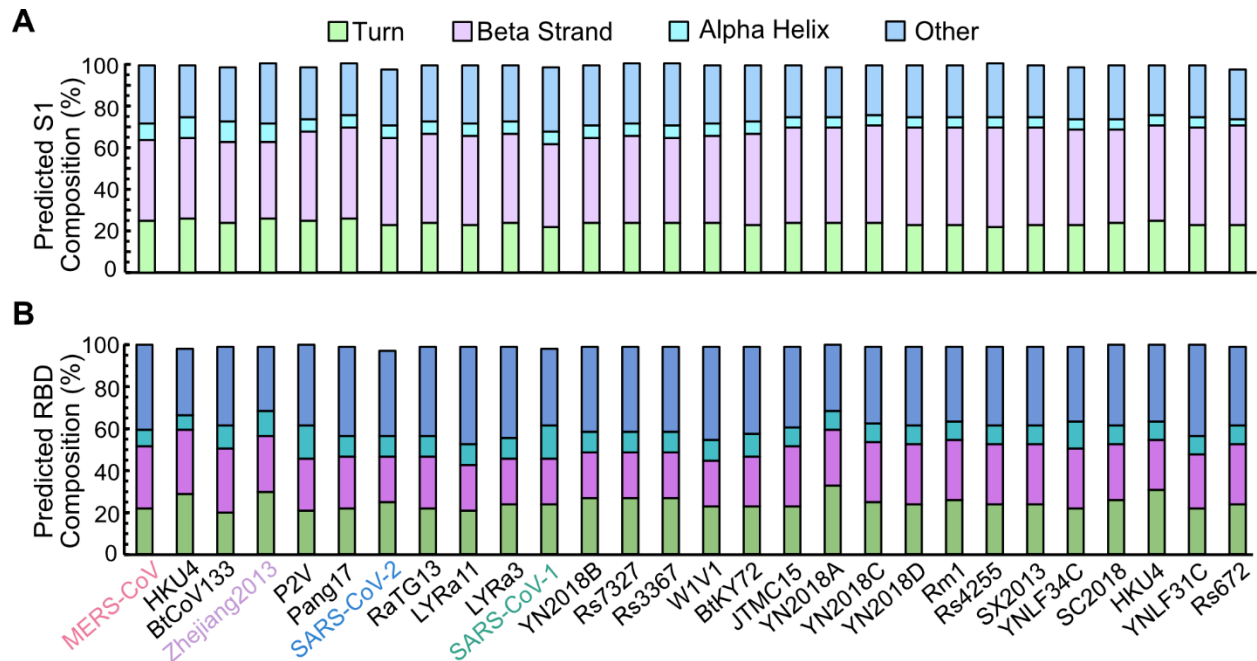


Figure B-1. Secondary structure composition of A) S1 and B) RBD using STRIDE to calculate the percentage of secondary structure.

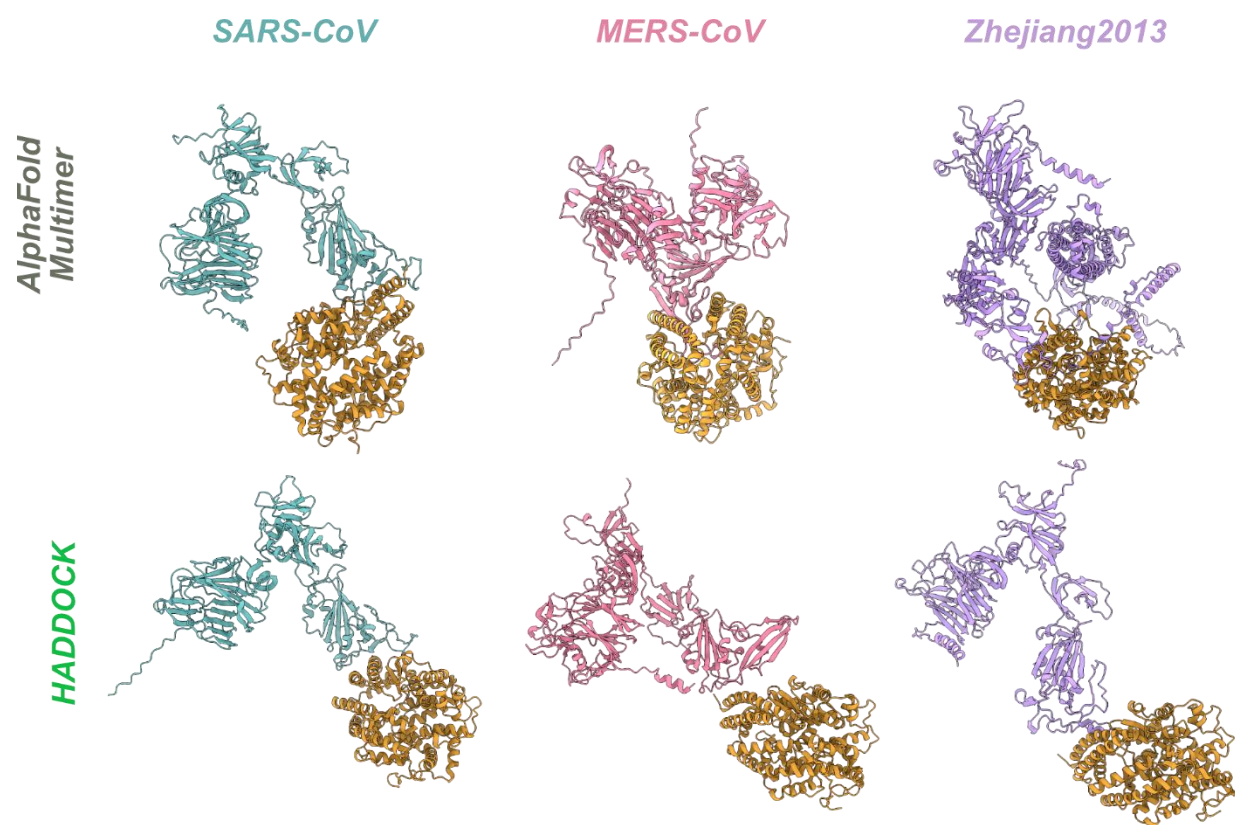


Figure B-2. S1-ACE2 static docking predictions of AlphaFold Multimer and HADDOCK.

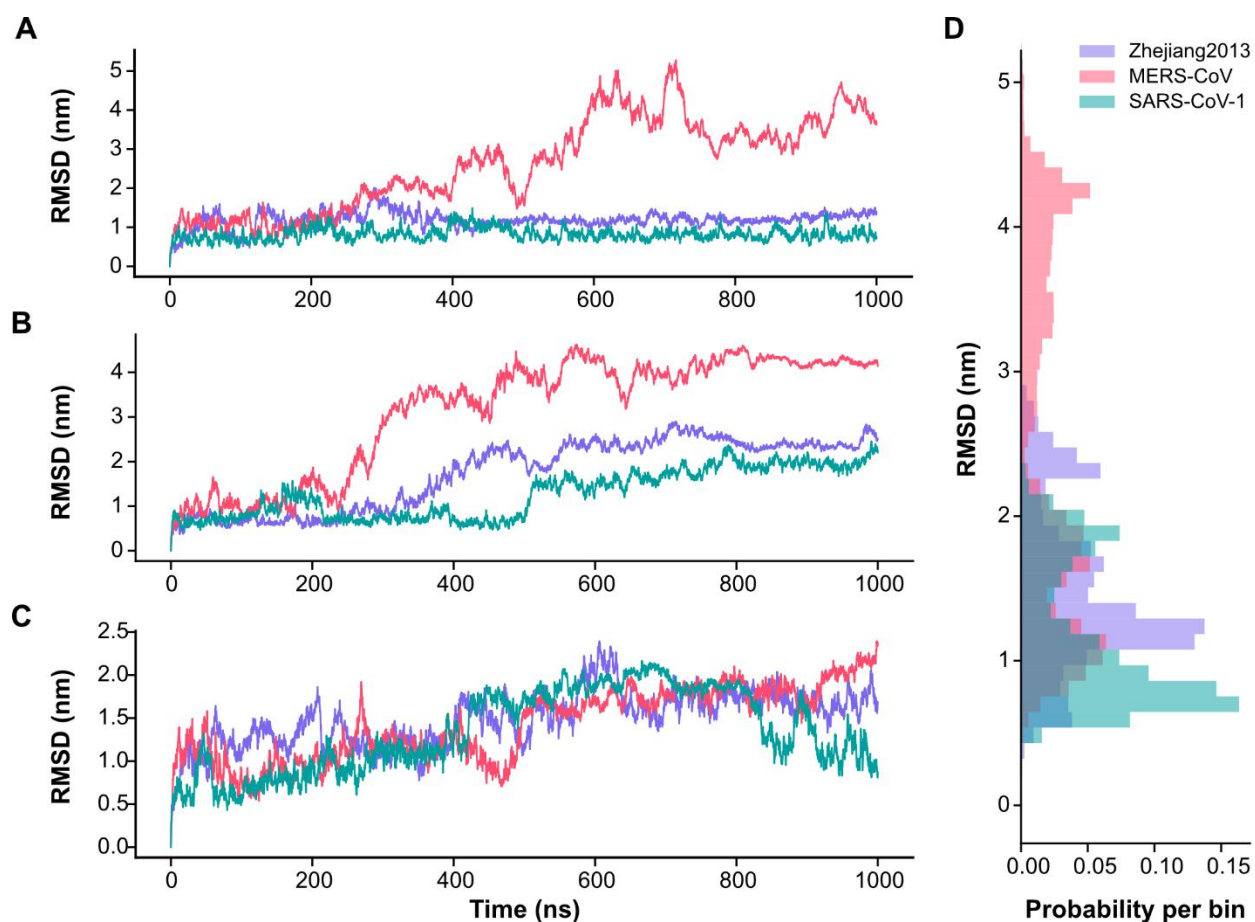


Figure B-3. Root mean squared deviation (RMSD) of coarse-graining simulations. A-C) RMSD plotted as a function of time in nanoseconds across three replicates. D) RMSD plotted against probability highlights the consistency of behavior across replicates. The RMSD of MERS-CoV is variable with low probability indicating its variable behavior over the simulations.

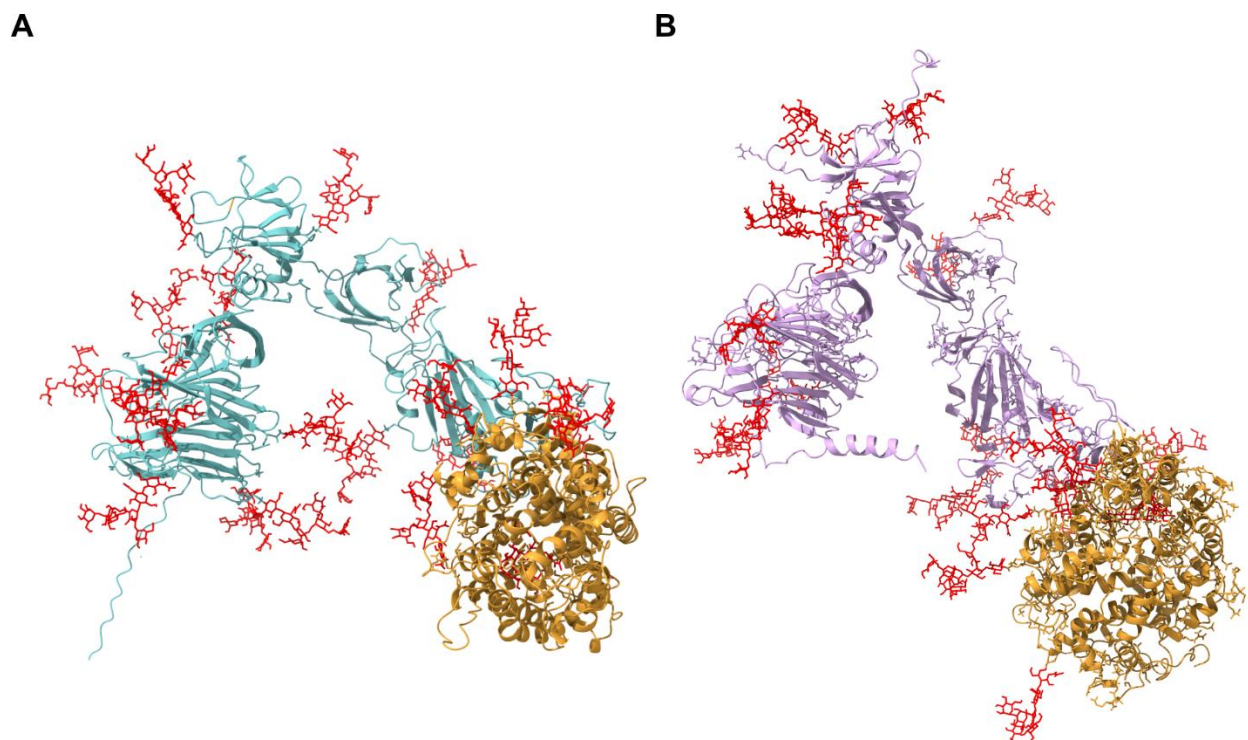


Figure B-4. Glycosylated structures of S1 bound to ACE2 for A) SARS-CoV-1 and B) Zhejiang2013. Glycans, shown in red, were predicted using GLYCAM.

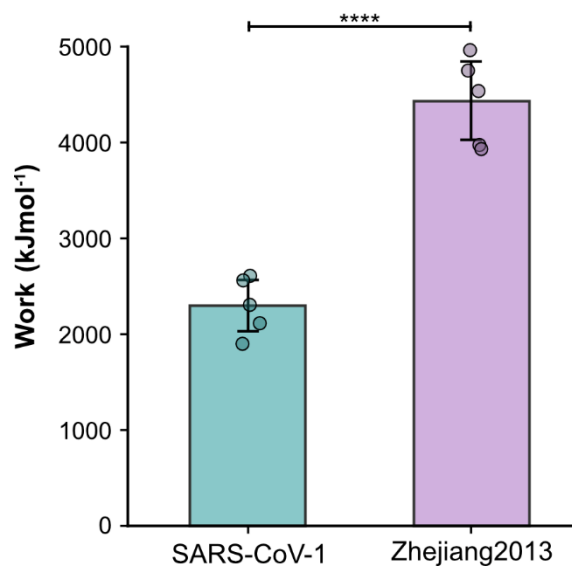


Figure B-5. Work acted upon the pulling simulations at a pulling rate of 1 nm/ns. Data represents the mean of five replicates and error bars represent standard deviation. Significance was calculated with a unpaired t-test, **** $p < 0.0001$.

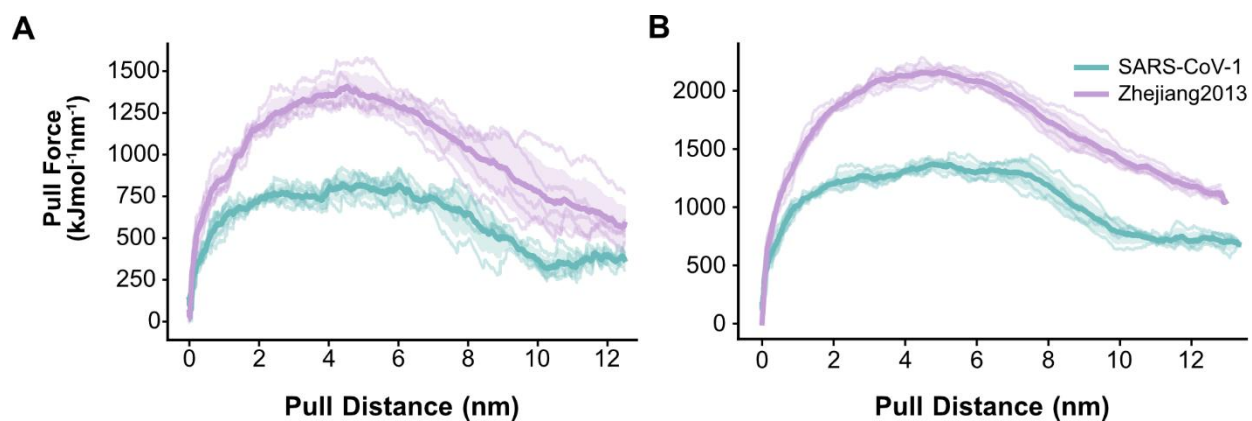


Figure B-6. Traces of pull force as a function of pull distance for all-atom steered MD simulations comparing Zhejiang2013 and SARS-CoV-1. A) Shows a pulling rate of 5 nm/ns and B) 10 nm/ns. We plot five replicates per system and highlight the mean of the replicates and shade the standard deviation region.

Table B-1. Sequence identity matrix of pairwise comparisons of S1 proteins (%)

	MERS	HKU4	BtCoV13 3	Zhejiang201 3	P2V	Pang17	SARS- CoV-2
MERS	100	61.67	61.81	21.45	19.84	19.91	20.34
HKU4	61.67	100	95.59	20.06	22.08	22.15	21.45
BtCoV133	61.81	95.59	100	19.32	22.08	22.15	21.29
Zhejiang201 3	21.45	20.06	19.32	100	26.09	26.28	28.04
P2V	19.84	22.08	22.08	26.09	100	99.85	87.22
Pang17	19.91	22.15	22.15	26.28	99.85	100	86.97
SARS-CoV- 2	20.34	21.45	21.29	28.04	87.22	86.97	100
RaTG13	20	21.14	20.98	28.05	87.96	88.22	95.89
LYRa11	19.75	21.27	21.43	27.05	65.01	65.36	65.31
LYRa3	19.91	21.27	21.43	27.05	65.01	65.36	65.01
SARS-CoV- 1	20.19	21.73	21.88	27.98	66.52	66.87	65.91
YN2018B	19.84	21.21	21.21	28.24	66.82	67.02	67.27
Rs7327	19.84	21.37	21.37	28.24	66.82	67.02	67.27
Rs3367	19.84	21.53	21.53	28.09	66.97	67.17	67.57
W1V1	19.84	21.53	21.53	27.54	66.82	67.15	67.02
BtKY72	19.34	21.75	21.75	25.81	61.09	61.31	61.47
JTMC15	21.2	22.79	22.79	28.55	60.13	61.15	59.59
SC2018	20.03	22.01	22.17	27.41	63.43	64.18	63.04
SX2013	19.71	21.39	21.39	28.1	63.06	63.98	62.67

YNLF31C	20.16	21	21	28.01	62.71	63.64	62.33
YNLF34C	20.16	21	21	28.01	62.71	63.64	62.33
Rs4255	19.36	21.49	21.32	27.56	62.71	63.64	63.09
Rs672	19.36	21.49	21.32	27.56	62.4	63.34	62.79
YN2018D	19.36	21.32	21.32	27.26	62.71	63.64	63.09
Rm1	19.52	21.32	21.16	27.41	63.02	63.79	62.79
HKU3/12	19.81	22.26	22.1	27.52	62.62	63.39	63
YN2018A	20.13	21.77	21.61	27.52	62.92	63.84	63.3
YN2018C	20.16	21.65	21.49	27.86	62.1	63.04	63.09
	RaTG1 3	LYRa1 1	LYRa3	SARS-CoV- 1	YN2018 B	Rs7327	Rs3367
MERS	20	19.75	19.91	20.19	19.84	19.84	19.84
HKU4	21.14	21.27	21.27	21.73	21.21	21.37	21.53
BtCoV133	20.98	21.43	21.43	21.88	21.21	21.37	21.53
Zhejiang201 3	28.05	27.05	27.05	27.98	28.24	28.24	28.09
P2V	87.96	65.01	65.01	66.52	66.82	66.82	66.97
Pang17	88.22	65.36	65.36	66.87	67.02	67.02	67.17
SARS-CoV- 2	95.89	65.31	65.01	65.91	67.27	67.27	67.57
RaTG13	100	66.37	66.06	67.12	67.72	67.72	68.17
LYRa11	66.37	100	98.51	83.96	84.58	83.83	83.53
LYRa3	66.06	98.51	100	83.96	84.13	83.38	83.08
SARS-CoV- 1	67.12	83.96	83.96	100	86.21	86.36	86.06

YN2018B	67.72	84.58	84.13	86.21	100	98.8	97.9
Rs7327	67.72	83.83	83.38	86.36	98.8	100	98.8
Rs3367	68.17	83.53	83.08	86.06	97.9	98.8	100
W1V1	68.02	83.38	82.93	85.91	97.75	98.65	99.85
BtKY72	61.54	63.37	63.83	64.89	64.13	64.44	64.59
JTMC15	59.65	62.44	62.75	62.28	62.44	62.13	62.44
SC2018	63.43	67.08	67.55	67.86	67.24	67.24	67.24
SX2013	62.91	65.89	66.2	65.58	66.05	65.89	66.36
YNLF31C	62.71	65.27	65.58	65.74	66.67	66.82	67.29
YNLF34C	62.71	65.27	65.58	65.74	66.67	66.82	67.29
Rs4255	63.94	66.51	66.82	67.29	66.98	67.13	67.13
Rs672	63.64	66.05	66.2	66.67	66.36	66.51	66.82
YN2018D	63.94	66.2	66.51	66.98	66.67	66.82	67.13
Rm1	63.79	66.82	66.98	66.51	66.98	66.98	67.29
HKU3/12	63.85	67.03	67.18	66.41	67.49	67.34	67.34
YN2018A	64.15	66.41	66.56	65.94	66.56	66.25	66.41
YN2018C	63.94	66.98	67.13	66.2	66.67	66.36	66.67
	W1V1	BtKY7 2	JTMC15	SC2018	SX2013	YNLF31 C	YNLF34C
MERS	19.84	19.34	21.2	20.03	19.71	20.16	20.16
HKU4	21.53	21.75	22.79	22.01	21.39	21	21
BtCoV133	21.53	21.75	22.79	22.17	21.39	21	21
Zhejiang201 3	27.54	25.81	28.55	27.41	28.1	28.01	28.01
P2V	66.82	61.09	60.13	63.43	63.06	62.71	62.71

Pang17	67.15	61.31	61.15	64.18	63.98	63.64	63.64
SARS-CoV-2	67.02	61.47	59.59	63.04	62.67	62.33	62.33
RaTG13	68.02	61.54	59.65	63.43	62.91	62.71	62.71
LYRa11	83.38	63.37	62.44	67.08	65.89	65.27	65.27
LYRa3	82.93	63.83	62.75	67.55	66.2	65.58	65.58
SARS-CoV-1	85.91	64.89	62.28	67.86	65.58	65.74	65.74
YN2018B	97.75	64.13	62.44	67.24	66.05	66.67	66.67
Rs7327	98.65	64.44	62.13	67.24	65.89	66.82	66.82
Rs3367	99.85	64.59	62.44	67.24	66.36	67.29	67.29
W1V1	100	64.66	62.9	67.66	66.67	67.56	67.56
BtKY72	64.66	100	59.67	66.27	65.33	64.69	64.69
JTMC15	62.9	59.67	100	70.44	75.19	74.14	74.14
SC2018	67.66	66.27	70.44	100	78.45	77.19	77.19
SX2013	66.67	65.33	75.19	78.45	100	95.56	95.7
YNLF31C	67.56	64.69	74.14	77.19	95.56	100	99.85
YNLF34C	67.56	64.69	74.14	77.19	95.7	99.85	100
Rs4255	67.41	65.43	72.03	81.48	81.93	81.39	81.39
Rs672	67.12	65.13	72.03	81.19	81.93	81.54	81.54
YN2018D	67.41	65.58	71.88	81.48	82.07	81.54	81.54
Rm1	67.71	65.28	70.98	81.33	86.07	85.23	85.23
HKU3/12	67.76	65.93	71.58	82.4	87.26	86.12	86.12
YN2018A	66.72	65.48	71.58	82.54	87.41	86.12	86.12
YN2018C	66.97	64.99	71.43	80.44	87.41	86.12	86.12

	Rs4255	Rs672	YN2018 D	Rm1	HKU3/1 2	YN2018 A	YN2018C
MERS	19.36	19.36	19.36	19.52	19.81	20.13	20.16
HKU4	21.49	21.49	21.32	21.32	22.26	21.77	21.65
BtCoV133	21.32	21.32	21.32	21.16	22.1	21.61	21.49
Zhejiang2013	27.56	27.56	27.26	27.41	27.52	27.52	27.86
P2V	62.71	62.4	62.71	63.02	62.62	62.92	62.1
Pang17	63.64	63.34	63.64	63.79	63.39	63.84	63.04
SARS-CoV-2	63.09	62.79	63.09	62.79	63	63.3	63.09
RaTG13	63.94	63.64	63.94	63.79	63.85	64.15	63.94
LYRa11	66.51	66.05	66.2	66.82	67.03	66.41	66.98
LYRa3	66.82	66.2	66.51	66.98	67.18	66.56	67.13
SARS-CoV-1	67.29	66.67	66.98	66.51	66.41	65.94	66.2
YN2018B	66.98	66.36	66.67	66.98	67.49	66.56	66.67
Rs7327	67.13	66.51	66.82	66.98	67.34	66.25	66.36
Rs3367	67.13	66.82	67.13	67.29	67.34	66.41	66.67
W1V1	67.41	67.12	67.41	67.71	67.76	66.72	66.97
BtKY72	65.43	65.13	65.58	65.28	65.93	65.48	64.99
JTMC15	72.03	72.03	71.88	70.98	71.58	71.58	71.43
SC2018	81.48	81.19	81.48	81.33	82.4	82.54	80.44
SX2013	81.93	81.93	82.07	86.07	87.26	87.41	87.41
YNLF31C	81.39	81.54	81.54	85.23	86.12	86.12	86.12

YNLF34C	81.39	81.54	81.54	85.23	86.12	86.12	86.12
Rs4255	100	99.11	98.97	86.85	86.85	88.33	88.92
Rs672	99.11	100	99.26	87	87.15	88.63	88.92
YN2018D	98.97	99.26	100	87.15	87	88.48	88.48
Rm1	86.85	87	87.15	100	93.35	93.5	92.91
HKU3/12	86.85	87.15	87	93.35	100	96.46	94.53
YN2018A	88.33	88.63	88.48	93.5	96.46	100	97.19
YN2018C	88.92	88.92	88.48	92.91	94.53	97.19	100

Table B-2. Sequence identity matrix of pairwise comparisons of RBD proteins (%)

	MERS-CoV	HKU4	BtCoV133	Zhejiang2013	JTMC15	SX2013	YNLF31C
MERS-CoV	100	56.62	57.53	19.34	18.08	18.64	18.64
HKU4	56.62	100	94.58	18.13	22.4	22.4	21.88
BtCoV133	57.53	94.58	100	17.58	22.92	22.92	22.4
Zhejiang2013	19.34	18.13	17.58	100	29.11	29.11	29.11
JTMC15	18.08	22.4	22.92	29.11	100	95.1	94.61
SX2013	18.64	22.4	22.92	29.11	95.1	100	98.04
YNLF31C	18.64	21.88	22.4	29.11	94.61	98.04	100
YNLF34C	18.64	21.88	22.4	29.11	94.61	98.04	100
HKU3	20.22	24.35	24.87	27.04	88.24	88.73	88.24
SC2018	20.79	23.83	24.35	27.67	88.73	89.71	88.73
YN2018A	20.79	23.83	24.35	27.04	88.24	89.22	88.24
Rm1	19.77	22.4	22.92	27.22	86.76	87.75	86.76
YN2018C	19.77	23.44	23.96	27.85	88.24	90.2	88.73
Rs672	19.21	23.96	24.48	26.58	87.75	88.73	87.75
YN2018D	19.21	23.96	24.48	26.58	87.75	88.73	87.75
Rs4255	19.21	23.96	24.48	27.22	87.75	89.22	87.75
BtKY72	19.89	22.89	23.38	25.71	67.49	70.44	69.95
SARS-CoV-1	18.09	23.65	24.14	26.7	69.12	69.61	69.61
LYRa11	18.09	23.15	23.65	27.27	68.63	70.59	70.1
LYRa3	18.09	23.15	23.65	27.27	68.63	70.59	70.1
Rs3367	18.62	23.65	24.14	26.14	69.12	71.08	71.57
W1V1	18.62	23.65	24.14	26.14	69.12	71.08	71.57

Rs7327	18.62	23.15	23.65	26.7	69.12	70.59	71.08
YN2018B	18.62	23.15	23.65	26.7	69.12	70.59	71.08
P2V	18.09	23.15	23.65	25.99	70.1	70.59	70.1
Pang17	18.09	23.15	23.65	25.99	70.59	71.08	70.59
SARS-CoV-2	17.55	21.18	21.67	27.68	70.1	70.59	70.59
RaTG13	18.62	21.18	21.67	25.99	69.12	69.61	69.61
	YNLF34C	HKU3	SC2018	YN2018A	Rm1	YN2018C	Rs672
MERS-CoV	18.64	20.22	20.79	20.79	19.77	19.77	19.21
HKU4	21.88	24.35	23.83	23.83	22.4	23.44	23.96
BtCoV133	22.4	24.87	24.35	24.35	22.92	23.96	24.48
Zhejiang2013	29.11	27.04	27.67	27.04	27.22	27.85	26.58
JTMC15	94.61	88.24	88.73	88.24	86.76	88.24	87.75
SX2013	98.04	88.73	89.71	89.22	87.75	90.2	88.73
YNLF31C	100	88.24	88.73	88.24	86.76	88.73	87.75
YNLF34C	100	88.24	88.73	88.24	86.76	88.73	87.75
HKU3	88.24	100	96.59	97.56	92.65	93.14	93.63
SC2018	88.73	96.59	100	98.54	93.63	93.63	93.14
YN2018A	88.24	97.56	98.54	100	94.61	93.63	94.12
Rm1	86.76	92.65	93.63	94.61	100	94.61	94.12
YN2018C	88.73	93.14	93.63	93.63	94.61	100	96.08
Rs672	87.75	93.63	93.14	94.12	94.12	96.08	100
YN2018D	87.75	93.63	93.14	94.12	94.12	96.08	100
Rs4255	87.75	93.63	93.14	94.12	94.61	97.06	99.02
BtKY72	69.95	70.59	71.57	71.08	70.44	71.43	69.95
SARS-CoV-1	69.61	68.78	68.78	68.29	69.12	71.08	68.63

LYRa11	70.1	69.76	69.76	69.27	69.61	72.06	69.61
LYRa3	70.1	69.27	69.76	69.76	69.61	72.06	69.61
Rs3367	71.57	70.24	70.24	69.76	70.59	72.55	70.1
W1V1	71.57	70.24	70.24	69.76	70.59	72.55	70.1
Rs7327	71.08	69.76	69.76	69.27	69.61	71.57	69.12
YN2018B	71.08	69.76	69.76	69.27	69.61	71.57	69.12
P2V	70.1	70.24	72.2	71.71	71.57	71.08	70.1
Pang17	70.59	70.73	72.68	72.2	72.06	71.57	70.59
SARS-CoV-2	70.59	70.73	70.73	70.24	70.59	71.57	69.12
RaTG13	69.61	70.73	70.73	70.24	71.08	71.57	69.12
	YN2018D	Rs4255	BtKY72	SARS-CoV-1	LYRa11	LYRa3	Rs3367
MERS-CoV	19.21	19.21	19.89	18.09	18.09	18.09	18.62
HKU4	23.96	23.96	22.89	23.65	23.15	23.15	23.65
BtCoV133	24.48	24.48	23.38	24.14	23.65	23.65	24.14
Zhejiang2013	26.58	27.22	25.71	26.7	27.27	27.27	26.14
JTMC15	87.75	87.75	67.49	69.12	68.63	68.63	69.12
SX2013	88.73	89.22	70.44	69.61	70.59	70.59	71.08
YNLF31C	87.75	87.75	69.95	69.61	70.1	70.1	71.57
YNLF34C	87.75	87.75	69.95	69.61	70.1	70.1	71.57
HKU3	93.63	93.63	70.59	68.78	69.76	69.27	70.24
SC2018	93.14	93.14	71.57	68.78	69.76	69.76	70.24
YN2018A	94.12	94.12	71.08	68.29	69.27	69.76	69.76
Rm1	94.12	94.61	70.44	69.12	69.61	69.61	70.59
YN2018C	96.08	97.06	71.43	71.08	72.06	72.06	72.55
Rs672	100	99.02	69.95	68.63	69.61	69.61	70.1

YN2018D	100	99.02	69.95	68.63	69.61	69.61	70.1
Rs4255	99.02	100	70.94	69.61	70.59	70.59	71.08
BtKY72	69.95	70.94	100	75.45	75.91	75.91	76.36
SARS-CoV-1	68.63	69.61	75.45	100	94.14	93.69	94.59
LYRa11	69.61	70.59	75.91	94.14	100	99.1	96.4
LYRa3	69.61	70.59	75.91	93.69	99.1	100	95.5
Rs3367	70.1	71.08	76.36	94.59	96.4	95.5	100
W1V1	70.1	71.08	76.36	94.59	96.4	95.5	100
Rs7327	69.12	70.1	76.36	95.5	97.3	96.4	98.2
YN2018B	69.12	70.1	76.36	95.5	97.3	96.4	98.2
P2V	70.1	71.08	72.85	74.77	72.97	72.97	74.77
Pang17	70.59	71.57	73.3	75.23	73.42	73.42	75.23
SARS-CoV-2	69.12	70.1	74.21	73.42	73.87	73.42	75.68
RaTG13	69.12	70.1	73.3	75.23	75.68	75.23	76.58
	W1V1	Rs7327	YN2018B	P2V	Pang17	SARS-CoV-2	RaTG13
MERS-CoV	18.62	18.62	18.62	18.09	18.09	17.55	18.62
HKU4	23.65	23.15	23.15	23.15	23.15	21.18	21.18
BtCoV133	24.14	23.65	23.65	23.65	23.65	21.67	21.67
Zhejiang2013	26.14	26.7	26.7	25.99	25.99	27.68	25.99
JTMC15	69.12	69.12	69.12	70.1	70.59	70.1	69.12
SX2013	71.08	70.59	70.59	70.59	71.08	70.59	69.61
YNLF31C	71.57	71.08	71.08	70.1	70.59	70.59	69.61
YNLF34C	71.57	71.08	71.08	70.1	70.59	70.59	69.61
HKU3	70.24	69.76	69.76	70.24	70.73	70.73	70.73

SC2018	70.24	69.76	69.76	72.2	72.68	70.73	70.73
YN2018A	69.76	69.27	69.27	71.71	72.2	70.24	70.24
Rm1	70.59	69.61	69.61	71.57	72.06	70.59	71.08
YN2018C	72.55	71.57	71.57	71.08	71.57	71.57	71.57
Rs672	70.1	69.12	69.12	70.1	70.59	69.12	69.12
YN2018D	70.1	69.12	69.12	70.1	70.59	69.12	69.12
Rs4255	71.08	70.1	70.1	71.08	71.57	70.1	70.1
BtKY72	76.36	76.36	76.36	72.85	73.3	74.21	73.3
SARS-CoV-1	94.59	95.5	95.5	74.77	75.23	73.42	75.23
LYRa11	96.4	97.3	97.3	72.97	73.42	73.87	75.68
LYRa3	95.5	96.4	96.4	72.97	73.42	73.42	75.23
Rs3367	100	98.2	98.2	74.77	75.23	75.68	76.58
W1V1	100	98.2	98.2	74.77	75.23	75.68	76.58
Rs7327	98.2	100	100	74.32	74.77	75.23	75.68
YN2018B	98.2	100	100	74.32	74.77	75.23	75.68
P2V	74.77	74.32	74.32	100	99.55	86.1	87.44
Pang17	75.23	74.77	74.77	99.55	100	86.55	87.89
SARS-CoV-2	75.68	75.23	75.23	86.1	86.55	100	90.13
RaTG13	76.58	75.68	75.68	87.44	87.89	90.13	100

Table B-3. Pairwise comparison of S1 TM-scores

Protein 1	Protein 2	TM-score
BtCoV133	BtKY72	0.75329
BtCoV133	HKU3	0.74683
BtCoV133	HKU4	0.99472

BtCoV133	JTMC15	0.74163
BtCoV133	LYRa11	0.71625
BtCoV133	LYRa3	0.7109
BtCoV133	MERS-CoV	0.94858
BtCoV133	P2V	0.72734
BtCoV133	Pang17	0.72403
BtCoV133	RM1	0.71927
BtCoV133	RaTG13	0.72616
BtCoV133	Rs3367	0.71841
BtCoV133	Rs4255	0.71525
BtCoV133	Rs672	0.72771
BtCoV133	Rs7327	0.71476
BtCoV133	SARS-CoV-1	0.69108
BtCoV133	SARS-CoV-2	0.70943
BtCoV133	SC2018	0.74739
BtCoV133	SX2013	0.72899
BtCoV133	W1V1	0.7202
BtCoV133	YN2018A	0.72895
BtCoV133	YN2018B	0.71003
BtCoV133	YN2018C	0.73852
BtCoV133	YN2018D	0.7405
BtCoV133	YNLF31C	0.73689
BtCoV133	YNLF34C	0.73532
BtCoV133	Zhejiang2013	0.76697
BtKY72	HKU3	0.94511

BtKY72	HKU4	0.75576
BtKY72	JTMC15	0.94028
BtKY72	LYRa11	0.93143
BtKY72	LYRa3	0.92678
BtKY72	MERS-CoV	0.75668
BtKY72	P2V	0.94881
BtKY72	Pang17	0.94471
BtKY72	RM1	0.92699
BtKY72	RaTG13	0.94955
BtKY72	Rs3367	0.93172
BtKY72	Rs4255	0.92613
BtKY72	Rs672	0.93629
BtKY72	Rs7327	0.9306
BtKY72	SARS-CoV-1	0.90396
BtKY72	SARS-CoV-2	0.92912
BtKY72	SC2018	0.95151
BtKY72	SX2013	0.93711
BtKY72	W1V1	0.9357
BtKY72	YN2018A	0.94427
BtKY72	YN2018B	0.92505
BtKY72	YN2018C	0.95283
BtKY72	YN2018D	0.95104
BtKY72	YNLF31C	0.95103
BtKY72	YNLF34C	0.94433
BtKY72	Zhejiang2013	0.82189

HKU3	HKU4	0.74815
HKU3	JTMC15	0.94553
HKU3	LYRa11	0.93408
HKU3	LYRa3	0.93304
HKU3	MERS-CoV	0.74253
HKU3	P2V	0.94285
HKU3	Pang17	0.94118
HKU3	RM1	0.96515
HKU3	RaTG13	0.9431
HKU3	Rs3367	0.93905
HKU3	Rs4255	0.96088
HKU3	Rs672	0.9718
HKU3	Rs7327	0.93461
HKU3	SARS-CoV-1	0.90319
HKU3	SARS-CoV-2	0.91674
HKU3	SC2018	0.96893
HKU3	SX2013	0.97696
HKU3	W1V1	0.94149
HKU3	YN2018A	0.96021
HKU3	YN2018B	0.93305
HKU3	YN2018C	0.96546
HKU3	YN2018D	0.96435
HKU3	YNLF31C	0.96482
HKU3	YNLF34C	0.98109
HKU3	Zhejiang2013	0.80801

HKU4	JTMC15	0.74212
HKU4	LYRa11	0.71661
HKU4	LYRa3	0.71146
HKU4	MERS-CoV	0.94966
HKU4	P2V	0.72833
HKU4	Pang17	0.72486
HKU4	RM1	0.71999
HKU4	RaTG13	0.72761
HKU4	Rs3367	0.71901
HKU4	Rs4255	0.71733
HKU4	Rs672	0.72929
HKU4	Rs7327	0.71541
HKU4	SARS-CoV-1	0.69312
HKU4	SARS-CoV-2	0.70997
HKU4	SC2018	0.74779
HKU4	SX2013	0.73088
HKU4	W1V1	0.72216
HKU4	YN2018A	0.73068
HKU4	YN2018B	0.71054
HKU4	YN2018C	0.73976
HKU4	YN2018D	0.74135
HKU4	YNLF31C	0.73833
HKU4	YNLF34C	0.73685
HKU4	Zhejiang2013	0.76819
JTMC15	LYRa11	0.92994

JTMC15	LYRa3	0.92902
JTMC15	MERS-CoV	0.73883
JTMC15	P2V	0.93907
JTMC15	Pang17	0.93701
JTMC15	RM1	0.93994
JTMC15	RaTG13	0.93736
JTMC15	Rs3367	0.9324
JTMC15	Rs4255	0.93933
JTMC15	Rs672	0.94881
JTMC15	Rs7327	0.92646
JTMC15	SARS-CoV-1	0.9013
JTMC15	SARS-CoV-2	0.91992
JTMC15	SC2018	0.95922
JTMC15	SX2013	0.94922
JTMC15	W1V1	0.93314
JTMC15	YN2018A	0.95593
JTMC15	YN2018B	0.92851
JTMC15	YN2018C	0.96154
JTMC15	YN2018D	0.96307
JTMC15	YNLF31C	0.96182
JTMC15	YNLF34C	0.94969
JTMC15	Zhejiang2013	0.81177
LYRa11	LYRa3	0.98578
LYRa11	MERS-CoV	0.71383
LYRa11	P2V	0.95333

LYRa11	Pang17	0.95514
LYRa11	RM1	0.94319
LYRa11	RaTG13	0.95184
LYRa11	Rs3367	0.98338
LYRa11	Rs4255	0.94162
LYRa11	Rs672	0.94315
LYRa11	Rs7327	0.9801
LYRa11	SARS-CoV-1	0.95876
LYRa11	SARS-CoV-2	0.93718
LYRa11	SC2018	0.94262
LYRa11	SX2013	0.94559
LYRa11	W1V1	0.98092
LYRa11	YN2018A	0.94316
LYRa11	YN2018B	0.98253
LYRa11	YN2018C	0.94615
LYRa11	YN2018D	0.94689
LYRa11	YNLF31C	0.9448
LYRa11	YNLF34C	0.94125
LYRa11	Zhejiang2013	0.77328
LYRa3	MERS-CoV	0.70889
LYRa3	P2V	0.9525
LYRa3	Pang17	0.95402
LYRa3	RM1	0.94523
LYRa3	RaTG13	0.95176
LYRa3	Rs3367	0.98162

LYRa3	Rs4255	0.94315
LYRa3	Rs672	0.94514
LYRa3	Rs7327	0.98185
LYRa3	SARS-CoV-1	0.96328
LYRa3	SARS-CoV-2	0.93487
LYRa3	SC2018	0.94178
LYRa3	SX2013	0.94838
LYRa3	W1V1	0.97944
LYRa3	YN2018A	0.94659
LYRa3	YN2018B	0.98743
LYRa3	YN2018C	0.94435
LYRa3	YN2018D	0.94475
LYRa3	YNLF31C	0.9454
LYRa3	YNLF34C	0.94258
LYRa3	Zhejiang2013	0.76942
MERS-CoV	P2V	0.72794
MERS-CoV	Pang17	0.72409
MERS-CoV	RM1	0.71411
MERS-CoV	RaTG13	0.72637
MERS-CoV	Rs3367	0.71458
MERS-CoV	Rs4255	0.7103
MERS-CoV	Rs672	0.72593
MERS-CoV	Rs7327	0.71107
MERS-CoV	SARS-CoV-1	0.69333
MERS-CoV	SARS-CoV-2	0.70919

MERS-CoV	SC2018	0.74363
MERS-CoV	SX2013	0.72977
MERS-CoV	W1V1	0.71957
MERS-CoV	YN2018A	0.72854
MERS-CoV	YN2018B	0.70719
MERS-CoV	YN2018C	0.73695
MERS-CoV	YN2018D	0.73813
MERS-CoV	YNLF31C	0.73562
MERS-CoV	YNLF34C	0.73417
MERS-CoV	Zhejiang2013	0.76808
P2V	Pang17	0.99306
P2V	RM1	0.93803
P2V	RaTG13	0.9952
P2V	Rs3367	0.95646
P2V	Rs4255	0.93647
P2V	Rs672	0.94136
P2V	Rs7327	0.9532
P2V	SARS-CoV-1	0.9252
P2V	SARS-CoV-2	0.97638
P2V	SC2018	0.96054
P2V	SX2013	0.94355
P2V	W1V1	0.95841
P2V	YN2018A	0.95706
P2V	YN2018B	0.95475
P2V	YN2018C	0.95767

P2V	YN2018D	0.95622
P2V	YNLF31C	0.95979
P2V	YNLF34C	0.94346
P2V	Zhejiang2013	0.79731
Pang17	RM1	0.93887
Pang17	RaTG13	0.99072
Pang17	Rs3367	0.95897
Pang17	Rs4255	0.93674
Pang17	Rs672	0.9394
Pang17	Rs7327	0.95621
Pang17	SARS-CoV-1	0.92559
Pang17	SARS-CoV-2	0.96985
Pang17	SC2018	0.96091
Pang17	SX2013	0.94234
Pang17	W1V1	0.96043
Pang17	YN2018A	0.95597
Pang17	YN2018B	0.95719
Pang17	YN2018C	0.95678
Pang17	YN2018D	0.95552
Pang17	YNLF31C	0.95952
Pang17	YNLF34C	0.94116
Pang17	Zhejiang2013	0.79469
RM1	RaTG13	0.93678
RM1	Rs3367	0.94684
RM1	Rs4255	0.98798

RM1	Rs672	0.9812
RM1	Rs7327	0.94902
RM1	SARS-CoV-1	0.92586
RM1	SARS-CoV-2	0.92466
RM1	SC2018	0.95634
RM1	SX2013	0.97298
RM1	W1V1	0.94513
RM1	YN2018A	0.9637
RM1	YN2018B	0.94708
RM1	YN2018C	0.96452
RM1	YN2018D	0.96493
RM1	YNLF31C	0.96348
RM1	YNLF34C	0.97631
RM1	Zhejiang2013	0.77889
RaTG13	Rs3367	0.95306
RaTG13	Rs4255	0.93548
RaTG13	Rs672	0.94311
RaTG13	Rs7327	0.94936
RaTG13	SARS-CoV-1	0.92935
RaTG13	SARS-CoV-2	0.97482
RaTG13	SC2018	0.95742
RaTG13	SX2013	0.94674
RaTG13	W1V1	0.95578
RaTG13	YN2018A	0.96061
RaTG13	YN2018B	0.95297

RaTG13	YN2018C	0.95798
RaTG13	YN2018D	0.95582
RaTG13	YNLF31C	0.96202
RaTG13	YNLF34C	0.94555
RaTG13	Zhejiang2013	0.79564
Rs3367	Rs4255	0.94495
Rs3367	Rs672	0.94557
Rs3367	Rs7327	0.98873
Rs3367	SARS-CoV-1	0.95529
Rs3367	SARS-CoV-2	0.93811
Rs3367	SC2018	0.94768
Rs3367	SX2013	0.94696
Rs3367	W1V1	0.99671
Rs3367	YN2018A	0.94501
Rs3367	YN2018B	0.98303
Rs3367	YN2018C	0.94692
Rs3367	YN2018D	0.94776
Rs3367	YNLF31C	0.94622
Rs3367	YNLF34C	0.94352
Rs3367	Zhejiang2013	0.77657
Rs4255	Rs672	0.98934
Rs4255	Rs7327	0.94671
Rs4255	SARS-CoV-1	0.92378
Rs4255	SARS-CoV-2	0.92451
Rs4255	SC2018	0.95288

Rs4255	SX2013	0.97626
Rs4255	W1V1	0.94282
Rs4255	YN2018A	0.96158
Rs4255	YN2018B	0.94473
Rs4255	YN2018C	0.96253
Rs4255	YN2018D	0.96383
Rs4255	YNLF31C	0.9604
Rs4255	YNLF34C	0.97869
Rs4255	Zhejiang2013	0.77808
Rs672	Rs7327	0.94256
Rs672	SARS-CoV-1	0.92405
Rs672	SARS-CoV-2	0.92455
Rs672	SC2018	0.95958
Rs672	SX2013	0.98708
Rs672	W1V1	0.94596
Rs672	YN2018A	0.96993
Rs672	YN2018B	0.94425
Rs672	YN2018C	0.97011
Rs672	YN2018D	0.97044
Rs672	YNLF31C	0.96954
Rs672	YNLF34C	0.98674
Rs672	Zhejiang2013	0.79076
Rs7327	SARS-CoV-1	0.95949
Rs7327	SARS-CoV-2	0.93754
Rs7327	SC2018	0.9439

Rs7327	SX2013	0.94204
Rs7327	W1V1	0.98503
Rs7327	YN2018A	0.94041
Rs7327	YN2018B	0.98105
Rs7327	YN2018C	0.94152
Rs7327	YN2018D	0.94205
Rs7327	YNLF31C	0.94103
Rs7327	YNLF34C	0.94363
Rs7327	Zhejiang2013	0.7721
SARS-CoV-1	SARS-CoV-2	0.94072
SARS-CoV-1	SC2018	0.91224
SARS-CoV-1	SX2013	0.92685
SARS-CoV-1	W1V1	0.95525
SARS-CoV-1	YN2018A	0.92756
SARS-CoV-1	YN2018B	0.95854
SARS-CoV-1	YN2018C	0.91716
SARS-CoV-1	YN2018D	0.91796
SARS-CoV-1	YNLF31C	0.92034
SARS-CoV-1	YNLF34C	0.91976
SARS-CoV-1	Zhejiang2013	0.74254
SARS-CoV-2	SC2018	0.9376
SARS-CoV-2	SX2013	0.9223
SARS-CoV-2	W1V1	0.9384
SARS-CoV-2	YN2018A	0.93819
SARS-CoV-2	YN2018B	0.93827

SARS-CoV-2	YN2018C	0.93641
SARS-CoV-2	YN2018D	0.93593
SARS-CoV-2	YNLF31C	0.93844
SARS-CoV-2	YNLF34C	0.92416
SARS-CoV-2	Zhejiang2013	0.77318
SC2018	SX2013	0.96345
SC2018	W1V1	0.94972
SC2018	YN2018A	0.97745
SC2018	YN2018B	0.9439
SC2018	YN2018C	0.98513
SC2018	YN2018D	0.98388
SC2018	YNLF31C	0.98131
SC2018	YNLF34C	0.96421
SC2018	Zhejiang2013	0.81026
SX2013	W1V1	0.9489
SX2013	YN2018A	0.97251
SX2013	YN2018B	0.94648
SX2013	YN2018C	0.97171
SX2013	YN2018D	0.97077
SX2013	YNLF31C	0.97384
SX2013	YNLF34C	0.99225
SX2013	Zhejiang2013	0.79155
W1V1	YN2018A	0.94642
W1V1	YN2018B	0.98025
W1V1	YN2018C	0.94789

W1V1	YN2018D	0.94824
W1V1	YNLF31C	0.94761
W1V1	YNLF34C	0.94507
W1V1	Zhejiang2013	0.78034
YN2018A	YN2018B	0.94667
YN2018A	YN2018C	0.98641
YN2018A	YN2018D	0.98388
YN2018A	YNLF31C	0.99436
YN2018A	YNLF34C	0.97029
YN2018A	Zhejiang2013	0.79568
YN2018B	YN2018C	0.94413
YN2018B	YN2018D	0.94526
YN2018B	YNLF31C	0.94512
YN2018B	YNLF34C	0.94095
YN2018B	Zhejiang2013	0.76787
YN2018C	YN2018D	0.99564
YN2018C	YNLF31C	0.98968
YN2018C	YNLF34C	0.97124
YN2018C	Zhejiang2013	0.80618
YN2018D	YNLF31C	0.98629
YN2018D	YNLF34C	0.97072
YN2018D	Zhejiang2013	0.80603
YNLF31C	YNLF34C	0.9726
YNLF31C	Zhejiang2013	0.80514
YNLF34C	Zhejiang2013	0.7969

Table B-4. Pairwise comparison of RBD TM-scores

Protein1	Protein2	TM-score
BtCoV133	BtKY72	0.69037
BtCoV133	HKU3	0.72603
BtCoV133	HKU4	0.98921
BtCoV133	JTMC15	0.71868
BtCoV133	LYRA11	0.69272
BtCoV133	LYRA3	0.69228
BtCoV133	MERS-CoV	0.92478
BtCoV133	P2V	0.68388
BtCoV133	Pang17	0.68419
BtCoV133	RATG13	0.68582
BtCoV133	Rm1	0.72725
BtCoV133	Rs3367	0.69474
BtCoV133	Rs4255	0.72733
BtCoV133	Rs672	0.72748
BtCoV133	Rs7327	0.69351
BtCoV133	SARS-CoV-1	0.69403
BtCoV133	SARS-CoV-2	0.68722
BtCoV133	SC2018	0.72536
BtCoV133	SX2013	0.72191
BtCoV133	W1V1	0.69355
BtCoV133	YN2018A	0.72245
BtCoV133	YN2018B	0.69356

BtCoV133	YN2018C	0.72066
BtCoV133	YN2018D	0.72458
BtCoV133	YNLF31C	0.72114
BtCoV133	YNLF34C	0.72106
BtCoV133	Zhejiang2013	0.56052
BtKY72	HKU3	0.92911
BtKY72	HKU4	0.68442
BtKY72	JTMC15	0.91961
BtKY72	LYRA11	0.97807
BtKY72	LYRA3	0.97802
BtKY72	MERS-CoV	0.67496
BtKY72	P2V	0.96722
BtKY72	Pang17	0.96713
BtKY72	RATG13	0.9665
BtKY72	Rm1	0.92859
BtKY72	Rs3367	0.9771
BtKY72	Rs4255	0.93069
BtKY72	Rs672	0.9311
BtKY72	Rs7327	0.97702
BtKY72	SARS-CoV-1	0.97668
BtKY72	SARS-CoV-2	0.9696
BtKY72	SC2018	0.9282
BtKY72	SX2013	0.92181
BtKY72	W1V1	0.97782
BtKY72	YN2018A	0.92917

BtKY72	YN2018B	0.97758
BtKY72	YN2018C	0.93018
BtKY72	YN2018D	0.93182
BtKY72	YNLF31C	0.91462
BtKY72	YNLF34C	0.93103
BtKY72	Zhejiang2013	0.59719
HKU3	HKU4	0.72168
HKU3	JTMC15	0.97604
HKU3	LYRA11	0.93136
HKU3	LYRA3	0.93163
HKU3	MERS-CoV	0.70275
HKU3	P2V	0.92716
HKU3	Pang17	0.92771
HKU3	RATG13	0.92847
HKU3	Rm1	0.99488
HKU3	Rs3367	0.93196
HKU3	Rs4255	0.99506
HKU3	Rs672	0.99466
HKU3	Rs7327	0.93184
HKU3	SARS-CoV-1	0.9321
HKU3	SARS-CoV-2	0.93401
HKU3	SC2018	0.9983
HKU3	SX2013	0.98169
HKU3	W1V1	0.93099
HKU3	YN2018A	0.99552

HKU3	YN2018B	0.93158
HKU3	YN2018C	0.98463
HKU3	YN2018D	0.99182
HKU3	YNLF31C	0.9753
HKU3	YNLF34C	0.99287
HKU3	Zhejiang2013	0.61841
HKU4	JTMC15	0.71443
HKU4	LYRA11	0.68716
HKU4	LYRA3	0.68735
HKU4	MERS-CoV	0.91809
HKU4	P2V	0.67874
HKU4	Pang17	0.67892
HKU4	RATG13	0.67992
HKU4	Rm1	0.72285
HKU4	Rs3367	0.68922
HKU4	Rs4255	0.72188
HKU4	Rs672	0.72231
HKU4	Rs7327	0.68856
HKU4	SARS-CoV-1	0.68842
HKU4	SARS-CoV-2	0.68338
HKU4	SC2018	0.72158
HKU4	SX2013	0.71816
HKU4	W1V1	0.68785
HKU4	YN2018A	0.71979
HKU4	YN2018B	0.68784

HKU4	YN2018C	0.71482
HKU4	YN2018D	0.72032
HKU4	YNLF31C	0.71664
HKU4	YNLF34C	0.71643
HKU4	Zhejiang2013	0.55357
JTMC15	LYRA11	0.92177
JTMC15	LYRA3	0.92196
JTMC15	MERS-CoV	0.69513
JTMC15	P2V	0.91904
JTMC15	Pang17	0.91964
JTMC15	RATG13	0.92042
JTMC15	Rm1	0.97676
JTMC15	Rs3367	0.9222
JTMC15	Rs4255	0.97818
JTMC15	Rs672	0.97806
JTMC15	Rs7327	0.92216
JTMC15	SARS-CoV-1	0.92242
JTMC15	SARS-CoV-2	0.92613
JTMC15	SC2018	0.97548
JTMC15	SX2013	0.98062
JTMC15	W1V1	0.92115
JTMC15	YN2018A	0.97296
JTMC15	YN2018B	0.92182
JTMC15	YN2018C	0.98329
JTMC15	YN2018D	0.97551

JTMC15	YNLF31C	0.99531
JTMC15	YNLF34C	0.97666
JTMC15	Zhejiang2013	0.6289
LYRA11	LYRA3	0.9999
LYRA11	MERS-CoV	0.67387
LYRA11	P2V	0.98278
LYRA11	Pang17	0.98285
LYRA11	RATG13	0.98299
LYRA11	Rm1	0.9298
LYRA11	Rs3367	0.99955
LYRA11	Rs4255	0.93231
LYRA11	Rs672	0.93326
LYRA11	Rs7327	0.99946
LYRA11	SARS-CoV-1	0.99899
LYRA11	SARS-CoV-2	0.9834
LYRA11	SC2018	0.93044
LYRA11	SX2013	0.92457
LYRA11	W1V1	0.99968
LYRA11	YN2018A	0.93063
LYRA11	YN2018B	0.99978
LYRA11	YN2018C	0.93118
LYRA11	YN2018D	0.93217
LYRA11	YNLF31C	0.9159
LYRA11	YNLF34C	0.93024
LYRA11	Zhejiang2013	0.60749

LYRA3	MERS-CoV	0.67359
LYRA3	P2V	0.98296
LYRA3	Pang17	0.98306
LYRA3	RATG13	0.98352
LYRA3	Rm1	0.93012
LYRA3	Rs3367	0.99963
LYRA3	Rs4255	0.93239
LYRA3	Rs672	0.93353
LYRA3	Rs7327	0.99961
LYRA3	SARS-CoV-1	0.99919
LYRA3	SARS-CoV-2	0.9837
LYRA3	SC2018	0.93098
LYRA3	SX2013	0.9255
LYRA3	W1V1	0.99946
LYRA3	YN2018A	0.93148
LYRA3	YN2018B	0.99961
LYRA3	YN2018C	0.93105
LYRA3	YN2018D	0.93257
LYRA3	YNLF31C	0.91596
LYRA3	YNLF34C	0.93049
LYRA3	Zhejiang2013	0.60759
MERS-CoV	P2V	0.66737
MERS-CoV	Pang17	0.66701
MERS-CoV	RATG13	0.66628
MERS-CoV	Rm1	0.70111

MERS-CoV	Rs3367	0.6742
MERS-CoV	Rs4255	0.70256
MERS-CoV	Rs672	0.70118
MERS-CoV	Rs7327	0.67276
MERS-CoV	SARS-CoV-1	0.67369
MERS-CoV	SARS-CoV-2	0.66788
MERS-CoV	SC2018	0.70012
MERS-CoV	SX2013	0.69478
MERS-CoV	W1V1	0.67474
MERS-CoV	YN2018A	0.69813
MERS-CoV	YN2018B	0.67371
MERS-CoV	YN2018C	0.69981
MERS-CoV	YN2018D	0.69893
MERS-CoV	YNLF31C	0.69773
MERS-CoV	YNLF34C	0.70006
MERS-CoV	Zhejiang2013	0.5615
P2V	Pang17	0.99993
P2V	RATG13	0.99848
P2V	Rm1	0.9256
P2V	Rs3367	0.98201
P2V	Rs4255	0.92783
P2V	Rs672	0.929
P2V	Rs7327	0.98194
P2V	SARS-CoV-1	0.98174
P2V	SARS-CoV-2	0.98912

P2V	SC2018	0.92653
P2V	SX2013	0.92216
P2V	W1V1	0.98242
P2V	YN2018A	0.92751
P2V	YN2018B	0.98213
P2V	YN2018C	0.92739
P2V	YN2018D	0.92735
P2V	YNLF31C	0.91305
P2V	YNLF34C	0.92633
P2V	Zhejiang2013	0.60511
Pang17	RATG13	0.99872
Pang17	Rm1	0.92627
Pang17	Rs3367	0.98219
Pang17	Rs4255	0.92834
Pang17	Rs672	0.92958
Pang17	Rs7327	0.9822
Pang17	SARS-CoV-1	0.98192
Pang17	SARS-CoV-2	0.98954
Pang17	SC2018	0.92726
Pang17	SX2013	0.92301
Pang17	W1V1	0.98233
Pang17	YN2018A	0.92835
Pang17	YN2018B	0.98216
Pang17	YN2018C	0.92774
Pang17	YN2018D	0.92799

Pang17	YNLF31C	0.91354
Pang17	YNLF34C	0.92683
Pang17	Zhejiang2013	0.60512
RATG13	Rm1	0.92764
RATG13	Rs3367	0.98334
RATG13	Rs4255	0.92927
RATG13	Rs672	0.93094
RATG13	Rs7327	0.98361
RATG13	SARS-CoV-1	0.98302
RATG13	SARS-CoV-2	0.99083
RATG13	SC2018	0.9287
RATG13	SX2013	0.92479
RATG13	W1V1	0.98266
RATG13	YN2018A	0.92935
RATG13	YN2018B	0.98292
RATG13	YN2018C	0.92776
RATG13	YN2018D	0.92949
RATG13	YNLF31C	0.9142
RATG13	YNLF34C	0.92722
RATG13	Zhejiang2013	0.60645
Rm1	Rs3367	0.93107
Rm1	Rs4255	0.99828
Rm1	Rs672	0.9983
Rm1	Rs7327	0.93099
Rm1	SARS-CoV-1	0.93086

Rm1	SARS-CoV-2	0.93402
Rm1	SC2018	0.99545
Rm1	SX2013	0.98478
Rm1	W1V1	0.92935
Rm1	YN2018A	0.99409
Rm1	YN2018B	0.93021
Rm1	YN2018C	0.98648
Rm1	YN2018D	0.99716
Rm1	YNLF31C	0.97411
Rm1	YNLF34C	0.9957
Rm1	Zhejiang2013	0.62056
Rs3367	Rs4255	0.93302
Rs3367	Rs672	0.93444
Rs3367	Rs7327	0.9999
Rs3367	SARS-CoV-1	0.9992
Rs3367	SARS-CoV-2	0.98393
Rs3367	SC2018	0.93184
Rs3367	SX2013	0.92619
Rs3367	W1V1	0.99933
Rs3367	YN2018A	0.93202
Rs3367	YN2018B	0.99961
Rs3367	YN2018C	0.93085
Rs3367	YN2018D	0.93326
Rs3367	YNLF31C	0.91607
Rs3367	YNLF34C	0.93047

Rs3367	Zhejiang2013	0.60982
Rs4255	Rs672	0.99927
Rs4255	Rs7327	0.93273
Rs4255	SARS-CoV-1	0.93307
Rs4255	SARS-CoV-2	0.93594
Rs4255	SC2018	0.99438
Rs4255	SX2013	0.9853
Rs4255	W1V1	0.93214
Rs4255	YN2018A	0.992
Rs4255	YN2018B	0.93269
Rs4255	YN2018C	0.99078
Rs4255	YN2018D	0.99722
Rs4255	YNLF31C	0.97518
Rs4255	YNLF34C	0.9966
Rs4255	Zhejiang2013	0.62146
Rs672	Rs7327	0.93424
Rs672	SARS-CoV-1	0.93432
Rs672	SARS-CoV-2	0.9371
Rs672	SC2018	0.99513
Rs672	SX2013	0.98668
Rs672	W1V1	0.93296
Rs672	YN2018A	0.99329
Rs672	YN2018B	0.93367
Rs672	YN2018C	0.98969
Rs672	YN2018D	0.99767

Rs672	YNLF31C	0.97395
Rs672	YNLF34C	0.9955
Rs672	Zhejiang2013	0.62076
Rs7327	SARS-CoV-1	0.99916
Rs7327	SARS-CoV-2	0.98383
Rs7327	SC2018	0.93191
Rs7327	SX2013	0.92652
Rs7327	W1V1	0.99901
Rs7327	YN2018A	0.93226
Rs7327	YN2018B	0.99948
Rs7327	YN2018C	0.93058
Rs7327	YN2018D	0.93327
Rs7327	YNLF31C	0.91586
Rs7327	YNLF34C	0.93036
Rs7327	Zhejiang2013	0.60756
SARS-CoV-1	SARS-CoV-2	0.98385
SARS-CoV-1	SC2018	0.9317
SARS-CoV-1	SX2013	0.92686
SARS-CoV-1	W1V1	0.99874
SARS-CoV-1	YN2018A	0.93214
SARS-CoV-1	YN2018B	0.99901
SARS-CoV-1	YN2018C	0.93128
SARS-CoV-1	YN2018D	0.93325
SARS-CoV-1	YNLF31C	0.91706
SARS-CoV-1	YNLF34C	0.93139

SARS-CoV-1	Zhejiang2013	0.61225
SARS-CoV-2	SC2018	0.9345
SARS-CoV-2	SX2013	0.93067
SARS-CoV-2	W1V1	0.98313
SARS-CoV-2	YN2018A	0.93431
SARS-CoV-2	YN2018B	0.98348
SARS-CoV-2	YN2018C	0.93507
SARS-CoV-2	YN2018D	0.93595
SARS-CoV-2	YNLF31C	0.92057
SARS-CoV-2	YNLF34C	0.93439
SARS-CoV-2	Zhejiang2013	0.60864
SC2018	SX2013	0.98457
SC2018	W1V1	0.9297
SC2018	YN2018A	0.99764
SC2018	YN2018B	0.93061
SC2018	YN2018C	0.98328
SC2018	YN2018D	0.99225
SC2018	YNLF31C	0.97281
SC2018	YNLF34C	0.99183
SC2018	Zhejiang2013	0.61876
SX2013	W1V1	0.92359
SX2013	YN2018A	0.9832
SX2013	YN2018B	0.92452
SX2013	YN2018C	0.99111
SX2013	YN2018D	0.98441

SX2013	YNLF31C	0.97936
SX2013	YNLF34C	0.98638
SX2013	Zhejiang2013	0.62217
W1V1	YN2018A	0.92947
W1V1	YN2018B	0.99979
W1V1	YN2018C	0.93099
W1V1	YN2018D	0.93165
W1V1	YNLF31C	0.91579
W1V1	YNLF34C	0.92993
W1V1	Zhejiang2013	0.61255
YN2018A	YN2018B	0.93034
YN2018A	YN2018C	0.98035
YN2018A	YN2018D	0.99273
YN2018A	YNLF31C	0.96884
YN2018A	YNLF34C	0.99038
YN2018A	Zhejiang2013	0.61939
YN2018B	YN2018C	0.93136
YN2018B	YN2018D	0.9325
YN2018B	YNLF31C	0.91617
YN2018B	YNLF34C	0.93045
YN2018B	Zhejiang2013	0.60754
YN2018C	YN2018D	0.9875
YN2018C	YNLF31C	0.98192
YN2018C	YNLF34C	0.98948
YN2018C	Zhejiang2013	0.62389

YN2018D	YNLF31C	0.97096
YN2018D	YNLF34C	0.99504
YN2018D	Zhejiang2013	0.62139
YNLF31C	YNLF34C	0.9762
YNLF31C	Zhejiang2013	0.62559
YNLF34C	Zhejiang2013	0.62012

Table B-5. Secondary structure composition of S1 and RBD calculated by STRIDE

	S1				RBD			
	Turn (%)	Beta Strand (%)	Alpha Helix (%)	Other (%)	Turn (%)	Beta Strand (%)	Alpha Helix (%)	Other (%)
SX2013	23	47	5	25	24	29	9	38
YNLF31C	23	47	5	25	22	26	9	44
YNLF34C	23	46	5	25	22	29	13	36
SC2018	24	45	5	26	26	27	9	39
Rs4255	22	48	5	26	24	29	9	38
Rs672	23	48	3	24	24	29	9	38
YN2018D	23	47	5	25	24	29	9	38
Rm1	23	47	5	25	26	29	9	36
HKU3	25	46	5	24	31	24	9	37
YN2018A	24	46	5	24	33	27	9	32
YN2018C	24	47	5	24	25	29	9	37
Rs3367	24	41	6	30	27	22	10	41

W1V1	24	42	6	28	23	22	10	45
YN2018B	24	41	6	29	27	22	10	41
Rs7327	24	42	6	29	27	22	10	41
SARS-CoV	22	40	6	31	24	22	16	37
LYRa11	23	43	6	28	21	22	10	47
LYRA3	24	43	6	27	24	22	10	44
P2V	25	43	6	25	21	25	16	39
Pang17	26	44	6	25	22	25	10	43
SARS-CoV-2	23	42	6	27	25	22	10	41
RATG13	24	43	6	27	22	25	10	43
BtKY72	23	44	6	27	23	24	11	42
JTMC15	24	46	5	25	23	29	9	39
Zhejiang2013	26	37	9	29	30	27	12	31
MERS-CoV	25	39	8	28	22	30	8	41
HKU4	26	39	10	25	29	31	7	32
BtCoV133	24	39	10	26	20	31	11	38

Table B-7. Residues used for static docking restraints

Name	ACE2-16	ACE2-16	ACE2-336
MERS-CoV	Y499	G538	V561
HKU4	Y503	G543	V566
BtCoV133	Y503	G543	V566
Zhejiang2013	Y477	N535	F546
P2V	Y453	E493	Y505

Pang17	Y451	E491	Y503
SARS-CoV-2	Y453	Q493	Y505
RATG13	Y453	Y493	H505
LYRa11	Y444	N483	Y495
LYRa3	Y444	N483	Y495
SARS-CoV-1	Y440	N479	Y491
YN2018B	Y441	N480	Y492
Rs7327	Y441	N480	Y492
Rs3367	Y441	N480	Y492
W1V1	Y441	N480	Y492
BtKY72	Y442	K 482	Y 494
JTMC15	Y434	S460	Y472
SC2018	Y438	S465	Y477
SX2013	Y437	S463	Y475
YNLF31C	Y439	Y440	Y477
YNLF34C	Y439	S465	Y477
Rs4255	Y439	S465	Y477
Rs672	Y439	S465	Y477
YN2018D	Y439	S465	Y477
Rm1	Y439	S465	Y477
HKU3/12	Y439	S466	Y478
YN2018A	Y439	S466	Y478
YN2018C	Y439	S465	Y477

Table B-8. Number of contacts calculated by protein docking platform

Name	ClusPro	Haddock	Multimer
Zhejiang2013	178	81	168
MERS-CoV	82	99	586
HKU4	100	90	1837
BtCoV133	62	101	434
P2V	120	73	98
Pang17	159	87	66
SARS-CoV-2	121	68	82
RATG13	147	65	149
LYRa11	142	89	73
LYRa3	107	135	86
SARS-CoV-1	147	90	88
YN2018B	156	109	93
Rs7327	132	112	96
Rs3367	120	98	84
W1V1	193	69	61
BtKY72	124	103	118
JTMC15	155	87	58
Rs4255	78	75	1782
YN2018D	97	74	57
Rm1	111	101	66
YN2018A	97	62	163
YN2018C	83	95	783
SC2018	95	69	39
SX2013	119	83	179

YNLF31C	76	60	288
YNLF34C	89	50	425
Rs672	109	81	93
HKU3/12	85	44	87