Title
Studies in Computational Biophysics: SARS-CoV-2 and Plant Cell Plate Maturation

Author
Jawaid, Muhammad Zaki

Publication Date
2022

Peer reviewed|Thesis/dissertation

Studies in Computational Biophysics: SARS-CoV-2 and Plant Cell Plate Maturation

By

Muhammad Zaki Jawaid

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

Daniel L. Cox, Chair

Georgia Drakakaki

Rajiv R. P. Singh

Committee in Charge

2022

For Ammi, Abbu,
Adeel, Mony, and Sidra.

Contents

List of Figures

List of Tables

Abstract

Studies in Computational Biophysics: SARS-CoV-2 and Plant Cell Plate Maturation

This dissertation is split into two parts. In part 1, we discuss plant cytokinesis, a fundamental process of plant life which involves de novo formation of a "cell plate" partitioning the cytoplasm of dividing cells. Cell plate formation is directed by orchestrated delivery, fusion of cytokinetic vesicles, and membrane maturation to form a nascent cell wall by timely deposition of polysaccharides. During cell plate maturation, the fragile membrane network transitions to a fenestrated sheet and finally a young cell wall. Here, we approximated cell plate sub-structures with testable shapes and adopted the Helfrich-free energy model for membranes, including a stabilizing and spreading force, to understand the transition from a vesicular network to a fenestrated sheet and mature cell plate. Regular cell plate development in the model was possible, with suitable bending modulus, for a two-dimensional late stage spreading force of 2–6pN/nm, an osmotic pressure difference of 2–10kPa, and spontaneous curvature between 0 and 0.04nm^{-1}. With these conditions, stable membrane conformation sizes and morphologies emerged in concordance with stages of cell plate development. To reach a mature cell plate, our model required the late-stage onset of a spreading/stabilizing force coupled with a concurrent loss of spontaneous curvature. Absence of a spreading/stabilizing force predicts failure of maturation. The proposed model provides a framework to interrogate different players in late cytokinesis and potentially other membrane networks that undergo such transitions. Callose, is a polysaccharide that accumulates transiently during cell plate maturation. Callose-related observations were consistent with the proposed model's concept, suggesting that it is one of the factors involved in establishing the spreading force.

In part 2, we discuss three different in-silico studies of SARS-CoV-2. These studies involve the use of molecular dynamics simulations, endpoint free-energy estimates, as well as predictive neural networks (AlphaFold). In Chapter 4, we present binding strength estimates of three critical fitness parameters (RBD/ACE2 binding, furin enzyme binding,

antibody escape) of the SARS-CoV-2 omicron variant. We show that our results align with the preliminary observations noted with the variant, i.e. weakened RBD/ACE2 binding, but increased antibody escape. In Chapter 5, we present an in-depth analysis of the most commonly observed sequences in the Furin Cleavage Domain (FCD) and their interaction with the furin enzyme. We show that the Delta variant exhibits the strongest possible binding with the furin enzyme, and we identify key observed and unobserved sequences that could exhibit the same binding strength. In Chapter 6, we present a computational design of a humanized ACE2 decoy to be used as a possible therapeutic or diagnostic agent based on the principles of competitive binding. We show that our design binds favorably well to multiple SARS-CoV-2 target RBDs, including the delta and omicron variants.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Professor Daniel Cox, for his selfless support and consideration. He has been an exemplary human being, an outstanding mentor, and an even kinder teacher. I feel incredibly privileged to have had this opportunity to be a part of his research group, and to have learned from his experience in both life and physics.

I am deeply indebted to my co-advisor, Professor Georgia Drakakaki. She has always been a strong advocate for my success and for my work, and has treated me with utmost kindness and respect. I am also extremely grateful to her for graciously including me in the National Science Foundation grant (1818219) awarded to her research group. I am thankful for the opportunity to work with her and her wonderful group.

I would like to extend my gratitude to my qualifying exam chair and thesis committee member, Professor Rajiv Singh, who has been extremely patient, kind, and prompt in helping me with this endeavor. I am also grateful to Professor Warren Pickett, Professor Valentin Taufour, Professor John Conway, and Professor Steven Thegg, for generously agreeing to be a part of my qualifying exam committee, and for their invaluable comments and suggestions.

I would also like to thank everyone in my research group, especially two of my senior members, Rachel Baarda and Amanda Parker, who taught me much of what I know now. I would also like to thank Rosalie Sinclair for patiently working with me despite my minimal knowledge of plant biology, as well as Avinash Baidya and for sharing his wonderful insight in neural networks.

I would like to express my sincere gratitude to the senior and founding members of the `Protein Architects` group for generously supporting my summer research in 2021. In particular, I'd like to thank Dr. Richard Davis for his strong support and belief in my efforts. On that note, I'd also like to thank all of my teachers, from Lyceum to LUMS, and from Beaconhouse to Davis.

This endeavor would not have been possible without the prayers, love, support, and affection of my family. To Ammi and Abbu, *Jazak Allah Khair*, I am infinitely grateful

to have you as my parents and there is little I could have done without your guidance. To Adeel and Aiman, Mony, Sidra and Danish, Sameen, Aaliya, Bilal, and Mikael, thank you, sincerely, for all that you've brought to my life and continue to bring. I am extremely lucky to have you. To my extended family, thank you for giving me a home away from home, for those wonderful times at the million or so family gatherings we've had, and for the countless laughs and joyful memories we've shared.

I would also like to extend my deepest gratitude to my beloved friends, roommates, seniors, and teaching assistants who are too numerous to name. Thank you for all the *chai*, FIFA, cricket, football, cornhole, the much needed breaks at Khoka and PDC, the so-called SSE physics TA room, the countless roadtrips and vacations in both the US and in Pakistan, the insane happenings at M4, the terrible birthdays at the volleyball court, and for all the wonderful laughs, memories, and conversations we've shared together. I am deeply grateful for all of you, and you all remembered extremely fondly.

Allah Kareem

Chapter 1

Introduction to Part 1: Physical Models of Biological Membranes

1.1 Background and Motivation

Biological membranes exhibit a wide variety of physical characteristics that are of special interest to soft matter physicists. They are stable against most thermal fluctuations, yet are soft enough to deformed by embedded protein action or external environmental factors. They serve as the main barrier to entry to a plethora of cells, yet are capable of self assembly to macroscopic length scales [1]. They also serve as an anchor for a number of proteins, including antigens, ion channels, receptors, and many others. Membrane fission and fusion events are both incredibly common yet extraordinarily complex, involving fascinating mechanisms that are studied by a variety of analytical, experimental, and computational methods. Needless to say, they also attract a significant amount of attention from biologists due to their prevalence in almost every biological system.

Despite significant research efforts being redirected into both animal and plant cell cytokinesis, there are still gaps in the understanding of the exact interplay of the biophysical forces that lead to cell fission [2]. In animal cells, cell fission occurs due to the cell membrane being 'pinched' along a division plane via the action of a contractile ring (Fig. 1.1, which ultimately results in the parent cell dividing into two separate daughter cells. The molecular mechanisms behind this pinching, in addition to the pathways that organize the proteins involved in the assembly and disassembly of the contractile ring

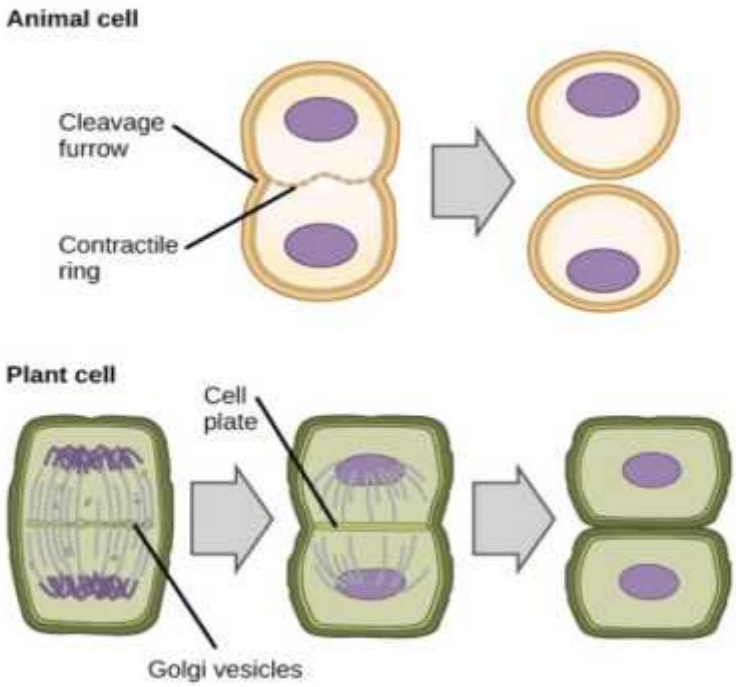


Figure 1.1: Comparing animal cell and plant cell cytokinesis (Image Credit [5], Creative Commons License). Animal cell cytokinesis involves pinching and reforming the cell membrane boundary via the action of a contractile ring. Plant cell cytokinesis requires orchested vesicle transport and polysaccharide deposition at the division plane due to the needs of the development of a new cell wall [6].

are still questions open to research [2]. However, several biophysical models have been proposed to explain the physics behind animal cell cytokinesis, some of which that have been successfully compared to experimental data [3, 4].

Since plant cells need to form a rigid cell wall, the contractile ring and membrane fission model no longer applies. Rather, scores of vesicles accumulate and fuse at the division plane to form a structure called the cell plate which then expands centrifugally [7, 8, 9, 10]. The development of the cell plate takes place in morphologically defined stages (Fig. 2.1). Eventually, as the gaps (resulting from a network of fused vesicles) in the cell plate sheet are gradually closed, a young cell wall which is originally sandwiched between two parallel plasma membranes fuses with the parental cell wall (Fig.2.1D) [7, 8], thereby resulting in two daughter cells. Electron tomography studies are used to study the transport of membrane material, as well as the evolution of different stages in the development of the

cell plate [8]. These cell walls are vitally important to the plant's survival. They are required for mechanical support with the added responsibility of providing resistance to excessive turgor pressure, as well as regulating diffusion and limiting excessive water loss. In the seed phase, they also serve as an important source of carbohydrates [7].

As in animal cells, a number of studies have investigated membrane dynamics in plant cells [11]. However, few reports exist on polysaccharide deposition and its explicit role during cell plate maturation. The delivery and deposition of cell wall materials to the cell plate has been primarily studied with electron and fluorescence microscopy utilizing polysaccharide-specific antibodies. It is currently believed that, matrix polysaccharides such as hemicellulose and pectins are transported through unidentified secretory vesicles to the expanding and maturing cell plate [12, 13, 6]. One such polysaccharide, a $\beta-1-3$ polymer called callose, is transiently synthesized (at the plasma membrane) and incorporated for mechanical support during the middle and late stages of cell plate formation, but ultimately replaced by cellulose, for a more rigid luminal network [7]. However, neither the temporal incorporation of callose or cellulose into the cell plate, nor their biosynthetic pathways during cell plate maturation have been fully elucidated [11, 14]. Furthermore, from a biophysical perspective, the process of incorporating callose only to be replaced by cellulose in the later stages is energetically expensive. The current hypothesis for callose deposition is that it provides an essential intermediate spreading and stabilizing force that is necessary for cell plate maturation [7, 8]. However, it is not clear whether there is a need of this spreading and stabilizing force in the first place. To our knowledge, Chapter 2 of this work is one of the first attempts at a biophysical model that shows an explicit need for this force. Furthermore, we reinforce our predictions for both the absence and the presence of this force by comparing our results with microscope images of plant cells undergoing cytokinesis in the presence of chemical inhibitors of cellulose, callose, and myosin. We also make this comparison with past studies of electron microscopy studies [8, 7] that shows detailed vesicular shapes and structures at various cell plate development stages.

1.2 Understanding membrane biophysics using analytical and computational tools

Landmark works in the physics of biological membranes were published in the early 1970's by Helfrich and Canham [15, 16]. They were able to provide a theory of membrane behaviour and elasticity based on geometric free energy descriptions while bolstering their claims with experimental evidence. Canham was able to show that his free energy description could explain the biconcave shape of a red blood cell by focusing on membrane bending deformations [16]. Helfrich expanded on that work with a more general free energy that addressed membrane tilt, stretching, curvature (including the effects of local curvature), and membrane responses to osmotic pressure differences and external magnetic fields [15]. The Helfrich free energy was adopted and aptly named the "shape equation". Helfrich derived the Euler-Lagrange equations with axisymmetric constraints, which were non-linear, fourth-order, partial differential equations that were incredibly difficult to solve. Given the complexity of solving these equations for the axisymmetric case and, especially, the general case, there were several attempts to provide a family of solutions using numerical and perturbation theory methods. In the modern computational age, finite-element methods and monte-carlo simulations are frequently used to study the global behavior of membranes and often provide great insights [1].

Briefly, the Helfrich Free Energy is commonly expressed as:

$$E = \frac{K_B}{2} \oint (H_1 + H_2 - c_o)^2 dA + \int \Delta p dV + \int \gamma dA + 4\pi K_G (1 - g) \tag{1.1}$$

The first term defines the energy penalty due to bending deformations of the membrane. K_B is the bending modulus for the membrane surface, H_1 and H_2 are the principal curvatures at a point on the surface, and c_o is the spontaneous curvature, or the preferred curvature for the membrane. The next term is the pressure energy which results from the difference in osmotic pressure between the inside and the outside of the cell plate, such that $\delta p = p_{out} - p_{in}$. The third term is the energy associated with the surface tension of the membrane which is given by the integral of the surface tension γ over the membrane surface. The fourth term is the Gaussian bending energy term which is simplified here due

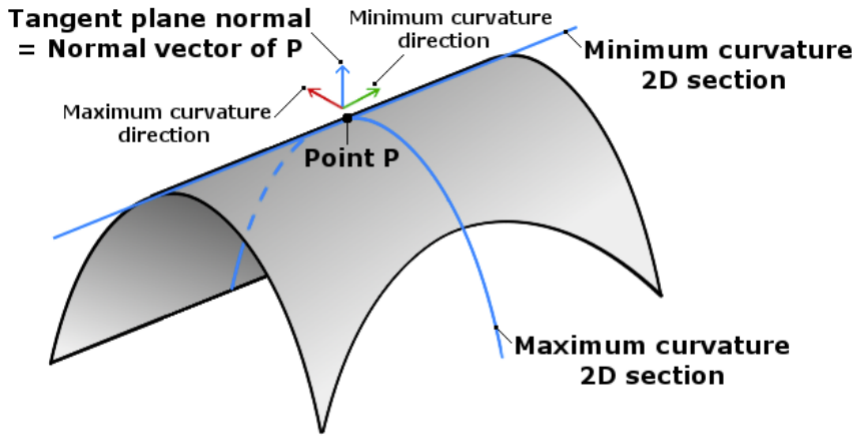


Figure 1.2: Describing principal curvatures using maximum and minimum curvature sections [17] (Open Access). The principal curvatures are the reciprocal of the local radii of curvature of the lines of intersection between the surface and the minimum and maximum curvature planes as shown.

to the Gauss-Bonnet Theorem. Fig. 1.2 describes how principle curvatures are calculated on a general membrane surface. The spontaneous curvature c_o can be non-zero due to multiple reasons, including protein action, membrane lipid profile, and non-homogenous bilayer distributions.

Further analytical work included the intensive derivations of membrane torque and stress tensors from first principles [18], which enabled predictions about membrane responses to an applied force (one that is often experimentally explored via atomic force microscopy), or the behaviour of two separate membranes adhering to one another [1].

Examples of computational approaches to descriptive membrane physics include monte-carlo simulations [19], local stress and torque tensor analysis, finite element and surface triangulation methods [20], as well as course-grained molecular dynamics simulations. These approaches, while computationally expensive, provide remarkable insight to the behavior of membranes in a variety of physical circumstances.

1.3 Objectives of Part 1

In Chapter 2, we explore the hypothesis that a spreading force is needed for successful cell plate maturation to a complete cell wall via a semi-quantitative approach. We model

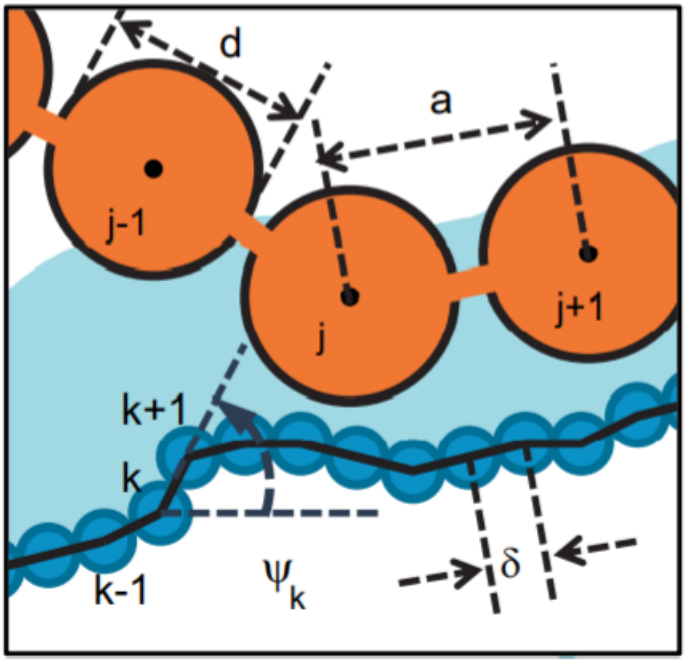


Figure 1.3: Example of a course grained membrane parameterization in a two-dimensional Monte-Carlo simulation used illustrate interactions between a membrane (blue) and a self-avoiding polymer (orange) [19] (Reprinted with Permission from APS, License No. RNP/22/MAY/053417).

this force by adding a phenomenological "areal pressure" term to the Helfrich model free energy [15] for the cell plate surface. We looked for energy minima within a parameterized restrictive geometry basis set, thereby adopting a restricted variational approach within testable approximated structures. The quasi-equilibrium is constantly redefined as vesicles are added at the cell plate boundary. The advantages of this approach are two-fold. It simplifies the modeling of a rapidly changing geometry that would be exhausting to define via traditional surface triangulation or course-grained molecular dynamics simulation. It also circumvents the demands of expensive computational time that the aforementioned systems demand. Finally, implementing a spreading force via a phenomenological "areal pressure" term within a parameterized restrictive geometry basis set is significantly easier to implement. Similar examples of restricted variational approaches are explored in [21, 22]

In addition, we provide the range of polysaccharide synthesis rates required to generate

the speculated required spreading force. These rates are postulated by using results from Flory self-avoiding polymer theory in two-dimensions [23]. We also show, compared to in-vitro results, that these rates are within reasonable bounds [24]. Finally, we find that for the spreading force hypothesis to work within our model, successful cell plate maturation also requires a concurrent decrease in spontaneous (or preferred) curvature up to a threshold value. In conclusion, we explore the possible biological mechanisms behind the required spreading force and concurrent decrease of spontaneous curvature.

Chapter 2

A Biophysical Model for Plant Cell Plate Maturation Based on the Contribution of a Spreading Force

This chapter appears as an article published in Plant Physiology *in collaboration with Rosalie Sinclair, Vincent Bulone, Daniel Cox, and Georgia Drakakaki [25]*

2.1 Introduction

Cytokinesis is a fundamental process of plant life that is different from animal cell cytokinesis. In plants, formation of a cell plate develops into the new cell wall, partitioning the cytoplasm of the dividing cell. Cell plate formation involves highly orchestrated vesicle accumulation, fusion, and membrane transformation concurrent with the time-specific deposition of polysaccharides such as callose, cellulose, and cross-linking glycans along with glycoproteins (Fig. 2.1). This development requires choreographed accumulation of post-Golgi vesicles via the phragmoplast, an assembly of microtubules and microfilaments that help organize vesicle delivery to the cell plate assembly matrix, at the division plane [10].

Cell plate expansion is centrifugal, led by the accumulation and fusion of cytokinetic vesicles to the leading edge and maturation of the membrane network from the center [7, 8, 9, 10]. Cell plate development takes place in morphologically defined stages (Fig. 2.1). It starts with the initial fusion of Golgi vesicles stage, in which cytokinetic vesicles are

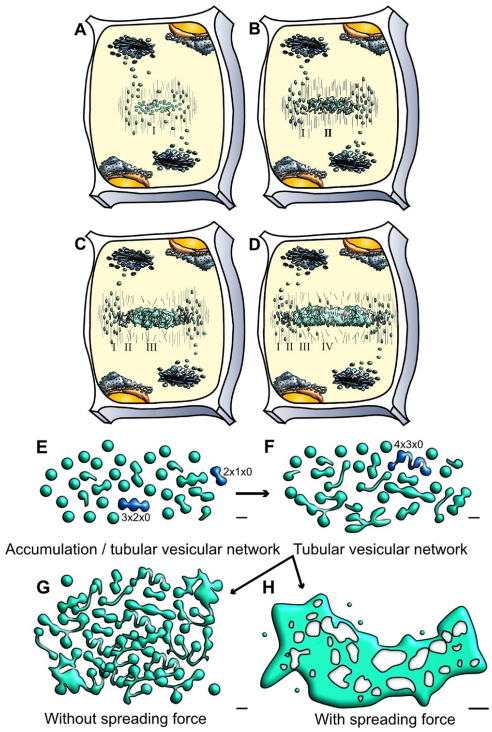


Figure 2.1: Schematic representation of cell plate development stages and the potential role of a spreading force in cell plate maturation. **A-D**, Cell plate development occurs centrifugally in multiple stages. **A**, During the first stage (I), cytokinetic vesicles guided by the phragmoplast accumulate at the center of the dividing cells, at the cell plate assembly matrix. **B**, Vesicles undergo fusion and fission and conformational changes resulting in TVN (Stage II). **C**, Interconnected membrane structures transition to a TN. At this stage high callose deposition occurs (Stage III). **D**, The membrane network further expands to an almost continuous fenestrated membrane sheet (PFS) (Stage IV). Deposition of additional polysaccharides helps transition to a new cell wall, separating the two daughter cells. Note that different stages can occur simultaneously, images are not to scale. This simplified representation emphasizes on cell plate membranes [8, 7]. **E**–**H**, Schematic representation of cell plate development describing the role of a spreading force. **E**, Early stages of vesicle accumulation and fusion and F, TVN and TN structures are shown. Two different possibilities are projected for stage transition (1) Incomplete/arrested cell plate **G**. In the absence of a spreading force **G**, tubular and fenestrated structures accumulate, and there is a lack of maturation towards a single, complete cell plate structure. (2) Normal cell plate transition **H**. In our calculations, we discover that for expansion/maturation to occur as in D, the presence of a spreading force is required, along with the decrease of spontaneous curvature to a threshold value. This allows for a sheet-like cell plate (SCP) structure to form. The structures in this schematic description are adapted from data collected from EM tomography [8] with bars in E–G=50nm, H=0.25μm. Dark blue vesicles denote those labeled by the mathematical naming schema as described in Fig.2.2. Whereas in E, 2×1×0 denotes two oblate spheroids, one tubular connection, and zero holes.

guided by the phragmoplast to the cell plate assembly matrix (Fig.2.1A). Fused vesicles are transformed into dumbbells that undergo conformational changes to form a tubulo-vesicular network (TVN) (Fig.2.1B), which transitions to a tubular network (TN) (2.1C). The TN expands into a planar fenestrated sheet (PFS). As the gaps in the fenestrated sheet are gradually closing, this leads to the formation of the young cell wall sandwiched between two parallel plasma membranes that fuses with the parental cell wall (Fig.2.1D) [7, 8]. Excess membrane material is recycled concurrently, along with the accumulation of different polysaccharide materials. Based on elegant electron tomography studies, it is estimated that 70% of membrane material is removed during the transition of the cell plate from TVN to TN and PFS [8]. Analysis of vesicle dynamics support electron microscopy studies showing an initial vesicle delivery with fast expansion, followed by slower expansion phase [26]. It is notable that the multiple stages exist simultaneously (Fig.2.1), adding complexity in dissecting them [6].

During cell plate expansion and maturation, membrane remodeling and network expansion are highly coordinated with the deposition of polysaccharides, providing an opportunity to study membrane morphology changes. The molecular basis of vesicle delivery at the cell plate has been extensively studied [27, 28, 29] with key components such as RAB GTPases, soluble N-ethylmaleimide-sensitive factor attachment protein receptor, tethering complexes, dynamin rings, and accessory proteins receiving attention [27, 9, 29]. However, the factors contributing to stage transition from a vesicular network to a fenestrated sheet, leading to cell plate maturation, are largely unknown. Dynamin rings and clathrin coated vesicles contribute to recycling of excess material [27], while the deposition of polysaccharides likely contributes to transition into a mature cell plate. Hemicelluloses and pectins are deposited via Golgi derived vesicles. Callose and cellulose are directly synthesized at the cell plate [11, 30]. Callose, a $\beta-1-3$ glucan is a dominant polysaccharide transiently synthesized at the cell plate. Structural glycoproteins such as extensins are part of the newly formed cell plate [31] and can contribute to cell plate maturation. Given the complexity of cell plate development and the concurrent presence of different stages, a biophysical model can be used as a framework for interrogation of

individual components that can provide insights and guide future research.

In this study, we used biophysical modeling to dissect the transition between the vesicular network stage to a fenestrated sheet and a mature cell plate. We tested the hypothesis that a time-dependent spreading and stabilizing force is necessary for cell plate maturation. We could model this force by adding a phenomenological "areal pressure" term to the Helfrich model free energy [15] for the cell plate surface. Furthermore, we monitored its influence by adopting a variational approach to locally minimize the model free energy in time, assuming the process is sufficiently slow to consider the system close to thermodynamic equilibrium. The quasi-equilibrium is constantly redefined as vesicles are added at the cell plate boundary. This enables us to use the total cell plate surface area as a proxy for time. We demonstrate semi-quantitatively that by assuming a late time onset of this spreading and stabilizing force, followed by the reduction of membrane spontaneous curvature, we can reproduce the observed morphological time dependent transition of the cell plate morphology.

2.2 Results

We took a modeling approach to generate tools to dissect better membrane network transition during cell plate maturation. Due to the complexity of cell plate development, we decided to look for energy minima within a parameterized restrictive geometry basis set, thereby adopting a restricted variational approach within testable approximated structures. We found that existing general adaptive mesh approaches, such as Surface Evolver, while in principle more accurate, were not amenable for application in our study, due to their inability to incorporate the spreading/stabilizing force into such a large-scale system [20].

2.2.1 Shape Approximation

First, we approximated subcellular structures with testable shapes that could be used in a model. As shown in 2.1, the cell plate, during its different transitional stages, contains cytokinetic vesicles, fused vesicles stretched to dumbbells, TVNs and fenestrated structures that finally mature to a complete cell plate and a new cell wall. These structures can be

modeled using a combination of oblate spheroids and elliptical hyperboloids (Fig.2.1 and 2.2). Namely, cytokinetic vesicles can be approximated using oblate spheroids, where the two defining radii can be used as variational parameters as shown in Fig.2.2A. The oblate spheroid can also be used to model the expanded/late-stage cell plate close to completion, as a very large oblate spheroid with a c.

Similarly, structures found within the fenestrated sheet and the TVN stages can be approximated using a combination of elliptic hyperboloids and oblate spheroids, such that the hyperboloids are continuous at the oblate spheroid boundaries. These elliptic hyperboloids can be parameterized by their length, hereafter referred to as l, and their skirt radius in the equatorial plane, hereafter referred to as a_h. The other parameters needed to define an elliptic hyperboloid can be written as a function of these two parameters due to boundary conditions that arise from mandating continuity. Fig.2.2B shows a cross-section of an elliptic hyperboloid, while Fig.2.2C shows an example of two vesicles joined by a single tube.

2.2.2 Naming convention of different approximated conformations

The naming convention for different conformations follows the example: A shape that is labeled as 6×7×2, represents a conformation that has six oblate spheroids (or vesicles), seven hyperboloids (or tubes), and two gaps (or fenestrations, with $g = 2$). 2×1×0 conformation, can approximate dumbbell structures found in early TNs, where two vesicles join via a single tube (2.2C). An example of a 4×4×1 conformation with a single fenestration, which can be used to approximate early fenestrated structures, is shown in Fig.2.2D.

2.2.3 Energy minimization

The area of any given conformation was calculated by numerical integration methods, and a corresponding parameter space of a given area was found. For simplicity, we analyzed conformations that shared the same parameters for each of the oblate spheroids, and each of the hyperboloids with examples shown in Fig. 2.2. We found a four-dimensional parameter space (a, c, a_h, l) that corresponded to a given area up to an error tolerance

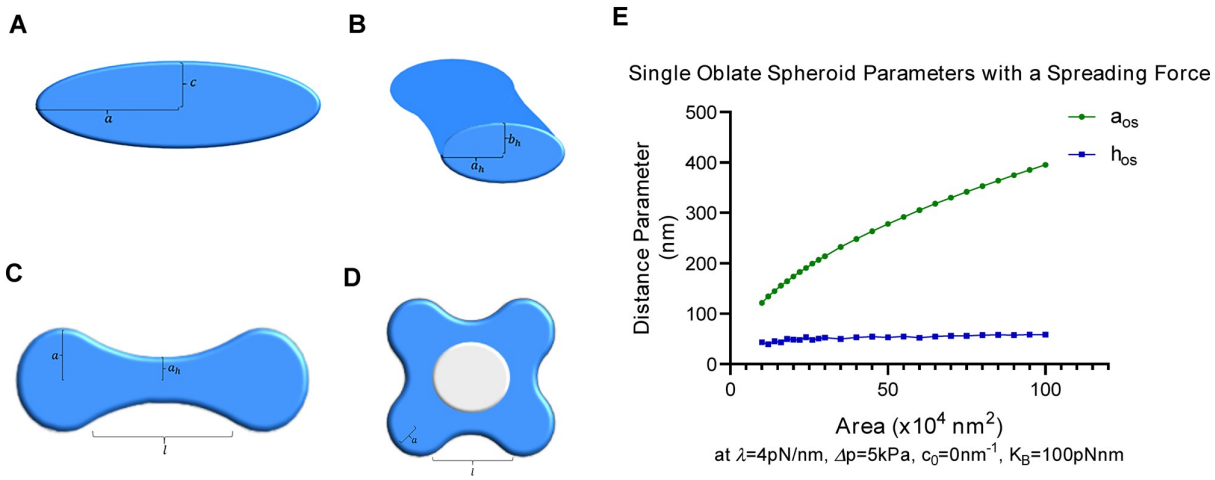


Figure 2.2: Approximating cell plate structures using a variational approach. **A**–**D**, Examples of membrane structure parameterizations used for modeling. **A**, Cross-section of an oblate spheroid through the polar axis. The major axis radius is labeled a and the minor axis radius is labeled c. This structure is used to model vesicles, or mature cell plate structures in the case where a c .**B**, Cross-section of an elliptic hyperboloid at its center, showing the skirt radii. The hyperboloid can be parameterized by its length l and its skirt radius in the equatorial plane ah, the skirt radius in the axial plane is given by b_h, which can be written as a function of the other parameters listed as shown in Eq.2.4. **C**, An example of a tubulo-vesicular structure parameterized by two oblate spheroids connected by a single elliptic hyperboloid (referred to as a 2x1x0 structure). Only the top view is shown. **D**, An example of a 4x4x1 conformation that models a transition to a fenestrated network with genus g = 1 (one gap).**E**, Evolution of single oblate spheroid parameters in the presence of a spreading force. In the presence of a spreading force, the thickness of the oblate spheroid remains in the 40–80nm range despite the increase in area. This reflects the thicknesses and growth patterns found in intermediate cell plate stages [7]. Here, $h_{os} = 2c(a_{os}, c$ shown in **A**), represents the overall height, or thickness, of the oblate spheroid. In the absence of a spreading force, h_{os}, or the thickness, is estimated to grow in values that are not observed experimentally. For reference, an area of $10^4 nm^2$ is roughly equal to that of a single vesicle.

($< 0.01\%$) for each conformation of interest, and then calculated the energies of Eq.2.1 within that parameter space. The energy minimum was then extrapolated from that parameter space. Additional information about the four-dimensional parameter space as well as the calculation of the area elements used in the numerical integration methods is given in the supplemental information (Eq.2.2 –2.5, 2.11, 2.12), and the full list of the parameters involved are shown in Fig. 2.6.

2.2.4 Modified Helfrich energy

In order to identify contributing factors for cell plate maturation beyond the vesicle network stage, we modeled the free energy for the cell plate surface by adopting the Helfrich energy [15] with the addition of a novel term to model the presence of a spreading/stabilizing force. The free energy is defined as follows:

$$E = E_{bending} + E_{pressure} + E_{tension} + E_{gaussian} + E_{spreading} \tag{2.1}$$

Each of the components is described in detail in the supplemental information corresponding to Eqs.2.6–2.10.

The terms describe (1) the bending energy over the closed membrane surface(s) of the cell plate, which depends on the local curvature or the spontaneous curvature of the membrane, given by c_o, and the bending modulus, given by K_B, (2) the pressure energy which results from the difference in osmotic pressure between the inside and the outside of the cell plate, represented by Δp, (3) the energy associated with the surface tension of the membrane, which depends on the local surface tension, given by γ, (4 the Gaussian bending energy, and (5 the novel term of spreading/stabilizing force.

The novel spreading force term is analogous to a two-dimensional pressure acting against the periphery of the cell plate structure along the equatorial plane. It is dependent on λ, which parameterizes the spreading/stabilizing force, having units of force/length. We allow for λ to be time-dependent, which would represent the "turning on" of polymer production in an expanding plate. We also allow for c_o to be time-dependent, accounting for differences in spontaneous curvature that may arise from changes in membrane composition during cell plate evolution.

This methodology can also be used to provide a basis for the quantitative assessment of membrane structures found in the endoplasmic reticulum and the Golgi apparatus which are so far limited in a large-scale view using Helfrich theory. Earlier work [32] examined possible morphologies as a function of curvature modifying proteins using full minimization of the free energy via the Surface Evolver finite element approach [20]. However, these finite element methods were unable to consider the spreading/stabilizing force required in Eq. 2.7 - 2.10 in any such available code. To establish a testable and functional model, we adopted the variational approach including multiple connected surfaces with negative curvature tubulations as a reasonable compromise approach to explore the quasi-equilibrium stabilities of different morphologies that are fully representative of the observed structures.

2.2.5 Model parameter ranges and the need of a spreading force

We first minimized the modified Helfrich energy (Eq. 2.1) for multiple conformations (vesicular, tubular, and fenestrated) to determine a range of parameters that would match the experimentally observed cell plate sizes/thicknesses. From electron tomography cryo-EM images of developing cell plates, we determined that the thickness of a cell plate in various stages of development was 40–120nm [8]. Therefore, we tuned the free parameters in our energy model such that conformations' thickness across the equatorial plane was in the range of 40–120nm. We determined that, depending on the choice of the bending modulus, the allowed values of the spreading/stabilizing force parameter should be between 0.0 and 6.0pN/nm, the spontaneous curvature c_o between 0 and 0.04nm^{-1}, and a finite pressure differenceΔp around 2–10kPa. A deviation from these ranges results in structures that are either too thick or too thin to exist in intermediate stages of cell plate development based on literature. An example of how we tuned the parameters to fit the experimental sizes and shapes is given in 2.2E, in which a single oblate spheroid evolves with the area increase with given parameter values while maintaining the experimentally observed thickness. It is notable that a spreading force is necessary to achieve the desired values. A summary of the full range of parameter values are given in 2.1 and a full description of these parameters is provided in the supplemental information.

Parameter	Value
Bending modulus K_B	$62.5 - 200pNnm$
Spontaneous curvature c_o	$0 - 0.04nm^{-}1$
Pressure difference ΔP	$2 - 10kPa$
Spreading force parameter λ	$2 - 6pN/nm$
Surface tension parameter γ	$1.6pN/nm$
Gaussian bending modulus K_G	$-0.8K_B$

Table 2.1: Model parameter ranges

Although the range of values for the pressure difference and the planar spreading force parameter were phenomenologically determined, they are within reasonable bounds. For instance, the solute concentration difference between the interior and the exterior of the cell plate by employing the van't Hoff equation, which yields a solute concentration difference between $8{\times}10^{-4}mol/l$ and $4{\times}10^{-3}mol/l$, comparable to protein solute concentration differences in higher plant cells. The spreading force required for cell plate maturation over a length of a nanometer is around 2–$6pN$, which is comparable to the polymerization ratchet forces of a microtubule [33].

2.2.6 A spreading force is required for cell plate maturation while its absence energetically favors the accumulation of tubular and vesicular networks

Our goal was to assess within the modeled free energy of Eq. 2.1, whether a spreading force is essential for the necessary transitions from a combination of TVN to a fenestrated sheet and finally to a single mature cell plate structure. Applying Equation 2.1, we compared the energy minima compared to that of a mature cell plate (Fig.2.3A). From an energy perspective, we identified that in the absence of a spreading force, tubulo-vesicular and fenestrated structures have a lower value of energy at minima and are more stable than a single late-stage cell plate structure (resembled by a single oblate spheroid) of the same area. Fig.2.3A shows the energy minima of tubular and fenestrated structures (7x6x0,…)

compared to those of a single oblate spheroid (1x0x0). ΔE_{min} represents the difference of the energy minima value of the labelled structure with that of a single oblate spheroid, so that ΔE_{min}(2x1x0) = E_{min}(1x0x0)$-E_{min}$(2x1x0). Thus, positive ΔE_{min} indicate the relative stability of the labelled conformation. Our simulations indicate that in the absence of a spreading force, increased tubularity is preferred with the increase in area (Fig. 2.1G and 2.3A). Furthermore, in the absence of a spreading force, some fenestrated structures (4x4x1, 6x7x2) are also energetically stable and are therefore likely to accumulate.

We then examined the possibility of a transition from TVNs to a single oblate spheroid or a fully mature cell plate in the presence of a spreading force (Fig.2.3B). Within the theory and the variational approach, we find that this is possible if the spontaneous curvature decreases to a threshold value ($0.015nm^{-1}$ with larger cell plate area, that is, in the presence of a spreading force. From an energy perspective, this suggests that TVNs, TNs, as well as fenestrated sheets, should be unstable as compared to a single oblate spheroid, and thus transition of their morphology to one without tubes or fenestrations. While a finite spontaneous curvature is necessary to explain the origin of stability of the incoming vesicles [34], a change in the spontaneous curvature of the membrane is predictable due to the expected changes in membrane composition and protein activity that occurs during cell plate development [35].

Fig.2.3B shows the relative instability of selected tubular and fenestrated structures compared to a single oblate spheroid in the presence of a spreading force and zero spontaneous curvature. Less tubular structures are now energetically favorable than highly tubular or fenestrated structures, with a single, complete structure being the most favorable. For structures without fenestrations or gaps (such as 2x1x0 or 3x2x0), we can also map a path to a single oblate spheroid if we relax the parameter restrictions that were initially imposed during the variational calculation. Fig.2.3B shows results with the parameter restrictions in place. Fig.2.7 shows data for a fenestrated structure in the absence (Fig.2.7A) and the presence (Fig.2.7B) of a spreading/stabilizing force, leading to gap shrinkage with the parameter restrictions in place.

To better represent a biological system, we compared an ensemble of 2x1x0 structures

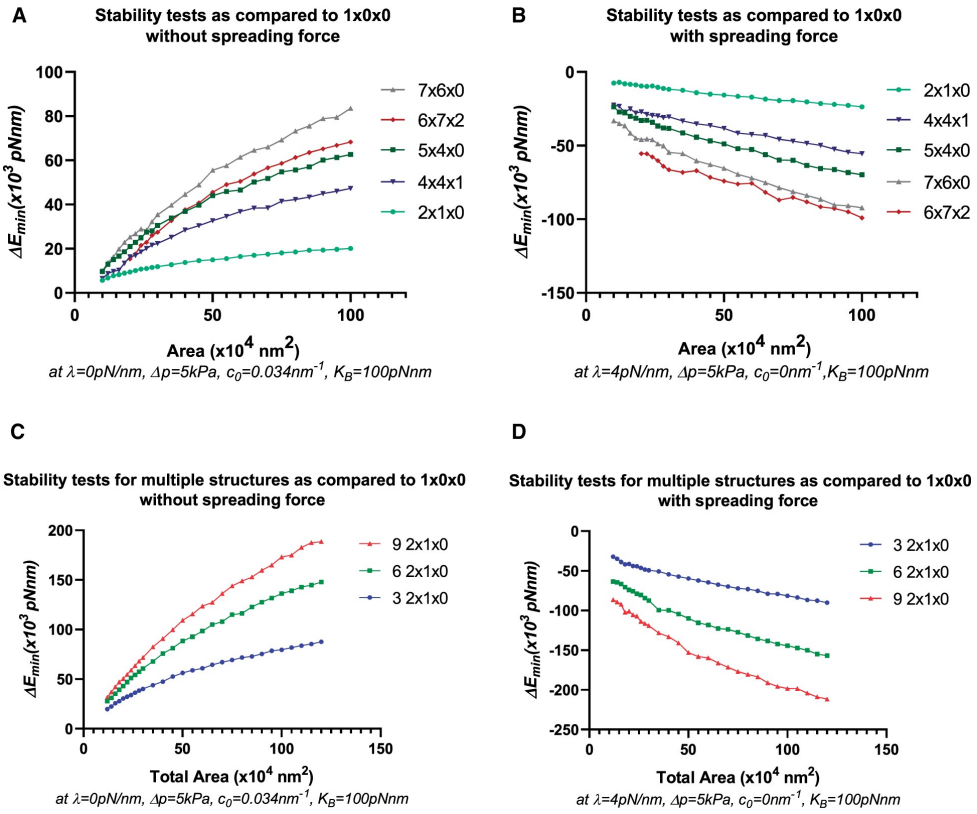


Figure 2.3: Stability tests to determine the role of a spreading force in different shape conformations. A and B, Stability tests determined by ΔE_{min} versus area for different conformations compared to a single oblate spheroid at the labelled area. A positive value of ΔE_{min} indicates relative stability of the labelled conformation as compared to a single oblate spheroid (1×0×0). A, Relative stability of tubular (2×1×0, 5×4×0, 7×6×0) and fenestrated (4×4×1, 6x7x2) structures in the absence of a spreading force with a finite spontaneous curvature. B, Stability of a single oblate spheroid over tubular and fenestrated structures in the presence of a spreading force and with zero spontaneous curvature. Note that in (B) a decrease of spontaneous curvature to a threshold value close to $0.015nm^{-1}$ yields similar results. C and D, Stability test for multiple 2x1x0 structures compared to a single oblate spheroid at the labeled area. C, Relative stability of multiple 2x1x0 structures compared to a single oblate spheroid in the absence of a spreading force. At a labeled area, a larger number of structures have collectively a higher, more positive value of ΔE_{min}, thereby indicating that in the absence of a spreading force, tubular, as well as emerging fenestrated/network structures (as inferred by the results of A and B) are energetically favorable and tend to accumulate as shown in Fig.2.1G. D, Stability of a single oblate spheroid compared to multiple 2×1×0 structures in the presence of a spreading force and with zero spontaneous curvature. In the presence of a spreading force, at a labeled area, a larger number of structures have a lower, more negative value of ΔE_{min} collectively, thereby indicating the energetic favorability of structures fusing to form larger, more mature structure(s).

(approximating accumulated fused vesicles forming a network) to a single oblate spheroid of the same combined area. Similar to our earlier calculations, we find that in the absence of a spreading force, a single oblate spheroid is less stable, as shown in Fig.2.3C. The relative instability is magnified with the increase of area, and with the increase in the number of tubular structures. In the presence of a spreading force and a decreased spontaneous curvature, as in Fig.2.3D, the inverse is true, favoring fewer complex structures. When comparing multiple structures to a single mature structure of the same area, there is no need to enforce the decrease in spontaneous curvature. However, for consistency, results with a zero spontaneous curvature in the presence of a spreading force, and a finite spontaneous curvature in the absence of a spreading force are shown.

To better represent a biological system, we compared an ensemble of 2x1x0 structures (approximating accumulated fused vesicles forming a network) to a single oblate spheroid of the same combined area. Similar to our earlier calculations, we find that in the absence of a spreading force, a single oblate spheroid is less stable, as shown in Fig.2.3C. The relative instability is magnified with the increase of area, and with the increase in the number of tubular structures. In the presence of a spreading force and a decreased spontaneous curvature, as in Fig.2.3D, the inverse is true, favoring fewer complex structures. When comparing multiple structures to a single mature structure of the same area, there is no need to enforce the decrease in spontaneous curvature. However, for consistency, results with a zero spontaneous curvature in the presence of a spreading force, and a finite spontaneous curvature in the absence of a spreading force are shown.

2.2.7 Exploring polysaccharide deposition as a contributing factor to cell plate maturation and model prediction

There are several potential contributing factors during cell plate maturation including cell wall polysaccharides. Given that the model examines the specific transition between TN to a fenestrated sheet and a mature cell plate, the timing of the contributing sources at the lagging zone is critical. Among the different polysaccharides we first examined callose.

Live staining of callose in dividing Arabidopsis thaliana roots showed a prominent and transient accumulation of the polysaccharide at the lagging zone, starting from the

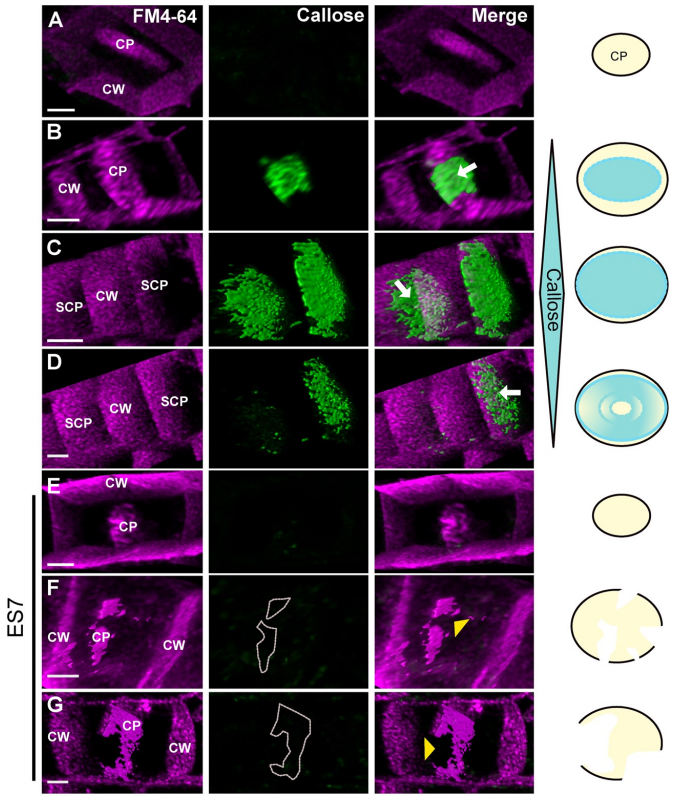


Figure 2.4: Progression of the cell plate in the presence and absence of callose. **A**–**D**, Cell plate progression in the presence of callose. (A), It shows an early stage cell plate before the accumulation of callose, while (B–D) represent later cell plate stages including SCP as indicated in Fig.2.1D. FM4–64 staining (magenta) is used to stain plasma membrane and the cell plate, while Aniline Blue fluorochrome (green) staining shows callose accumulation. Note the transient accumulation of callose in later stages leading to the maturation of cell plate during normal cytokinesis (B–D). **C** and D represent two snapshots of a time series. **C**, Two cell plates can be observed, and as maturation continues to **D**, callose is eliminated from one cell plate indicating its transient nature. Arrows indicate callose accumulation at the cell plate. **E**–**G**, Progression of cytokinesis under ES7 treatment for 2 h that inhibits callose deposition. Note that early cell plate development is not affected with ES7 treatment as shown in earlier studies [14] (E). However, in late stages of cell plate development under ES7 treatment, the absence of callose prevents the transition into a stable mature single structure, leading to characteristic "cell plate stubs" (**F** and **G**). CP indicates cell plate, SCP indicates SCP as depicted in 2.4. CW indicates cell wall. Yellow arrowheads denote lack of callose at cell plate breakage points. Dotted lines in F, G outline the position where callose should be deposited. Images are 3D reconstructions from Z-stacks of live confocal imaging and show single timepoints. **C** and **D** are snapshots of a time series. Figures are representative of root tips from a minimum of 10 Arabidopsis seedlings. A schematic representation on the right indicates the accumulation of callose in relation to cell plate development. White gaps at the bottom indicate cell plate fragmentation. Bars=3μm.

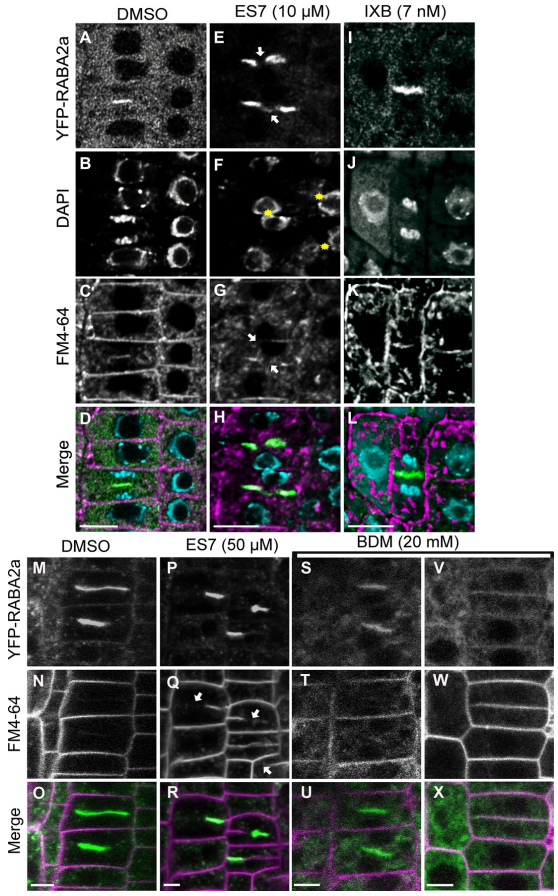


Figure 2.5: Comparison of chemical inhibitors of cellulose, callose and myosin on cytokinesis. **A**–**L**, Evaluation of cytokinesis inhibition under 5 day chemical treatment in Arabidopsis root tips. Under control DMSO treatment normal cytokinesis is observed (A–D). Under ES7 treatment typical cytokinesis defects are observed with the cytokinesis marker RABA2a (E), multinucleate cells (F) are shown by DAPI staining. Under IXB treatment cell plate progression was observed (I) without discernable cytokinetic defects in the form of binucleate cells (J) or cell plate stubs (**I**, **K**, and **L**). Please note cell swelling under IXB treatment. The cytokinesis marker RABA2a is shown in green, while FM4–64 staining of plasma membrane is shown in magenta. Nuclei staining by DAPI are indicated in blue. Samples were stained with FM4–64FX, fixed and stained post fixation for DAPI. Results were observed in at least six roots for each drug treatment. Samples are single scans of fixed cells. Bars=10μm. **M**–**X**, Effect of 2 h short-term (50μM) ES7 and the putative myosin inhibitor 2,3-butanedione 2-monoxime (20 mM BDM) treatment in cytokinesis. Under DMSO control treatment normal progression of cytokinesis is observed (M–O). Under ES7 treatment, characteristic cell plate stubs were observed with RABA2a and the plasma membrane stain FM4–64 (P–R). Under BDM treatment, a reduction of RABA2a signal was observed with increase in cytoplasmic pattern (S–X). The cytokinesis marker RABA2a is shown in green, while FM4–64 staining of plasma membrane is shown in magenta. Samples are single scans of live cell confocal imaging. Results were observed in at least six roots for each drug treatment. Bars=5μm.

center (2.4, B–D). Treatment with Endosidin7 (ES7), which inhibits cytokinetic callose deposition [14], caused failure of the cell plate to mature into a cross wall. 2.4, E–G shows an example of arrested cell plate development in the absence of callose, in contrast to normal gradual cell plate maturation concomitant with callose deposition as in 2.4, A–D. This is consistent with the model's prediction in the absence of a spreading force, where planar fenestrated and tubular structures accumulate, but do not mature into a stable cross-wall like structure.

Given the loadbearing role of cellulose, we then examined the effect of cellulose compared to callose inhibition in our experimental conditions. Cellulose inhibition by isoxaben (IXB) treatment led to strong reduction of root growth [36, 37] and a root swollen phenotype compared to ES7 (Supplemental 2.17). However, while cytokinesis defects in the form of cell plate stubs, were observed with ES7 treatment (2.5, E and Supplemental 2.17, E and J), this effect was not detectable in IXB treatment (2.5 and Supplemental 2.17, G–J). ES7 treatment caused binucleate cells as a result of failed cytokinesis (2.5F); however, this phenotype was not pronounced in the IXB treatment (2.5J).

We then included a treatment with the myosin inhibitor 2,3-butanedione monoxime (BDM), interfering with actin-based organelle transport [38, 39, 40]. Unlike ES7 (2.5, P–R), a $20mMBDM$ treatment for 2hrs led to inhibition of RABA2a trafficking 2.5, S–X, but not cell plate fragmentation, showing an effect on endomembrane trafficking, that in turn could impact cell plate development.

2.2.8 Estimating the required polysaccharide synthesis rates for a stabilizing/spreading force

We note that the spreading/stabilizing force can be described by the mean square end-to-end excursion within the Flory self-avoiding polymer theory in two-dimensions [23]. If we assume reasonable values of polymer persistence length and areal density, we require a rate of polysaccharide synthesis close to $dN/dt \sim 1.75x10^6 s^{-1}$ to obtain the spreading force parameter of $\lambda = 4pN/nm$. The modeled dN/dt is biologically achievable given an estimated cellular value of $1.8{\times}10^5 s^{-1}$ based on in vitro callose synthase activity ([24]) and an average protein concentration in eukaryotic cells ([41, 42]). However, in-vitro ex-

perimental systems of the relevant polymer(s) synthase(s) in artificial vesicles are required to test this hypothesis. A full description of this derivation is given in the supplemental information corresponding to Eqs. 2.13 - 2.12.

2.3 Discussion

Although several proteins have been identified that regulate cell plate formation [27, 9, 43], the mechanisms underlying the complex transition from a vesicle membrane network (TN) to a fenestrated sheet and a mature cell plate are not well understood.

To circumvent these hurdles and to better dissect cell plate maturation, we used biophysical modeling. We developed a model based on the Helfrich free energy for the cell plate surface with the incorporation of a spreading/stabilizing force as an "areal pressure" (force per unit length). From an energy minimization analysis, we have shown that a planar spreading/stabilizing force is vital for cell plate to transition from vesicle membrane network to a fenestrated sheet and late stage/mature cell plate. We also show that in the absence of a spreading/stabilizing force, the addition of membrane material yields stable TN structures, but that those structures are unable to mature beyond this stage. As shown by different simulations, the need for this spreading/stabilizing force is magnified when we compare a single mature cell plate to multiple smaller vesicle network structures of the same total area. We do not have the detailed molecular scale mechanisms behind the spreading/stabilizing force, but we show that a simple model based upon the expansion of a quasi-two-dimensional self-avoiding polymer captures the correct form.

To reach a mature cell plate, our model requires the late-stage onset of the spreading force coupled with a concurrent loss of spontaneous curvature. This raises the intriguing possibility of a common origin to the decrease in spontaneous curvature and onset of a spreading/stabilizing force. In the model, the spreading force is relevant when there is sufficient connection of individual oblate spheroidal vesicles, and it is at this stage that we shut off the spontaneous curvature. The nanoscale surface topography can potentially serve as a direct biochemical signal to activate this process [44]. The possible tethering of polysaccharides or glycoproteins to the membrane could concomitantly induce spread-

ing and reduce spontaneous curvature by modifying the membrane mechanics. Notably, inhibition of long chain fatty acid affects cell plate maturation [45]. It is plausible that membrane microdomains control both spontaneous curvature and the onset of a spreading force at the cell plate.

Curvature-stabilizing proteins are active at the cross-sections of tubules and sheet edges of endoplasmic reticulum [32, 46]. While force generating proteins are involved in the tubulation and membrane material recycling processes [47, 48], no proteins have been identified with properties of membrane expansion at the cell plate.

The phragmoplast-driven vesicle delivery is a dynamic and complex process [49] that with the aid of motor proteins can be considered as a spreading/stabilizing force during cell plate maturation. For example, Myosin VIII plays a role in guiding phragmoplast expansion [50], while several kinesins are involved in the functional organization of the phragmoplast [49]. Microtubule directed vesicle delivery occurs at the leading edge; however, it is followed by microtubule depolymerization at the lagging zone, which is the transitional stage that the model describes [51]. Furthermore, inhibition of myosin causes a broader effect on cell plate expansion, as it is involved in general vesicle delivery (2.5). Therefore, it is challenging to assign a specific function of motor proteins to cell plate maturation at the lagging zone. Time-lapse experiments directed at the role of motor proteins at the lagging zone will shed light on their contribution to the stabilizing and spreading force that the model predicts.

It is plausible that polysaccharide deposition serves as this stabilizing and spreading role. The matrix polysaccharides hemicellulose and pectin are synthesized in the Golgi apparatus and delivered via vesicles from the beginning of cytokinesis [13, 12, 52]. Thus, these classes of polysaccharides are unlikely the major players as they do not overlap with the predicted onset of the spreading/stabilizing force, although experimental verification awaits. Callose and cellulose are synthesized directly at the plasma membrane and are excellent candidates for exploration. Our data showed that pharmacological inhibition of cellulose at the root tip inhibited cell elongation in general, while inhibition of callose deposition led to cytokinesis defects consistent with the conformations predicted by

the model in the absence of a spreading force. Callose accumulation peaks at the intermediate TN stage, a transitional stage that coincides with loss of membrane volume [7, 8]. The timing of callose deposition in late stages when the overall cell plate membrane network "flattens" [7, 8] is consistent with the need of callose in providing a lateral spreading/stabilizing force. Furthermore, the predicted required values of callose deposition are within biological thresholds [24, 53]. Notably, a study by Thiele et al. [54] indicates that callose is required to establish the connection between the nascent cross-wall and the parental cell wall, rather than stabilizing the young cell plate, so that further analysis on the role of callose in the proposed model awaits verification. It is plausible that callose could serve as a scaffold into which other more permanent polysaccharides and proteins are later deposited [55, 24]. Potential transient interaction with cellulose or other glucans [11, 56, 57] can contribute to a composite that supports the stability of the cell plate and helps the attachment to the parental cell wall. Structural glycoproteins such as extensins [31] can provide a scaffold for polysaccharide deposition, and these altogether can generate the desired spreading/stabilizing force proposed by the model. Further (challenging) experiments are necessary to determine how the possible conformations of different polysaccharides and proteins or their combinations, synthesized in vitro in an artificial membrane setup, can contribute to different magnitudes of spreading/stabilizing force in lipid vesicle networks.

A unique element in the study was the approximation of cellular compartments with testable shapes such as vesicles and complete cell plates with oblate spheroids, fused vesicles and tubular structures with elliptical hyberboloids and their combination in a network. Approximating vesicles, tubulations and their networks in the current model has the potential of a wider application and can be adopted during quantitative assessment of membrane dynamics. It can be used as a basis for addressing the equilibrium of vesiculation (oblate spheroids) and tubulation (elliptic hyperboloids) and applied to ER-intermediate compartments, Golgi, and endosomes in all eukaryotic cells.

In conclusion, our model provides a framework for understanding how membrane structures evolve in the presence of a spreading/stabilizing force and will likely shed light

in such transitions that occur beyond cytokinesis.

2.4 Supplemental Information

2.4.1 Materials and Methods

2.4.1.1 Parameter Setup

To implement energy minimization on a parameterized basis set defined at a particular surface area, we first identified the parameter space for a given conformation that yielded the desired area. For a single oblate spheroid, only two parameters are needed, namely, the radius along the major axis, $'a'$, and the radius along the minor axis, $'c'$. However, for the consolidated tubular networks and emerging fenestrated structures, one must account for continuity between the oblate spheroid and the elliptic hyperboloid, while also making the relevant corrections in area. This continuity can be achieved by matching the slopes of the elliptical hyperboloid with the oblate spheroid along the primary axes, while enforcing contact.A normal one-sheeted elliptical hyperboloid centered at a distance $'d'$ along the x axis can be described by the following equation:

$$\frac{y^2}{a_h^2} + \frac{z^2}{b_h^2} - \frac{(x-d)^2}{c_h^2} = 1 \tag{2.2}$$

Here, a_h is the skirt radius along the xy- plane, b_h is the skirt radius on the xz-plane, and c_h describes the elongation along the y-axis, where the xy-plane defines the equatorial plane, and the z-axis is the polar axis. Fig. 2.6 shows these parameters for a hyperboloid in a 2x1x0 structure. By enforcing continuity with an oblate spheroid at the origin, we can derive the following relationships for d and b_h such that:

$$\begin{aligned} d &= \frac{1}{2}\sqrt{a^2 - a_h^2}\sqrt{a^2 - c_h^2} \\ b &= a_h \frac{c}{a} \end{aligned} \tag{2.3}$$

Using this, the length of the hyperboloid l is given by:

$$l = 2d(1 - \frac{a_h^2}{a_h^2 + c_h^2}) \tag{2.4}$$

Using these restrictions, we can form connected structures of oblate spheroids and hyperboloids to form approximate representations of tubular and fenestrated structures. Additional demands that are placed by specific conformations such as in complex emerging fenestrated structures (for example 6x9x3) were taken into account by eliminating choices of hyperboloids that would cause clashes. Thus, our complete parameter space for a given conformation at a particular area is given by a list of values of (a, c, a_h, l) where the corresponding area is calculated by numerical integration methods up to an error tolerance of 0.01%, taking into account spatial constraints.

2.4.1.2 Energy Minimization

For our energy minimization calculations, we treat the membrane boundary of the cell plate as an incompressible two-dimensional surface to a first approximation, with certain defined characteristics. In reality, each lipid bilayer has a finite thickness of 4-6 nm [58], however, it is justified to treat the surface of the cell plate as two-dimensional due to the smallness of its thickness compared to the sizes of the overall structures in the cell plate. It can also be shown that lipid bilayers present a high level of incompressibility due to the energy penalty associated with areal stretching being significantly higher compared to membrane deformations due to bending [59]. This is similar to the approach taken by Choksi et al. and Sarasij et al. [22, 21]. At a specified area, we define a free energy of our membrane surface that is essentially the Helfrich energy (Helfrich, 1973) with the addition of a novel term to model the spreading force, possibly due to polysaccharide deposition, as given in 2.1. We discuss the possible origin of this spreading force within a simple model elsewhere. Finally, we consider the surface area of the cell plate as a proxy for cell plate development stage/time and then we minimize this energy for a given surface area.

2.4.1.3 Modified Helfrich Energy

The modified Helfrich energy [15] is described as follows:

$$E = E_{bending} + E_{pressure} + E_{tension} + E_{gaussian} + E_{spreading} \tag{2.5}$$

The first term describes the bending energy over the closed membrane surface(s) of the cell plate. It is given by:

$$\frac{K_B}{2}\oint(H_1+H_2-c_o)^2dA \tag{2.6}$$

Where K_B is the bending modulus for the membrane surface, H_1 and H_2 are the principal curvatures at a point on the surface, and c_o is the spontaneous curvature, or the preferred curvature for the membrane. For simplicity, we assume that the bending modulus is time independent. We allow for the spontaneous curvature, c_o, to be time dependent, which reflects potential differences in membrane composition during cell plate evolution.

The next term is the pressure energy which results from the difference in osmotic pressure between the inside and the outside of the cell plate, such that $\Delta p = p_{out} - p_{in}$. It is given by:

$$E_{pressure} = \int \Delta p dV \tag{2.7}$$

The third term is the energy associated with the surface tension of the membrane, given by:

$$E_{tension} = \int \gamma dA \tag{2.8}$$

We assume that the surface tension given by γ is a constant. Since we assume that our system locally equilibrates in time at a constant area, this term only adds a constant energy equal to the area at that time, multiplied with the surface tension.

The fourth term is the Gaussian bending energy term, given by:

$$E_{gaussian} = 4\pi K_G(1-g) \tag{2.9}$$

Here, K_G is the Gaussian bending modulus, and g is the genus of the surface. This is a result of the Gauss-Bonnet theorem [60]. We consider the surface area of the cell plate as a proxy for cell plate development time and then we minimize this energy for a given surface area.

Finally, we introduce the novel element of a spreading force, which is analogous to a two-dimensional pressure that acts against the periphery of the cell plate structure along

the equatorial plane. In this interpretation, λ as units of force/length, with the energy representation as follows:

$$E_{spreading} = \int \lambda dA \tag{2.10}$$

It is important to note that the integral in Eq. 2.10 is over the equatorial plane of closed cell plate surfaces. Therefore, fenestrations are not integrated over.

As with spontaneous curvature, we allow for the spreading force coefficient λ to be time dependent, representing, e.g., the "turning on" of callose production in an expanding plate.

We use established values for the bending modulus K_B [61], as well as the Gaussian bending modulus K_G [62], and the surface tension γ [63]. However, the pressure difference Δp was phenomenologically determined. The bending modulus is sensitive to the environment of the cell as well as the membrane type and conformation [61], and therefore there is a range of available literature values. We decided to test over the full range of literature values ranging from about $62.5pNnm$ to $200pNnm$ (roughly corresponding to a range of $15k_bT$ to $50k_bT$. Note that we have excluded a line tension term from Eq. 2.1. Such a term gives the wrong morphological sequence in time for the plate, and requires unphysically large values to stabilize the mature, $genus0$ plate structure.

2.4.1.4 Calculation of area elements

To calculate the mean curvature and the area elements of the parameterized structures, we employed the following methods:

$$\begin{aligned} H &= \frac{1}{2}(H_1 + H_2) = \frac{eG - 2fF + gE}{2(EG - F^2)} \\ dA &= \sqrt{EG - F^2} \end{aligned} \tag{2.11}$$

Where E, F, and G are coefficients of the first fundamental form (line element) and e, f, and g are coefficients of the second fundamental form (shape tensor), and u and g parameterize the surface. The selection of E, F, G are based standard practices as described in [64]. Thus, we can calculate the $E_{bending}$ from Eq. 2.1 for any conformation given the conformation type and its corresponding parameter space.

The code used to calculate the bending energies and the surface areas of the aforementioned conformations are given in the following public github repository:

https://github.com/zaki92/Bending_Energy_Calculations

2.4.1.5 Plant Growth

Arabidopsis thaliana seedlings of Col-0 were used in this study. Seeds were sterilized using 30% (v/v) sodium chlorate in ethanol (absolute) with 0.06% (v/v) of Triton X-100 (Sigma-Aldrich). Seeds were plated on 0.25 Murashige and Skoog medium (1.15 g L−1 Murashige and Skoog minimal organics salt, 10 gL−1 Suc, 5 g L−1 Phytagel (Sigma-Aldrich), and cold vernalized for 48 h at 4°C in the dark, after which plates were transferred to a plant growth chamber for seedling growth. Plants were grown in temperature and photoperiod-controlled environments, set to long-day (16-h-light/8-h-dark cycle) conditions, using fluorescent light (at 100 to 150 mmol quanta photosynthetically active radiation (PAR) $m^{-2}s^{-1}$) at 22 to 24°C.

2.4.1.6 Chemical treatment and Imaging

Four day old Arabidopsis seedlings were treated with 50 μM Endosidin7 (ES7), 20 mM 2,3-Butanedione monoxime (BDM) and Dimethylsulfoxide (DMSO) in 0.25 MS medium for two hours as previously described [14, 26]. A Leica SP8 or Zeiss 710 confocal microscope was used for imaging. Aniline blue fluorochrome (Biosupplies Australia) was used to detect callose deposition, while the lipophilic membrane dye FM4-64 (10 μM) (ThermoFisher Scientific) was used to stain the plasma membrane and the developing cell plate and DAPI was used for nuclei staining. For DAPI staining FM4-64FX stained seedlings were fixed in 4% PFA [14] and post fixation was stained with1 μg mL−1 4 ,6-diamidino-2-phenylindole (DAPI). 6-10 seedlings were imaged per individual treatment. Fluorescence signals of callose stained by Aniline blue fluorochrome and DAPI (excitation 405 nm, emission 415-500 nm), FM4-64 (excitation 510nm, emission 620-759nm), YFP-RABA2a (excitation 510nm, emission 520-570nm) were collected with 40x (water), 63x (oil) objectives. Z stacks were generated across the volume of full cell and were subsequently deconvolved with Huygens (SVI). 3D reconstructions were prepared using Imaris, Bitplane and figures were assembled using Affinity Designer. Four day- chemical treat-

ments were performed on seedling germinated in 0.25 MS agar media supplemented with the indicated inhibitors:7 or 10 nM isoxaben (IXB), 10 μM ES7 and DMSO. Root length was quantified using Image J [65] as previously described [37]. Multi- factor analysis of variance (ANOVA) was performed using R x 64 version 4.0.3 (R Core Team, 2017) in Rstudio (RStudio, PBC), version 1.3.1093 with the basic ANOVA function. Least square means (LS means) analysis was performed using the emmeans package (version 1.5.3) and the multcomp package (version 1.4- 15) in R and the p-value adjusted for multiple comparison using the Tukey method. The graphs were generated using ggplot2 package (version 3.3.3) and the emmeans package. Letters assigned by LS means, p = .05.

2.4.2 Supplemental Results

2.4.2.1 Results with full range of bending moduli

The data in Supplemental Fig.2.8-2.12 show additional calculations in a range of bending moduli. A bending modulus of $62.5pN-nm$ (about $15k_BT$) corresponds to the lower range of bending moduli, consistent with published data (Dimova, 2014), while a bending modulus of $200pN-nm$ (about $50k_BT$) corresponds to the higher range. It is important to note that the key outcome remains the same, a finite spreading force coupled with a decrease in spontaneous curvature is essential for a transition to a single, complete cell plate structure, regardless of the choice of bending modulus.

2.4.2.2 Fenestrated structures data at higher areas

The supplemental data in Fig.2.13, 2.14, show ΔE_{min} calculations for different types of emerging fenestrated structure conformations at larger cell plate areas. In the absence of a spreading force, and with finite spontaneous curvature, larger, more tubulated fenestrated structures are more stable than a single cell plate. In the presence of a spreading force and with decreased spontaneous curvature, a transition to a single, mature cell plate structure is energetically favorable. Fig.2.7 shows the effect of the spreading force when restricted to a parameter basis set for a chosen structure, which in Fig.2.7 is 8x10x3. When these parameter restrictions are removed, we see structures change conformation types, as seen in Supplemental Videos S1, S2 [25].

2.4.2.3 Emergence of Spreading Force from Two-Dimensional Self-Avoiding Polymer Physics

As a finite thickness polymer, the mean square extent of a polysaccharide polymer is subject to the law of self-avoiding polymers, *viz.*

$$< R^2 >= b^2 N^{2\nu} \tag{2.12}$$

with, as per flory theory [23]

$$\nu = \frac{3}{2+d} \tag{2.13}$$

obtained from a balancing of two entropic effects of self-avoidance and entropic springiness (high probability of zero to end-to-end distance). Here, b is the size of a polymer link, of the order of the persistence length. In two dimensions, $\nu = \frac{3}{4}$. Hence, as a potential polymer r is deposited forming a network with other polysaccharides or glycoproteins in the lumen of the cell plate and potentially tethers to each membrane side of the inner cell plate, a radial pressure can be exerted at the edge (Fig.2.15). The radial growth speed v_F is given by:

$$v_F = \frac{1}{2\pi R}\frac{d < R^2 >}{dt} = \frac{3}{2}\frac{b^2}{2\pi R}N^{\frac{1}{2}}(\frac{dN}{dt}) \tag{2.14}$$

and the rate of change of area is

$$\frac{dA}{dt} = \pi\frac{d < R^2 >}{dt} = \frac{3\pi}{2}b^2 N^{\frac{1}{2}}(\frac{dN}{dt}) \tag{2.15}$$

Hence, the magnitude of the radial force acting on the edge of the plate is (m_c is the mass of a constituent monomer i.e., glucose, and σ_c is the areal density of the polymer in the plate):

$$\frac{dp}{dt} = m_c\sigma_c\frac{dA}{dt}v_F = \frac{9\pi}{8}b^4\frac{m_c]sigma_c N}{\pi R}(\frac{dN}{dt})^2 \tag{2.16}$$

The spreading force λ or areal pressure is the total work done per unit area in expanding the plate a radial distance dR so:

$$\lambda = \frac{d|W|}{dA} = \frac{dp}{dt}\frac{1}{R} = \frac{9\pi}{8} b^4 m_c \sigma_c^2 (\frac{dN}{dt})^2 \tag{2.17}$$

Hence, in this simple model, the spreading force is directly related to the production rate of a polymer in the cell plate. Interestingly callose deposition with the size of a polymer $b \sim 200nm$ [24, 53] and a reasonable value for σ_c, provides an estimate of $\frac{dN}{dt} \sim 1.75X10^6 s^{-1}$ for $\lambda = 4\frac{pN}{nm}$ which aligns well with the model.

We recognize two important assumptions entering this two-dimensional picture of the spreading force: 1) the only way to generate such a quasi-two-dimensional force from a self-avoiding polymer network is by breaking the symmetry. There is already a symmetry breaking to the nascent cell plate via the phragmoplast guided vesicle delivery which aligns arriving vesicles and vesicle diameter sets the basic cell plate thickness. To prevent a 3D self-avoiding polymer growth, we must assume there is confinement of the growth which necessitates tethering of the polymer network to each side of the new cell wall. 2) Because of the out of equilibrium dynamics of polysaccharide production and the absence of equilibration from the cytosol to the interior of the cell plate, we are approximating the areal pressure at the edge of the cell plate using Newton's second law and relating that to potential polysaccharide production rates. This contrasts the usual energy supplied polymer ratchet model of actin or microtubules which assumes energy supply by ATP or GTP hydrolysis and a steady state of monomers and polymers, with force generation arising from monomers added at the leading edge of the polymer and removed from the trailing end [66].

2.4.2.4 Line Tension

We additionally considered the notion of a line tension at the membrane boundary lying on the equatorial plane. The energy of such a line tension would be $E_{spreading} = -intboundarytdl$, where t would have units of force, and dl is the line element along the boundary of a cell plate structure. However, we found that a line tension was unable to reflect the tendency of the spreading force to widen and expand tubular networks and it rather tended to proliferate the length of the membrane boundary instead. This resulted in the increase of the sizes of the fenestrations in fenestrated structures. Attempts

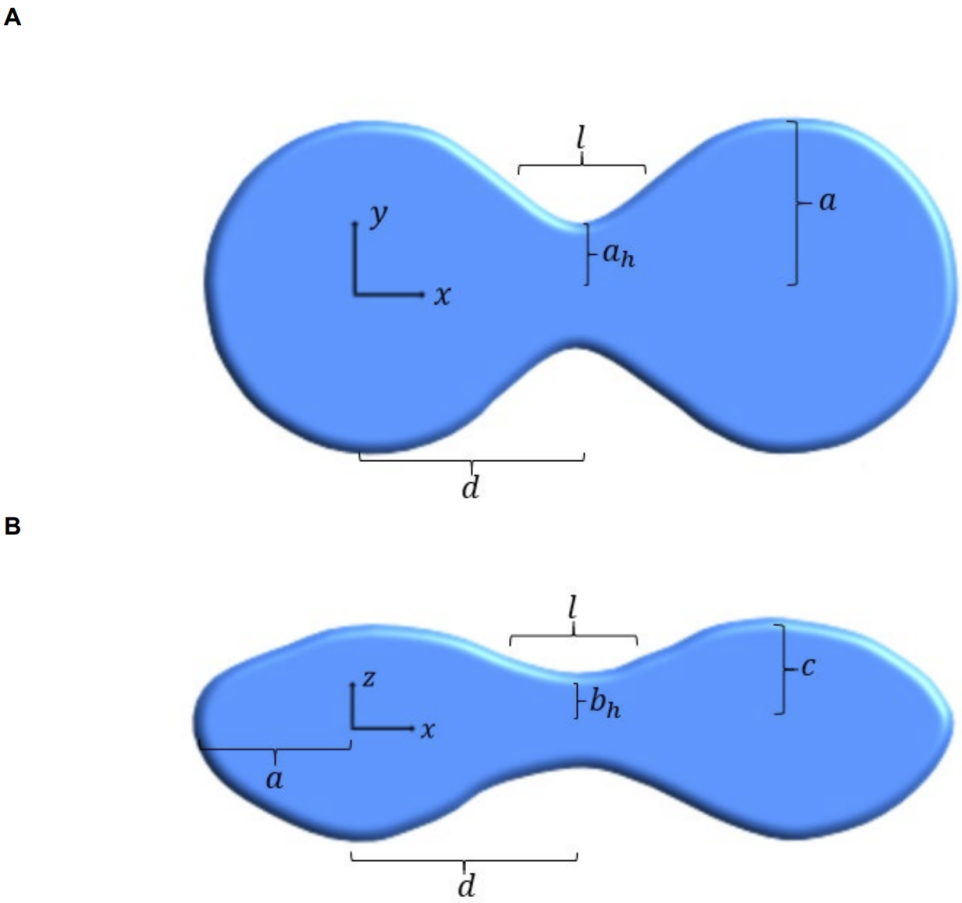


Figure 2.6: Parameters visualized on a representative 2x1x0 structure. To enforce continuity between an oblate spheroid of given parameters (a, c) and an elliptic hyperboloid with parameters (a_h, b_h, c_h), we can calculate d and b_h since they are dependent variables, a full conformation can be described by the type of conformation and the parameter set (a, c, a_h, l), or equivalently (a, c, a_h, c_h). The perpendicular arrows show the respective axes of the conformation. **A**, shows the top view of the conformation, while **B**, shows the side view of the same conformation.

to recreate experimentally observed structures of genus zero using this form of the line tension such as those found in Fig.2.2E resulted in unrealistic values for t (upwards of $200pN$). Additionally, this would be nonzero for a tangential force at the boundary, where the spreading force is anticipated to be normal to the boundary. We concluded, therefore, that a spreading force that is analogous to a two-dimensional pressure acting at the boundary of the equatorial plane is best represented by Eq.2.10.

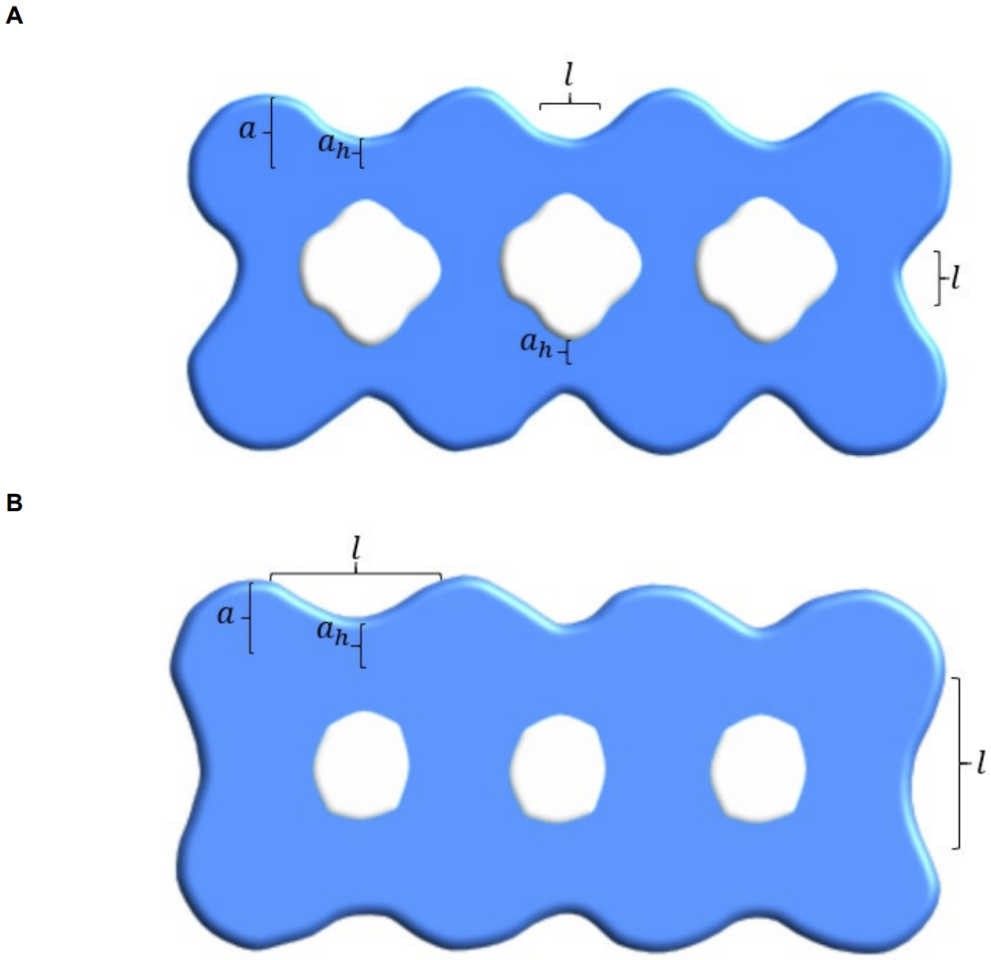


Figure 2.7: Effect of a spreading force visualized in a 8x10x3 conformation. The parameters shown here were extracted after energy minimization calculations on an 8x10x3 structure with parameter restrictions in place. In the absence of a spreading force, larger fenestrations, and narrower tubular connections are predicted, as shown in a top view in A. This structure has an area of $2x10^5 nm^2$, while the parameters (a, c, a_h, l) are given by $(52, 31.5, 20, 25.8)nm$. As a spreading force is turned on and the spontaneous curvature is decreased, the tubular connections widen, thereby shrinking the fenestration sizes, as shown in a top view in B. For the same area, the parameters change to $(58, 24, 33.5, 25.41)nm$. If we relax the imposed parameter restrictions in the presence of a spreading force, the resulting structure would reach a single oblate spheroid with $a = 173nm, c = 25nm$.

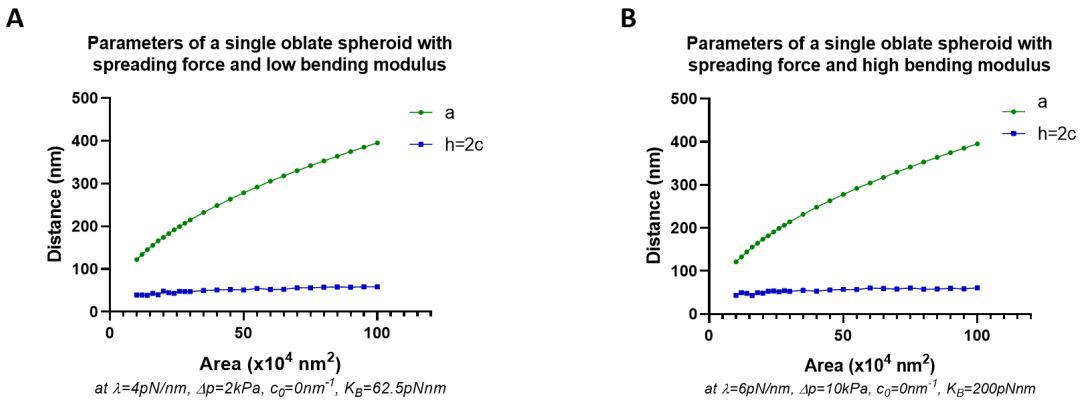


Figure 2.8: Evolution of single oblate spheroid parameters in the presence of a spreading force. Results with extremal values of the bending modulus are shown in **A** and **B**. Despite the increasing area, the height (h) remains in the 40-80nm region. With a smaller bending modulus, as in A, a smaller value of the spreading force parameter λ and pressure difference ΔP is required to maintain the height within the desired region for the specified areas. With a larger bending modulus, as in B, larger values of λ and ΔP are required.

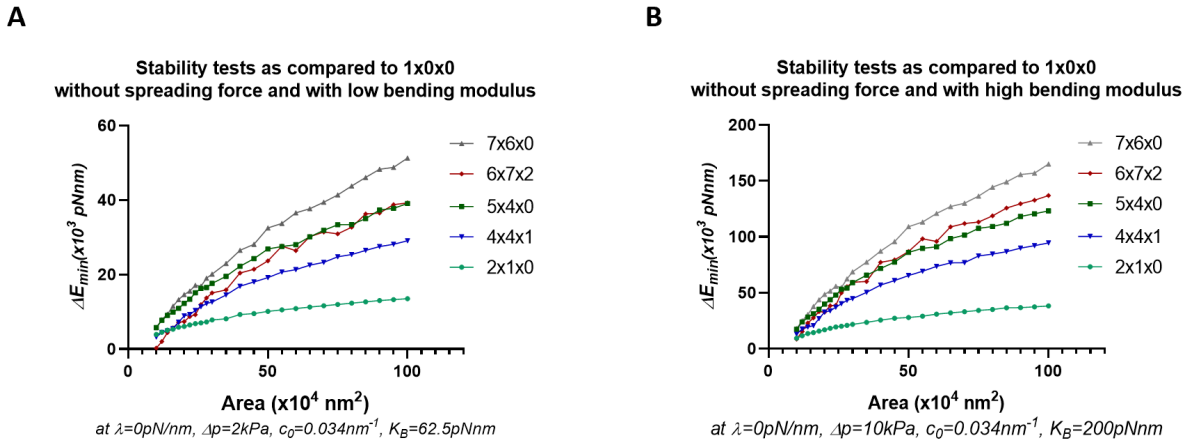


Figure 2.9: Stability tests of various configurations under different bending modulus in the absence of a spreading force. **A**, Stability tests for a small value of bending modulus while **B** shows calculations for a larger value of bending modulus. A positive value of ΔE_{min} indicates relative stability of the labelled conformation as compared to a single oblate spheroid. Note that in the absence of a spreading force and finite spontaneous curvature, increasingly tubular and fenestrated structures are more stable as compared to a single oblate spheroid. The different values of Δp and λ arise due to the constraints on structure thickness as shown in Fig. 2.2 and Fig.2.3.

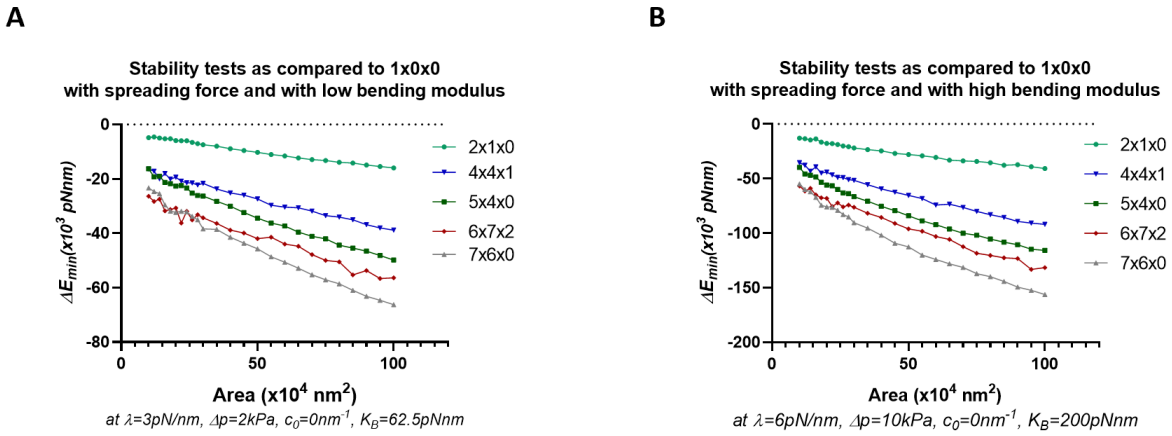


Figure 2.10: Stability tests of various configurations under different bending modulus in the presence of a spreading force and with zero spontaneous curvature. **A**, Stability tests for a lower value of bending modulus and **B** higher value of bending modulus. Note that with the presence of a spreading force and with zero spontaneous curvature, increasingly tubular and fenestrated structures (i.e.7x6x0) are increasingly unstable as compared to a single oblate spheroid, indicating the energetic favorability for cell plate structures to mature to a disk like shape. A positive value of ΔE_{min} indicates relative stability of the labelled conformation as compared to a single oblate spheroid. The different values of Δp and λ arise due to the constraints on structure thickness as shown in Fig.2.2 and Supplemental Fig.2.8.

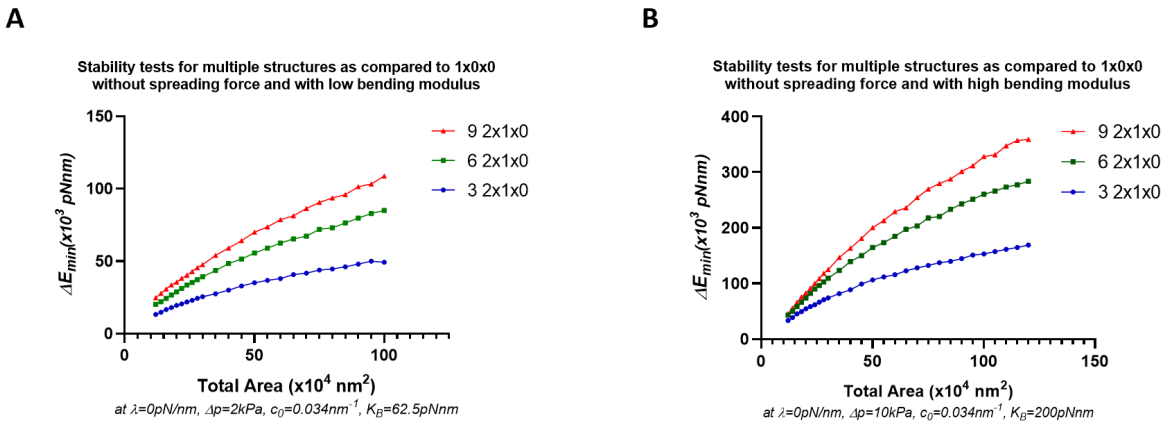


Figure 2.11: Stability tests of multiple 2x1x0 structures as compared to a single oblate spheroid in the absence of a spreading force and with finite spontaneous curvature. ΔE_{min} of multiple 2x1x0 structures as compared to a single oblate spheroid under for a low value of bending modulus **A** and high value of bending modulus **B** are shown. Note that in the absence of a spreading force and with finite spontaneous curvature, tubular structures are energetically favorable in these conditions, thereby modeling a membrane network stage. A positive value of ΔE_{min} indicates relative stability of the labelled conformation as compared to a single oblate spheroid. The different values of ΔP and λ arise due to constrains on structure thickness as shown in Fig.2.7 and 2.8

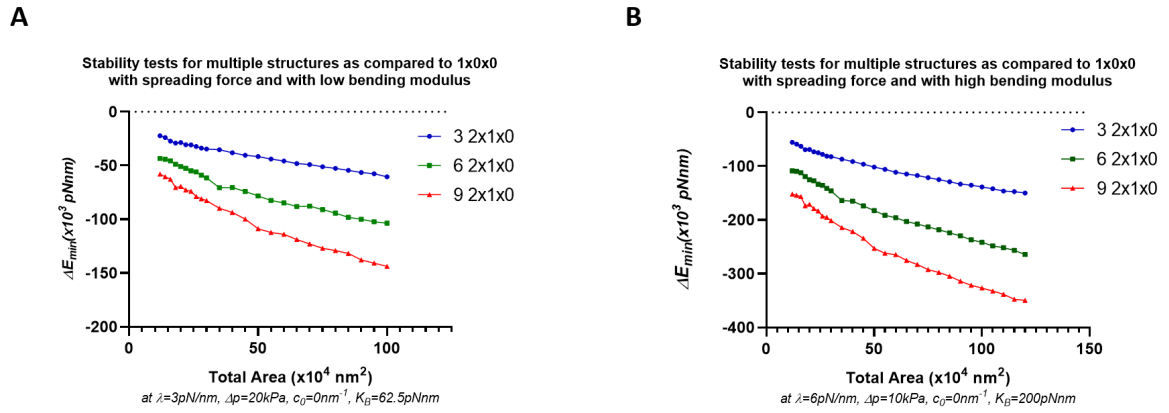


Figure 2.12: Stability tests of multiple 2x1x0 structures as compared to a single oblate spheroid in the presence of a spreading force and with zero spontaneous curvature. In the presence of a spreading force and with zero spontaneous curvature, tubular structures are unstable compared to a single oblate spheroid, thereby indicating the energetic favorability of structures fusing to form larger, more mature structure(s). **A**, shows results for a small value of bending modulus while **B**, shows results for a larger value of bending modulus. A positive value of ΔE_{min} indicates relative stability of the labelled conformation as compared to a single oblate spheroid. The different values of Δp and λ arise due to the limitations on structure thickness as shown in Fig.2.2 and 2.8.

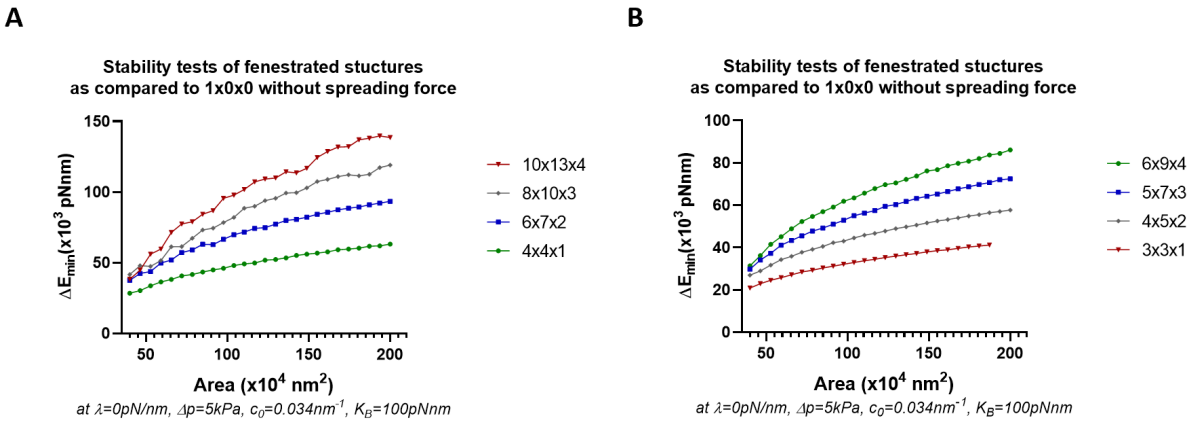


Figure 2.13: Stability tests of tubular/fenestrated structures as compared to a single oblate spheroid in the absence of a spreading force. In the absence of a spreading force, and with finite spontaneous curvature, fenestrated and tubular structures are, in general, more stable than a single oblate spheroid. This relative stability is magnified with the increase in area particularly for heavily tubular structures (10x13x4 in **A**, 6x9x4 in **B**), consistent with observations at tubular network/very early fenestrated sheet stages.

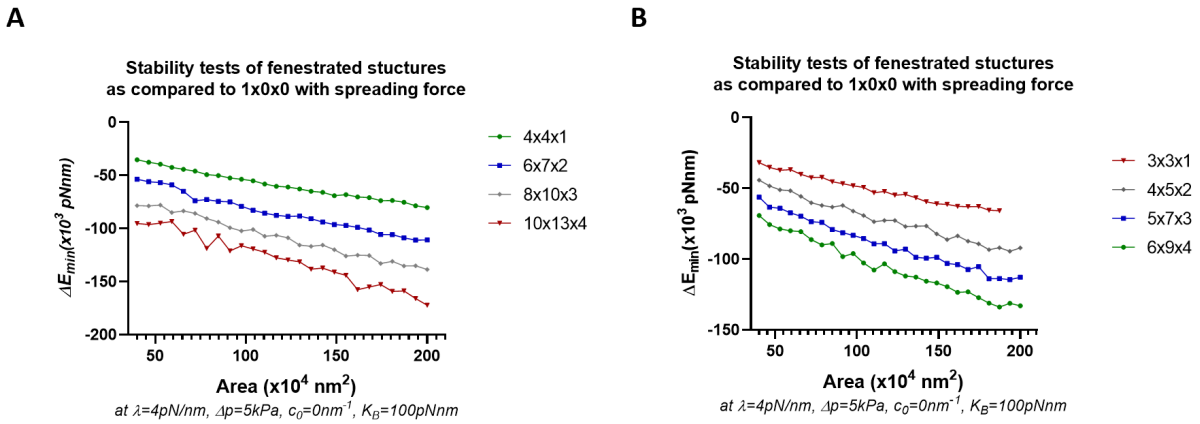


Figure 2.14: Stability tests of tubular/fenestrated structures as compared to a single oblate spheroid in the presence of a spreading force. **A, B,** In the presence of a spreading force, and with decreased spontaneous curvature, a single oblate spheroid is more stable compared to larger, tubular, fenestrated structures. This indicates the necessity of a spreading force to incur a transition from a tubular/ fenestrated sheet stage to a single mature cell plate structure.

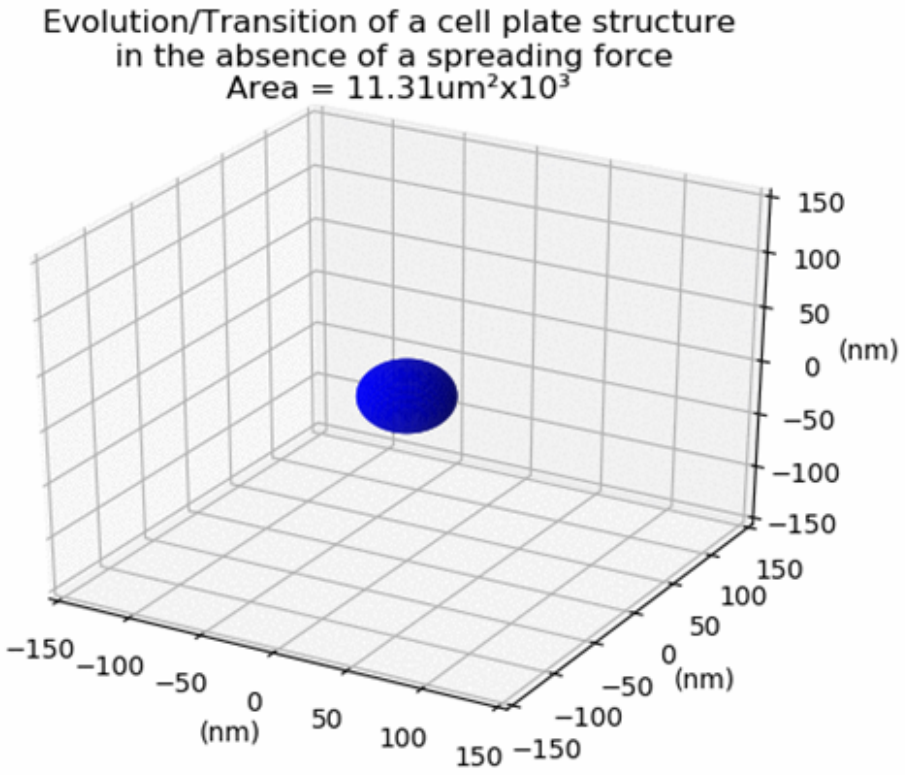


Figure 2.15: Evolution/transition of a cell plate structure in the absence of a spreading force as predicted by the model. Still image from Supplemental Video S1 [25]. As membrane area increases in the absence of a spreading force (with the same parameters as in Fig.2.3A), a vesicle eventually transitions to an oblate until a tube (shown in red, modeled by elliptical hyperboloids as shown in Fig.2.2B) forms in between, after which the tubular regions grows longer and narrower, taking away membrane material from the oblate regions (blue, modeled by oblate spheroids as in Fig.2.2A). Here, we see a transition from a 1x0x0 structure to a 2x1x0 structure in the absence of a spreading force. If the area were to continue increasing in the absence of a spreading force, we would likely see the formation of more tubes (structures like 3x2x0, 4x3x0) as well as fenestrations in some cases (4x4x1, 6x7x2..), as predicted in Fig.2.3B. Scale (x,y,z) (300nmx300nmx300nm). For full animation please see supplemental video S1 [25]

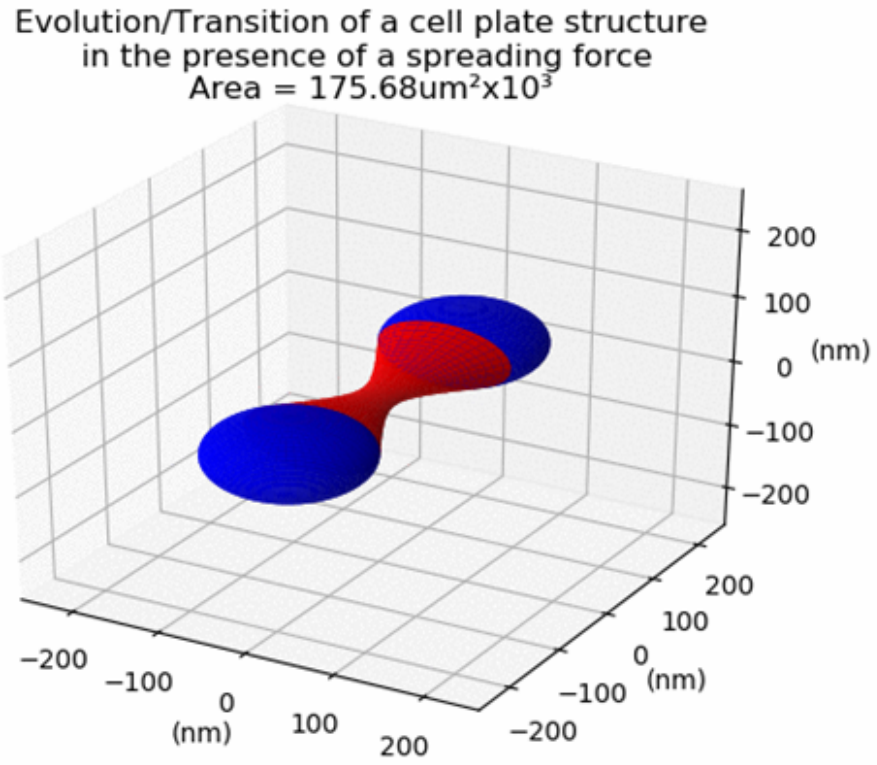


Figure 2.16: Evolution/transition of a final cell plate structure from Fig. S10 in the presence of a spreading force as predicted by the model. Still Image from Supplemental Video S2 [25]. As membrane area increases in the presence of a spreading force (starting from the shape in Fig.2.16, now using the same parameters as in Fig. 2.3B), the tubular regions (shown in red, modeled by elliptical hyperboloids as shown in 2.2B) fatten and widen until the structure is fully oblate. Here, we show a transition from a 2x1x0 to a 1x0x0 structure. We note that such a transition is not possible without a planar spreading/stabilizing force. If we were to start from a more tubular or fenestrated configuration with the same area, we would ultimately arrive at the same final shape (1x0x0). Scale (500nmx500nmx500nm). For full animation please see supplemental video S2 [25]

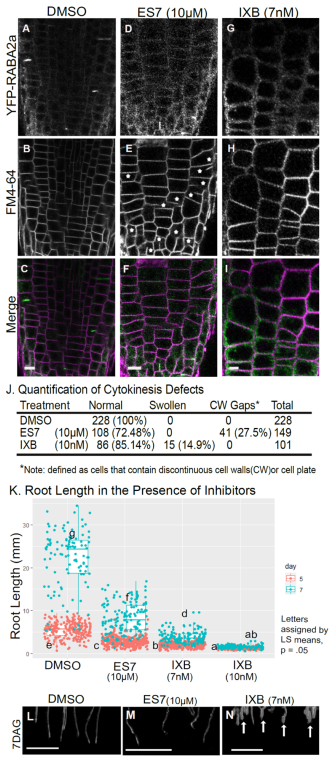


Figure 2.17: Effect of isoxaben (IXB) and Endosidin ES7 on cellular organization and root growth **A** − **I**, Cellular organization of Arabidopsis root tips in 7 nM IXB and 10 μM ES7. Arabidopsis root tips from 5 day old seedlings grown under chemical treatments. **A -C**, DMSO treated seedlings display regular cellular organization without cytokinetic defects. **D** − **F**, Treatment with ES7 leads to cytokinetic defects in the form of discontinuous cell walls, indicated by a star. **G** − **I**, Treatment with IXB leads to a swollen cell phenotype. The cytokinesis marker RABA2a is shown in green, while FM4-64 staining of plasma membrane is shown in magenta. Samples are single scans of live cell confocal imaging. Bars = 10μm. J. Quantification of the discontinuous cell wall phenotype showed a 27% in ES7 treatment with no discernable phenotype in DMSO or IXB treatment. Data represent quantification of 5-10 seedlings per treatment **K**, Germination of Arabidopsis seedlings in 7 nM IXB and 10 μmM ES7 5 and 7 Days after germination (DAG). The root growth inhibition is significantly higher under IXB treatment compared to ES7. Letters assigned by LS means, p = 0.05. (5 day/ 7 days). DMSO 7DAG n = 212, DMSO5DAG n =110, ES7 5DAG n = 356, ES7 7DAG n = 145, IXB 7nM 5DAG n = 324, IXB 7nM 5DAG n =136, IXB 5DAG 10nM n = 289,IXB 7DAG n=125. Individual data points of root length were plotted in box-whisker plot. Boxes indicate the median and interquartile range. Whiskers show 1.5 times the interquartile range. **L- N**, Root tips of Arabidopsis seedlings 7 days after germination in media supplemented with DMSO (L), 10μm ES7 (M) and 7 nM IXB (N). Note the prominent root swelling in IXB compared to the other treatments.

Chapter 3

Introduction to Part 2: *In Silico* Studies of SARS-CoV-2

3.1 Background and Motivation

In December 2019, a novel virus that is now referred as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was detected in Wuhan, China. The virus triggered a global pandemic which, at the time of writing, resulted in close to half a billion reported cases, in addition to about six and a half million deaths [67, 68]. The virus triggered several lockdowns that led to severe social and economic disruption. The resulting stoppages in supply chain networks caused severe food and resource shortages, triggering extraordinary ripple effects that the global economy has yet to recover from [69].

Scientists around the world rallied together to coordinate a joint response (Fig. 3.1. A complete genome was published within two months of detection [70, 71]. X-ray diffraction and electron microscope studies revealing the virus's spike protein crystal structure were published shortly after [72, 73], followed by crystal structures of bound antibody-spike protein complexes [74, 75, 76]. Extensive databases for reporting cases, genome data, and phylogenetic analysis were set up [77], and finally, mRNA and traditional vaccine candidates were available to the public in record time [78].

The aforementioned experimental advances were vital to the global response to the pandemic. However, as with any experimental approach, the time taken to collect and analyze experimental data continues to be a limiting factor. The immobilization of man-

power caused by countrywide lockdowns further exasperated experimental efforts to study the virus. *In silico* approaches to study molecular interactions are often used to guide experiments and thereby address experimental limiting factors. These approaches have become more and more relevant due to the exponential rise in computational power, as well as data management. These advances have been vital in our attempts to understand large scale molecular interactions, for example, protein-folding [79].

Together with experimental approaches, a tremendous amount of resources were allocated to studying the virus *in silico*. Countless important developments were made, including but not limited to, *in silico* studies of possible therapeutic agents [81], drug-discovery screens [82], model vaccine efficacy testing [83], molecular dynamics and binding strength estimates of critical receptor-ligand sites [84] among many others.

In this work, we will explore, using existing genomic and predicted data [71], a critical cleavage event that significantly increases the probability of the SARS-CoV-2 primary virus-cell entry mechanism [85]. We will also study receptor and antibody interactions of the SARS-CoV-2 omicron variant [84].

3.2 Methods: Studying molecular interactions using *in silico* approaches

In chapters 4, 5, and 6 of our work, we primarily use data from classical molecular dynamics simulations and we make binding strength estimates using a variety of statistical methods. A brief overview of the techniques used in chapters 4-6 is given below.

3.2.1 Classical Molecular Dynamics

Classical molecular dynamics (MD), as the name suggests, are a class of simulation techniques that are used to study the dynamics of systems with a large number of atoms. This is primarily done by numerically solving for the coordinates of atoms in discrete time steps using their equations of motions. The resulting trajectory coordinates are obtained by using an integrator that takes as an input the initial coordinates of the system, as well as a general expression for the potential energy of the system, through which the equations of motion are obtained.

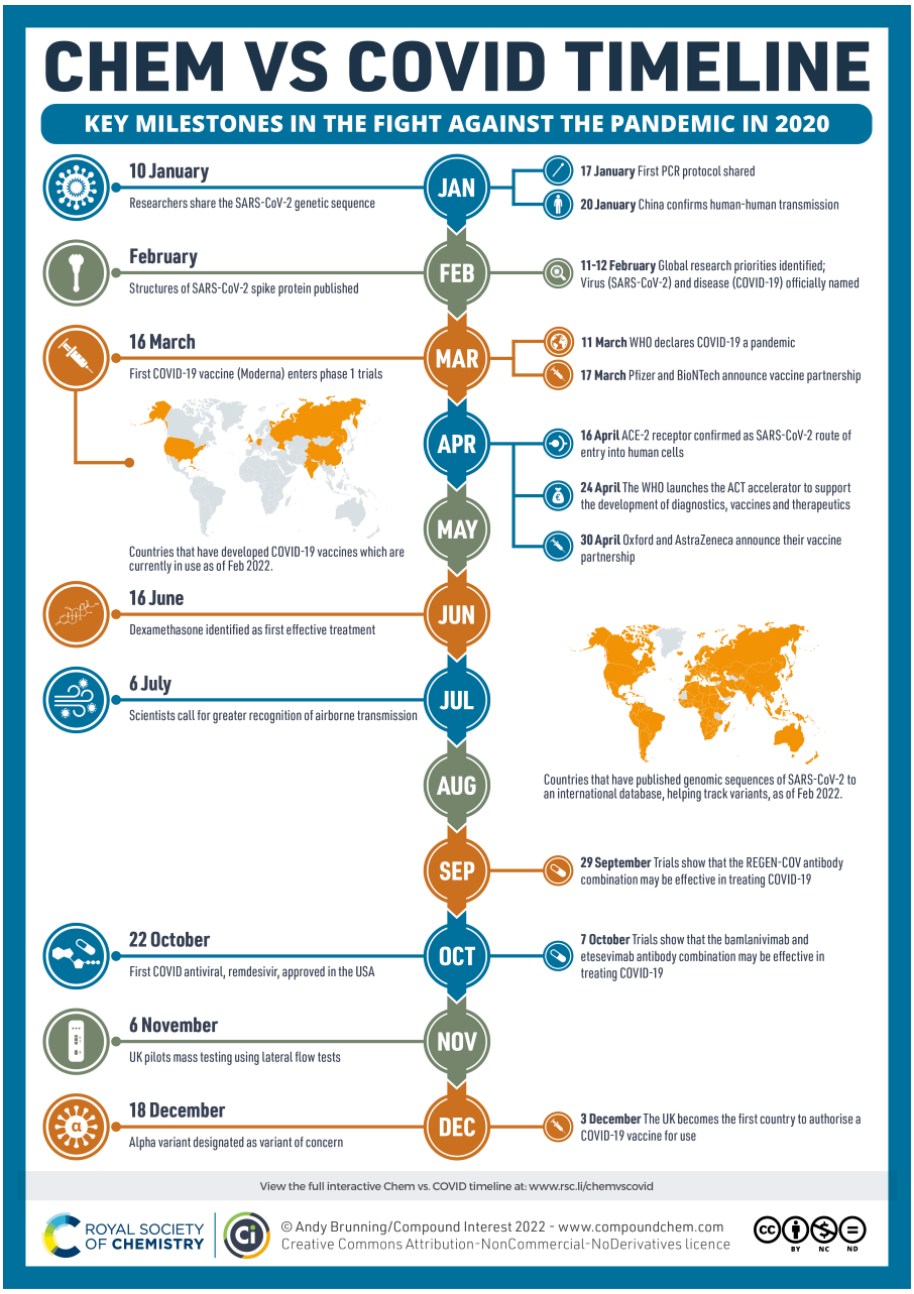


Figure 3.1: A timeline of COVID-19 research milestones [80] (Creative Commons License)

MD simulations depend on a vast array of initial conditions. These include, in no particular order of importance, the potential energy description of a general atom (the choice of this description is colloquially known as a force-field), the initial coordinates of the system, the choice of defining the solvent explicitly or implicitly, the time-step used, the ensemble conditions (such as temperature, pressure, volume, and energy), as

well as a model for the solvent. They are run for biologically relevant timescales (often $ns - \mu s$) depending on system size and equilibrium configurations, and therefore generate a tremendous amount of data which is then used to make conclusions for the system of interest.

A general potential energy for an atom in a system comprises of contributions resulting from non-bonded interactions (such as leonard-jones potentials, electrostatic coloumb potentials) and bonded interactions (bonded potentials, stretching/twisting, dihedral, and angular potentials). The exact choice of the parameters, cut-off distances, approximations, and atomistic models used in these potentials constitute a 'force-field'. Force-field developers obtain and tune these parameters by comparing the results of their simulations to vast libraries of experimental data. The primary force field used in this work is AMBER14 [86].

$$V_{\mathrm{AMBER}} = \sum_{i}^{n_{\mathrm{bonds}}} b_i(r_i - r_{i,\mathrm{eq}})^2 + \sum_{i}^{n_{\mathrm{angles}}} a_i(\theta_i - \theta_{i,\mathrm{eq}})^2 + \sum_{i}^{n_{\mathrm{dihedrals}}} \sum_{n}^{n_{i,\max}} (V_{i,n}/2)[1 + \cos(n\phi_i - \gamma_{i,n})]$$
$$+ \sum_{i<j}^{n_{\mathrm{atoms}}}{}' \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \sum_{i<j}^{n_{\mathrm{atoms}}}{}' \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}},$$

Figure 3.2: AMBER force-field potential proposed in 1995 [87]. The first term approximates the energy of covalent bonds using a harmonic potential. This assumes that covalently bonded atoms are close to equilibrium bond length. The second term represents the energy due to the geometry of electron orbitals involved in covalent bonding. The third term, also known as the torsion energy, represents the energy due to bond twisting. The fourth term models van der Waals interactions using a functional form similar to the Lennard Jones potential. The fifth term represents the energy due to electrostatics. The last two terms are referred to as non-bonded interactions.

Algorithms such as the particle mesh Ewald method [88], and use of cut-off distances for fast diminishing energies together with the implementation of efficient leap-frog integrators [89] are used to decrease computational load. An important consideration is the integration timestep used. A timestep that is too long risks exorbitant changes in energies and velocities that may be unphysical, while a timestep that is too short results in a higher computational load. Most MD simulations will employ a timestep that is at least a fifth of the the smallest period of oscillation which for most molecular interactions, tends to

be about 10-12fs for a single hydrogen bond [90]. Therefore, most MD simulations tend to use a timestep between 1-2fs.

3.2.2 Estimating binding free energy using MM/GBSA

A cornerstone of our work is using data from molecular dynamics simulations to predict the binding strength of protein-protein interactions. Multiple techniques can be used to achieve these estimates computationally. These include alchemical methods, free-energy perturbation methods, poisson-boltzmann surface area approximations, and Generalized Born surface area approximations [91]. The first two methods, while in principle more accurate, are very taxing computationally. We ultimately chose the Generalized Born surface area (MM/GBSA) method due to the relative speed of computation and flexibility.

The binding free energy change for a protein receptor and ligand to form a complex is usually expressed as [91]:

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \tag{3.1}$$

$-T\Delta S$ is the conformational entropy change upon binding. ΔE_{MM} is referred to as the gas phase molecular mechanics energy and can be further decomposed into:

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw} \tag{3.2}$$

Here, $\Delta E_{internal}$ refers to energy changes due to bond, angle, and dihedral energies. $\Delta E_{electrostatic}$ and ΔE_{vdw} refer to electrostatic and van der Waals energies respectively. The calculation of $\Delta E_{electrostatic}$ and ΔE_{vdw} is dependant on the choice of force field used and so will not be discussed here.

$\Delta E_{internal}$ is normally ignored because when snapshots are taken from molecular dynamics simulations of the complex directly, the contributions from the receptor, ligand, and the receptor-ligand complex cancel out to zero in principle [91].

Finally, ΔG_{solv} can be broken down as the sum of changes due to nonpolar (ΔG_{SA} and electrostatic contributions to energy (ΔG_{el}).

$$\Delta G_{sol} = \Delta G_{SA} + \Delta G_{el} \tag{3.3}$$

ΔG_{SA} is approximated by using the change in solvent accessible surface area in a method known as the LCPO method [92]:

$$\Delta G_{SA} = 0.0072X(SA_complex - (SA_receptor + SA_ligand)) \tag{3.4}$$

Finally, (ΔG_{el} is approximated using Generalized Born methods. A common functional form approximating (ΔG_{el} is [92]:

$$\Delta G_{el} \approx \Delta G_{GB} = -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f^{GB}(r_{ij}, R_i, R_j)}(1 - \frac{e^{-\kappa f_{ij}^{GB}}}{\epsilon_w}) \tag{3.5}$$

Where r_{ij} is the distance between atoms i and j, and R_i and R_j are the effective Born radii of atoms i and j. f^{GB} is a scaling factor, a common choice of which is expressed as:

$$f^{GB} = [r_{ij}^2 + R_i R_j exp(-\frac{r_{ij}^2}{4R_i R_j})]^{0.5} \tag{3.6}$$

The effective Born radius is interpreted as the degree of burial of an atom in a molecule. For reference, a common choice of a Born radius of an isolated atom is its van der waal radius. For sufficiently buried atoms, the choice of how to calculate the Born radius is sufficiently model and assumption dependent. A detailed review of commonly used models is given in [91, 92, 93, 94].

Normally, one stays consistent with the choice of Born radius model used between a set of simulations. Furthermore, MM/GBSA is averaged over a large number of snapshots taken across an MD simulation in an attempt to reduce noise.

3.2.3 Hbond Count

In chapter 4, we show that there is a strong correlation between the GBSA binding energy and the HBond count for cases which are generally well shielded from water molecules (Fig. 5.5). This Hbond count is obtained using a distance and angle approximation described in [95] (example shown in Fig. 3.3. This simplification means that salt bridges of proximate residues are effectively counted as H-bonds between basic side chain amide groups and acidic side chain carboxyl groups. Thus, for bound complexes, we will refer to the Hbond count as a proxy for the binding strength of a complex.

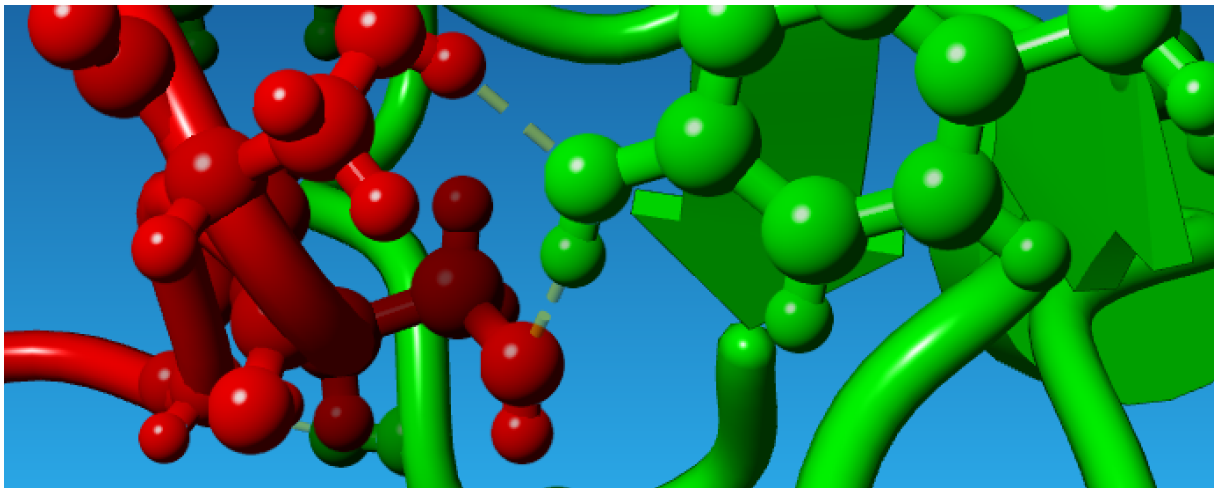


Figure 3.3: Example of detected Hbonds (yellow) between receptor (green) and ligand (red) using a distance and angle approximation [95].

3.3 Objectives of Part 2

In Chapter 4, we present binding strength estimates of three critical fitness parameters of the SARS-CoV-2 omicron variant [96] . These parameters are 1) Spike protein and human receptor binding (RBD/ACE2), 2) Furin binding to the virus's furin cleavage domain, and 3) key antibody interactions. We show that our results align with the preliminary observations noted with the variant, i.e. weakened RBD/ACE2 binding, but increased antibody escape.

In Chapter 5, we present an in-depth analysis of the most commonly observed sequences in the Furin Cleavage Domain (FCD) and their interaction with the furin enzyme. We show that the Delta variant exhibits the strongest possible binding with the furin enzyme, and we identify key observed and unobserved sequences that could exhibit the same binding strength. We also use the AlphaFold [97, 98] suite to predict binding modes between the furin cleavage domain and the furin enzyme, while verifying our predictions with observed crystal structures of furin inhibitors and furin complexes.

In Chapter 6, we present a computational design of a humanized ACE2 decoy to be used as a possible therapeutic or diagnostic agent based on the principles of competitive binding. We show that our design binds favorably well to multiple SARS-CoV-2 target RBDs, including the newer delta and omicron variants.

Chapter 4

SARS-CoV-2 Omicron simulations: broad antibody escape, weakened ACE2 binding, modest Furin cleavage

This chapter appears as a preprint at biorxiv [84] and is undergoing the peer review process at the time of writing. This work was in collaboration with Avinash Baidya, Rustin Mahboubi-Ardakani, Richard Davis, and Daniel Cox.

4.1 Introduction

The omicron variant of the SARS-CoV-2 virus was first detected publicly in Nov. 2021 [99], and traced back to variants which appeared in mid 2020. Because the variant contains a large number of mutations relative to the original strain, including three relevant regions of the viral surface spike protein (the receptor binding domain (RBD), the furin cleavage domain (FCD), and the n-terminal domain (NTD)), the variant is of great concern. There is preliminary evidence that it has overtaken the predominant delta variant in South Africa where it was first detected [96].

The fitness of a particular variant depends upon several factors. First, strong binding to surface receptors is of critical importance, and the SARS-CoV-2 RBD binds with high affinity to the ACE2 protein on human cells [100]. This contrasts with likely weaker

binding of coronaviruses associated with the common cold such as OC43 which binds more weakly to sialic acid groups on the cell [101]. Second, escaping the background antibody (Ab) spectrum can confer relative fitness to the dominant variant. Third, efficient membrane fusion and transmission is apparently strongly regulated by the FCD. It has been shown, for example, that ferrets inoculated with a WT SARS-CoV-2 with the FCD deleted can become infected but fail to transmit to other ferrets [102]. The high viral load of the delta variant has been clearly associated with the mutation P681R of the FCD [103] and has led to the current dominance of 96% of SARS-CoV-2 sequences worldwide [104].

Given the time lag in carrying out protein synthesis, structure determination of bound complexes, determining protein binding affinities, and measuring viral neutralization by Abs for new variants, there is clearly a role for rapid computational studies that can assess the differences of new variants relative to background variants as they arise.

In this Letter, we point out here that computational *ab initio* molecular dynamics studies of omicron RBD-ACE2, RBD-antibody (AB), FCD-Furin, and NTD-antibody are consistent with: 1) robust antibody escape in all regions compared to wild type (WT) and delta, 2) FCD binding to Furin intermediate between WT and delta, and 3) weaker binding to the ACE2 than WT or delta. The Ab escape can confer transmissibility advantages for a population with a prevalent delta variant Ab spectrum, but the weaker binding to ACE2 and modest enhancement of furin binding are likely to lead to weaker transmissibility than delta. This work uses ColabFold's [105] implementation of AlphaFold-Multimer [98] to generate structures for FCD-furin binding.

4.2 Materials and Methods

4.2.1 Molecular Models

We drew our starting structures for RBD-ACE2 binding from PDB file 7A94 [72] reference for the WT and delta variants, and PDB file 6M0J [73] for the omicron variant. For Class I ABs, which bind in the same region of the RBD as the ACE2, we used 7CJF [75] and 7KVF [74], while as a representative class III Ab that binds to the RBD away from the

ACE2 interface, we used 6YOR [76]. For an NTD-Ab we used 7C2L [106].

The antibodies chosen do not comprehensively portray all neutralizing Abs for the SARS-CoV-2 spike protein, but are representative of the spectrum of antibodies that neutralize the SARS-CoV-2 virus. This study does not account for t-cell binding sites [107]. Fig. 4.1 shows the structures of the different complexes studied in this paper.

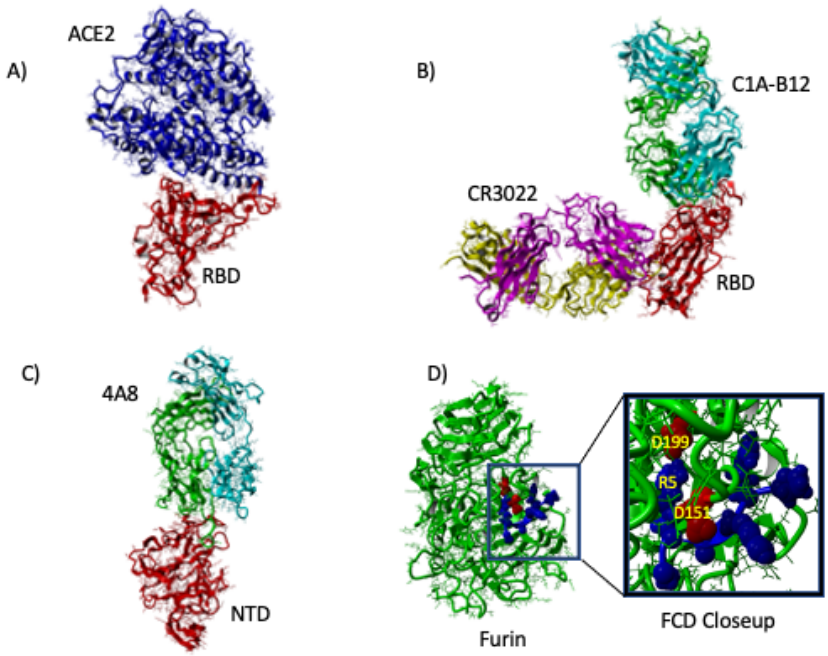


Figure 4.1: **Structures of WT spike protein complexes studied** A) ACE2(red)-RBD(blue) binding. B) Binding of RBD(red) to Class I Ab C1A-B12 (binds in ACE2 interface region; heavy chain green, light chain cyan) and Class III Ab CR3022 (binds away from ACE2; heavy chain magenta, light chain yellow). C) Binding of NTD to 4A8 Ab (heavy chain green, light chain cyan). D) Binding of FCD (blue) to furin (red). Blowup highlighting position of fifth residue R5 (R685 for WT SARS-CoV-2) with proximate aspartic acid residues D151, D199 of the furin enzyme.

4.2.2 Molecular Dynamics

To simulate the protein-protein interactions, we used the molecular-modelling package `YASARA` [89] to substitute individual residues and to search for minimum-energy conformations on the resulting modified structures of the complexes listed in Table 4.1. For all of the structures, we carried out an energy-minimization (EM) routine, which includes steepest descent and simulated annealing (until free energy stabilizes to within 50 J/mol)

minimization to remove clashes. All molecular-dynamics simulations were run using the AMBER14 force field with [108] for solute, GAFF2 [109] and AM1BCC [110] for ligands, and TIP3P for water. The cutoff was 8 Å for Van der Waals forces (AMBER's default value [111]) and no cutoff was applied for electrostatic forces (using the Particle Mesh Ewald algorithm [112]). The equations of motion were integrated with a multiple timestep of 1.25 fs for bonded interactions and 2.5 fs for non-bonded interactions at $T = 298$ K and $P = 1$ atm (NPT ensemble) via algorithms described in [95]. Prior to counting the hydrogen bonds and calculating the free energy, we carry out several pre-processing steps on the structure including an optimization of the hydrogen-bonding network [113] to increase the solute stability and a pK_a prediction to fine-tune the protonation states of protein residues at the chosen pH of 7.4 [95]. Insertions and mutations were carried out using `YASARA`'s BuildLoop and SwapRes commands [95] respectively. Simulation data was collected every 100ps after 1-2ns of equilibration time, as determined by the solute root mean square deviations (RMSDs) from the starting structure.

The hydrogen bond (HBond) counts were tabulated using a distance and angle approximation between donor and acceptor atoms as described in [113]. Note that in this approach, salt bridges of proximate residues, are effectively counted as H-bonds between basic side chain amide groups and acidic side chain carboxyl groups.

4.2.3 Endpoint Free Energy Analysis

We calculated binding free energy for the energy-minimized structure using the molecular mechanics/generalized Born surface area approximation [91, 94, 114], which is implemented by the `HawkDock` server [93]. While the MM/GBSA approximations overestimate the magnitude of binding free energy relative to *in-vitro* methods, the obtained values correlate well with H-bond counts. For each RBD-ACE2, RBD-AB, and NTD-Ab binding pair we average over five snapshots of equilibrium conformations. For each FCD-furin pair, we average over ten snapshots of equilibrium conformations.

4.2.4 Use of ColabFold/AlphaFold for Furin binding domain

We will present a complete description of the use of ColabFold/AlphaFold for modeling the FCD-furin binding separately [85]. Full details of this method are provided in [97, 98, 105]. In brief, we used the heterocomplex prediction method known as AlphaFold-Multimer [97, 98] as implemented within ColabFold[105] to predict the best bound structure to the furin enzyme of the six residue FCD from the WT protein. The delta and omicron structures were then obtained by mutation from the predicted WT FCD-furin structure.

4.3 Results

4.3.1 Binding Strengths: HBond and Binding Free Energy

Our main results for interfacial HBonds are summarized in Table 4.1. We find weaker binding to the ACE2 receptor in contrast to a recent quantum calculation[115], which should moderate infectivity, and significant antibody escape of the omicron for all three regions (Class I, Class III, and NTD) considered. This escape is measured by the reduction in hydrogen bond count between the antibodies and the spike protein. For the FCD-Furin binding, six residues fit into the binding pocket, which we argue elsewhere to begin with residue 681 for WT, alpha, and delta. For omicron, we consider the possibility of leading with the N679K mutation or P681H mutation. The latter is the same as the alpha variant. We see that the expected binding to the FCD is at best the same as the alpha variant, and significantly less than the delta variant.

The binding energies from the GBSA analysis of molecular dynamics equilibrium conformations are shown in Table 4.2. The same PDBs are utilized. Evidently the trend of binding energies tracks well with the easier to estimate interfacial HBond count, with the lone exception of the Class III Ab binding to the delta variant.

4.3.2 Mutations leading to Ab Escape and weaker ACE2 binding

Fig. 4.2 illustrates the key mutations leading to differences in binding for the delta and omicron variants relative to WT.

ACE2 For ACE2 binding, four key mutations weaken the ACE2 binding for omicron: 1) K417N removes the K417(RBD)-D30(ACE2) salt bridge. 2) Q493K removes hydrogen

Bound Pair	WT	delta	omicron	PDB
RBD-ACE2	8.43±1.69	8.32±1.46	5.06±1.20	7A94(WT,delta) 6M0J
RBD-P4A1 (Class I)	17.76±2.0	17.61±2.03	13.9±1.89	7CJF
RBD-C1A-B12 (Class I)	19.23±1.87	14.94±2.66	11.41±2.49	7KFV
RBD-CR3022 (Class III)	12.59±1.82	12.0±1.53	10.85±1.43	6YOR
NTD-4A8	9.62±1.78	7.48±1.49	4.73±1.63	7C2L
FCD-Furin	10.93±1.72	15.36±1.86	8.65±1.59 (K679) 12.05±1.68	NA

Table 4.1: Interfacial hydrogen bonds between proteins for WT, delta, and omicron. Last column: reference PDB structure

Bound Pair	WT	delta	omicron
RBD-ACE2	-62.9±6.8	-64.3±3.6	-55.5±4.5
RBD-P4A1 (Class I)	-115.9±7.0	-110.2±7.5	-75.6±9.5
RBD-C1A-B12 (Class I)	-77.9±8.0	-70.3±8.4	-50.8±5.6
RBD-CR3022 (Class III)	-95.4±2.5	-111.7±7.8	-84.1±13.3
NTD-4A8	-88.9±17.9	-81.1±5.7	
FCD-Furin	-83.6±8.4	-117.3±4.8	-63.5±3.6 (K679) -93.2±3.8

Table 4.2: GBSA Binding free energy estimate in kcal/mole

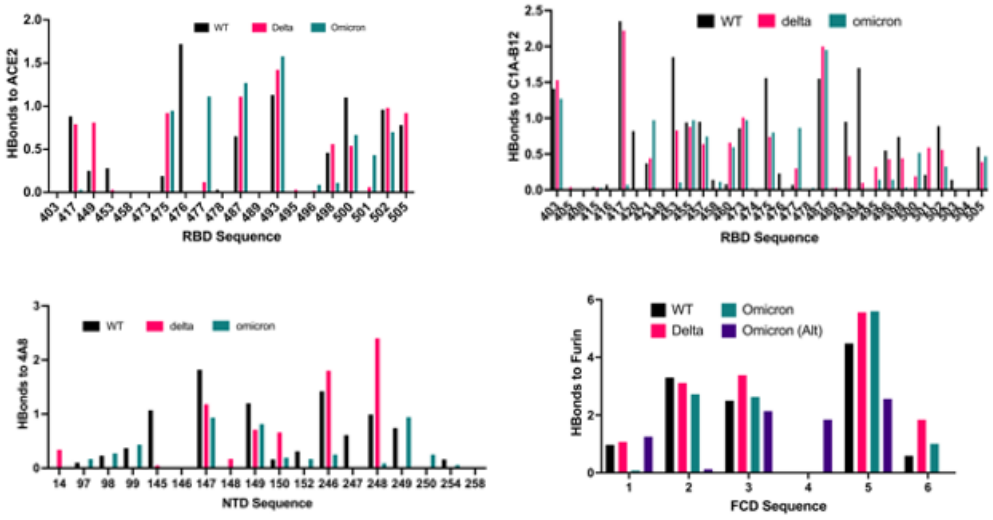


Figure 4.2: **Overview of binding changes for delta and omicron variants relative to WT** Color coding is the same for all charts. For the FCD to furin binding, R1-R6 correspond to 681-686, except for the alternate omicron sequence 679-684. For clarity, RBD binding to P4A1 and CR3022 Abs are not shown.

bonding between the glutamine side chain and residues K31 and E35 on the ACE2 driven by K-K coulomb repulsion. 3) Q498R removes hydrogen bonding between the glutamine side chain and K353 of the ACE2 driven by R-K Coulomb repulsion. 4) Y505H removes hydrogen bonding between the Y505 sidechain where the O acts as a donor and the E37 sidechain of the ACE2.

Class I Abs For Class I antibodies, the following mutations are critical to reducing binding strength: For binding to P4A1, 1) the Y455 binding to Y33.HC of the Ab heavy chain (HC) is removed. 2) The Q493K, G496S, and Q498R mutations lead to removal of bonds with E101.HC, W32.LC of the Ab light chain, and S67.LC. 3) The Y505H mutation removes bonds to S93.LC. For binding to C1A-B12, 1) the K417N mutation removes a salt bridge to D96.HC, a side chain bond to S98.HC, and weakens a side chain bond to Y52.HC. 2) The mutations Q493K, G496S, and Q498R remove bonds to R100.HC, S30.HC, and S67.HC. 3) The N501Y and Y505H mutations weaken bonds in the 501-505 region to G28.LC, S30.LC, and S93.LC.

Class III Ab For the Class III antibody CR3022, the most noticeable differences compared to WT are 1) the absence of binding at N370 to Y27.HC. This appears to be driven by the hydrophobic substitution S371L, which pulls the asparagine at 370 out of bonding distance from Y27.HC. 2) Weakened bonding of T385 to S100.HC.

NTD Ab For the NTD Ab 4A8, we find that the notable differences of omicron compared to WT are 1) weakened binding at 145-152 presumably due to the deletion at 142-145 relative to WT, and 2) significantly weakened bonding at 246-254 driven by the EPE insertion at 214 and the deletion at 211. Both the 142-145 deletion and the 211 deletion with EPE insertion disrupt the epitope positionings at 145-152 and 246-254 respectively.

4.3.3 Mutations in the FCD

As we will discuss in more detail elsewhere, for the generic 681-686 sequence of the FCD, the most critical residue apears to be the 685. In the WT, the arginine is able to form a salt bridge in the interior pocket with D199 of the furin, and bond additionally with S146, W147, D151, A185, and S261. This tendency is illustrated in Fig. 4.2. These bonds are all strengthened for delta and omicron. For the alternate KSHRRA sequence of the omicron, beginning at 679, the position of the arginine in the binding pocket allows only the salt bridge formation with D199.

As shown in Fig. 4.3, we observe that the binding strength, which is determined to a large degree by the binding of the fifth residue of the FCD, correlates inversely with the root mean square fluctuation (RMSF) of the backbone Cα of the first FCD residue at 681. This suggests that locking the 681 Cα as happens for P681R is a key to lowering the fluctuation spectrum of the 685 residue allowing for stronger binding at this site. Evidently, the gain in binding enthalpy offsets any advantages in conformational entropy for the FCD.

4.4 Discussion

The binding strength of Furin to the FCD appears to correlate well with the fluctuations of the initial residue at 681. The lower the fluctuation of the backbone carbon, the lower

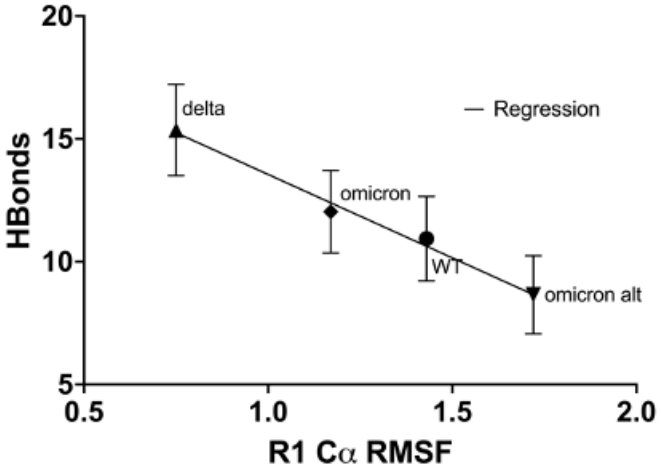


Figure 4.3: **Correlation of FCD-furin interfacial HBond count with RMSF of first residue in FCD** The higher the RMSF of the first residue in the FCD, the harder it is to bind to the furin, especially for the critical fifth residue which inserts into the furin pocket as shown in Fig. 4.1D. R1 is residue 681 for all but the alternate omicron sequence which starts at residue 679.

the fluctuation of the backbone carbon for residue 685, which dominates the bonding to the furin. The P681R mutation provides the lowest Cα RMSF observed amongst the four FCD examples considered here, and the alternate K679 starting point for omicron provides the largest Cα RMSF.

The lower severity of omicron versus delta is related to the Furin Cleavage Domain. Once a corona virus is bound to a human cell, disease severity is regulated by furin cleavage of the spike protein. After initial binding to the human ACE2 protein, Furin protease cleavage breaks the spike to facilitate cell wall fusion[103] and viral reproduction. The higher the furin-FCD HBond binding count, the more efficient the fusion at the molecular level, and ultimately, higher viral load on the host.

In summary, a consistent picture of omicron in comparison to the delta strain is emerging. Hospitalization data points to higher disease transmissibility but lower severity for the omicron strain compared to delta[116]. Our simulations see lower interfacial HBond counts for omicron for known RBD and NTD binding regions consistent with this, as well

as weaker ACE2 binding and furin binding than the delta variant. Against an immunity background tuned to the delta variant, omicron will be more transmissible. (Note that two other studies suggest stronger ACE2 binding from omicron[115, 117].) Experimental studies of the binding of the RBD to ACE2 and the FCD to furin will be needed to confirm these predictions.

4.5 Acknowledgments

We acknowledge useful conversations with Javier Arsuaga, Victor Muñoz, and Mariel Vazquez.

Chapter 5

Computational study of the furin cleavage domain of SARS-CoV-2: delta binds strongest of extant variants

This chapter appears as a preprint at biorxiv [85] and is undergoing the peer review process at the time of writing. This work was in collaboration with Avinash Baidya, Sofia Jakovcevic, Jacob Lusk, Rustin Mahboubi-Ardakani, Nathan Solomon, Georgina Gonzalez, Javier Arsuaga, Mariel Vazquez, Richard Davis, and Daniel Cox.

5.1 Introduction

While the spike protein of the SARS-CoV-2 virus is similar to SARS-CoV-1, a key difference is a polybasic insertion beginning at P681 in the spike protein[68]. It has been shown that this insertion is critical to the higher transmissibility of SARS-CoV-2[118, 102] over SARS-CoV-1, and that the mutations P681H for the alpha and omicron variants and P681R for the delta variant play a large role in increased transmissibility of the variants over the wild type (WT)[103]. Similar polybasic furin cleavage domains (FCDs) occur in other human coronaviruses OC43, HUK1, 229E, MERS, and NL63[119], and in many other viruses including H5N1 influenza[120].

The FCD of SARS-CoV-2 has not been well studied for at least two reasons. First,

the FCD belongs to a rapidly fluctuating random coil region of the protein which has not been resolved by structural probes (see, e.g., Ref. [72], PDB structure 7A94, for which residues 677-688 are unresolved) . Second, because the furin rapidly cleaves the protein at this domain, there are no bound structures available. The absence of structural data has limited computational studies of the binding domain.

A number of small peptides that can act as furin inhibitors have been studied elsewhere. It is known that the four amino acid inhibitor RVKR, suitably terminated, is a potent inhibitor of furin activity[121]. Right handed hexa-arginine and nona-arginine peptides are potent inhibitors of furin also[122]. Additionally, the peptide Arg-Arg-Arg-Val-Arg-4-aminomethyl-benzamidine (RRRVR-Amba, I1 peptide)[123], is similar to the delta variant FCD RRRARS, and binds to furin with pM affinity. This leads to the conjecture that the FCDs of SARS-CoV-2 and other coronaviruses may bind in similar fashion to the furin enzyme. For SARS-CoV-2, the insertion that begins with P681 for the WT, alpha, delta, and omicron variants commences a six residue sequence (through 686) that hijacks the furin enzyme from its useful physiological functions to assisting the virus. We have focused on several six amino acid FCDs for SARS-CoV-2 and other viruses.

In the absence of structural data for the FCD, we turned to the deep learning based AlphaFold program[97]. We used AlphaFold Multimer[98], as implemented within the ColabFold environment[105], to generate candidate structures for the FCD-furin complexes. We find that AlphaFold accurately predicts the furin structure and the backbone of the bound furin-I1 structure (Figs. 5.1A,B), so it is natural to attempt binding to the FCD, shown for WT in Fig 5.4C. We have used AlphaFold Multimer as the only way to generate a *de novo* structure for the WT FCD to furin binding. With this hypothesis, we can either generate *de novo* structures from AlphaFold Multimer, or assume the WT is well represented by the AlphaFold candidate structure, and mutations from that structure can be used to assess the binding of the FCDs for variants and other viruses. We simulate these structures with molecular dynamics to assess equilibrium binding strength, characterized by two quantities, interfacial hydrogen bonds between the furin and FCD (FCD-furin HBonds), and Generalized Born Surface Area (GBSA) binding energies. Details of the

simulation protocols are in the Supplementary Materials and Methods section 0.4. The AlphaFold approach reaches different conclusions about the FCD-furin bound structure than an earlier approach based upon docking[124].

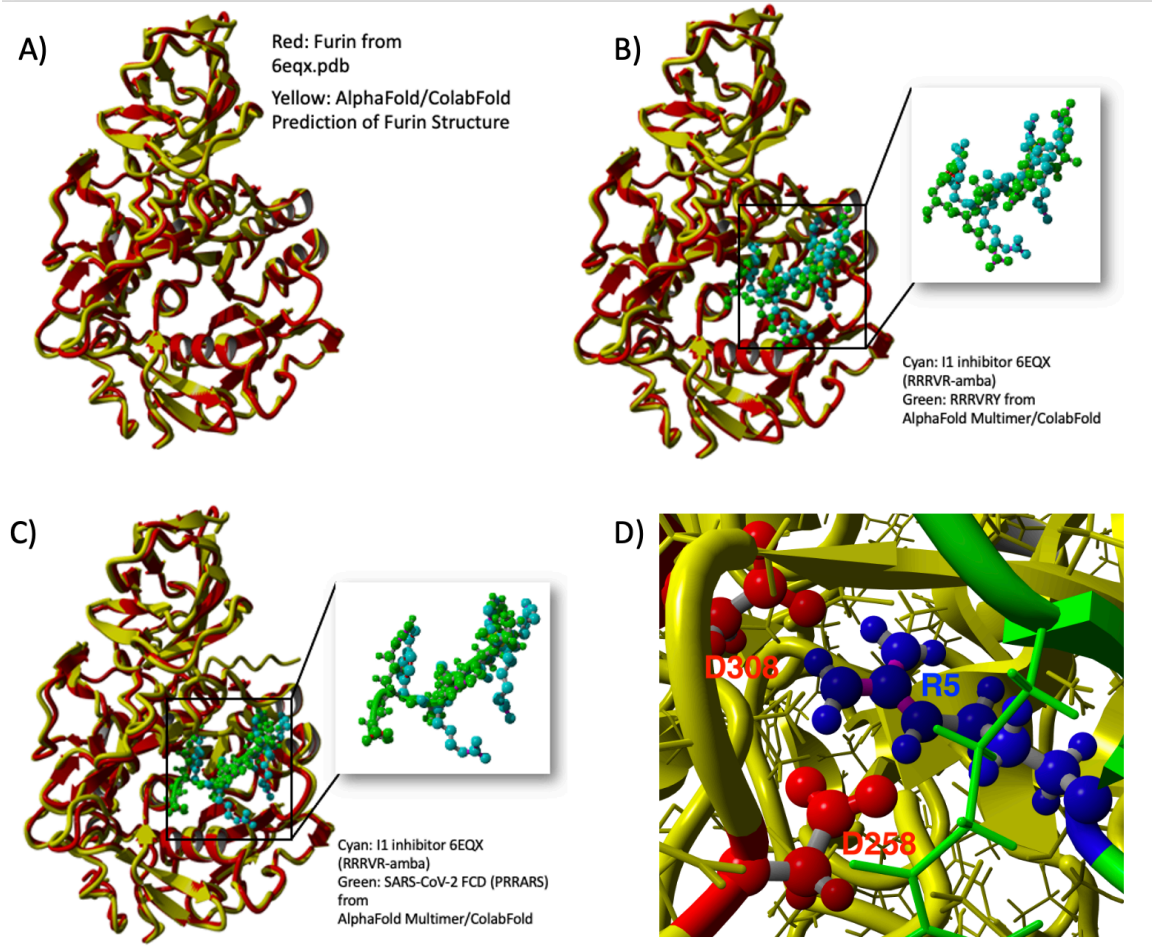


Figure 5.1: A) Comparison of structure of furin from Ref. [124] and PDB file 6EQX with the structure from AlphaFold[97] using the ColabFold environment[105]. Clearly, the agreement is excellent (RMSD of 1.79Å). B) Comparison of structure of furin with RRRVR-Amba inhibitor from Ref. [123] with structure generated for the similar sequence RRRVRY by AlphaFold Multimer[98] using ColabFold[105]. The Amba is buried in the furin S1 pocket[123] for the inhibitor, while AlphaFold predicts burial of the R at position 5. The backbone RMSD between the I1 and RRRVRY peptides is 2.77Å . C) Predicted structure by AlphaFold Multimer[98] for the WT PRRARS sequence of SARS-CoV-2 compared to Furin-I1 structure. D) Close up of binding pocket for fifth residue of PRRARS (WT FCD). Furin backbone in yellow, FCD backbone in green, R5 from FCD is blue, D258,D306 from furin in red.

In the I1 sequence, the sixth (Amba) residue, a nonstandard amino acid, binds most

strongly to furin as we discuss below. When we mutate that nonstandard residue to the structurally similar tyrosine, the binding pocket is occupied by the arginine at sequence position 5. Accordingly, we hypothesize that insertion of the residue at position 5 into the furin S1 pocket is the most important for FCD binding to furin. We confirm this hypothesis by simulating dozens of observed sequences. In 93% of observed SARS-CoV-2 FCD sequences starting from aligned position 681, the fifth amino acid is arginine.

We obtain a number of important results. First, per Fig. 5.2, the delta variant has the strongest binding of existing extant SARS-CoV-2 variants, and only two rare or unobserved FCD sequences bind as strongly within statistical accuracy. This dominance of the delta variant FCD extends to other coronaviruses and the H5N1 influenza virus. Second, as made clear in the heat map of Fig. 5.3, the most important residue is the fifth residue, which binds in the S1 pocket of furin[123] containing two aspartic acid residues. In particular, this pocket matches structurally to arginine better than lysine as discussed below. Third, we find that there is mechanistic predictive power in three quantities that help explain the differences between delta and other variants and viruses: (1) the strength of the binding strongly correlates inversely with the root mean square fluctuation (RMSF) of the first residue. This suggests that the more the backbone outside the pocket fluctuates, the less likely the arginine at position 5 can bind well to the furin. (2) The number of FCD-furin hydrogen bonds between residue 5 and the furin strongly predicts the total binding strength, even though it only represents a plurality of the HBonds. (3) The maximum mean number of FCD-furin HBonds for a given number of basic residues peaks at 15.7 hydrogen bonds for 4.06 basic residues.

5.2 Results

To avoid confusion between the conventional N-to-C terminal sequence numbering of peptides and proteins vs. the reverse numbering used in the Furin Data Base (FurinDB)[125] and other references[122, 121, 123], we will refer to the FCD residues as positions 1-6, which for all viral sequences considered here will correspond to the FurinDB notation P5-P4-P3-P2-P1-P1', with the cleavage site between P1 and P1'. For example, in the WT

SARS-CoV-2 FCD sequence PRRARS, the R at position 5 corresponds to P1, the S to position P1'. We will note the FurinDB identification parenthetically.

We first applied AlphaFold[97] through the ColabFold[105] environment to examine how well we could match the folded structure of furin. The result is shown in Fig. 5.1A. The AlphaFold structure for furin matches that from the PDB entry 6EQX[123] to within a root mean square deviation of 1.79Å. Next, we included the furin inhibitor RRRVR-4-aminomethyl-benzamidine (RRRVR-Amba) from Ref. [123] into the AlphaFold Multimer program[98], but because we could not enter the nonstandard residue Amba into the AlphaFold search we substituted tyrosine, which is similar to Amba away from the side chain terminus. As shown in Fig. 5.1b, this produces a structure substantially similar to furin with bound RRRVR-Amba, except that the S1 pocket, which binds the position 6(P1) Amba nonstandard residue, accepts the position 5(P2) arginine for RRRVRY. In essence, the Y for Amba substitution shifts the sequence to P5-P4-P3-P2-P1-P1'. The RMSD deviation of the RRRVR-Amba backbone from the RRRVRY backbone in the binding position is 2.77Å, which is relatively small and reasonable given the different placement of the Amba vs. arginine in the S1 pocket.

This sequence is very similar to the SARS-CoV-2 delta sequence commencing with arginine at 681, namely, RRRARS. It is known that the furin cleaves between the arginine at 685 and serine at 686. Hence, we hypothesize that the fifth residue(P1) enters the furin S1 pocket. When we utilize AlphaFold Multimer to explore the binding of the WT sequence (beginning at 681) PRRARS, we do find that the fifth arginine(P1) enters the furin S1 pocket, and binds strongly to two aspartic acid residues at positions 258 near the pocket entry, and 306 at the interior end of the pocket (Fig. 5.1D). We note that the last arginine in the RVKR sequence of Ref. [121] also has close proximity to D258 and D306. The RMSD deviation of the RRRVR-Amba backbone from the PRRARS backbone of the FCD for SARS-CoV-2 from AlphaFold Multimer is 4.1Å. This is not surprising given the large sequence difference.

Arginine is particularly well suited for this binding, with its three side chain nitrogens in contrast to the single nitrogen in lysine. A lysine at the fifth position (P1) is only able

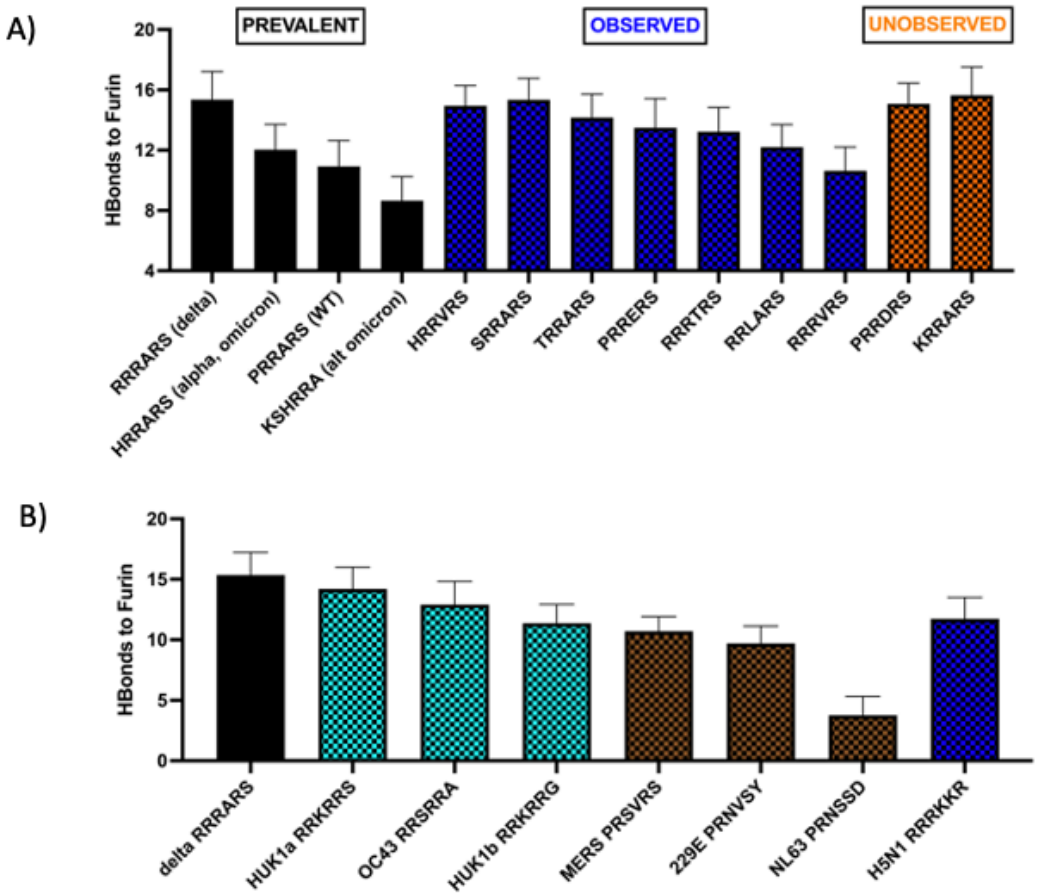


Figure 5.2: A) FCD-furin hydrogen bond counts between furin and SARS-CoV-2 binding sequences at 681-686 of the spike protein. The first four bars are prevalent forms (WT, delta, omicron/alpha, and alt omicron where we assume the sequence starts at position 679. The blue sequences are rare but observed in the GISAID[71] database; of these HRRARN and SRRARS bind as strongly to furin with in statistical accuracy as the delta sequence (RRRARS). The two unobserved sequences require double base mutations from existing extant codons, but bear watching because of their strong binding to furin. B) FCD-furin hydrogen bond counts between furin and other viruses. The SARS-CoV-2 delta variant shows the strongest binding of any human coronavirus and exceeds the H5N1 influenza cleavage site.

to bind to the D306. Of all 62 observed sequences identified from GISAID for SARS-CoV-2, 58 have an arginine at position 5 (P1). For the other human coronaviruses, four (MERS, OC42, HUK1a,b) have an arginine at this position. NL63 and 229E have serines at this position, and the H5N1 flu has lysine.

There is also a strong bias towards a hydrophobic residue at position 4(P2) in the

SARS-CoV-2 sequences. Alanine arises there in 46 of 62 sequences, and valine in 4 of 62. The alanine side chain carbon is within 5Å of side chain carbons on W147 and L120 from the furin in the delta structure. Of the other viruses, MERS and 229E have valine at position 4, while the others have arginine (OC43, HUK1a,b) or serine (NL63) at this position. The H5N1 flu has lysine at this position.

Given a starting structure, we can simulate and measure characteristics of the binding, such as counting FCD-furin HBonds, calculating the binding energy of the complex, or measuring the interfacial surface area, defined as half the difference between the solvent accessible surface area of the separated furin and FCD vs the solvent accessible surface area of the complex. We utilize the `YASARA` molecular modeling program[89], simulating each bound FCD-furin complex for at least 10 ns past energy minimization and equilibration. We then count interfacial protein hydrogen bonds using the criteria outlined for `YASARA`[113]. For computing the binding free energy, we use the Generalized Born Surface Area (GBSA) endpoint free energy calculation from the `HawkDock` server [93]. Because the binding interface is tight, there is essentially no water entry between the peptide and furin. As shown in the supplemental information, we obtain a strong correlation between the GBSA binding energy and the FCD-furin HBond count (Fig. 5.5). For the rest of this paper, we shall use the FCD-furin HBond count as a proxy for binding strength. Note that in this approach, salt bridges of proximate residues, are effectively counted as H-bonds between basic side chain amide groups and acidic side chain carboxyl groups. Hence, the R685 residue of the spike protein FCD forms a salt bridge with the D306 residue of the furin protein, but this is counted in FCD-furin HBonds in this approach.

We have used AlphaFold as the only way to generate a candidate structure for the binding of the WT peptide to furin. With the other sequences we have a choice of using AlphaFold or using the mutation approach within YASARA. We generally find that there are small differences in favor of the mutation approach as we discuss in detail in the supplemental information (Fig. 5.2).

We have surveyed a total of 62 observed six member SARS-CoV-2 furin FCD sequences at 681-686 for this paper drawn from from the GISAID database[71, 126, 127], out of which

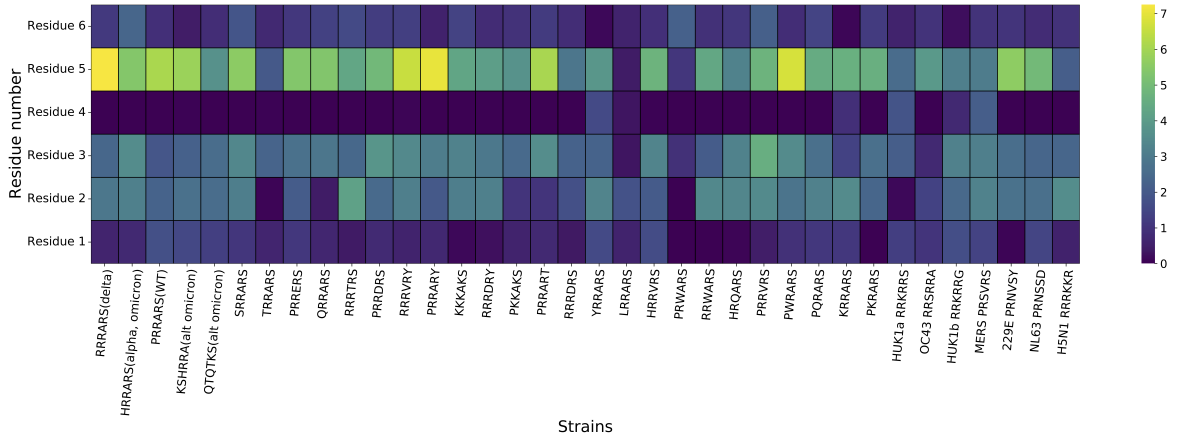


Figure 5.3: Heat map of interfacial hydrogen bonds from furin to the six residue peptide by residue number (vertical) for various observed SARS-CoV-2 along with two unobserved, and for other human coronaviruses and H5N1. Clearly, the key residue for binding is the fifth.

the delta sequence RRRARS is the top binder to within statistical significance. Those used in this paper are acknowledged in Supplemental Table 1. Fig. 2A shows the FCD-furin HBond counts for the prevalent 681-686 sequences WT(PRRARS), delta (RRRARS), and omicron/alpha (HRRARS). We have also included KSHRRA as an alternate omicron sequence in view of the N679K mutation. Additionally, we include seven observed but rare sequences found from GISAID chosen either for their frequency of occurrence or their high FCD-furin HBond count. Finally, we include two sequences (PRRDRS and KRRARS) which can be arrived at by two base mutations from either the WT or delta variants. The prevalent codon at position 684 cannot swap by a single base to obtain D, and the prevalent codon at position 681 cannot swap by a single base to obtain K. By performing pairwise t-tests within GraphPad, we find that the FCD-furin HBond count for the delta variant sequence binding to furin exceeds all but one of the observed sequences with statistical significance ($p < 0.05$) or extreme ($p < 0.0001$) statistical significance, and differences are statistically insignificant in comparison to the observed SRRARS and unobserved PRRDRS and KRRARS sequences ($p > .1$ for each).

A similar picture emerges compared to other human coronaviruses and the H5N1 flu as shown in Fig. 5.2B. The candidate sequences for OC43, NL63, HUK1a,b, 229E, and MERS were obtained by homology alignment of the spike proteins using BLAST[128]. The

FCD-furin HBond count difference between the delta variant and these viral sequences is extremely significant ($p < 0.0001$). We note that the binding is strongest for the cold viruses HUK1a,b and OC43.

To assess the importance of the different residues in the six member peptide to binding strength, we analyzed the hydrogen bonding patterns in detail. We display a heat map in Fig. 5.3 for many of the sequences shown in Fig. 5.2. We find in nearly every case that the strongest binding, representing a significant plurality of the binding strength, is for the position 5(P1) residue, with arginine the preferred amino acid there. Notably, the H5N1 sequence with a K at position 5, and the trial sequence PKKAKS where all arginines are replaced by lysines, fare poorly at position 5(P1) compared to the other sequences.

In searching for an understanding for these observations, we have uncovered three correlations, two of which that can independently explain nearly 50% of the variation between SARS-CoV-2 sequences and separately between viruses. First, by examining the root mean square fluctuation of the backbone C-alpha of the first residue (P5), compared to the FCD-furin HBond counts of the observed sequences with at least 50 appearances in the GISAID tables for SARS-CoV-2, we see in Fig. 5.4A that this backbone fluctuation correlates inversely with the binding strength with a linear regression coefficient of $R^2 = 0.53$. Fig. 5.4B shows the correlation between the FCD-furin HBond count and the RMSF of the first alpha carbon (CA) atom for delta, the six other human coronaviruses with homology in this region, and the H5N1 flu virus. The linear regression coefficient is $r^2 = 0.49$. The best fit slope of -2.74±1.25 FCD-furin HBonds/Å is less than that for SARS-CoV-2 (-4.33± 1.54 FCD-furin Hbonds/Å), but the difference is statistically insignificant. Second, by examining the number of FCD-furin HBonds associated with the residue at position 5(P1), we observe (Figs. 5.4C,D) that there is a high degree of correlation with the total FCD-furin HBond count. For the observed higher frequency furin binding sequences, the best fit slope is 1.41±.38 with $R^2 = 0.59$, and for comparison of delta to other viruses, the slope is similar 1.77±.51 with $R^2 = 0.66$. Third, as shown in Fig. 5.4E, the number of basic residues (H,K, or R) in the six residue sequence helps determine the maximum number of FCD-furin HBonds. Fitting the maximum envelope

of the plot to a quadratic, as in Fig. 5.4F, gives

$$Hbonds = -0.66(N_B - 4.06)^2 + 15.7 \tag{5.1}$$

where N_B is the number of bases. The nonlinear regression coefficient is R^2=0.98. This suggests that the maximum number of FCD-furin HBonds is 16, for four basic residues (as per the delta variant), and to within statistical accuracy, no sequence exceeds delta in the number of FCD-furin HBonds to furin.

5.3 Discussion

The most important results of this paper are: 1) by using AlphaFold Multimer[98] we have validated by comparison to the binding of furin with a known six residue inhibitor, we are able to predict bound structures for over 60 observed FCD sequences of SARS-CoV-2 (at residues 681-686 of the spike protein and two alternate sequences for omicron) and eight other viruses (six human coronaviruses (OC43, HUK1a, HUK1b, MERS, NL63, 299E), the H5N1 influenza, and Epstein-Barr virus). From among these, the delta variant FCD of SARS-CoV-2 has the strongest binding to furin within statistical accuracy, with 15.3 mean FCD-furin hydrogen bonds. 2) Within these sequences we find selection for arginine at position 5 (P1), which fits into a furin S1 pocket having aspartic acids at the entrance and within. The structure of arginine allows binding to both aspartic acids, while lysine's structure does not. 3) There is also bias towards a hydrophobic residue at position 4(P2) of the six residue FCD, which appears to interface favorably with W147 and L120 of the furin. 4) We find that two features of the sequences each predict about half of the binding strength: (i) the backbone fluctuation of the first residue in the binding sequence correlates inversely with the overall binding strength as measured by FCD-furin HBonds, and (ii) the number of hydrogen bonds associated with the binding of residue 5(P1) in the furin S1 pocket correlates positively with FCD-furin HBond count. This residue never accounts for more than a plurality of the FCD-furin HBonds so it is somewhat surprising that it correlates with the observed trend of binding. (iii) By considering the variation of the FCD sequence with the number of basic residues, we conclude that no more than 16 FCD-furin HBonds are possible, and within statistical accuracy delta achieves the

maximum value. We conjecture that the physical basis for this is a tradeoff between binding efficacy of the basic residues (especially arginine) and Coulomb repulsion as more are added.

In conclusion, we find that spike FCD-furin binding depends critically upon insertion of arginine in the fifth position (P1) of the FCD in a furin pocket that includes D258 at the opening and D306 at the interior end. This prediction emerges uniquely from the application of AlphaFold Multimer[98] to predict the bound structure, and contrasts with earlier work that employed a docking program for interface prediction[124]. It is therefore critical to have experimental structural biology test of this prediction.

Note that the omicron FCD sequence is the same as alpha, and alternate FCD sequences (KSHRRA, beginning at K679, or QTQTKS, with K679 at position 5(P1)) have fewer FCD-furin HBonds than any observed variants, consistent with the observed milder impact of omicron on the lungs[130, 131, 84].

We conclude that it is quantitatively unlikely that any SARS-CoV-2 variant, or any other virus can bind significantly more strongly to the furin protease than the delta variant. This is based on a survey of a large number of observed SARS-CoV-2 spike sequences new SARS-CoV-2 spike sequences not yet observed, other human coronaviruses, H5N1 influenza, and Epstein-Barr virus. The basic model for viral infection is that after spike RBD binding to ACE2, furin cleavage at the FCD regulates fusogenicity leading to syncytia and viral reproduction. Our theoretical studies indicate that furin-FCD driven fusogenicity is at its worst with the delta variant among all observed SARS-CoV-2 variants of interest or concern. Of concern and cause for caution are some rarely observed or unobserved FCD sequences which could be just as consequential for furin cleavage as delta (observed: SRRARS, RRRARN,HRRVRS; unobserved: PRRDRS, KRRARS).

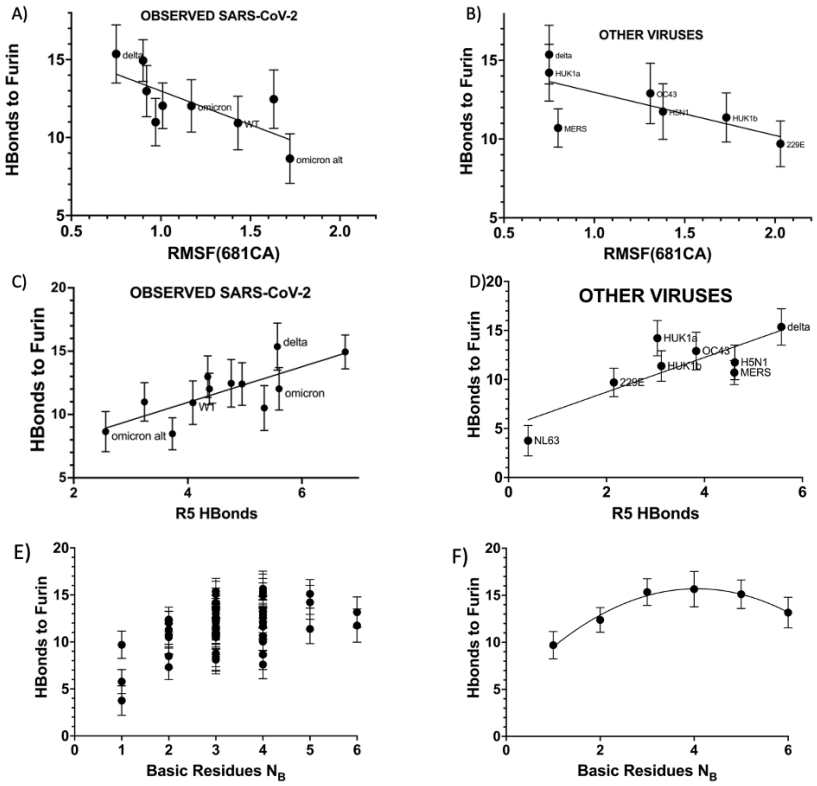


Figure 5.4: A) Correlation of the backbone fluctuation from Residue 1 of the sequence with the total number of FCD-furin HBonds between the binding sequence and furin for SARS-CoV-2 sequences observed at least 50 times. B) Correlation of the residue 1 backbone fluctuation for delta and other viruses. C) Correlation of the interfacial HBonds for Residue 5 with the total number of HBonds for observed SARS-CoV-2 sequences of A). D) Correlation of the FCD-furin HBonds for Residue 5 with total number of FCD-furin HBonds for delta and other viruses. E) The number of HBonds for a given number of basic FCD residues plotted for 56 sequences. F) The maximum FCD-furin HBonds envelope as a function of the number of basic residues. This is fit with R^2=0.98 by Eq. (1) of the main text. The sequences for the peak values are for 1-6 respectively: PRNSVY (229E coronavirus), PRQARS (SARS-CoV-2), SRRARS (SARS-CoV-2), KRRARS (SARS-CoV-2, unobserved), RRRRRD (Epstein-Barr, ref. [129]), RRRRRR (unobserved).

5.4 Supplementary Information

5.4.1 Materials and Methods

5.4.1.1 Molecular Models

For the furin structure and for furin binding to the inhibitor RRRVR-Amba, we used PDB entry 6EQX[123].

5.4.1.2 Sequence Alignment for Other Coronaviruses

To identify homology in the furin cleavage domain for the other coronaviruses OC43, NL63, 229E, MERS, HUK1a, HUK1b, we utilized BLAST[128] at the National Center for Biotechnology Information. We compared entire spike sequences and zoomed in on the furin cleavage domain based upon the PRRARS sequence for SARS-CoV-2.

5.4.1.3 Identification of other SARS-CoV-2 sequences using GISAID: Genomic Data Set and Sequence Pre-Processing

We obtained SARS-CoV-2 sequences for this study from the GISAID database on Nov 11, 2020 [132]. Our data set contains FASTA files for every complete human SARS-CoV-2 nucleotide sequence (from all geographical locations) available in GISAID between and 12/1/19 and 7/11/2021. The sequences were then aligned using `ClustalOmega` with the default parameters [133]. We found that `ClustalOmega` ran faster on our data set than common alternatives like `ClustalW` [134] and `MUSCLE` [135].

After aligning the sequences, we extracted the spike protein by comparing the aligned sequence with the NCBI's SARS-CoV-2 reference sequence (NC_045512.2; "WT") [50] and tabulated the frequencies of different furin binding domain inserts.

Table ST1 includes accession numbers and acknowledgements for the first of each 111 unique nucleotide sequences referenced in this paper as they appear in GISAID.

5.4.1.4 Molecular Dynamics

To simulate the protein-protein interactions, we used the molecular-modelling package `YASARA` [89] to substitute individual residues and to search for minimum-energy conformations on the resulting modified structures of the FCD-furin. For all of the structures, we carried out an energy-minimization (EM) routine, which includes steepest descent and simulated annealing (until free energy stabilizes to within 50 J/mol) minimization to re-

move clashes. All molecular-dynamics simulations were run using the AMBER14 force field with [108] for solute, GAFF2 [109] and AM1BCC [110] for ligands, and TIP3P for water. The cutoff was 8 Å for Van der Waals forces (AMBER's default value [111]) and no cutoff was applied for electrostatic forces (using the Particle Mesh Ewald algorithm [112]). The equations of motion were integrated with a multiple timestep of 1.25 fs for bonded interactions and 2.5 fs for non-bonded interactions at $T = 298$ K and $P = 1$ atm (NPT ensemble) via algorithms described in [95]. Prior to counting the FCD-furin hydrogen bonds and calculating the free energy, we carry out several pre-processing steps on the structure including an optimization of the hydrogen-bonding network [113] to increase the solute stability and a pK_a prediction to fine-tune the protonation states of protein residues at the chosen pH of 7.4 [95]. Insertions and mutations were carried out using `YASARA`'s BuildLoop and SwapRes commands [95] respectively. Simulation data was collected every 100ps after 1-2ns of equilibration time, as determined by the solute root mean square deviations (RMSDs) from the starting structure. For all bound structures, we ran for at least 10 ns post equilibrium, and verified stability of time series for FCD-furin hydrogen bond counts and root mean square deviation (RMSD) from these starting structure. Because of concerns about the validity of short time simulations, and more variability for the weaker binding for the omicron RBD-ACE2 complex, we ran for 40 ns postequilibration in that case.

The FCD-furin hydrogen bond (HBond) counts were tabulated using a distance and angle approximation between donor and acceptor atoms as described in [113]. Note that in this approach, salt bridges of proximate residues, are effectively counted as H-bonds between basic side chain amide groups and acidic side chain carboxyl groups. Hence, the R685 residue of the spike protein FCD forms a salt bridge with the D306 residue of the furin protein, but this is counted in HBonds in this approach.

Note that in view of the likely ambient pH for cell surface or endosomal furin cleavage, and the polybasic environment of the FCD, we have assumed all histidines to be singly protonated at the delta site. Choosing the epsilon site makes little difference. For the alpha sequence, doubly protonated histidine binds more strongly, but for the alternate

omicron sequence, there is little difference among the three protonation states.

5.4.1.5 Endpoint Free Energy Analysis

We calculated binding free energy for the energy-minimized structure using the molecular mechanics/generalized Born surface area (MM/GBSA) method [91, 94, 114], which is implemented by the `HawkDock` server [93]. While the MM/GBSA approximations overestimate the magnitude of binding free energy relative to *in-vitro* methods, the obtained values correlate well with H-bond counts. For each RBD-ACE2, RBD-AB, and NTD-Ab binding pair we average over five snapshots of equilibrium conformations. For each FCD-furin pair, we average over ten snapshots of equilibrium conformations.

5.4.1.6 Use of ColabFold/AlphaFold for Furin binding domain

Full details of this method are provided in [97, 98, 105]. In brief, we used the heterocomplex prediction method known as AlphaFold-Multimer[97, 98] as implemented within ColabFold[105] to predict the best bound structure to the furin enzyme of the six residue FCD from the WT protein. We inferred the ordering of this sequence by comparison with a very similar six residue peptide inhibitor of furin with the sequence RRRVR-aminomethyl-benzamidine (RRRVR-Amba) [123]. In this case the backbone of the WT FCD aligns well with that of the inhibitor, but the arginine at residue 5 enters the furin S1 pocket[123] while the Amba enters the furin S1 pocket for the inhibitor. The serine is in proper cleavage position for furin. Most other structures were then obtained by mutation from the predicted WT FCD-furin structure.

5.4.1.7 Statistical Analysis and Graphics

We computed the statistical significance of pairwise differences using the GraphPad unpaired t-test calculator. Regression analysis for Fig. 5.4, Fig. 5.5 was carried out using the GraphPad Prism (v. 9) package. Structural images for Fig. 5.1 were created in YASARA. Figs. 2,4, S1, S2 were created with GraphPad Prism (v. 9). Fig. 5.2 was created with Seaborn (v. 0.11.2), a Python data visualization library.

5.4.2 Supplementary Text

5.4.2.1 Correlation of FCD-furin HBonds and MM/GBSA binding energy estimates

Fig. 5.5 summarizes the correlation between FCD-furin HBonds and the MM/GBSA estimates for binding energy in kcal/mole. Note that MM/GBSA usually overestimates binding energy strength significantly but is good for producing binding energy trends. The regression coefficient is $R^2 = 0.61$, and the best fit slope is -.107 Hbonds/(Kcal/mole) with 95% confidence intervals of -.110 to -.1042. Clearly, the correlation is strong between FCD-furin HBonds and binding energy.

5.4.2.2 Differences between simulations with AlphaFold and mutation

While we must take the WT FCD-furin structure from AlphaFold[97, 98], we can mutate using the Swap command in YASARA from there to obtain other starting structures for molecular dynamics simulations. In general, AlphaFold produces structures with slightly less binding strength than mutating from the WT, with a few exceptions, the delta variant being one. This is demonstrated for five sequences in Fig. 5.6. Accordingly, because the resultant binding is stronger we have used the mutant results where possible to provide a more accurate starting point for the equilibration runs in molecular dynamics.

5.4.2.3 Examples of sequence frequency and codons

Fig. 5.7 shows a table of FCD sequences used in the figures as well as one synonymous/silent mutation based upon the consensus codons for WT, alpha, and delta. The last two entries are for unobserved but potentially potent FCD sequences. Mutation to those would require two base swaps from either WT or delta.

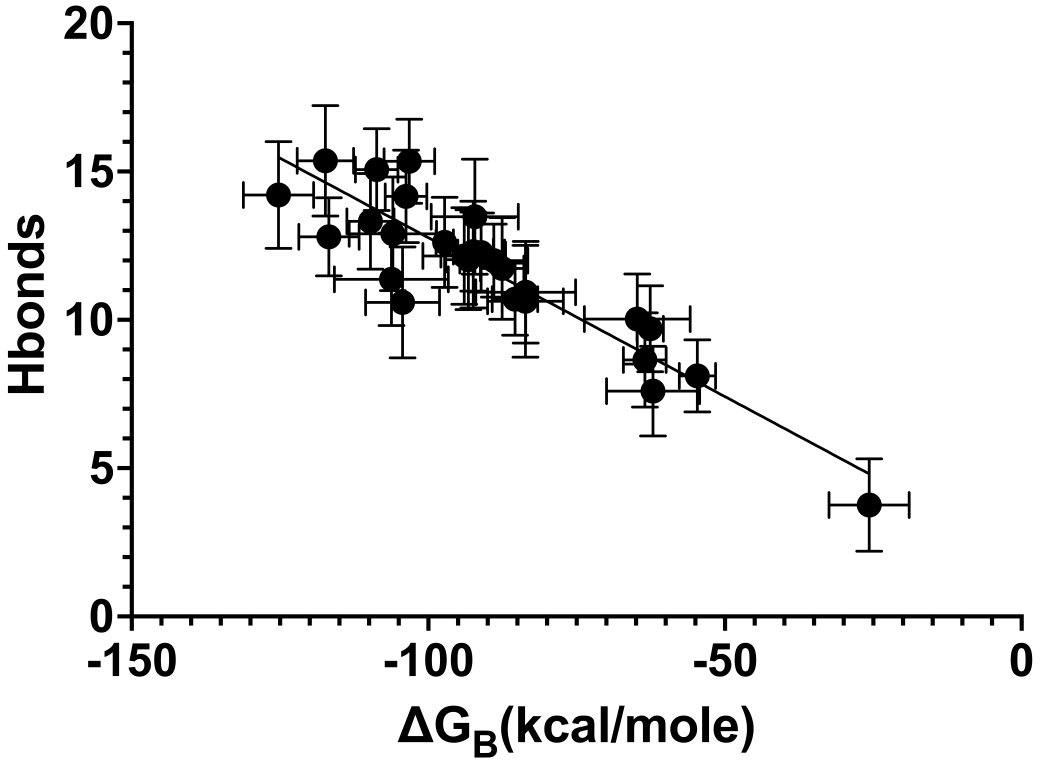


Figure 5.5: **Correlation of FCD-furin HBond counts with MM/GBSA Binding Energy** FCD-furin HBond counts are estimates from YASARA[113, 89] simulations, while MM/GBSA Binding Energy comes from the HawkDock server[93]. Regression analysis using GraphPad Prism 9 provides the straight line fit (see text for details).

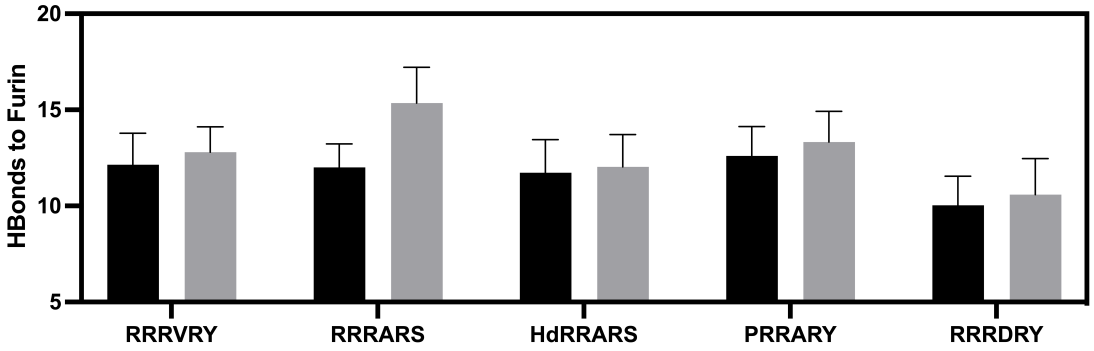


Figure 5.6: **Differences in equilibrated FCD-furin HBond counts between AlphaFold generated starting FCD-furin structures and starting structures mutated from the AlphaFold WT structure** In general, after equilibration the AlphaFold structures have slightly less binding strength, with a few exceptions such as the delta variant where AlphaFold misses dramatically. For comparison, the p-values for AlphaFold vs mutant in this plot are RRRVRY- p=0.0035 (very significant); RRRARS (delta)- p<0.000001 (extremely signficant); HRRARS (alpha/omicron)- p=0.181 (not significant); PRRARY - p=0.00094 (very significant); RRRDRY - p=0.0164 (very significant)

FREQUENCY(%)	P1	P2	P3	P4	P5	P6
47.37203884	H(CAT)	R(CGG)	R(CGG)	A(GCA)	R(CGT)	S(AGT)
5.94182775	R(CGT)	R(CGG)	R(CGG)	A(GCA)	R(CGT)	S(AGT)
45.63914496	P(CCT)	R(CGG)	R(CGG)	A(GCA)	R(CGT)	S(AGT)
0.07247210	H(CAT)	R(CGG)	R(CGT)	A(GCA)	R(CGT)	S(AGT)
0.04053524	H(CAT)	R(CGG)	R(CGG)	V(GTA)	R(CGT)	S(AGT)
0.02800617	P(CCT)	R(CGG)	R(CGG)	V(GTA)	R(CGT)	S(AGT)
0.02712176	L(CTT)	R(CGG)	R(CGG)	A(GCA)	R(CGT)	S(AGT)
0.00422549	Y(TAT)	R(CGG)	R(CGG)	A(GCA)	R(CGT)	S(AGT)
0.00397982	S(TCT)	R(CGG)	R(CGG)	A(GCA)	R(CGT)	S(AGT)
0.00014740	P(CCT)	R(CGG)	R(CGG)	E(GAA)	R(CGT)	S(AGT)
0.00000000	P(CCT)	R(CGG)	R(CGG)	D(GAC)	R(CGT)	S(AGT)
0.00000000	K(AAA)	R(CGG)	R(CGG)	A(GCA)	R(CGT)	S(AGT)

Figure 5.7: **Examples of FCD sequences from GISAID for analysis here with codons** The observed frequencies of sequences between 12/1/19 and 7/11/21 appear at left, and the predominant codons for each position are tabulated. Row 4 shows a synonymous/silent mutation to the alpha variant, while the rest show missense mutations. The last two sequences are unobserved (requiring double codon swaps relative to either WT or delta) but bind as strongly to furin as the delta FCD. Note that over this entire pre-omicron time frame that delta (RRRARS) has less accumulated percentage of the sequences than WT (PRRARS) or alpha (HRRARS).

Chapter 6

Computational design of a novel decoy therapeutic agent for SARS-CoV-2 using truncated human angiotensin converting enzyme 2.

6.1 Introduction

The surge of recent SARS-CoV-2 infections, due in part to the emergence of recent variants such as omicron [99] and delta [104], render the need for safe and potent therapeutic agents. Furthermore, due to the diversity of observed sequences and the potential of new variants [85], the need for the development of a variant-resistant therapeutic agent is high.

All variants of SARS-CoV-2 feature a receptor binding domain (RBD) that binds with high affinity to the angiotensin converting enzyme 2 (ACE2) protein on human cells [100]. This binding is of critical importance for virion cellular entry, and its inhibition is a potential defense against infection. A possible route to a variant-resistant therapeutic agent is a decoy protein that effectively models the interaction region of the ACE2. The decoy competitively binds to the RBD, which decreases the number of available RBDs that can bind to human cells. Since ACE2 is always the target of the spike RBD regardless of the SARS-CoV-2 variant, the decoy would, therefore, be just as effective against newer and potentially different variants. Decoy proteins also need to be safe to administer due

to the possibility of immunogenicity.

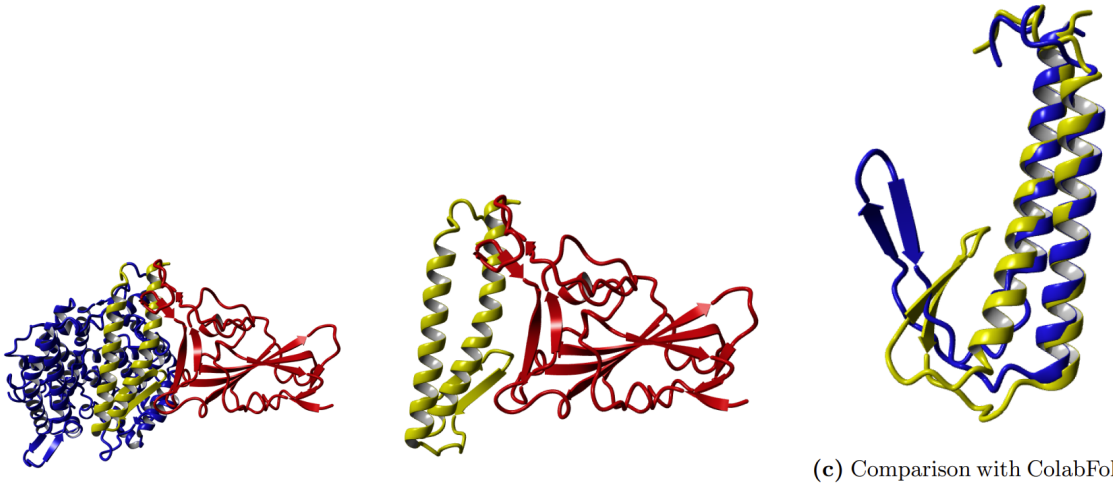
Previous computational designs of potential protein decoy therapeutic agents feature the helix-turn-helix region (HTH) of the ACE-2 [136], which is studied extensively due to the large number of residues that participate in binding from the ACE2 [100]. Other computational approaches include engineered proteins that replicate the RBD-ACE2 binding interface [137], in addition to approaches which use other fragments of the ACE2 as decoys with minor modifications [138].

Rapid computational studies provide an avenue to study potential decoy designs and their interactions with the target protein. Here, we draw our motivation for the decoy design using computational binding data from wildtype (WT) ACE2-RBD simulations. Since most of the residues that take part in interfacial bonds between the ACE2 and the RBD lie on the HTH region of the ACE2 and the beta loop that carries K353, we engineered a HTH decoy with a cut at the hairpin that makes space for the K353 loop as shown in Fig.6.1. We then independently test possible folded structures of the decoy using AlphaFold, a deep learning approach used to predict folded protein structures implemented within the ColabFold suite, and we found that the desired decoy shapes were reasonably attainable. We then present binding data from molecular dynamics simulations of the decoy while bound to RBD configurations from 6vw1 [139]. We show that the decoy binds to the RBDs with strength comparable to the ACE2-RBD interaction, as measured by interfacial hydrogen bonds, in addition to binding free energy estimates using Generalized Born Surface Area approximations. By mutating target residues from pdb files, we are also able to present results from potential omicron and delta RBD-decoy interactions.

Materials and Methods

6.1.1 Molecular Models

We drew our starting structures for RBD-ACE2 binding from PDB file 6vw1 [139]. We computationally built the decoy from the PDB file using the molecular modelling package `YASARA` [95], where we used the SwapRes, DelRes, and `BuildLoop` functions to build our



(a) ACE2-RBD **(b)** Decoy-RBD **(c)** Comparison with ColabFold Prediction

Figure 6.1: (a) Shows the ACE2 (blue)- RBD (red) complex from 6vw1[139]. The residues that were selected for computational decoy design are highlighted in yellow. This design includes all but one of the key residues in RBD-ACE2 bind as shown in Fig.6.1(a). (b) Shows the decoy (yellow) in complex with the RBD (red), this was done by superposing the decoy on ACE2 using the `MUSTANG` algorithm [140] and was used as the starting structure for all decoy-RBD simulations pre-equilibration. (c) Shows the alignment between the computational decoy design (yellow) and the top ColabFold prediction (blue) [105] for the same sequence.

decoy and to find the best docking configuration that closely matches the ACE2-RBD configuration. We also used the `MUSTANG` algorithm for structural alignments [].

The decoy is built using truncated sections from ACE2. Namely, the first helix from residues 19-52, which contains the residues that participate in the majority of the bonds between the ACE2 and the RBD [100]. At the hairpin between the two helices at residue 52, we attach the beta loop from residues 340-366, which contains the majority of the remaining ACE2-RBD bonds. To this, we then attach the second helix from residue 54-91 for decoy stability.

6.1.2 Molecular Dynamics

To simulate the protein-protein interactions, we used the molecular-modelling package `YASARA` [89] to substitute individual residues and to search for minimum-energy conformations on the resulting modified structures of the complexes listed in 6.2. For all of the

structures, we carried out an energy-minimization (EM) routine, which includes steepest descent and simulated annealing (until free energy stabilizes to within 50 J/mol) minimization to remove clashes. All molecular-dynamics simulations were run using the AMBER14 force field with [108] for solute, GAFF2 [109] and AM1BCC [110] for ligands, and TIP3P for water. The cutoff was 8 Å for Van der Waals forces (AMBER's default value [111]) and no cutoff was applied for electrostatic forces (using the Particle Mesh Ewald algorithm [112]). The equations of motion were integrated with a multiple timestep of 1.25 fs for bonded interactions and 2.5fs for non-bonded interactions at $T = 298$ K and $P = 1$ atm (NPT ensemble) via algorithms described in [95]. Prior to counting the hydrogen bonds and calculating the free energy, we carry out several pre-processing steps on the structure including an optimization of the hydrogen-bonding network [113] to increase the solute stability and a pK_{a} prediction to fine-tune the protonation states of protein residues at the chosen pH of 7.4 [95]. Insertions and mutations were carried out using `YASARA`'s BuildLoop and SwapRes commands [95] respectively. Simulation data was collected every 100ps after an appropriate amount of equilibration time, as determined by the solute root mean square deviations (RMSDs) from the starting structure. For all bound structures, we ran for at least 25ns post equilibrium, and verified stability of time series for hydrogen bond counts and root mean square deviation (RMSD) from the starting structure.

The hydrogen bond (HBond) counts were tabulated using a distance and angle approximation between donor and acceptor atoms as described in [113]. Note that in this approach, salt bridges of proximate residues are effectively counted as H-bonds between basic side chain amide groups and acidic side chain carboxyl groups.

6.1.3 Endpoint Free Energy Analysis

We calculated binding free energy for the energy-minimized structure using the molecular mechanics/generalized Born surface area (MM/GBSA) method [91, 94, 114], which is implemented by the `HawkDock` server [93]. While the MM/GBSA approximations overestimate the magnitude of binding free energy relative to *in-vitro* methods, the obtained values correlate well with H-bond counts as shown in a previous publication [85]. For each RBD-ACE2 and RBD-decoy binding pair, we average over twenty snapshots of equilib-

rium conformations.

6.1.4 Use of ColabFold/AlphaFold for decoy structure prediction

While we designed the decoy computationally from experimentally determined structures of ACE2, we also predicted the folded states of the decoy using AlphaFold as implemented within ColabFold to provide additional evidence for the stability of the decoy within the desired shape. Full details of this method are provided in [97, 98, 105].

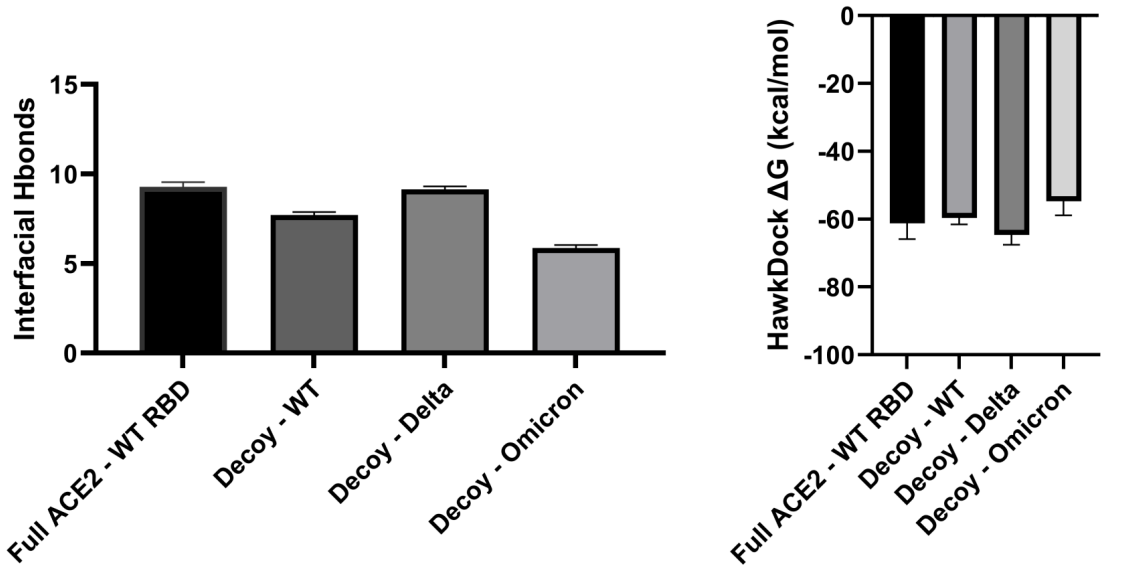
6.2 Results

6.2.1 Binding Strengths: HBond and Binding Free Energy

We find that the number of interfacial hydrogen bonds in the WT decoy-RBD binding pair is comparable to the ACE2-RBD binding pair. These results are summarized in Fig.2. We also include binding energy estimates from GBSA analysis of molecular dynamics equilibrium conformations in Fig.6.2(a). The comparable strength of the decoy-RBD binding pairs as compared to the ACE2-WT RBD pair suggests that this decoy is a strong candidate for competitive binding. With the exception of E329 on the ACE2, the decoy includes all of the residues that are involved in the ACE2 - WT RBD binding interface. The trend of binding energies tracks well with the easier to estimate interfacial HBond count. The specific binding pairs as calculated by `YASARA`'s Hbond algorithm [113] are given in the supplementary information. We also summarize the key residue positions involved in Fig. 6.3.

We find the binding strength of the decoy to the omicron RBD to be slightly weaker compared to delta and WT RBDs. This is in agreement with our previous computational work [84] where we show that the omicron spike exhibits weaker receptor binding but with significant reduction of antibody binding strength corresponding to escape.

In 6.3 we see that the number and strength of the interactions between the WT RBD and ACE2 (6.3(a)) as compared to WT RBD and decoy (6.3(b)) are roughly similar with the exception of E329 interactions due to its absence in the decoy design. We observe that most of the interactions involve the helix-turn-helix region (residues 19-85 in ACE2) and



(a) Interfacial Hbond Count

(b) ΔG values calculated using MM/GBSA

Figure 6.2: (a) Average total interfacial hbonds post equilibration for the complexes shown with 95% confidence intervals. The decoy is within 2 hbonds of the WT-ACE2 interactions in all simulations except for the Decoy-Omicron simulation. (b) Average $\Delta G(kcal/mol)$ values calculated using an MM/GBSA method implemented within HawkDock [91, 114].

that the K353 beta hairpin plays an important role as shown by the variety and number of interactions by K353, D355, and N330. For the delta variant (6.3(c)) we see a slight improvement in the D38-Y449 interaction, which is possibly introduced by the L452R mutation. Furthermore, we also see S445 and T446 from the RBD interacting with Q42 from the decoy, which was unobserved in the ACE2/Decoy-RBD WT simulations. For the omicron variant (6.3(d)), we see that the K353-R498 interaction completely disappears. This is explained due to the Q498R mutation which leads to electrostatic repulsion. The K493 mutation is also important as it interacts with the D38 from the decoy. These variations are expected and in line with our current understanding of ACE2-RBD interactions.

6.2.2 Decoy structure predictions using AlphaFold

In 6.3, we show the AlphaFold's top ranking predicted folded structure superposed to the desired equilibrium structures obtained using `YASARA` [97, 140]. The main features of the

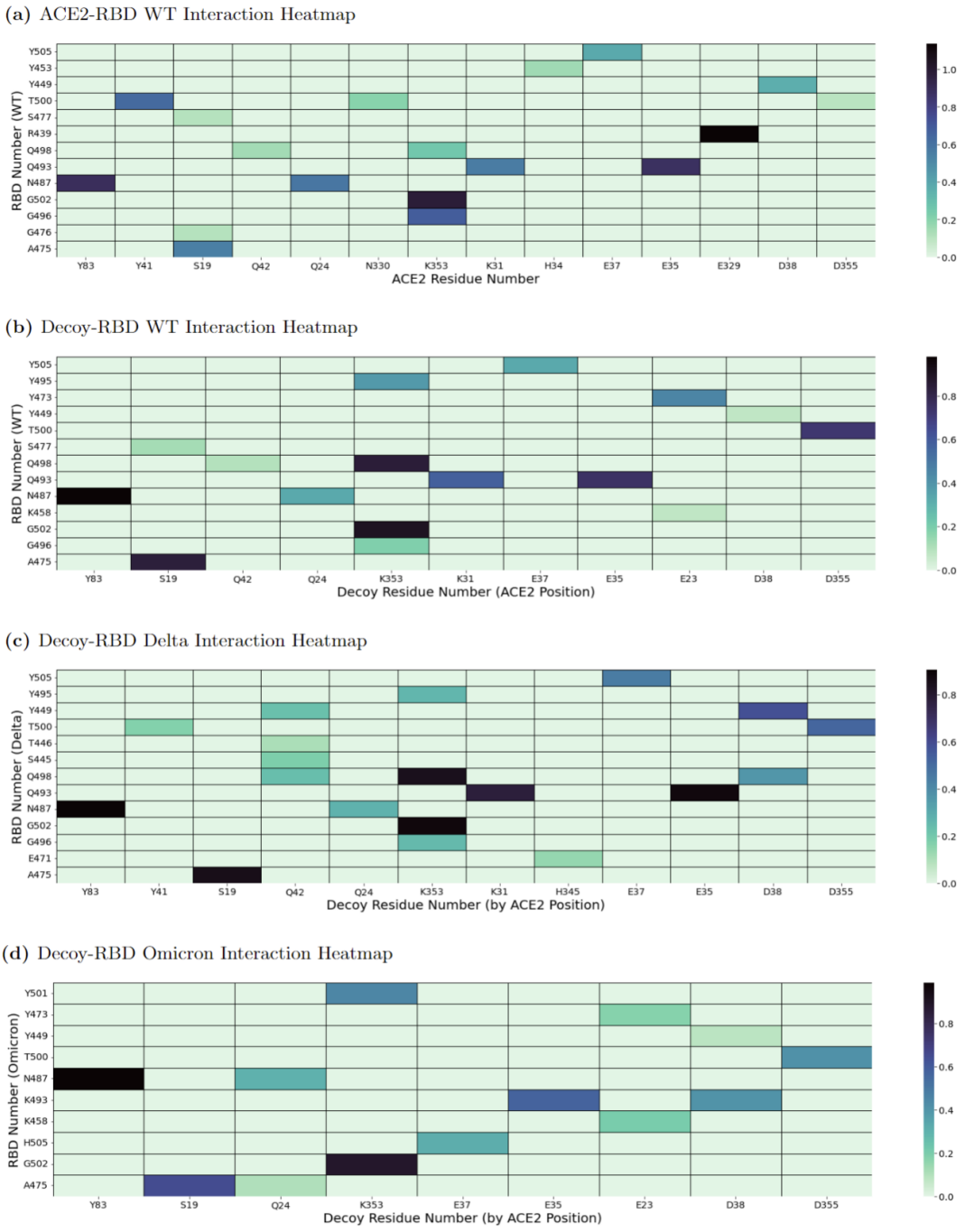


Figure 6.3: Heat map of interfacial hbonds averaged over the full simulations from ACE2 to RBD (vertical) (a), and decoy (b-d) to RBD as labelled by default residue positions in 6vw1 [139]. The key residues identified in (a) are mostly involved in decoy-RBD binding as shown in (b-d).

decoy are intact with the main fluctuations originating from the linkers (residues 340-350, 346-366) connecting the beta hairpin to the helices.

6.3 Discussion

We present a computational design of a novel decoy therapeutic agent that has the potential to limit the severity of infection via competitive binding to the receptor binding domain. Compared to helix based decoys mentioned in [136, 137], this decoy includes all the major residues involved in RBD-ACE2 binding, in particular the lysine in the beta hairpin at position 353 of the ACE2.

A key advantage of this computational design is the usage of truncated ACE2 sequences. This confers an advantage in the lowered chance of a human immune response due to the truncated ACE2 sequences having human origins [141]. The work of Basit et al. prefers a similar approach in the usage of the HTH sequence [136], albeit missing the beta hairpin that includes the key lysine residue at position 353. The usage of primarily human sequences are hoped to be effective against future variants due to their ability to mimic the binding pocket of the ACE2. Recent work demonstrating the efficacy of this approach includes the use of s-ACE2 molecules with engineered mutations [142], where experimental evidence of improved survival rates and reduced symptoms was obtained against four variants of concern in mice.

Chapter 7

Conclusion

In part 1 of our work, we make progress on a number of questions that are important in plant cytokinesis and the biophysical understanding of membrane networks. We show, using a modified Helfrich free energy, that a spreading force coupled with a decrease in the membrane's spontaneous curvature is essential for cell plate maturation, and we reinforce our findings with experimental data from light microscopy. Using our experimental data, We suggest possible sources of this spreading force and concurrent decrease in spontaneous curvature, whilst noting that it is highly probabable that callose synthesis at the cell plate is a key determinant to both effects. We also provide theoretical synthesis rates for polysaccharides using flory's self avoiding polymer theory in two dimensions. We also introduced membrane shape approximations using a parameterized geometric basis set (vesicles and complete cell plates with oblate spheroids, fused vesicles and tubular structures with elliptical hyberboloids and their combination in a network). As noted in Chapter 2, this has the potential of wider applications pertaining to the quantitative assessment of membrane dynamics. It can also be used as a basis for addressing the equilibrium and quantitative analysis of vesiculation, tubulation, ER-intermediate compartments, Golgi, and endosomes in all eukaryotic cells. We also provide the python code that can calculate the relevant Helfrich free energies in a fast, computationally efficient manner.

Possible future directions include a microscopic model with a mechanistic understanding of the spreading force and concurrent decrease in spontaneous curvature. One possible

way to understand this is by designing an appropriate Monte-Carlo simulation of a triangulated membrane surface coupled to a tethered, growing self avoiding polymer (motivated by works such as Chen et al [19]). We also indicate the need of time-lapse experiments directed at the role of motor proteins which can possibly shed light on their contribution to the stabilizing and spreading force that the model predicts.

In part 2 of our work, we study various aspects of SARS-CoV-2 *in silico*, primarily using a combination of molecular dynamics simulations, generalized Born surface area approximations, and predictive neural networks such as AlphaFold. In Chapter 4, we comment on three separate fitness parameters of the omicron variant. In Chapter 5, we present an in-depth analysis of the Furin Cleavage Domain, an area that is difficult to study due to lack of experimental data on the domain's crystal structure. In Chapter 6, we present a novel computational design of a humanized competitive inhibitor based on the virus's target receptor, ACE2.

At the time of writing, SARS-CoV-2 averages close to half a million cases a day. It remains the research focus of a significant percentage of scientists worldwide and will continue to do so until until it poses minimal threat to the human and wildlife population. As the first pandemic of the 21st century, one hopes that the lessons and research methodologies learned in response to SARS-CoV-2 are used to minimize the risk of potential future pandemics.

References

[1] Markus Deserno. Fluid lipid membranes: From differential geometry to curvature stresses. *Chemistry and Physics of Lipids*, 185:11–45, January 2015.

[2] Thomas D. Pollard. Nine unanswered questions about cytokinesis. *Journal of Cell Biology*, 216(10):3007–3016, August 2017.

[3] Thomas D. Pollard and Ben O'Shaughnessy. Molecular mechanism of cytokinesis. *Annual Review of Biochemistry*, 88(1):661–689, June 2019.

[4] Roie Shlomovitz and Nir S. Gov. Physical model of contractile ring initiation in dividing cells. *Biophysical Journal*, 94(4):1155–1168, February 2008.

[5] Ian Alexander. *Kilbaha biology*. Kew, Vic : Kilbaha Multimedia Publishing, 2018. For secondary school age.

[6] Georgia Drakakaki. Polysaccharide deposition during cytokinesis: Challenges and future perspectives. *Plant Science*, 236:177–184, July 2015.

[7] A L Samuels, T H Giddings, and L A Staehelin. Cytokinesis in tobacco BY-2 and root tip cells: a new model of cell plate formation in higher plants. *Journal of Cell Biology*, 130(6):1345–1357, September 1995.

[8] Jose M. Seguí-Simarro, Jotham Austin, Erin A. White, and L. Andrew Staehelin. Electron tomographic analysis of somatic cell plate formation in meristematic cells of arabidopsis preserved by high-pressure freezing. *The Plant Cell Online*, 16:836 – 856, 2004.

[9] Andrei Smertenko, Farhah Assaad, František Baluška, Magdalena Bezanilla, Henrik Buschmann, Georgia Drakakaki, Marie-Theres Hauser, Marcel Janson, Yoshinobu Mineyuki, Ian Moore, Sabine Müller, Takashi Murata, Marisa S. Otegui, Emmanuel Panteris, Carolyn Rasmussen, Anne-Catherine Schmit, Jozef Šamaj, Lacey Samuels, L. Andrew Staehelin, Daniel Van Damme, Geoffrey Wasteneys, and Viktor Žárský. Plant cytokinesis: Terminology for structures and processes. *Trends in Cell Biology*, 27(12):885–894, December 2017.

[10] Yuh-Ru Julie Lee and Bo Liu. The rise and fall of the phragmoplast microtubule array. *Current Opinion in Plant Biology*, 16(6):757–763, December 2013.

[11] Fabien Miart, Thierry Desprez, Eric Biot, Halima Morin, Katia Belcram, Herman Höfte, Martine Gonneau, and Samantha Vernhettes. Spatio-temporal analysis of cellulose synthesis during cell plate formation in arabidopsis. *The Plant Journal*, 77(1):71–84, November 2013.

[12] Kiminori Toyooka, Yumi Goto, Satoru Asatsuma, Masato Koizumi, Toshiaki Mitsui, and Ken Matsuoka. A mobile secretory vesicle cluster involved in mass transport

from the golgi to the plant cell exterior . *The Plant Cell*, 21(4):1212–1229, April 2009.

[13] Patricia J. Moore and L. Andrew Staehelin. Immunogold localization of the cell-wall-matrix polysaccharides rhamnogalacturonan i and xyloglucan during cell expansion and cytokinesis in trifolium pratense l.; implication for secretory pathways. *Planta*, 174(4):433–445, 1988.

[14] Eunsook Park, Sara M. Díaz-Moreno, Destiny J. Davis, Thomas E. Wilkop, Vincent Bulone, and Georgia Drakakaki. Endosidin 7 specifically arrests late cytokinesis and inhibits callose biosynthesis, revealing distinct trafficking events during cell plate maturation. *Plant Physiology*, 165(3):1019–1034, May 2014.

[15] W Helfrich. Elastic properties of lipid bilayers: Theory and possible experiments. *Zeitschrift für Naturforschung C*, 28(11-12):693–703, December 1973.

[16] P.B. Canham. The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell. *Journal of Theoretical Biology*, 26(1):61–81, January 1970.

[17] S. Gauthier, W. Puech, R. Bénière, and G. Subsol. Analysis of digitized 3d mesh curvature histograms for reverse engineering. *Computers in Industry*, 92-93:67–83, November 2017.

[18] Jean-Baptiste Fournier. On the stress and torque tensors in fluid membranes. *Soft Matter*, 3:883–888, 2007.

[19] Jeff Z. Y. Chen. Model of a polymer chain adsorbed to a soft membrane surface. *Physical Review E*, 82(6), December 2010.

[20] Kenneth A Brakke. The surface evolver and the stability of liquid surfaces. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 354(1715):2143–2157, September 1996.

[21] R.C. Sarasij, Satyajit Mayor, and Madan Rao. Chirality-induced budding: A raft-mediated mechanism for endocytosis and morphology of caveolae? *Biophysical Journal*, 92(9):3140–3158, May 2007.

[22] Rustum Choksi, Marco Morandotti, and Marco Veneroni. Global minimizers for axisymmetric multiphase membranes. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(4):1014–1029, 2013.

[23] N. Schulmann, H. Meyer, T. Kreer, A. Cavallo, A. Johner, J. Baschnagel, and J. P. Wittmer. Strictly two-dimensional self-avoiding walks: Density crossover scaling. *Polymer Science Series C*, 55(1):181–211, July 2013.

[24] Joséphine L.K. Him, Ludovic Pelosi, Henri Chanzy, Jean-Luc Putaux, and Vincent Bulone. Biosynthesis of (1→3)--scpd/scp-glucan (callose) by detergent extracts of a microsomal fraction fromiarabidopsis thaliana/i. *European Journal of Biochemistry*, 268(17):4628–4638, September 2001.

[25] Muhammad Zaki Jawaid, Rosalie Sinclair, Vincent Bulone, Daniel L Cox, and Georgia Drakakaki. A biophysical model for plant cell plate maturation based on the contribution of a spreading force. *Plant Physiology*, 188(2):795–806, 11 2021.

[26] Chloë van Oostende-Triplet, Dominique Guillet, Thomas Triplet, Elvis Pandzic, Paul W. Wiseman, and Anja Geitmann. Vesicle dynamics during plant cell cytokinesis reveals distinct developmental phases. *Plant Physiology*, 174(3):1544–1558, May 2017.

[27] Colleen M. McMichael and Sebastian Y. Bednarek. Cytoskeletal and membrane dynamics during higher plant cytokinesis. *New Phytologist*, 197(4):1039–1057, January 2013.

[28] Joanna Boruc and Daniel Van Damme. Endomembrane trafficking overarching cell plate formation. *Current Opinion in Plant Biology*, 28:92–98, December 2015.

[29] Gerd Jürgens, Misoon Park, Sandra Richter, Sonja Touihri, Cornelia Krause, Farid El Kasmi, and Ulrike Mayer. Plant cytokinesis: a tale of membrane traffic and fusion. *Biochemical Society Transactions*, 43(1):73–78, January 2015.

[30] Hsiang-Wen Chen, Staffan Persson, Markus Grebe, and Heather E. McFarlane. Cellulose synthesis during cell plate assembly. *Physiologia Plantarum*, 164(1):17–26, June 2018.

[31] Maura C. Cannon, Kimberly Terneus, Qi Hall, Li Tan, Yumei Wang, Benjamin L. Wegenhart, Liwei Chen, Derek T. A. Lamport, Yuning Chen, and Marcia J. Kieliszewski. Self-assembly of the plant cell wall requires an extensin scaffold. *Proceedings of the National Academy of Sciences*, 105(6):2226–2231, February 2008.

[32] Tom Shemesh, Robin W. Klemm, Fabian B Romano, Songyu Wang, Joshua C. Vaughan, Xiaowei Zhuang, Hanna Tukachinsky, Michael M. Kozlov, and Tom A. Rapoport. A model for the generation and interconversion of er morphologies. *Proceedings of the National Academy of Sciences*, 111:E5243 – E5251, 2014.

[33] Ian A. Kent and Tanmay P. Lele. Microtubule-based force generation. *WIREs Nanomedicine and Nanobiotechnology*, 9(3), August 2016.

[34] H. T. Jung, B. Coldren, J. A. Zasadzinski, D. J. Iampietro, and E. W. Kaler. The origins of stability of spontaneous vesicles. *Proceedings of the National Academy of Sciences*, 98(4):1353–1357, January 2001.

[35] Harvey T. McMahon and Emmanuel Boucrot. Membrane curvature at a glance. *Journal of Cell Science*, 128(6):1065–1070, March 2015.

[36] Wolf-Rüdiger Scheible, Ravit Eshed, Todd Richmond, Deborah Delmer, and Chris Somerville. Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in iarabidopsis ixr1/i mutants. *Proceedings of the National Academy of Sciences*, 98(18):10079–10084, August 2001.

[37] N. Worden, V. Esteva Esteve, D. S. Domozych, and G. Drakakaki. Using chemical genomics to study cell wall formation and cell growth in arabidopsis thaliana and penium margaritaceum. In *Methods in Molecular Biology*, pages 23–39. Springer New York, October 2014.

[38] J. Samaj, M. Peters, D. Volkmann, and F. Baluska. Effects of myosin ATPase inhibitor 2, 3-butanedione 2-monoxime on distributions of myosins, f-actin, microtubules, and cortical endoplasmic reticulum in maize root apices. *Plant and Cell Physiology*, 41(5):571–582, May 2000.

[39] M. Tominaga, E. Yokota, S. Sonobe, and T. Shimmen. Mechanism of inhibition of cytoplasmic streaming by a myosin inhibitor, 2, 3-butanedione monoxime. *Protoplasma*, 213(1-2):46–54, March 2000.

[40] Takumi Higaki, Natsumaro Kutsuna, Toshio Sano, and Seiichiro Hasezawa. Quantitative analysis of changes in actin microfilament contribution to cell plate development in plant cytokinesis. *BMC Plant Biology*, 8(1), July 2008.

[41] Kathy R. Albe, Margaret H. Butler, and Barbara E. Wright. Cellular concentrations of enzymes and their substrates. *Journal of Theoretical Biology*, 143(2):163–195, March 1990.

[42] Ron Milo. What is the total number of protein molecules per cell volume? a call to rethink some published values. *BioEssays*, 35(12):1050–1055, September 2013.

[43] Ying Gu and Carolyn G Rasmussen. Cell biology of primary cell wall synthesis in plants. *The Plant Cell*, 34(1):103–128, October 2021.

[44] Hsin-Ya Lou, Wenting Zhao, Xiao Li, Liting Duan, Alexander Powers, Matthew Akamatsu, Francesca Santoro, Allister F. McGuire, Yi Cui, David G. Drubin, and Bianxiao Cui. Membrane curvature underlies actin reorganization in response to nanoscale surface topography. *Proceedings of the National Academy of Sciences*, 116(46):23143–23151, October 2019.

[45] Liên Bach, Lionel Gissot, Jessica Marion, Frédérique Tellier, Patrick Moreau, Béatrice Satiat-Jeunemaître, Jean-Christophe Palauqui, Johnathan A. Napier, and Jean-Denis Faure. Very-long-chain fatty acids are required for cell plate formation during cytokinesis iniarabidopsis thaliana/i. *Journal of Cell Science*, 124(19):3223–3234, October 2011.

[46] Yonatan Schweitzer, Tom Shemesh, and Michael M. Kozlov. A model for shaping membrane sheets by protein scaffolds. *Biophysical Journal*, 109(3):564–573, August 2015.

[47] Marisa S. Otegui, David N. Mastronarde, Byung-Ho Kang, Sebastian Y. Bednarek, and L. Andrew Staehelin. Three-dimensional analysis of syncytial-type cell plates during endosperm cellularization visualized by high resolution electron tomography[w]. *The Plant Cell*, 13(9):2033–2051, September 2001.

[48] Gyeongik Ahn, Hyeran Kim, Dae Heon Kim, Hong Hanh, Youngdae Yoon, Indira Singaram, Kaveesha J. Wijesinghe, Kristen A. Johnson, Xiaohong Zhuang, Zizhen Liang, Robert V. Stahelin, Liwen Jiang, Wonhwa Cho, Byung-Ho Kang, and Inhwan Hwang. SH3 domain-containing protein 2 plays a crucial role at the step of membrane tubulation during cell plate formation. *The Plant Cell*, 29(6):1388–1405, June 2017.

[49] Henrik Buschmann and Sabine Müller. Update on plant cytokinesis: rule and divide. *Current Opinion in Plant Biology*, 52:97–105, December 2019.

[50] Shu-Zon Wu and Magdalena Bezanilla. Myosin VIII associates with microtubule ends and together with actin plays a role in guiding plant cell division. *eLife*, 3, September 2014.

[51] Yuh-Ru Julie Lee and Bo Liu. The rise and fall of the phragmoplast microtubule array. *Current Opinion in Plant Biology*, 16(6):757–763, December 2013.

[52] Katarzyna Rybak, Alexander Steiner, Lukas Synek, Susan Klaeger, Ivan Kulich, Eva Facher, Gerhard Wanner, Bernhard Kuster, Viktor Zarsky, Staffan Persson, and Farhah F. Assaad. Plant cytokinesis is orchestrated by the sequential action of the TRAPPII and exocyst tethering complexes. *Developmental Cell*, 29(5):607–620, June 2014.

[53] Ludovic Pelosi, Tomoya Imai, Henri Chanzy, Laurent Heux, Eric Buhler, and Vincent Bulone. Structural and morphological diversity of (1→3)--scpd/scp-glucans synthesized iin vitro/i by enzymes from isaprolegnia monoïca/i. comparison with a corresponding iin vitro/i product from blackberry (irubus fruticosus/i). *Biochemistry*, 42(20):6264–6274, April 2003.

[54] Knut Thiele, Gerhard Wanner, Viktoria Kindzierski, Gerd Jürgens, Ulrike Mayer, Fiona Pachl, and Farhah F. Assaad. The timely deposition of callose is essential for cytokinesis in arabidopsis. *The Plant Journal*, 58(1):13–26, December 2008.

[55] Bruce Stone and A. E. Clarke. *Chemistry and biology of (1 –> 3)-[beta]-glucans / Bruce A. Stone and Adrienne E. Clarke*. La Trobe University Press [Bundoora, Vic.], 1992.

[56] Fangwei Gu, Martin Bringmann, Jonathon Combs, Jiyuan Yang, Dominique Bergmann, and Erik Nielsen. The arabidopsis CSLD5 functions in cell plate formation in a cell cycle dependent manner. *The Plant Cell*, page tpc.00203.2016, June 2016.

[57] Radwa H. Abou-Saleh, Mercedes C. Hernandez-Gomez, Sam Amsbury, Candelas Paniagua, Matthieu Bourdon, Shunsuke Miyashima, Ykä Helariutta, Martin Fuller, Tatiana Budtova, Simon D. Connell, Michael E. Ries, and Yoselin Benitez-Alfonso. Interactions between callose and cellulose revealed through the analysis of biopolymer mixtures. *Nature Communications*, 9(1), October 2018.

[58] W. Rawicz, K.C. Olbrich, T. McIntosh, D. Needham, and E. Evans. Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophysical Journal*, 79(1):328–339, July 2000.

[59] Guillermo R. Lázaro, Ignacio Pagonabarraga, and Aurora Hernández-Machado. Phase-field theories for mathematical modeling of biological membranes. *Chemistry and Physics of Lipids*, 185:46–60, January 2015.

[60] John M. Lee. Model riemannian manifolds. In *Graduate Texts in Mathematics*, pages 55–83. Springer International Publishing, 2018.

[61] Rumiana Dimova. Recent developments in the field of bending rigidity measurements on membranes. *Advances in Colloid and Interface Science*, 208:225–234, 2014. Special issue in honour of Wolfgang Helfrich.

[62] Mingyang Hu, John J. Briguglio, and Markus Deserno. Determining the gaussian curvature modulus of lipid membranes in simulations. *Biophysical Journal*, 102(6):1403–1410, March 2012.

[63] Elisabeth Fischer-Friedrich, Anthony A. Hyman, Frank Jülicher, Daniel J. Müller, and Jonne Helenius. Quantification of surface tension and internal pressure generated by single mitotic cells. *Scientific Reports*, 4(1), August 2014.

[64] *Modern Differential Geometry of Curves and Surfaces with Mathematica, Third Edition*. Chapman and Hall/CRC, September 2017.

[65] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. NIH image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, June 2012.

[66] Alex Mogilner and George Oster. Force generation by actin polymerization II: The elastic ratchet and tethered filaments. *Biophysical Journal*, 84(3):1591–1605, March 2003.

[67] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5):533–534, May 2020.

[68] Edward C. Holmes, Stephen A. Goldstein, Angela L. Rasmussen, David L. Robertson, Alexander Crits-Christoph, Joel O. Wertheim, Simon J. Anthony, Wendy S. Barclay, Maciej F. Boni, Peter C. Doherty, Jeremy Farrar, Jemma L. Geoghegan, Xiaowei Jiang, Julian L. Leibowitz, Stuart J. D. Neil, Tim Skern, Susan R. Weiss,

Michael Worobey, Kristian G. Andersen, Robert F. Garry, and Andrew Rambaut. The origins of sars-cov-2: A critical review. *Cell*, 184(19):4848–4856, 2021.

[69] David Adam. COVID's true death toll: much higher than official records. *Nature*, 603(7902):562–562, March 2022.

[70] Fan Wu, Su Zhao, Bin Yu, Yan-Mei Chen, Wen Wang, Zhi-Gang Song, Yi Hu, Zhao-Wu Tao, Jun-Hua Tian, Yuan-Yuan Pei, Ming-Li Yuan, Yu-Ling Zhang, Fa-Hui Dai, Yi Liu, Qi-Min Wang, Jiao-Jiao Zheng, Lin Xu, Edward C. Holmes, and Yong-Zhen Zhang. A new coronavirus associated with human respiratory disease in china. *Nature*, 579(7798):265–269, February 2020.

[71] S. Khare, C. Gurry, L. Freitas, M.B. Schultz, Bach G., A. Diallo, N. Akite, Lee R.T.C., W. Yeo, GISAID Core Curation Team, and S. Maurer-Stroh. Gisaid's role in pandemic response. *China CDC Weekly*, 3:1049–1051, 2021.

[72] Donald J. Benton, Antoni G. Wrobel, Pengqi Xu, Chloë Roustan, Stephen R. Martin, Peter B. Rosenthal, John J. Skehel, and Steven J. Gamblin. Receptor binding and priming of the spike protein of sars-cov-2 for membrane fusion. *Nature*, 588(7837):327–330, 2020.

[73] Jun Lan, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, Linqi Zhang, and Xinquan Wang. Structure of the sars-cov-2 spike receptor-binding domain bound to the ace2 receptor. *Nature*, 581(7807):215–220, 2020.

[74] Sarah A. Clark, Lars E. Clark, Junhua Pan, Adrian Coscia, Lindsay G.A. McKay, Sundaresh Shankar, Rebecca I. Johnson, Anthony Griffiths, and Jonathan Abraham. Molecular basis for a germline-biased neutralizing antibody response to sars-cov-2. *bioRxiv*, page 2020.11.13.381533, 2020.

[75] Yu Guo, Lisu Huang, Guangshun Zhang, Yanfeng Yao, He Zhou, Shu Shen, Bingqing Shen, Bo Li, Xin Li, Qian Zhang, Mingjie Chen, Da Chen, Jia Wu, Dan Fu, Xinxin Zeng, Mingfang Feng, Chunjiang Pi, Yuan Wang, Xingdong Zhou, Minmin Lu, Yarong Li, Yaohui Fang, Yun-Yueh Lu, Xue Hu, Shanshan Wang, Wanju Zhang, Ge Gao, Francisco Adrian, Qisheng Wang, Feng Yu, Yun Peng, Alexander G. Gabibov, Juan Min, Yuhui Wang, Heyu Huang, Alexey Stepanov, Wei Zhang, Yan Cai, Junwei Liu, Zhiming Yuan, Chen Zhang, Zhiyong Lou, Fei Deng, Hongkai Zhang, Chao Shan, Liang Schweizer, Kun Sun, and Zihe Rao. A sars-cov-2 neutralizing antibody with extensive spike binding coverage and modified for optimal therapeutic outcomes. *Nature Communications*, 12(1):2623, 2021.

[76] Jiandong Huo, Yuguang Zhao, Jingshan Ren, Daming Zhou, Helen M. E. Duyvesteyn, Helen M. Ginn, Loic Carrique, Tomas Malinauskas, Reinis R. Ruza, Pranav N. M. Shah, Tiong Kit Tan, Pramila Rijal, Naomi Coombes, Kevin R. Bewley, Julia A. Tree, Julika Radecke, Neil G. Paterson, Piyada Supasa, Juthathip

Mongkolsapaya, Gavin R. Screaton, Miles Carroll, Alain Townsend, Elizabeth E. Fry, Raymond J. Owens, and David I. Stuart. Neutralization of sars-cov-2 by destruction of the prefusion spike. *Cell Host & Microbe*, 28(3):445–454.e6, 2020.

[77] James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, May 2018.

[78] Fernando P. Polack, Stephen J. Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L. Perez, Gonzalo Pérez Marc, Edson D. Moreira, Cristiano Zerbini, Ruth Bailey, Kena A. Swanson, Satrajit Roychoudhury, Kenneth Koury, Ping Li, Warren V. Kalina, David Cooper, Robert W. Frenck, Laura L. Hammitt, Özlem Türeci, Haylene Nell, Axel Schaefer, Serhat Ünal, Dina B. Tresnan, Susan Mather, Philip R. Dormitzer, Uğur Şahin, Kathrin U. Jansen, and William C. Gruber. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. *New England Journal of Medicine*, 383(27):2603–2615, December 2020.

[79] Heidi Ledford. Supercomputer sets protein-folding record. *Nature*, October 2010.

[80] Royal Society of Chemistry. Chem vs covid timeline, December 2020.

[81] Bhumi Shah, Palmi Modi, and Sneha R. Sagar. In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. *Life Sciences*, 252:117652, July 2020.

[82] Divya M. Teli, Mamta B. Shah, and Mahesh T. Chhabria. In silico screening of natural compounds as potential inhibitors of sars-cov-2 main protease and spike rbd: Targets for covid-19. *Frontiers in Molecular Biosciences*, 7, 2021.

[83] Giulia Russo, Marzio Pennisi, Epifanio Fichera, Santo Motta, Giuseppina Raciti, Marco Viceconti, and Francesco Pappalardo. In silico trial to test COVID-19 candidate vaccines: a case study with UISS platform. *BMC Bioinformatics*, 21(S17), December 2020.

[84] M. Zaki Jawaid, A. Baidya, R. Mahboubi-Ardakani, Richard L. Davis, and Daniel L. Cox. Simulation of the omicron variant of SARS-CoV-2 shows broad antibody escape, weakened ACE2 binding, and modest increase in furin binding. December 2021.

[85] M.Z. Jawaid, A Baidya, S Jakovcevic, R Mahboubi-Ardakani, J Lusk, N Solomon, G Gonzalez, J Arsuaga, M Vazquez, R.L. Davis, and D.L. Cox. Structures of the furin cleavage domain of coronaviruses studied using deep learning and molecular dynamics. *In preparation, to be submitted to Science*, 2021.

[86] Chuan Tian, Koushik Kasavajhala, Kellon A. A. Belfon, Lauren Raguette, He Huang, Angela N. Migues, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu,

and Carlos Simmerling. ff19sb: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of Chemical Theory and Computation*, 16(1):528–552, 2020. PMID: 31714766.

[87] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, May 1995.

[88] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: Anin/i·log(in/i) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, June 1993.

[89] Elmar Krieger and Gert Vriend. Yasara view—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics*, 30(20):2981–2982, 2014.

[90] K. Anton Feenstra, Berk Hess, and Herman J. C. Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798, June 1999.

[91] TJ Hou, JM Wang, YY Li, and W Wang. Assessing the performance of the mm/pbsa and mm/gbsa methods: I. the accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information & Modeling*, 51:69–82, 2001.

[92] J rg Weiser, Peter S. Shenkin, and W. Clark Still. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry*, 20(2):217–230, January 1999.

[93] G Q Weng, E C Wang, Z Wang, H Liu, D Li, F Zhu, and T J Hou. Hawkdock: a web server to predict and analyze the structures of protein-protein complexes based on computational docking and mm/gbsa. *Nucleic Acids Research*, 47:W322–W330, 2019.

[94] Huiyong Sun, Youyong Li, Sheng Tian, Lei Xu, and Tingjun Hou. Assessing the performance of mm/pbsa and mm/gbsa methods. 4. accuracies of mm/pbsa and mm/gbsa methodologies evaluated by various simulation protocols using pdbbind data set. *Physical Chemistry Chemical Physics*, 16:16719–16729, 2014.

[95] Elmar Krieger and Gert Vriend. New ways to boost molecular dynamics simulations. *Journal of computational chemistry*, 36(13):996–1007, 2015.

[96] Ingrid Torjesen. Covid-19: Omicron may be more transmissible than other variants and partly resistant to existing vaccines, scientists fear. *BMJ*, 375:n2943, 2021.

[97] J. Jumper, R. Evans, and A. et al. Pritzel. Highly accurate protein structure prediction with alphafold. *Nature*, 2021.

[98] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with alphafold-multimer. *bioRxiv*, page 2021.10.04.463034, 2021.

[99] Salim S. Abdool Karim and Quarraisha Abdool Karim. Omicron sars-cov-2 variant: a new chapter in the covid-19 pandemic. *The Lancet*, 2021.

[100] Alexandra C Walls, Young-Jun Park, M Alejandra Tortorici, Abigail Wall, Andrew T McGuire, and David Veesler. Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein. *Cell*, 181(2):281–292. e6, 2020.

[101] Ruben J. G. Hulswit, Yifei Lang, Mark J. G. Bakkers, Wentao Li, Zeshi Li, Arie Schouten, Bram Ophorst, Frank J. M. van Kuppeveld, Geert-Jan Boons, Berend-Jan Bosch, Eric G. Huizinga, and Raoul J. de Groot. 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, 2019.

[102] Thomas P. Peacock, Daniel H. Goldhill, Jie Zhou, Laury Baillon, Rebecca Frise, Olivia C. Swann, Ruthiran Kugathasan, Rebecca Penn, Jonathan C. Brown, Raul Y. Sanchez-David, Luca Braga, Maia Kavanagh Williamson, Jack A. Hassard, Ecco Staller, Brian Hanley, Michael Osborn, Mauro Giacca, Andrew D. Davidson, David A. Matthews, and Wendy S. Barclay. The furin cleavage site in the sars-cov-2 spike protein is required for transmission in ferrets. *Nature Microbiology*, 6(7):899–909, 2021.

[103] Akatsuki Saito, Takashi Irie, Rigel Suzuki, Tadashi Maemura, Hesham Nasser, Keiya Uriu, Yusuke Kosugi, Kotaro Shirakawa, Kenji Sadamasu, Izumi Kimura, Jumpei Ito, Jiaqi Wu, Kiyoko Iwatsuki-Horimoto, Mutsumi Ito, Seiya Yamayoshi, Samantha Loeber, Masumi Tsuda, Lei Wang, Seiya Ozono, Erika P. Butler-tanaka, Yuri L. Tanaka, Ryo Shimizu, Kenta Shimizu, Kumiko Yoshimatsu, Ryoko Kawabata, Takemasa Sakaguchi, Kenzo Tokunaga, Isao Yoshida, Hiroyuki Asakura, Mami Nagashima, Yasuhiro Kazuma, Ryosuke Nomura, Yoshihito Horisawa, Kazuhisa Yoshimura, Akifumi Takaori-Kondo, Masaki Imai, Mika Chiba, Hirotake Furihata, Haruyo Hasebe, Kazuko Kitazato, Haruko Kubo, Naoko Misawa, Nanami Morizako, Kohei Noda, Akiko Oide, Mai Suganami, Miyoko Takahashi, Kana Tsushima, Miyabishara Yokoyama, Yue Yuan, Shinya Tanaka, So Nakagawa, Terumasa Ikeda, Takasuke Fukuhara, Yoshihiro Kawaoka, Kei Sato, and Consortium The Genotype to Phenotype Japan. Enhanced fusogenicity and pathogenicity of sars-cov-2 delta p681r mutation. *Nature*, 2021.

[104] Adam Vaughan. Delta to dominate world. *New Scientist*, 250(3341):9, 2021.

[105] Milot Mirdita, Sergey Ovchinnikov, and Martin Steinegger. Colabfold - making protein folding accessible to all. *bioRxiv*, page 2021.08.15.456425, 2021.

[106] Xiangyang Chi, Renhong Yan, Jun Zhang, Guanying Zhang, Yuanyuan Zhang, Meng Hao, Zhe Zhang, Pengfei Fan, Yunzhu Dong, Yilong Yang, Zhengshan Chen, Yingying Guo, Jinlong Zhang, Yaning Li, Xiaohong Song, Yi Chen, Lu Xia, Ling Fu, Lihua Hou, Junjie Xu, Changming Yu, Jianmin Li, Qiang Zhou, and Wei Chen. A neutralizing human antibody binds to the n-terminal domain of the spike protein of sars-cov-2. *Science*, 369(6504):650–655, 2020.

[107] AJ Venkatakrishnan, Praveen Anand, Patrick Lenehan, Rohit Suratekar, Bharathwaj Raghunathan, Michiel J Niesen, and Venky Soundararajan. Omicron variant of sars-cov-2 harbors a unique insertion mutation of putative viral or human genomic origin, Dec 2021.

[108] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.

[109] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[110] Araz Jakalian, David B Jack, and Christopher I Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation. *Journal of computational chemistry*, 23(16):1623–1641, 2002.

[111] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

[112] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.

[113] Elmar Krieger, Roland L Dunbrack, Rob WW Hooft, and Barbara Krieger. Assignment of protonation states in proteins and ligands: Combining pk a prediction with hydrogen bonding network optimization. In *Computational Drug Discovery and Design*, pages 405–421. Springer, 2012.

[114] F Chen, H Liu, HY Sun, PC Pan, YY Li, D Li, and TJ Hou. Assessing the performance of the mm/pbsa and mm/gbsa methods. 6. capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. *Physical Chemistry Chemical Physics*, 18:22129–22139, 2016.

[115] Luigi Genovese, Marco Zaccaria, Michael Farzan, Welkin E Johnson, and Babak Momeni. Investigating the mutational landscape of the sars-cov-2 omicron variant via ab initio quantum mechanical modeling. *bioRxiv*, page 2021.12.01.470748, 2021.

[116] N. Chapeshamano. Discovery health, south africa's largest private health insurance administrator, releases at-scale, real-world analysis of omicron outbreak based on 211 000 covid-19 test results in south africa, including collaboration with the south africa, 2021.

[117] Suresh Kumar, Thiviya S Thambiraja, Kalimuthu Karuppanan, and Gunasekaran Subramaniam. Omicron and delta variant of sars-cov-2: A comparative computational study of spike protein. *bioRxiv*, page 2021.12.02.470946, 2021.

[118] Thomas P. Peacock, Daniel H. Goldhill, Jie Zhou, Laury Baillon, Rebecca Frise, Olivia C. Swann, Ruthiran Kugathasan, Rebecca Penn, Jonathan C. Brown, Raul Y. Sanchez-David, Luca Braga, Maia Kavanagh Williamson, Jack A. Hassard, Ecco Staller, Brian Hanley, Michael Osborn, Mauro Giacca, Andrew D. Davidson, David A. Matthews, and Wendy S. Barclay. The furin cleavage site of sars-cov-2 spike protein is a key determinant for transmission due to enhanced replication in airway cells. *bioRxiv*, page 2020.09.30.318311, 2020.

[119] Yiran Wu and Suwen Zhao. Furin cleavage sites naturally occur in coronaviruses. *Stem Cell Research*, 50:102115, 2021.

[120] Panita Decha, Thanyada Rungrotmongkol, Pathumwadee Intharathep, Maturos Malaisree, Ornjira Aruksakunwong, Chittima Laohpongspaisan, Vudhichai Parasuk, Pornthep Sompornpisut, Somsak Pianwanit, Sirirat Kokpol, and Supot Hannongbua. Source of high pathogenicity of an avian influenza virus h5n1: Why h5 is better cleaved by furin. *Biophysical Journal*, 95(1):128–134, 2008.

[121] Stefan Henrich, Angus Cameron, Gleb P. Bourenkov, Reiner Kiefersauer, Robert Huber, Iris Lindberg, Wolfram Bode, and Manuel E. Than. The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nature Structural & Molecular Biology*, 10(7):520–526, 2003.

[122] Magdalena M. Kacprzak, Juan R. Peinado, Manuel E. Than, Jon Appel, Stefan Henrich, Gregory Lipkind, Richard A. Houghten, Wolfram Bode, and Iris Lindberg. Inhibition of furin by polyarginine-containing peptides: Nanomolar inhibition by nona-d-arginine *. *Journal of Biological Chemistry*, 279(35):36788–36794, 2004.

[123] Sven O. Dahms, Kornelia Hardes, Torsten Steinmetzer, and Manuel E. Than. X-ray structures of the proprotein convertase furin bound with substrate analogue inhibitors reveal substrate specificity determinants beyond the s4 pocket. *Biochemistry*, 57(6):925–934, 2018.

[124] Naveen Vankadari. Structure of furin protease binding to sars-cov-2 spike glycoprotein and implications for potential targets and virulence. *Journal of Physical Chemistry Letters*, 11(16):6655–6663, 2020.

[125] Sun Tian, Qingsheng Huang, Ying Fang, and Jianhua Wu. Furindb: A database of 20-residue furin cleavage site motifs, substrates and their associated drugs. *International Journal of Molecular Sciences*, 12(2), 2011.

[126] Stefan Elbe and Gemma Buckland-Merrett. Data, disease and diplomacy: Gisaid's innovative contribution to global health. *Global Challenges*, 1(1):33–46, 2017.

[127] Yuelong Shu and John McCauley. Gisaid: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22(13):30494, 2017.

[128] Eric W. Sayers, Evan E. Bolton, J. Rodney Brister, Kathi Canese, Jessica Chan, Donald C Comeau, Ryan Connor, Kathryn Funk, Chris Kelly, Sunghwan Kim, Tom Madej, Aron Marchler-Bauer, Christopher Lanczycki, Stacy Lathrop, Zhiyong Lu, Francoise Thibaud-Nissen, Terence Murphy, Lon Phan, Yuri Skripchenko, Tony Tse, Jiyao Wang, Rebecca Williams, Barton W Trawick, Kim D Pruitt, and Stephen T Sherry. Database resources of the national center for biotechnology information. *Nucleic Acids Research*, page gkab1112, 2021.

[129] Elisabeth Braun and Daniel Sauter. Furin-mediated protein processing in infectious diseases and cancer. *Clinical & Translational Immunology*, 8(8):e1073, 2019.

[130] Bo Meng, Isabella Ferreira, Adam Abdullahi, Steven A. Kemp, Niluka Goonawardane, Guido Papa, Saman Fatihi, Oscar Charles, Dami Collier, Citiid-Nihr BioResource COVID-19 Collaboration, Consortium The Genotype to Phenotype Japan, Jinwook Choi, Joo Hyeon Lee, Petra Mlcochova, Leo James, Rainer Doffinger, Lipi Thukral, Kei Sato, and Ravindra K. Gupta. Sars-cov-2 omicron spike mediated immune escape, infectivity and cell-cell fusion. *bioRxiv*, page 2021.12.17.473248, 2021.

[131] Zeng Cong, John P. Evans, Panke Qu, Julia Faraone, Yi-Min Zheng, Claire Carlin, Joseph S. Bednash, Tongqing Zhou, Gerard Lozanski, Rama Mallampalli, Linda J. Saif, Eugene M. Oltz, Peter Mohler, Kai Xu, Richard J. Gumina, and Shan-Lu Liu. Neutralization and stability of sars-cov-2 omicron variant. *bioRxiv*, page 2021.12.16.472934, 2021.

[132] Y Shu and J McCauley. Gisaid: Global initiative on sharing all influenza data – from vision to reality. *EuroSurveillance*, 22, 2017.

[133] Fabian Sievers and Desmond G Higgins. Clustal omega for making accurate alignments of many protein sequences. *Protein Science*, 27:135–145, 2018.

[134] J D Thompson, D G Higgins, and T J Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673–4680, 1994.

[135] Robert C. Edgar. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32:1792–1797, 2004.

[136] Abdul Basit, Tanveer Ali, and Shafiq Ur Rehman. Truncated human angiotensin converting enzyme 2 a potential inhibitor of SARS-CoV-2 spike glycoprotein and potent COVID-19 therapeutic agent. *Journal of Biomolecular Structure and Dynamics*, 39(10):3605–3614, May 2020.

[137] Thomas W. Linsky, Renan Vergara, Nuria Codina, Jorgen W. Nelson, Matthew J. Walker, Wen Su, Christopher O. Barnes, Tien-Ying Hsiang, Katharina Esser-Nobis, Kevin Yu, Z. Beau Reneer, Yixuan J. Hou, Tanu Priya, Masaya Mitsumoto, Avery Pong, Uland Y. Lau, Marsha L. Mason, Jerry Chen, Alex Chen, Tania Berrocal, Hong Peng, Nicole S. Clairmont, Javier Castellanos, Yu-Ru Lin, Anna Josephson-Day, Ralph S. Baric, Deborah H. Fuller, Carl D. Walkey, Ted M. Ross, Ryan Swanson, Pamela J. Bjorkman, Michael Gale, Luis M. Blancas-Mejia, Hui-Ling Yen, and Daniel-Adriano Silva. De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science*, 370(6521):1208–1214, December 2020.

[138] Shiho Tanaka, Gard Nelson, C. Anders Olson, Oleksandr Buzko, Wendy Higashide, Annie Shin, Marcos Gonzalez, Justin Taft, Roosheel Patel, Sofija Buta, Ashley Richardson, Dusan Bogunovic, Patricia Spilman, Kayvan Niazi, Shahrooz Rabizadeh, and Patrick Soon-Shiong. An ACE2 triple decoy that neutralizes SARS-CoV-2 shows enhanced affinity for virus variants. *Scientific Reports*, 11(1), June 2021.

[139] Jian Shang, Gang Ye, Ke Shi, Yushun Wan, Chuming Luo, Hideki Aihara, Qibin Geng, Ashley Auerbach, and Fang Li. Structural basis of receptor recognition by SARS-CoV-2. *Nature*, 581(7807):221–224, March 2020.

[140] Arun S. Konagurthu, James C. Whisstock, Peter J. Stuckey, and Arthur M. Lesk. MUSTANG: A multiple structural alignment algorithm. *Proteins: Structure, Function, and Bioinformatics*, 64(3):559–574, May 2006.

[141] Lingfei Jia, Yilun Wang, and Cun-Yu Wang. circFAT1 promotes cancer stemness and immune evasion by promoting STAT3 activation. *Advanced Science*, 8(13):2003376, May 2021.

[142] Lianghui Zhang, Soumajit Dutta, Shiqin Xiong, Matthew Chan, Kui K. Chan, Timothy M. Fan, Keith L. Bailey, Matthew Lindeblad, Laura M. Cooper, Lijun Rong, Anthony F. Gugliuzza, Diwakar Shukla, Erik Procko, Jalees Rehman, and Asrar B. Malik. Engineered ACE2 decoy mitigates lung injury and death induced by SARS-CoV-2 variants. *Nature Chemical Biology*, 18(3):342–351, January 2022.