UNIVERSITY OF CALIFORNIA
RIVERSIDE

p38 MAPK Pathway Regulation By Resonance Selectivity, IS200/IS605 Family Transposon Excision Dynamics, and SARS-CoV-2 Structural Protein Purification

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Michael Worcester

March 2024

Dissertation Committee:

Dr. Thomas E. Kuhlman, Chairperson
Dr. Marcus Kaul
Dr. Roya Zandi

The Dissertation of Michael Worcester is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to thank my advisor Dr. Thomas E. Kuhlman for his guidance and, more importantly, patience along the way. Likewise I am grateful to my labmates who have assisted me, including Dr. Davneet Kaur, Femila Manoj, Aisa Anbir, Melissa Gomez, Pratyasha Mishra, Shayan Nejad, and Jackson Kishbaugh-Maish. I would be a much worse scientist without the help they've given.

I would like to thank Dr. Marcus Kaul, Dr. Nina Yuan, and Dr. Roya Zandi, whose professionalism has enriched our work immensely.

To my friends and family and Liz

ABSTRACT OF THE DISSERTATION

p38 MAPK Pathway Regulation By Resonance Selectivity, IS200/IS605 Family Transposon Excision Dynamics, and SARS-CoV-2 Structural Protein Purification

by

Michael Worcester

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, March 2024
Dr. Thomas E. Kuhlman, Chairperson

Oscillations in protein levels or activation state are ubiquitous in eukaryotic signaling pathways, but the function of these oscillations remains unclear. We find that p38α MAPK activation state oscillates at multiple frequencies in response to stress rather than at a singular frequency. Fourier analysis of p38α activation state measured with a novel FRET sensor shows that unique stressors including IL-1β, SARS-CoV-2 Spike protein, and sorbitol induce unique response patterns in the frequency domain. Analysis of statistically over represented frequencies suggests that frequency bins are fixed and organized as harmonics with a fundamental frequency of 0.09 hr^{-1}(p=0.04). Cross spectral dynamics between p38α and a substrate indicate that substrate activation behaves similarly to a forced oscillator system being driven at resonance. We have developed a model that characterizes the MAPK phosphorylation cascade as a chemical resonator that impart multiple frequencies on p38α, allowing the kinase to regulate its substrates through resonance selectivity. Stochastic simulations of substrates being driven on and off-resonance produce spectra comparable to experimental data.

Experiments on *E. coli* expressing a transposon of the IS200/IS5605 family reveal that expression of transposon protein TnpB leads to a higher excision rate. Furthermore, cells that excise possess TnpB at a significantly higher expression level than cells that don't, with TnpA expression levels making no discernible difference in excision dynamics.

We present a protocol for purification of SARS-CoV-2 structural proteins using a bacterial expression system, SUMO tags, and affinity purification methods. We further report a method to insert synthesized dimers into a suspended lipid membrane in a homogeneous orientation. AFM and Cryo-EM in tandem with molecular dynamics simulations show membrane thinning around the M protein and a propensity to induce local membrane curvature.

Contents

List of Figures

Chapter 1

Introduction

I never thought I'd see a resonance cascade, let alone create one!

Unnamed Black Mesa scientist

Experimental biophysics is the application of laboratory methods widely employed in physics to the study of living systems. In Cell Biology, researchers are generally more interested in the identification of proteins, molecules, genetic elements, etc. and don't tend to analytically characterize the mechanisms that have evolved naturally in organisms over the span of billions of years. As an analogy, experimental condensed matter physics is typically predicated on the process of fabricating a sample and measuring its physical properties in an attempt to understand underlying intrinsic phenomena. Therefore in biophysics, the sample is usually a culture that has been grown up and subjected to particular experimental conditions, such as stimulation by physical or chemical sources. Some examples of research

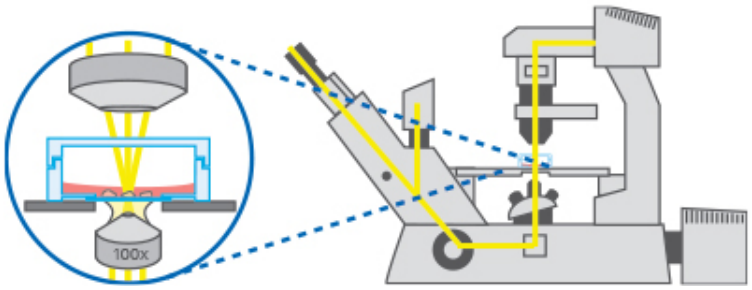

Figure 1.1: Inverted microscope that is set up to image an adherent cell culture[2]

interests for biophysics researchers include: the structure of DNA, ion channels, protein folding, and neural networks[1].

The primary experimental method by which I conduct my research is the employment of inverted fluorescence microscopy, or, microscopy where the target is stimulated by monochromatic light sources and imaged from below (Figure 1.1). This process is advantageous for samples that form monolayers, such as bacteria confined to lateral growth and adherent cell cultures like the cells that line our kidneys. Images taken at high magnification further allow for localization of fluorescent reporters in live cells or disambiguation of population dynamics, something that is typically glossed over in biology research.

Where biology in praxis is largely concerned with the better understanding and improvement of human health, research on human cell samples is one of the vessels by which this is accomplished. The bulk of my research in this field utilizes HEK293T cells, which are embryotic kidney cells that form monolayers and pack themselves tightly into lattices. In these cells I have researched oscillations in stress signaling pathways, an attempt to decode the presently not understood mechanism by which cells transduce simple extrcellular signals

into tightly coordinated response. In this research I have employed the use of the Fourier transform in tandem with images capturing thousands of cells to identify periodicities in these pathways. So far, implementation of Fourier transform in stress pathway research has almost never been done.

As physicists trying to uncover a naturally evolved biological mechanism for intracellular signaling, we have to be able to conceive of ways that such a method of communication can be achieved. Radios, for example, function by being able to exclude the multitudes of signals that float around in the airwaves in favor of focusing on just one frequency. We ask ourselves if cellular signaling is achieved by creating a dense ocean of instructions that can be individually isolated for by the process of resonance selectivity, exactly like a radio would.

Another area of interest for my research has been in the dynamics of transposition. Transposons are self-interested genetic elements capable of changing their position in the genome and do not serve any direct function to the host. Although transposons are not presently a hot topic of interest in biological research, there are a number of types that have potential applications as gene editing tools in a manner similar to CRISPR-associated protein 9[3][4]. In my experiments I have probed the dynamics of *Escherchia coli* undergoing the process of transposition in an attempt to derive the function of a mysterious transposon-related protein known as TnpB.

In order to reliably perform live experiments on human cells with minimal perturbation, the environmental conditions required for maintaining homeostasis must be met. For human cell cultures, this broadly means 5% atmospheric CO2 concentration, an ambient

temperature of 37°, and humidity >90%. Sealed environments that allow for simultaneous imaging from an inverted microscope are purchasable from manufacturers, but their price tags that dwell in the tens of thousands of dollars construct an insurmountable wall for some laboratories. We presently live in an era where single-board computers such as Arduino and Raspberry Pi allow for complex and programmable control of electronics systems, and the advent of rapid prototyping in 3D printing allows for low cost customizable geometries to be fabricated. I have therefore embarked on a mission to build a device that performs the same functions as a professionally manufactured stage-top incubator for an orders-of-magnitude price reduction.

Finally, the arrival and endemic status of COVID-19 has kindled interest in understanding how the associated virus assembles. The membrane protein is the most abundant of the four structural proteins and necessary for assembly, but it is difficult to purify and therefore not well characterized. On this project I assist in developing a high yield method of SARS-CoV-2 structural protein purification that employs a Small Ubiquitin-related Modifier tag-based expression system. Such a technique for purification thereby paves the way for more powerful experiments employing the likes of atomic force microscopy and cryogenic electron microscopy to probe the behavior of this elusive protein.

Chapter 2

Oscillations in the p38 MAPK Pathway

2.1 Introduction

The contents of this chapter are in preparation to be submitted as an article in a scientific journal under the title "Differential Human Stress Response Pathway Activation by p38α MAPK Resonance Selectivity". I am the primary author of this work, with my role on the project being sample preparation, microscope operation, and data analysis. Secondary authors include Aisa Anbir and Shayan Nejad who conducted additional experiments, Pratyasha Mishra who performed Western blot assays, Siyu Lu and Kevin Yang who did p38 pathway modeling, and Nina Yuan who performed Western blot assays and subsequent analysis. We are indebted to the labs of Roya Zandi and Marcus Kaul, whose

material and manual support made this project possible. Thomas E. Kuhlman will also act as the corresponding author for submission.

Stress, in biology, is an external influence that directly threatens the health of your cells. Lying in the sun for too long leaves your skin exposed to ultraviolet light, which damages the DNA in your skin cells and will invariably lead to cancer if sustained and repeated for years. After that DNA damage occurs, an apparatus is set into motion as each skin cell undergoes a repair response and begins coordinating with its neighbors about what to do. These apparati of communication are known as signaling pathways and DNA damage is just one stress that our bodies are evolved to handle.

The p38 MAPK signaling pathway is tied with a number of cellular functions, among them being inflammatory response, cell differentiation, and apoptosis. Functionally, the pathway exists to relay stress detection into a specific response that aims to preserve the health of the organism. Its dysregulation unfortunately is associated with health problems such as inflammation, immune disorders, tumors, and various cancers, making it a target of interest for therapeutics[5][6].

Activation of the p38 pathway typically begins with detection of an environmental stress via cell surface receptor binding followed by a three-tiered phosphorylation cascade, starting with MAP3Ks, then MAP2Ks, and finally MAPKs at the bottom[7]. While there are four known isoforms of p38 MAPK, p38α is the most numerously expressed in the majority of cell types. It is activated by dual phosphorylation of Thr and Tyr on the Thr-Gly-Tyr motif, allowing for the protein to take on a more open conformation. Through this simple mechanism of activation, p38α efficiently regulates over a hundred substrates

covering several important functions[8]. The process by which this happens is not presently known in detail.

Oscillations in the p38 pathway have already been detected in several instances. Tomida et al. tracked the periodic activation of a p38 substrate in HeLa cells following stimulation by pro-inflammatory cytokine IL-1β[9]. They constructed a p38 activity reporter known as "PerKy-38" (henceforth referred to as PerKy) from the docking domain of Mef2A, a protein that enhances transcriptional activity[8]. The reporter was also dual tagged with YPet and super-enhanced CFP on the activation loop to allow for Förster Resonance Energy Transfer (FRET) [Section 3.1] measurements to track phosphorylation state. Single cell tracking following IL-1β induction revealed asynchronous oscillations with an observed period of roughly 2 hours. Furthermore, they determined that pro-inflammatory genes IL-6 and IL-8 were most expressed when cells were driven at the observed experimental period, concluding that p38 oscillation is necessary for appropriate inflammatory gene expression.

Geva-Zatorsky et al. found that p53, a substrate of p38 that is instrumental in the DNA repair process, is periodically expressed in MCF7 cells in response to double-stranded DNA breaks due to gamma radiation exposure[10]. Multiple frequencies of oscillation were detected, forming harmonics with a fundamental frequency of 0.14 ± 0.02 hr^{-1}. Analysis of cells not exposed to gamma radiation similarly revealed a peak at 0.14 hr^{-1}, but no subsequent harmonics.

Sun et al. found that cytokine TNF induces oscillations in the expression of over 5000 genes[11]. Said oscillations were characterized as rapid, with periods only reaching as

high as one hour. Furthermore, oscillations in MAPK activation states were also detected in response to TNF.

Fourier analysis as an analytical tool in probing signaling pathways has virtually never been implemented. Our approach in this series of experiments is oriented toward evaluating p38 activity in the frequency domain, where periods of oscillation in principle can be extracted in addition to discerning significance derived from Fourier amplitude. Furthermore, taking numerous images of hundreds of cells allows for the resolution of robust ensemble averages that eliminate potentially misleading stochastic phenomena. Finally, whereas the experiments of Tomida et al. have indirectly measured p38 response over the course of 8 hours, our experiments were recorded over the course of 48 hours, capturing multiple cell cycles and allowing for the quantification of transientness (Figure 3.2.2) in response to various unique stimuli.

2.2 Results

2.2.1 IL-1β Induction Produces a Singular Peak In PerKy293 Cells

Tomida et al.'s experiments did not include spectral analysis, giving significant error to the observed period of roughly 2 hours. Furthermore, naked eye inspections of the time domain cell traces following IL-1β induction would not identify multiple frequencies hidden in the asynchronous oscillations.

We repeated the aforementioned experiment by inducing HEK-293T cells stably transfected with PerKy (henceforth referred to as PerKy293) with 300 ng of IL-1β and

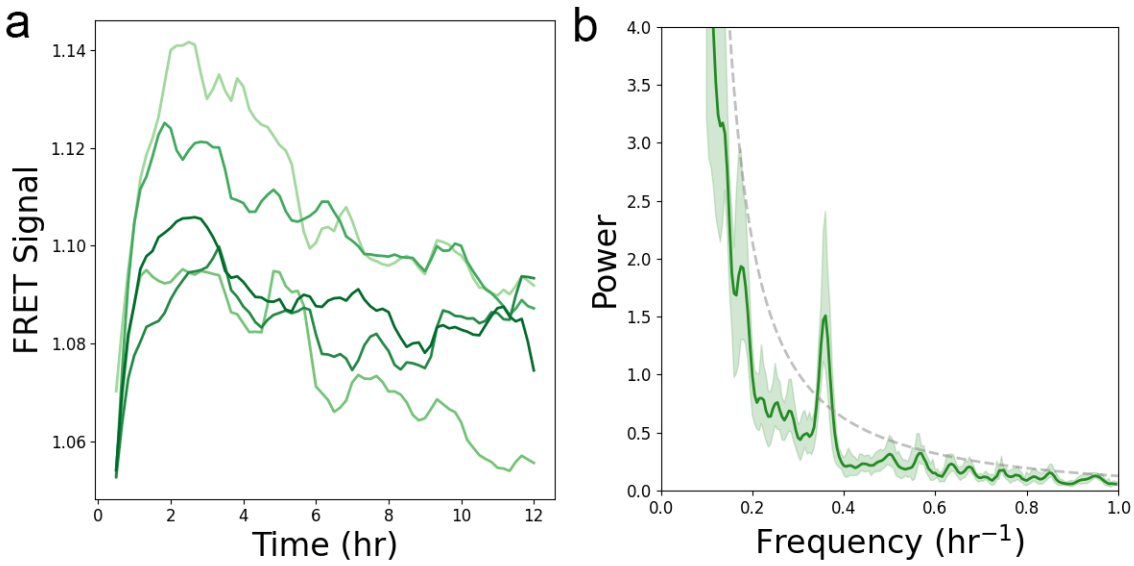

Figure 2.1: PerKy293 cells induced with 300 ng/mL IL-1β. **a)** FRET signal on a per-FOV basis, with each trace representing a single FOV. **b)** Ensemble averaged Fourier spectrum. Grey dotted curve indicates a 99% confidence interval derived from a control spectrum. Green shaded region represents a 95% confidence interval for individual spectra that comprise the ensemble average (49 spectra).

observing changes in the FRET signal over a period of 48 hours. Images were analyzed on a per-field-of-view (FOV) basis as opposed to on a single cell basis, due to the high motility and dynamic physiological changes of HEK293T cells making segmentation difficult. Induction by IL-1β immediately causes a sharp increase in the overall FRET signal followed by a gradual relaxation. Figure 2.1a qualitatively shows this in addition to potential oscillations around this trend. However, from visual inspection it is not conclusive that there is periodic behavior.

Fourier analysis of these FRET time traces reveals a strong peak at 0.36 hr^{-1}(Figure 2.1b), indicating a distance of 2.78 hours between peaks. Inspection of individual spectra also reveals that the peak at 0.36 hr^{-1} is broadly represented and not an artifact derived from a single spectrum (Figure 2.2). Confidence intervals as shown in the figure were gen-

erated from control data (Figure 3.3.1). Repeating this process with the addition of 10 μM potent p38 inhibitor SB203580 15 minutes prior to stress induction abrogates the main peak at 0.36 hr^{-1}, indicating that these oscillations coincide with p38 activation. Now that we had corroborated the findings of Tomida et al., we sought a more complete picture of p38 behavior with a directly labeled FRET reporter.

2.2.2 Directly Labeling p38 As A Novel FRET Sensor

Directly-labeled kinase FRET reporters are typically not employed due to the possibility of structural alterations interfering with the kinase's catalytic ability[12]. Instead, sensors are usually constructed downstream from substrates, bypassing this concern while still being able to report on kinase activity. The disadvantage to this method is that this insight is incomplete, due to there existing other substrates with a variety of functions and multiple phosphatases that down regulate p38. PerKy, more specifically, is also excluded from the nucleus, limiting potential future experiments to simultaneously track the activation state and nuclear translocation of p38 during stress induction.

Given the sensitivity of Fourier transform to periodic behavior, it is entirely possible to detect oscillations from time domain data that would otherwise appear uninteresting at a cursory glance. Hence, we developed VPC, a direct FRET-based reporter of p38α activation state. VPC consists of translational fusion of the N and C terminal ends of p38α to the monomeric fluorescent reporters mVenus[13] and mCerulean3[14] respectively. Western blot of lysate from IL-1β-stimulated HEK-293T cells stably transfected to express VPC

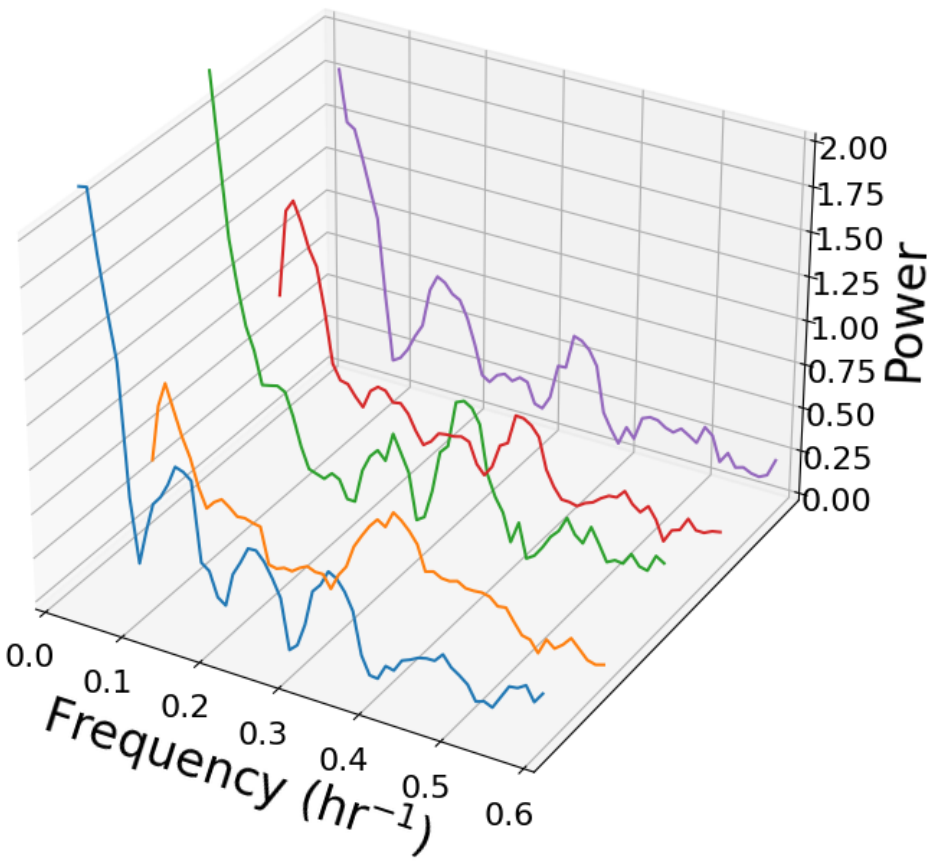

Figure 2.2: Traces of spectra show prominence of multiple peaks on a per-FOV basis.

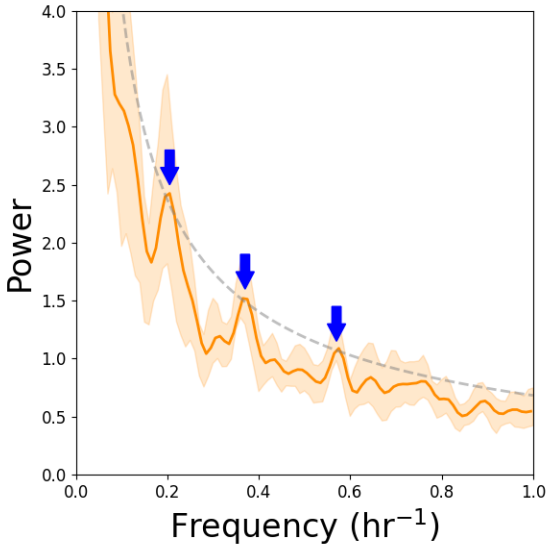

Figure 2.3: VPC293 cells induced with 300 ng/mL IL-1β show three statistically significant peaks in the ensemble-averaged Fourier spectrum. The peaks are located at 0.2 hr^{-1}, 0.35 hr^{-1}, and 0.57 hr^{-1}.

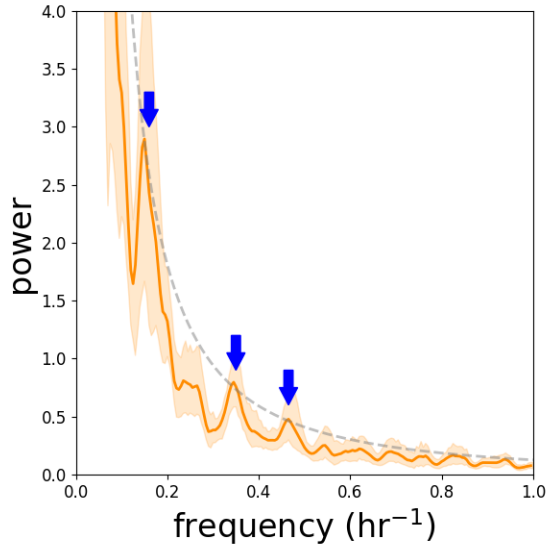

Figure 2.4: VPC293 cells induced with 300 ng/mL IL-1β and 10 ug/mL SB203580 show three statistically significant peaks in the ensemble-averaged Fourier spectrum. The peaks are located at 0.18 hr^{-1}, 0.35 hr^{-1}, and 0.46 hr^{-1}.

(henceforth referred to as VPC293) using antibodies specific to phospho-p38α demonstrate that VPC is phosphorylated as with native p38α (Figure 2.5A). Moreover, the dynamics of VPC phosphorylation state measured both by western blot and ensemble-averaged FRET recapitulate those of native p38α in untransfected HEK-293T cells (Figure 2.5B). This demonstrates that the translationally fused fluorescent tags do not affect the phosphorylation dynamics of VPC and that FRET accurately reports p38α activation state on the time scales measured here. Additionally, overexpression of VPC by transient transfection in HEK-293T renders cells extremely sensitive to a wide variety of weak perturbations, frequently resulting in cell death. This suggests that VPC remains catalytically active and capable of phosphorylating downstream targets, and conseqently all experiments reported here were performed with stably transfected VPC293 cells expressing low levels of VPC. Finally, VPC displays unique patterns of localization upon stimulation, consistent with observations of nuclear and cytoplasmic localization displayed by native p38α upon phosphorylation and activation [Figures 2.5C,D;[15]], with FRET state changing dynamically in time and amplitude in response to perturbations of various strength (Figure 2.5E).

IL-1β induction of VPC293 cells under the same experimental conditions as the previous subsection recapitulates the findings in addition to the presence of another peak at 0.54 hr^{-1} (Figure 2.3). The presence of multiple prominent peaks challenges the previous knowledge of single-frequency oscillations generated solely by the feedback loop between p38 and DUSP1/MKP1.

SB203580 inhibits p38 catalytic activity by binding to the ATP binding pocket. Since it does not affect the phosphorylation state, its use would not affect the FRET signal

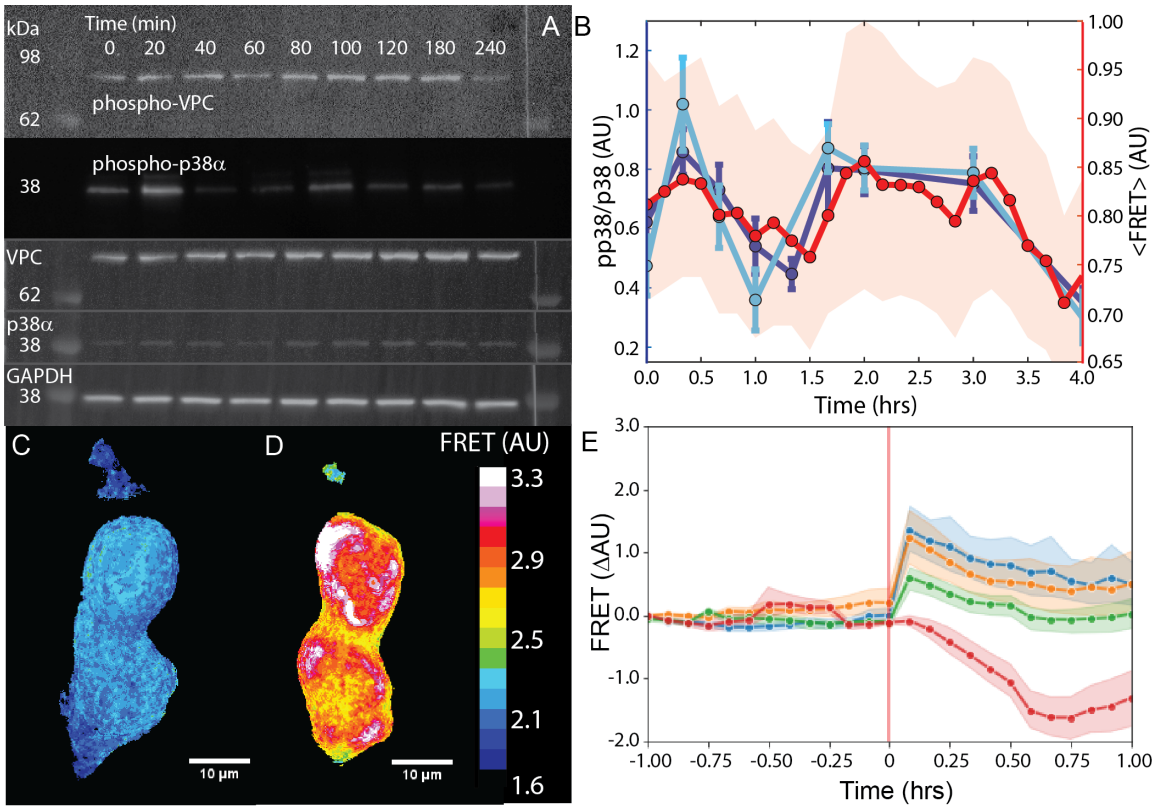

Figure 2.5: **A**) Western blot analysis of p38α and VPC phosphorylation. Example blots as a function of time after addition of 300 nM Il-1β of (from top): phospho-VPC; phospho-p38α; total VPC; total p38α; and GAPDH loading control. **B**) VPC (dark blue, from VPC293 cells) and p38α (light blue, from untransfected HEK-293T cells) phosphorylation dynamics measured using western blot and ensemble-averaged FRET (red, N≈1,500 cells). Shaded region is the 95% confidence bound for FRET data, while error bars are standard deviation of three biological western blot replicates. **C**) FRET state of two resting HEK293 cells expressing VPC (VPC293). **D**) The same two cells, immediately after hyperosmotic shock induced by the addition of 120 mM sorbitol. **E**) Dynamic ensemble averaged FRET response of VPC293 cells upon hyperosmotic (Green: 40 mM, Orange: 80 mM, Blue: 120 mM sorbitol) or hypoosmotic (2x dilution with ultrapure H_2O) shock. Shock was induced at t = 0, indicated by the vertical red line. Points: average over all cells, shaded region indicates 95% confidence bounds.

measured in VPC293 cells. This contrasts with PerKy293 cells, since inhibited catalytic activity of p38 would not allow for resultant substrate activation of PerKy. Figure 2.4 shows that VPC293 cells induced with IL-1β and SB203580 produce a similar multi-frequency spectrum compared to cells induced with only IL-1β. A key difference between the two is a third peak that exists at either 0.45 hr^{-1} or 0.54 hr^{-1}. The cause for this discrepancy is not currently known, but could possibly be attributed to SB203580's function as an inducer of autophagy, which is a function also loosely attributed to the p38 pathway[16][17].

2.2.3 Frequency Response is Unique Among Unique Inducers

p38 is typically activated by several unique stresses when receptor binding at the cell membrane initiates a kinase cascade that results in the activation of p38α. Transient activation occurs for the bulk of stressors, but whether or not some form of oscillatory behavior is imparted was unknown until now.

The SARS-CoV-2 Spike protein is a glycoprotein found on the virus surface that engages entry into host cells by process of receptor-mediated membrane fusion[18]. It is a trimeric class I protein that contains two subunits S1 and S2. Spike protein binds in humans to the ACE2 receptor, which is commonly expressed among tissues including heart, lung, brain, kidney, and vessels. The HEK293 cell line is known to express ACE2[19], and induction of human cells with Spike protein has been observed to activate MAPK pathways including p38 and JNK[20]. Pathway activation similar to that caused by infectious prions could induce neurotoxicity and lead to the development of various neurodegenerative diseases. Furthermore, the Spike protein has broadly been a major target for vaccines and antiviral treatments.

Induction of VPC293 cells with 2 nM of Spike protein reveals a prominent periodicity at 0.45 hr^{-1}, marking a difference from the 0.36 hr^{-1} main peak detected from IL-1β stimulation. The two stresses are qualitatively different and would in principle induce separate physiological responses in target cells, which would make this difference of result feasible. Similarly, sorbitol, a sugar that activates p38 by osmotic stress, produces a prominent periodicity at 0.54 hr^{-1}. The detection of periodicities as a result of these stresses is significant in that previously they were only observed to cause transient activation.

2.2.4 Statistically Significant Peaks Are Organized Into Harmonics

Having induced VPC293 cells with various distinct stresses, we observed commonalities across spectra. In addition to the recurrence of a periodicity at 0.36 hr^{-1}, we multiply observed periodicities at 0.18 hr^{-1} and 0.45 hr^{-1}. This suggests that a cell's innate frequencies of oscillation are fixed and that their relatively even grouping could make them multiples of a fundamental frequency, similar to the findings of Geva-Zatorsky et al.

To test this, we employed an algorithm to identify statistically over represented frequencies among our experiments. Figure 2.7 shows these frequencies in a manner analogous to an emission spectrum. After combining individual spectra into an ensemble we applied a "distance-to-tine" algorithm to determine the statistical significance of the grouping of our over represented frequencies (Section 3.3.2). We therefore determined the data to have a fundamental frequency of 0.09 hr^{-1} (p=0.04), corresponding to a period of $\sim$11 hr.

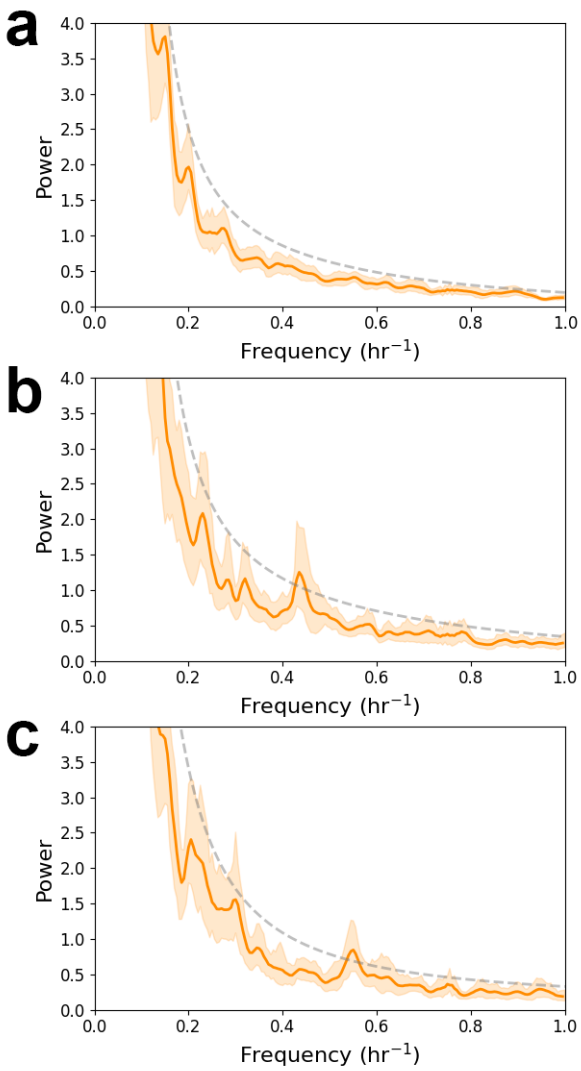

Figure 2.6: Fourier spectra of VPC293 cells. Shaded regions represent 95% confidence bounds and dotted gray lines represent 99% confidence bounds for statistically significant peaks. **a)** Unstimulated cells show no significant periodicities. **b)** Cells induced with 2 nM Spike protein show a prominent periodicity at 0.45 hr^{-1}. **c)** Cells induced with 120 mM sorbitol show a prominent periodicity at 0.55 hr^{-1}.

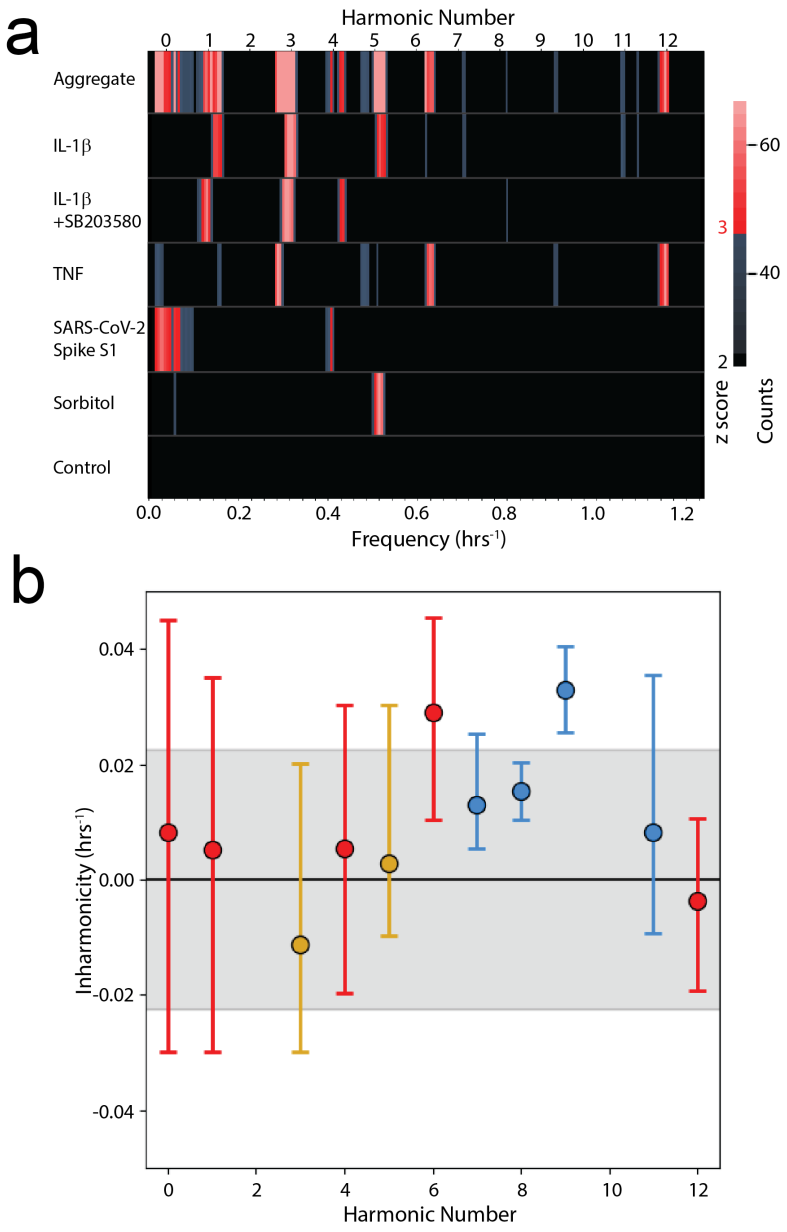

Figure 2.7: Harmonic organization of periodicities in VPC293 cells. **a)** Multiple unique stimulus spectra are combined into an ensemble spectrum. Red represents 3σ peaks, while blue represents 2σ. **b)** Inharmonicity is visualized by distance to closest harmonic. Each point represents a $>2\sigma$ band. Gray shaded region represents a distance range at which frequency groupings are unlikely.

2.2.5 Coupled Oscillations in Signaling Pathways

The p38 MAPK pathway characteristically features a complex cascade of kinases which include the MAP2K kinases MKK3 and MKK6[21], as well as members of the MAP3K family including MEKK4[22], DLK[23], TAK1[24], and ASK1[25]. Conversely, a compliment of phosphatases that can dephosphorylate p38α include members of the MAPK phosphatase (MKP)/dual specificity phosphatase (DUSP) family MKP1/DUSP1, MKP5/DUSP10, MKP8/DUSP26, DUSP8, and DUSP12 [8, 26, 27, 28]. It is not presently known why there are so many kinases upstream from p38α, nor why the cascade is organized into three tiers.

We present the p38 MAPK phosphorylation cascade as a chemical resonator capable of producing multiple frequencies to feed to p38α. This model demonstrates both the purpose of the extensive phosphorylation cascade and its necessity to be as multitudinous as possible to transduce a binary receptor signal into a complex cellular response.

Each oscillator in an activation pathway could represent a degree of freedom in the form of a unique oscillation frequency. Figure 2.8 shows the complexity of the phosphorylation cascades associated with MAPK signaling and still remains an incomplete picture of the system. The path taken to activate p38α is receptor-dependent, suggesting that in principle each spectrum should be unique for each stressor. Branching paths into the cell add additional possibilities and expand the capabilities of cell response. Furthermore, there exist noncanonical routes of p38 activation which include T cell receptor stimulation and intracellular infection[7].

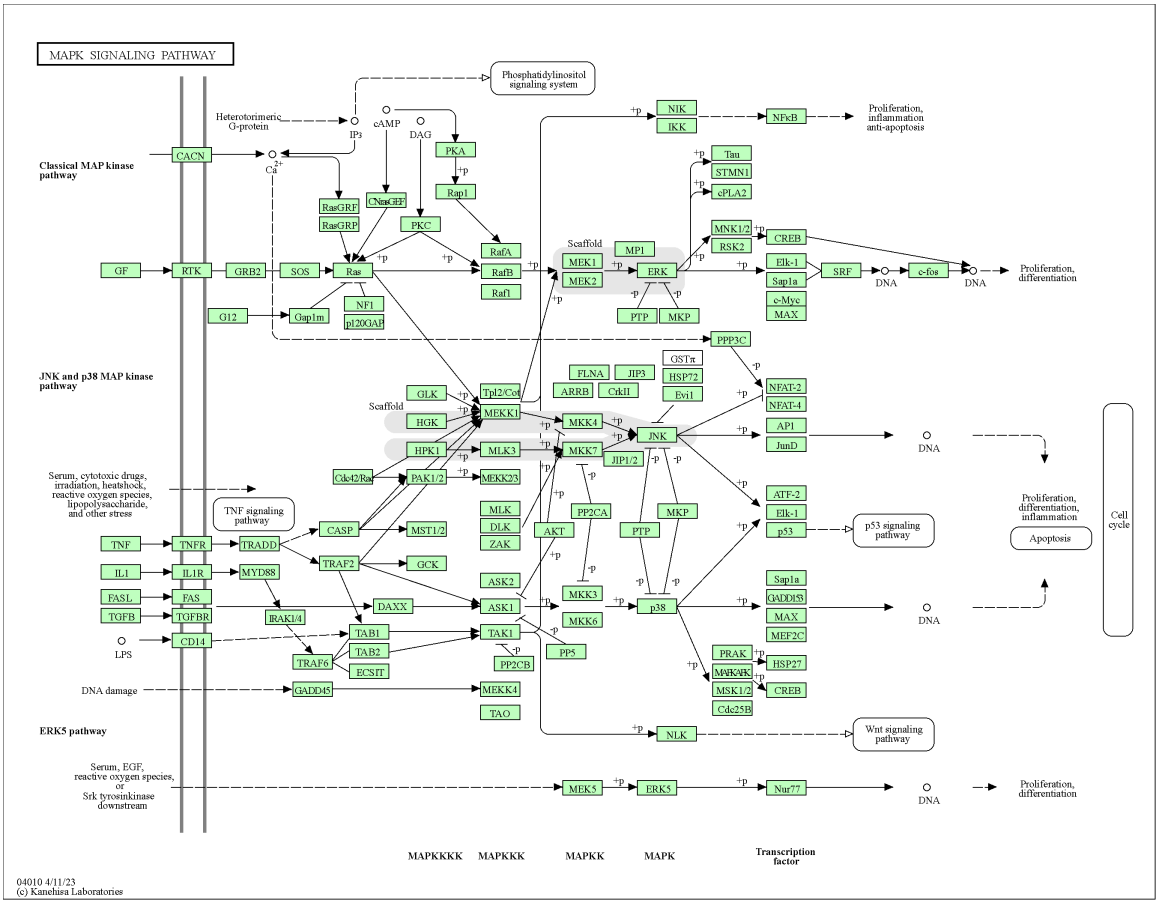

Figure 2.8: Phosphorylation cascade of the MAPK pathways[29]

p38α, whose activation state contains multiple frequencies at the terminus of the phosphorylation cascade, will thereby couple to downstream targets. Ideally these targets would each possess a resonant frequency that allows for efficient gene regulation. Damped oscillators are difficult to drive off-resonance, and therefore the power spectra of cells expressing PerKy would either show a peak at their resonant frequency (0.36 hr^{-1}) or no peaks at all. Figure 2.9 demonstrates that Spike protein induces oscillations in VPC293 cells but not in PerKy293.

Cross Power Spectral Density analysis (Section 3.2.1) between PerKy293 cells and VPC293 cells both stimulated by IL-1β shows a strong peak at the common frequency of 0.36 hr^{-1}(Figure 2.9). Moreover, there exists a characteristic discontinuity in the phase spectrum of the Cross Spectral Density that represents the $\frac{\pi}{2}$ phase shift between the driver and the target when the frequency of the driver is on-resonance with the target. Finally, the same Cross Power Spectral Density analysis for Spike shows nothing noteworthy. These results suggest that PerKy is a substrate that can only be driven at its natural frequency.

We hypothesized that the additional harmonics present in p38 activation state oscillations resonantly couple to other naturally occurring targets with corresponding natural frequencies. Of particular interest is the well-characterized tumor suppressor protein p53, which is a substrate of p38 and which has been shown to oscillate with a natural fundamental frequency centered at 0.14 ± 0.02 hr^{-1} (period T = 7 hrs) in response to stress[10]. Our observations therefore predict that, for example, stimulation of the cell with IL-1β, which strongly activates the 1st harmonic at a similar frequency of 0.18 ± 0.03 hr^{-1}, may resonantly couple to p53's fundamental natural frequency and drive strong resonance in

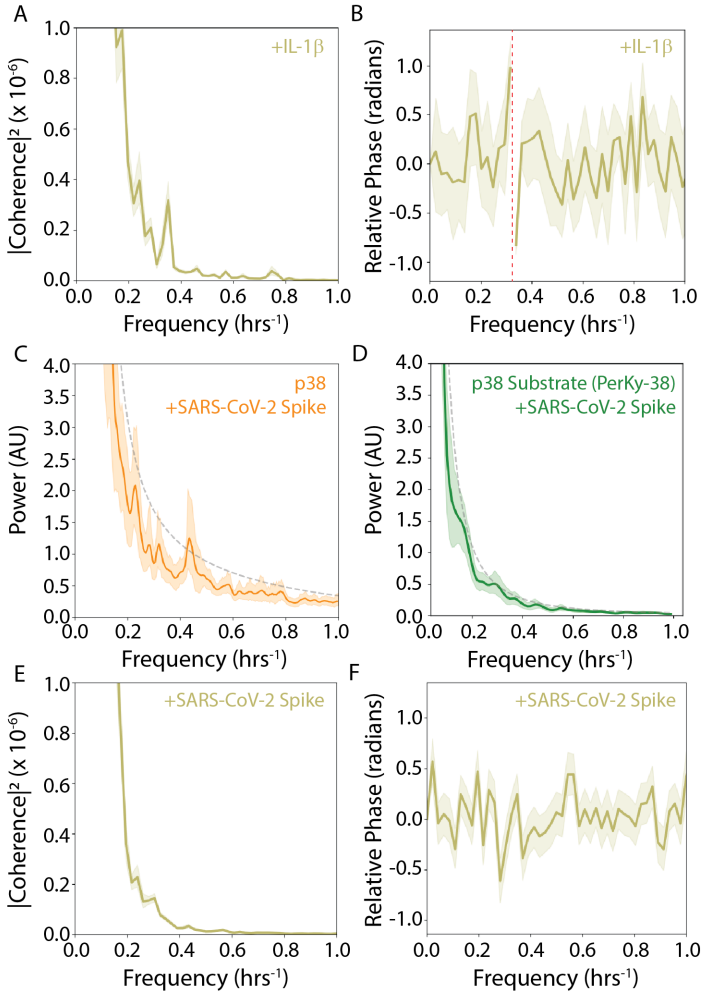

Figure 2.9: VPC and PerKy behave like a forced damped oscillator system. **a)** Magnitude-squared coherence spectrum between VPC293 and PerKy293 cells each stimulated by IL-1β. There is a prominent shared peak at 0.36 hr^{-1}. **b)** Relative phase spectrum calculated from a) shows a discontinuity at 0.36 hr^{-1}. **c)** Fourier spectrum of VPC293 cells stimulated with SARS-CoV-2 Spike protein. **d)** Fourier spectrum of PerKy293 cells stimulated with SARS-CoV-2 Spike protein. **e)** Magnitude-squared coherence spectrum between **c)** and **d)** displays no significant peaks, consistent with an oscillator being driven off-resonance. **f)** Relative phase spectrum displays no discontinuities.

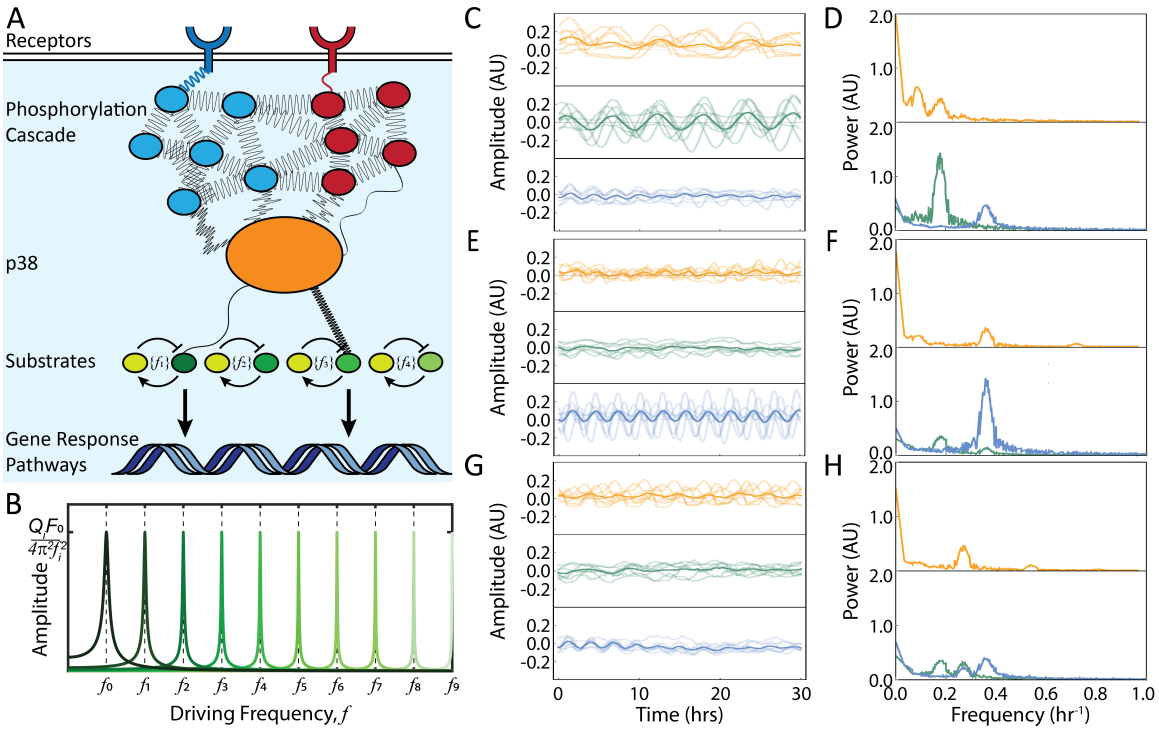

Figure 2.10: (**A**) Stimuli activate their receptors, which engage and activate the connected components of the phosphorylation cascade (red and blue ovals). The cascade, together with p38 (orange), generate a stimulus-specific collection of harmonics through yet unclear mechanisms. The harmonic frequncies with which p38 oscillates can resonantly drive oscillations of its substrates (green shades) appropriate to the received stimulus, which oscillate at some natural frequency or set of frequencies (f_n) determined by their interactions with other components (yellow), *e.g.*, the transcription factor (TF) substrate p53 and its inhibitor Mdm2[10]. Once resonantly activated, TFs can drive expression of appropriate stimulus-specific response pathways. (**B**) Selective amplification of susbtrate oscillations by resonance selectivity. Example amplitude response functions of ten substrates with evenly spaced natural frequencies f_i (green shades) as a function of the frequency of the driving oscillator, f. Changes of the driving frequency result in selection and amplification of oscillations of specific substrates when $f \approx f_i$, with the width and height of the peak characterized by the Quality factor, Q, of the system. Here the Q factor of each substrate is adjusted to yield equal amplitudes for simplicity. (**C-H**) Example FRET time traces and ensemble averaged Fourier spectra for simulations of p38 resonantly coupling to two substrates with natural frequencies $f_A = 0.18$ hr^{-1} (green) and $f_B = 0.36$ hr^{-1} (blue) for the cases: (C,D) on resonance with Substrate A; (E,F): on resonance with Substrate B; (G,H) off-resonance with both Substrates A and B. See Section 2.4.5 for details of simulations.

p53 activation. Conversely, stimuli such as SARS-CoV-2 Spike or hyperosmotic shock with sorbitol, which do not activate the 1st harmonic, should not strongly couple with p53 and therefore drive no or weak p53 response. These expectations are consistent with observations from the literature[30, 31, 32], and we tested these predictions by western blot of lysate from HEK-293T cells stimulated with Il-1β or SARS-CoV-2 Spike probing the expression and p38-mediated Ser46 phosphorylation[33] of p53. Consistent with the predictions of this model, we find that Il-1β stimulates $\sim$100x phosphorylation of p53 (Figure 2.11A-B), which increases over long-time scales. Conversely, stimulation of cells with SARS-CoV-2 Spike results in a shorter time scale and smaller amplitude $\sim$7x phosphorylation of p53, which decays back to baseline over four hours.

2.3 Discussion

The phenomenon of oscillations in gene expression and gene regulatory network components and the requirement of oscillations for appropriate gene regulation and expression has been thoroughly documented, however the mechanisms through which these oscillations are put to use or are interpreted by cells remains poorly understood. Here, we demonstrate that the human central stress response regulator, p38α, exhibits complex oscillations, and that these oscillations are used to couple to and drive downstream targets through forced resonant oscillations. Moreover, we demonstrate that different stimuli result in differential excitation of multiple harmonics. These different harmonics can be used to selectively couple to downstream pathways to drive responses appropriate to the experienced stimulus. The time dependence of the appearance of harmonics can be complex

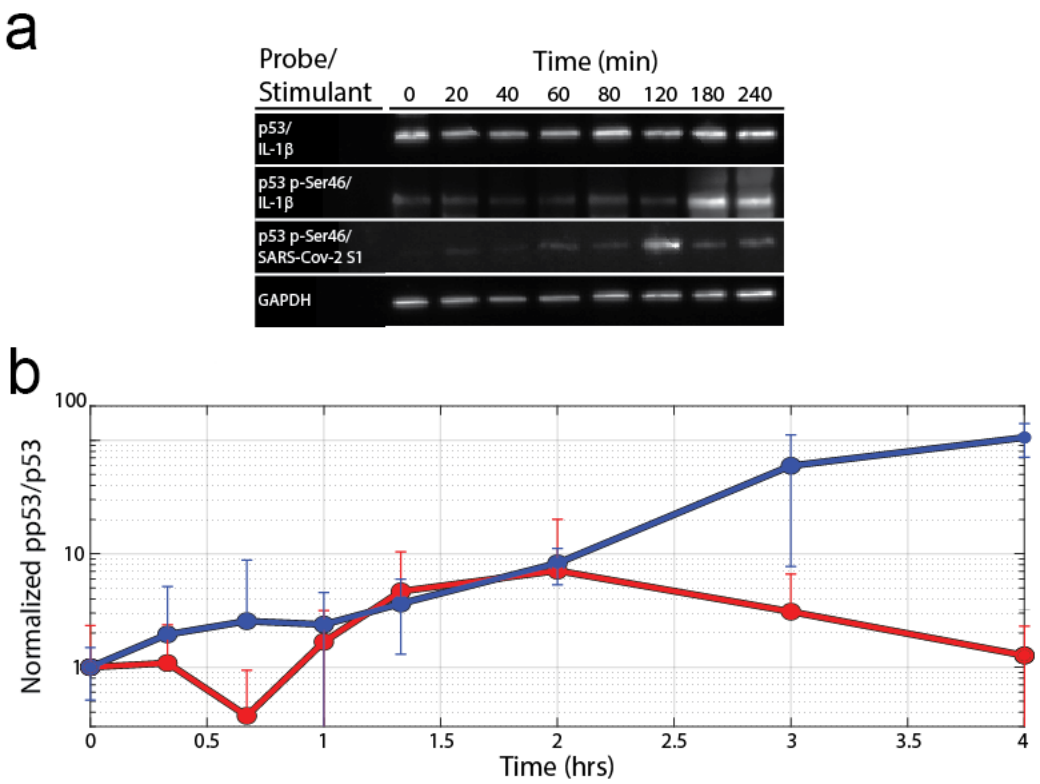

Figure 2.11: **a**) Western blots probing stimulated HEK-293T p53 expression and phosphorylation as a function of time. From top to bottom: (1) total p53 when stimulated with 300 ng/mL IL-1β, SARS-CoV-2 results are similar; (2) p53 phospho-Ser46 + 300 ng/mL IL-1β; (3) p53 phospho-Ser46 +2 nM SARS-CoV-2 Spike S1; (4) GAPDH loading control. **b**) Quantification of p53 western blots. Color indicates stimulant; Blue: 300 ng/mL IL-1β; red: 2 nM SARS-CoV-2 Spike S1.

(Figure 3.10), allowing for temporal coordination of gene response. Here, however, we focus on describing the time-average frequency response, leaving detailed analysis of temporal dependence for future investigation.

It remains unclear what mechanisms determine the fundamental frequency and generate the observed harmonics with which the activation state of p38α/VPC oscillates. However, such harmonics could arise from only a limited number of fundamental physical mechanisms. One possibility, as has previously been suggested for the p53 system, is that p38 behaves as an intrinsically nonlinear oscillator, or that the higher harmonics are induced by noise through nonlinear interactions of p38 with other cellular components[34]. Alternatively, multiple components of the MAPK cascade have been observed to exhibit oscillations in number, localization, or activation state, including ERK[35], NF-κB[36], p53[10], and more than 5000 genes within or related to the MAPK cascade when induced by TNF or RANKL[11, 37]. These cascade components can interact to form a network of coupled oscillators whose normal modes manifest as harmonics. Such a model also suggests a straightforward way for the system to differentially generate harmonics, where, in analogy to natural harmonics of a stringed instrument, inhibitors or phisphatases can impose oscillatory nodes on the system corresponding to specific harmonics. However, the spacing of the harmonics of such a system is a function of the connectivity, dimensionality, and coupling strengths and frequencies of the oscillators, and in general would require fine tuning to generate evenly spaced harmonics as observed here. Other potential mechanisms include nonlinearities in upstream components of the phosphorylation cascade, the cascade functioning as a collection of partially phase-locked Kuramoto oscillators with random, global

couplings[38], or systems of oscillators including time delays[39], all of which can also, under certain circumstances, generate evenly spaced harmonics, although these approaches also require some degree of fine tuning.

The use of resonant frequencies to drive specific responses is a control scheme frequently employed in electronic circuits referred to as resonance selectivity. A radio or television set receives signals via over-the-air electromagnetic radiation at specific frequencies. The desired frequency signal can then be chosen by tuning the capacitance, and hence the oscillatory frequency, of an LC circuit within the receiver to be in resonance with the frequency of the desired station. This results in large amplitude forced oscillations in the receiving circuit that are then decoded by the receiver as audio or visual information. We hypothesize that the p38α MAPK system, and other oscillatory systems in the cell, function in an analogous way, allowing p38 to encode and transmit far more information than would otherwise be possible in a simple two state system. Here, each pathway for responding to a specific stress would be represented by a protein or collection of proteins (e.g., the p53/Mdm2 or MEF2A/PerKy-38/PP2B systems) that naturally oscillate at a particular frequency or set of frequencies. Stimulation of the MAPK phosphorylation cascade by receptors for specific signals activate the cascade to serve as a resonator generating certain stress-specific harmonic frequencies. The resonator cascade drives oscillations in p38α MAPK at these specific harmonic frequencies, which then resonantly couple to the downstream pathways appropriate for the experienced stress.

The use of resonant oscillatory frequencies to select and drive specific responses is a control scheme referred to as resonance selectivity, frequently employed in electronic circuits

such as radio receivers. A desired frequency signal can be sleected from the multitude of transmitted signals by tuning the capacitance, and hence the oscillatory frequency, of an inductor-capacitor(LC) circuit within the receiver to be in resonance with the frequency of the desired signal. This results in large amplitude forced oscillations in the receiving circuit that are decoded by the receiver as audio or visual information. We hypothesize that the p38α MAPK system, and other such oscillatory systems in the cell, function in an analogous way, allowing p38 to encode and transmit far more information than would otherwise be possible in a simple two state system. Here, each pathway for responding to a specific stress would be represented by a protein or collection of proteins (*e.g.*, the p53/Mdm2 or MEF2A/c-Jun/PerKy-38 systems) that naturally oscillate at a particular frequency or set of frequencies. Stimulation of the MAPK phosphorylation cascade by receptors for specific signals activate the cascade and p38 MAPK to generate a spectrum of stress-specific harmonic frequencies, which can resonantly couple to and drive pathways appropriate for the experienced stress. This resonance selectivity model also suggests the hypothesis that some of the many disease states associated with p38 may be a consequence of inappropriate function of this mechanism, or mutations in component resulting in altered oscillatory frequencies that no longer couple to and resonate appropriately with p38, opening the possibility that such disease states could be rectified by externally driving p38 or its substrates at appropriate frequencies.

The employment of dual-labeled FRET reporters to detail activation states has potential to be a useful tool in the study of MAPK pathway oscillations. Conducting experiments at timescales longer than the doubling time of cells allows for Fourier analysis at

resolutions adequate enough to detect multiple periodicities. In principle, similarly labeled reporters can be constructed for the other major pathways such as ERKs and JNKs, of which cross-talk with p38 has been reported in the regulation of pro-inflammatory factors[40].

Tunable inhibition of p38α activity could provide further hypothesis testing for resonance selectivity. Use of an optogenetic JNK inhibitor constructed from JIP1 and LOV2Jα has shown that JNK is maximally inhibited when stimulated by a blue LED at two unique pulse frequencies[41]. Multiple experiments could be performed driving cells at fixed unique frequencies followed by qPCR to identify and quantify genes that are being expressed as a result. Cataloging of said genes based on frequency bins would allow us to decipher cellular function.

Alternatively, similar results could be derived by driving targets periodically with media containing stressor through use of a pump system[9]. Monitoring of p38α activation in tandem with a frequency sweep of induction/inhibition could then be used to identify characteristic frequencies and verify the results of prior experiments.

Our experiments have yielded relatively low signal-to-noise ratio (SNR) when comparing the amplitude of oscillations to the random fluctuations of our decoupled FRET sensor VP+PC. Higher SNR would open doors for analysis that utilize X-V phase plots (Figure 2.12), which visualize the characteristic qualities of an oscillator system, i.e. if they're linear or nonlinear. Unfortunately, the only way to do so would be through employing a more sensitive biosensor and/or resolving images at higher resolutions to track domains within each cell. FLINC biosensors could fill exactly that role, both with their sensitivity and their capacity for super resolution imaging[42].

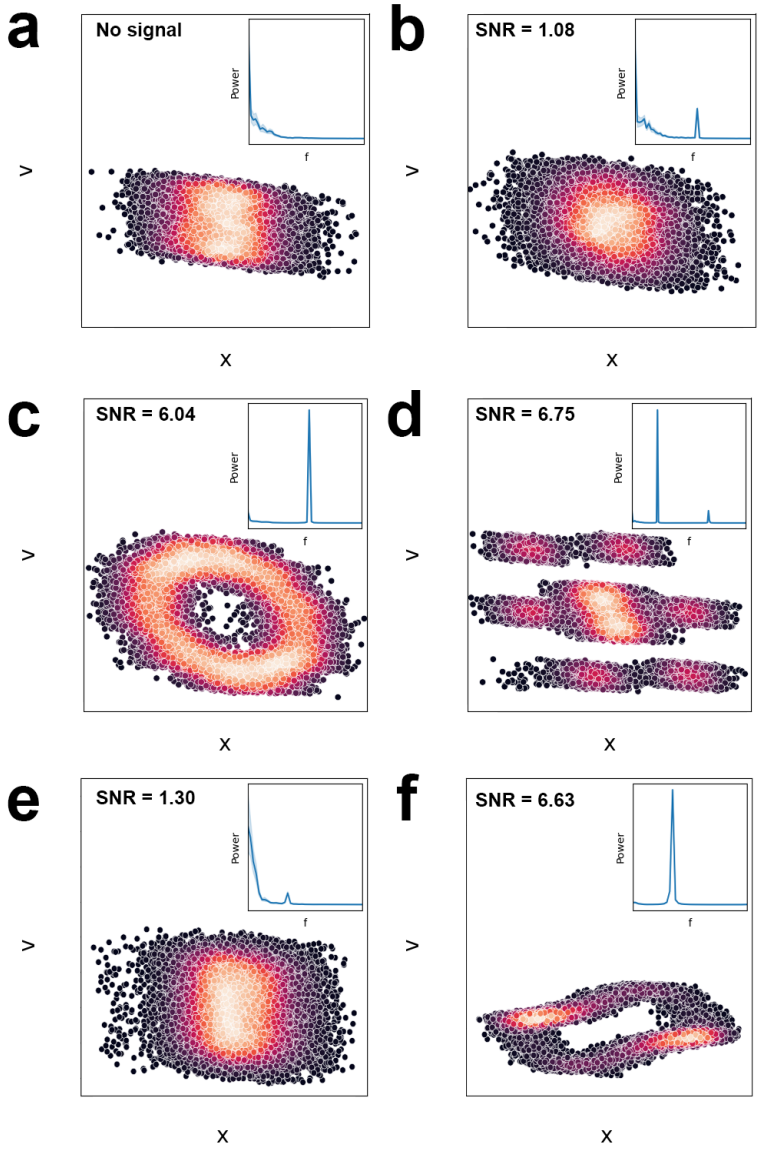

Figure 2.12: X-V phase diagrams that visualize the velocity (first time derivative) of a Brownian oscillator vs. its displacement. Each datum represents a time point for a simulated parametrically-fixed Brownian process. **Inset**: Ensemble-averaged Fourier transform for each simulation (N=50). **a)** Brownian noise with no added periodic signal. **b)** Weak periodic signal of frequency f_0 and SNR 1.08 displays little qualitative difference compared to **a**). **c**) Strong periodic signal of frequency f_0 and SNR 6.04 shows the X-V plot take on a characteristic elliptical shape. **d**) Strong periodic signal of frequencies f_0 and $3f_0$ and SNR 6.75 show a unique complex pattern. **e)** Weak periodic signal of frequency f_0 and SNR 1.30 of a nonlinear oscillator (Van der Pol) is qualitatively similar to **a)**. **f)** Strong periodic signal of frequency f_0 and SNR 6.63 of a Van der Pol oscillator.

The p38 pathway is atypically activated in microglial cells in response to HIV gp-120 proteins, leading to the overproduction of TNF-α and resultant neurotoxicity[43], a possible major contributor to the development of HIV-1 Associated Neurocognitive Disorders. A peak in p38 activation followed by a decline and subsequent peak 21 hours later could suggest that gp-120 induces periodic p38 activation. Furthermore, stable transfection of microglial cells such as HMC-3 with PerKy and VPC reporters could further elucidate the exact response. With the endemic status of Covid-19 and widespread detection of Postacute Neurologic Sequelae, long-term affects associated with infection by SARS-CoV-2 may further come to prominence[44].

2.4 Materials and Methods

2.4.1 Plasmids and Construction

The plasmid pB510B-1-PerKy-38 was the kind gift of Dr. Taichiro Tomida [9] . The coding sequence for VPC, expressed from a CMV promoter, was designed using VectorNTI software (Thermo Fisher Scientific) and synthesized de novo and cloned into the high copy number pUC57-kan (GENEWIZ/Azenta). VPC was then subcloned into the expression vector pB510B-1 by digestion of both pUC57-kan-VPC and pB510B-1 (System Biosciences) with NheI-HF and BamHI-HF (New England Biolabs), gel purification of the VPC brand from pUC57-kan (Qiagen QIAquick gel extraction kit) and ligated into digested pB510B-1. The resulting plasmid, pB510B-1-VPC, was verified by nanopore sequencing (Plasmidsaurus).

2.4.2 Cell Culturing

HEK-293T cells stably transfected with PerKy-38 or VPC were generated by PiggyBac Transposon Vector System (System Biosciences) according to the manufacturer's directions. Briefly, HEK-293T cells were doubly transfected with the pB510B-1 vectors along with PiggyBac Transposase vector pB210PA-1. Cells were allowed to recover for 48 hours, after which puromycin was added (5 µg/ml) and cultures were expanded. After 14 days, surviving cells lines exhibiting blue and yellow fluorescence were collected, and the sequence of the inserted cassette verified by PCR and sequencing.Transient transfection of HEK-293T cells with PerKy-38 or VPC was performed using FuGENE HD Transfection Reagent at a ratio of 3:1.

HEK-293T cells were cultured in complete media containing 90% DMEM with high glucose (Gibco), 10% heat-inactivated FBS (Gibco), 200 mM L-glutamine (100X), 10 mM MEM nonessential amino acids (100X), 100 mM MEM sodium pyruvate (100X), and concentrated puromycin (Gibco). Seeding was done at 30% confluency in 35 mm glass cover slip dishes (Cellvis) and cells were incubated at 37°C and 5% CO 2 for 24 hours prior to imaging.

SARS CoV-2 Spike protein was purified in house using immobilized metal ion affinity chromatography (IMAC). IL- 1β was purchased from Invitrogen (Waltham, MA). Sorbitol was purchased from Fisher Scientific. Kinase inhibitor SB203580 was purchased from Selleck Chemicals (Houston, TX).

2.4.3 Western Blot Analysis

For western blot analysis, cells were grown and treated as described above in 6 well plates (Corning Costar). Cells were washed with PBS and lysed with 100 μL ice cold Lysis buffer [1X RIPA buffer (Cell Signaling Technologies) + cOmplete Protease Inhibitor Cocktail (Roche) + Phosphatase Inhibitor Cocktail (EMD Millipore)]. This mixture was pulse sonicated to ensure complete lysis and then centrifuged at 21k g for 15 minutes in a refrigerated centrifuge, and subsequently stored at -80°C for use.

For western blot, the total protein concentration was determiend using the bicinchoninic acid (BCA) protein assay kit (Pierce/Thermo Fisher Scientific; Rockford, IL). 25 μg of protein samples were mixed with 4x LDS sample buffer (Life Technologies, CA, USA) and 10x reducing agents (Invitrogen; Carlsbad, CA) and heated for 5 min at 100°C. Samples were loaded in a 4-12% SDS-PAGE gel (Mini-PROTEAN TGX, Bio-Rad) for electrophoretic separation and subsequently electro-transferred to nitrocellulose membrane (Bio-Rad). Membranes were blocked with 5% bovine serum albumin (BSA; phospho blots; RPI Corp; 830075) or 5% milk (dry powdered milk; non-phospho blots; RPI Corp; M17200) in 1x TBST for 5 min each, and afterward incubated overnight at 4°C in primary antibodies. Following primary antibodies (rabbit) were used: p38 (Cell Signaling Technologies; 9212S; 1:1000 1x TBST + 5% milk powder), phospho-p38 (ProSci 9169; 1:1000 1x TBST + 5% BSA), p53 (Thermo Fisher, MA5-14516; 1:1000 1x TBST + 5% milk powder), phospho-p53 pSer46 (Proteintech; 28960-1-AP; 1:1000 1x TBST + 5% BSA), and GAPDH (Cell Signaling Technologies; 2118S; 1:1000 1x TBST + 5% milk powder). Membranes were then washed with 1x TBST (3 times for 15 minutes) and incubated with goat anti-rabbit-HRP

(Invitrogen; 31460; 1:2000 1x TBST 1% BSA or milk powder as appropriate) secondary antibody. Blots were then washed 3x with 1x TBST for 5 min each. Blots were then stained by exposure to SuperSignal West Dura Extended Duration Substrate (Thermo Fisher).

Membranes were imaged with a Bio-Rad ChemiDoc system and images were quantified using Fiji/ImageJ[45]. Values reported are the means of three biological replicates with error bars indicating the standard deviation. For those values reported as ratios, *e.g.*, pp53/p53, values and error bars were calculated as

$$z = \frac{\langle x \rangle}{\langle y \rangle} \pm \sigma_z \tag{2.1}$$

$$\sigma_z = \sqrt{\sum_i \left(\frac{\partial z}{\partial x_i} dx_i \right)^2} = \sqrt{\left(\frac{1}{\langle y \rangle} \sigma_x \right)^2 + \left(\frac{\langle x \rangle}{\langle y \rangle^2} \sigma_y \right)^2}, \tag{2.2}$$

where $\langle x_i \rangle$ and σ_{x_i} are the means and standard deviations of measurements taken for each component respectively.

2.4.4 Sample Microscopy and Analysis

Cells were imaged using a Nikon Ti-E Inverted Microscope and cultured in an Ibidi Silver Line stage top incubator at 37°C, 5% CO 2 , and 95% relative humidity. Fluorescent proteins were excited with a Nikon Intensilight epifluorescent illuminator, and both FRET reporters were imaged simultaneously using a Photometrics Dual View with YFP/CFP filter set on an Andor iXon DU897 Ultra camera. Separate fluorescent images were combined into FRET images using ImageJ[45] and CellProfiler[46] . Analysis of FRET images was performed using a custom Python script.

2.4.5 Stochastic Simulations

For simulations of p38 activity, we use the set of nonlinear differential equations describing p38 (MAPK) phosphorylation state developed by Tomida *et al.*[9],

$$\begin{aligned}
\frac{d[\text{MAP2K}]}{dt} &= k_0 S[t](1 - [\text{MAP2K}]) - k_1[\text{MAP2K}] \\
\frac{d[\text{MAPK}]}{dt} &= k_2[\text{MAP2K}](1 - [\text{MAPK}]) - k_3[\text{MKP} - 1]\frac{[\text{MAPK}]}{k_4 + [\text{MAPK}]} \\
\frac{d[\text{MKP1RNA}]}{dt} &= k_5[\text{MAPK}](1 - [\text{MKP1RNA}]) - k_6[\text{MKP1RNA}] \\
\frac{d[\text{MKP1}]}{dt} &= k_7[\text{MKP1RNA}](1 - [\text{MKP1}]) - k_8[\text{MKP1}],
\end{aligned} \tag{2.3}$$

with the addition of Brownian noise. Parameter values are as given in [9] scaled by a factor of 0.1175 to yield the experimentally observed fundamental frequency of 0.09 hr^{-1}. Note that the first equation for the time evolution of [MAP2K] is decoupled from and can be solved independently of the others. Tomida *et al.* used *S[t]* = 1 when p38 is activated by MAP2K and *S[t]* = 0 otherwise. Instead, we treat [MAP2K] as a driving oscillator representing p38's interactions with the phosphorylation cascade in the equation for the time evolution of [MAPK] (p38), setting [MAP2K] = $A[1 + B\sin(2\pi\text{ft})]$ with $A = 0.2857$ and B a random number drawn from a uniform distribution [0,1]. To generate the simulated FRET time traces and ensemble averaged Fourier spectra shown in Figure 2.10C-H, we solve the above equations using the 4^{th} order Runge-Kutta method with Figure 2.10C,D: f = 0.09 hr^{-1}; Figure 2.10E,F: f = 0.36 hr^{-1}; Figure 2.10G,H: f = 0.27 hr^{-1} to yield a solution $[\text{MAPK}_f(\text{t})]$.

To simulate resonance selectivity, we model two substrates, A and B, as driven two component linear oscillators with repressors R_i:

$$\begin{bmatrix} \dot{\mathrm{A}} \\ \dot{\mathrm{R}}_{\mathrm{A}} \end{bmatrix} = \begin{bmatrix} -\beta_{\mathrm{A1}} & \alpha_{\mathrm{A1}} \\ \alpha_{\mathrm{A2}} & -\beta_{\mathrm{A2}} \end{bmatrix} \begin{bmatrix} \mathrm{A} \\ \mathrm{R}_{\mathrm{A}} \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} [\mathrm{MAPK}_f(t)]$$

$$\begin{bmatrix} \dot{\mathrm{B}} \\ \dot{\mathrm{R}}_{\mathrm{B}} \end{bmatrix} = \begin{bmatrix} -\beta_{\mathrm{B1}} & \alpha_{\mathrm{B1}} \\ \alpha_{\mathrm{B2}} & -\beta_{\mathrm{B2}} \end{bmatrix} \begin{bmatrix} \mathrm{B} \\ \mathrm{R}_{\mathrm{B}} \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} [\mathrm{MAPK}_f(t)]. \tag{2.4}$$

We set $\zeta_2 = 0$ and use

$$\begin{aligned} &\alpha_{\mathrm{A1}} = 1.6 \\ &\alpha_{\mathrm{A2}} = 0.8 \\ &\beta_{\mathrm{A1}} = \beta_{\mathrm{A2}} = 1/13 \\ &f_{\mathrm{A0}} = \frac{\sqrt{\alpha_{\mathrm{A1}}\alpha_{\mathrm{A2}} + \beta_{\mathrm{A1}}\beta_{\mathrm{A2}}}}{2\pi} = 0.18 \text{ hr}^{-1} \\ &Q_{\mathrm{A}} = \frac{\sqrt{\alpha_{\mathrm{A1}}\alpha_{\mathrm{A2}} + \beta_{\mathrm{A1}}\beta_{\mathrm{A2}}}}{\beta_{\mathrm{A1}} + \beta_{\mathrm{A2}}} = 1.17 \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} &\alpha_{\mathrm{B1}} = 1.6 \\ &\alpha_{\mathrm{B2}} = 0.8 \\ &\beta_{\mathrm{B1}} = \beta_{\mathrm{B2}} = 1/13 \\ &f_{\mathrm{B0}} = \frac{\sqrt{\alpha_{\mathrm{B1}}\alpha_{\mathrm{B2}} + \beta_{\mathrm{B1}}\beta_{\mathrm{B2}}}}{2\pi} = 0.36 \text{ hr}^{-1} \\ &Q_{\mathrm{B}} = \frac{\sqrt{\alpha_{\mathrm{B1}}\alpha_{\mathrm{B2}} + \beta_{\mathrm{B1}}\beta_{\mathrm{B2}}}}{\beta_{\mathrm{B1}} + \beta_{\mathrm{B2}}} = 2.34. \end{aligned} \tag{2.6}$$

Chapter 3

Oscillations in the p38 MAPK Pathway - Supplemental Information

3.1 Förster Resonance Energy Transfer Sensors in Biology

3.1.1 Förster Resonance Energy Transfer

Phosphorylation of a protein is the process by which a covalently bound phosphate group is attached at regions known as phosphorylation sites. This alters the conformational state of the protein, typically affecting both its activity in the cell and structure[47]. A kinase such as p38α is activated by dual phosphorylation at Threonine and Tyrosine sites and subsequently is more energetically capable of phosphorylating other proteins.

Fluorescent proteins (FPs) as they're employed in experimental biology are genetically encodeable tags that can be excited by an incident photon and subsequently emit a photon of lesser energy[48]. Particular FP pairs in close proximity (1-10 nm) have a probability to undergo nonradiative dipole-dipole coupling when the more energetic fluorophore is excited. In this process, the more energetic FP acts as a donor and transfers its energy to an acceptor, thereby affecting the brightness of both FPs[49]. The spatial dependence for the probability of such a transition occurring is characterized by

$$E = \frac{1}{1 + (\frac{r}{R_0})^6}, \tag{3.1}$$

where R_0 is the distance at which E = 0.5. This makes the transfer efficiency highly sensitive to small positional changes. Typically, single proteins will be tagged with two distinct FPs as a means of tracking conformational changes, or individual proteins will be tagged with with one unique FP each as a means of tracking protein-protein interactions. VPC is a Förster Resonance Energy Transfer (FRET) sensor that employs the former method, where energy transfer between mVenus and mCerulean labels denotes catalytic activation by phosphorylation. Figure 3.1 shows the overlap between the mCerulean emission spectrum and mVenus excitation spectrum, detailing the need for spectral overlap when designing FRET sensors.

3.1.2 Properties of FRET Reporters PerKy-38 and VPC

"PerKy-38" is a FRET reporter that was developed from: YPET, super-enhanced CFP, the p38 docking domain of MEF2A, the c-Jun MAPK phosphorylation region ex-

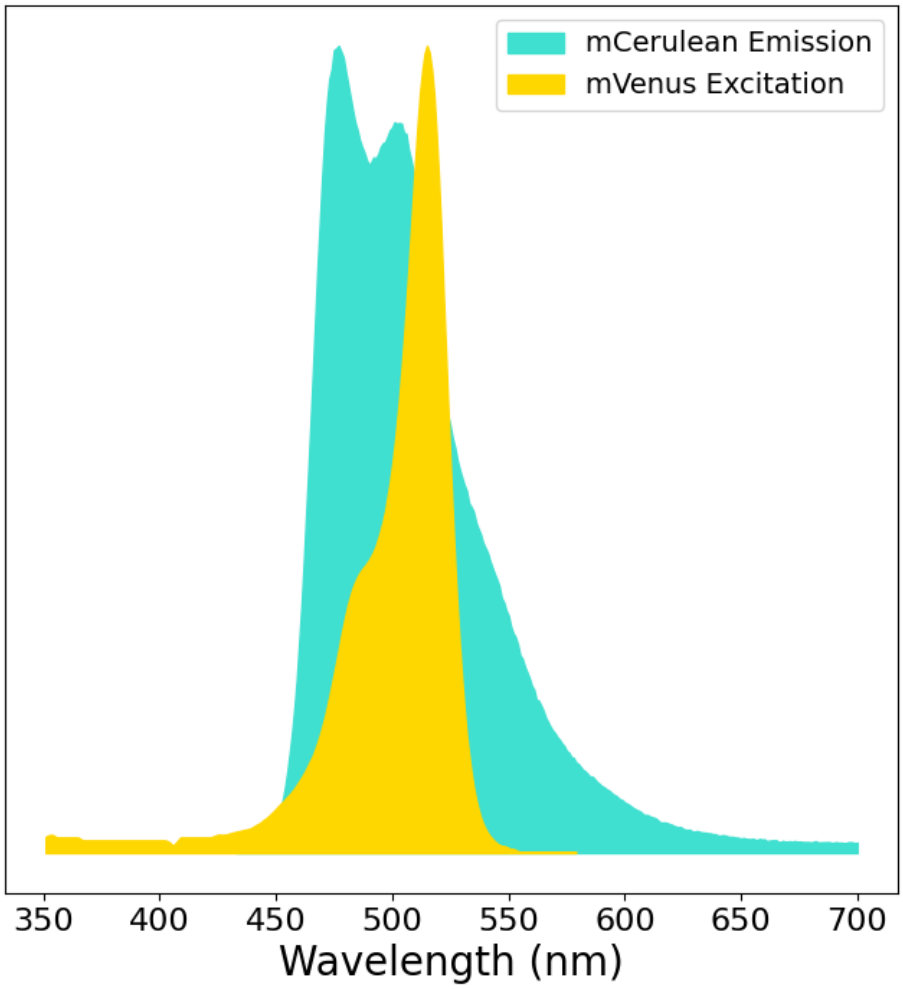

Figure 3.1: The excitation spectrum for mVenus is broadly contained by the emission spectrum of mCerulean.

cluding the JNK docking site, and the nuclear export signal of MEK1[9]. In principle, the docking site of p38 excludes phosphorylation by JNK and dephosphorylation by its associated downregulating proteins. However, p38 does have multiple upstream kinases and phosphatases that could complicate the regulation of PerKy[41][50].

"mVenus-p38-mCerulean" is a FRET reporter that we developed from p38α that has been translationally fused at the N and C terminals with mVenus and mCerulean3 respectively. Catalytically active FRET reporters can alter the health of a host cell, and so substrates are typically employed as reporters. HEK293T cells stably transfected with VPC plasmid via PiggyBac transposon are capable of being cultured indefinitely with a flat FRET signal, meaning that VPC does not pose a significant challenge to cell viability. Conversely, HEK293T cells transiently transfected with VPC plasmid via FuGENE HD, a non-liposomal transfection reagent, undergo a runaway increase in FRET signal inevitably leading to cell death.

Dynamic activation of p38 has been widely recorded across several cell types and unique stressors[51][9][52]. Sorbitol, among the earliest recorded activators of p38, induces hyperosmotic stress in human cells and activates the Fas receptor, which mediates programmed cell death (Apoptosis)[53]. Figure 3.2 shows the rapid dynamic activation of p38 in HEK293 cells transfected with both PerKy and VPC in response to sorbitol. Hypotonic stress induced by dilution of cell media with water causes a gradual drop in FRET signal followed by a steady rise above basal levels. Why the change in activation state is seemingly opposite for oppositely oriented changes in osmotic pressure is not known.

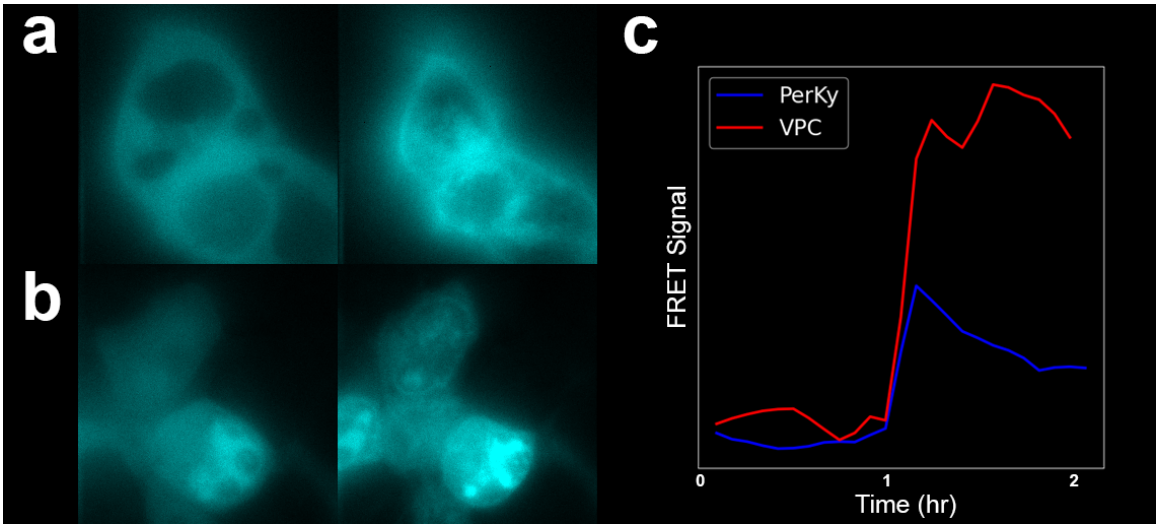

Figure 3.2: Cells transiently activated by 120 mM sorbitol. **a)** Fluorescent images of cells transfected with PerKy before (left) and after (right) induction by 120 mM sorbitol. **b)** Fluorescent images of cells transfected with VPC before and after induction by 120 mM sorbitol. Osmotic shock induces a unique p38 nuclear localization pattern. **c)** FRET signal of both regimes show a comparable rapid response to osmotic stress.

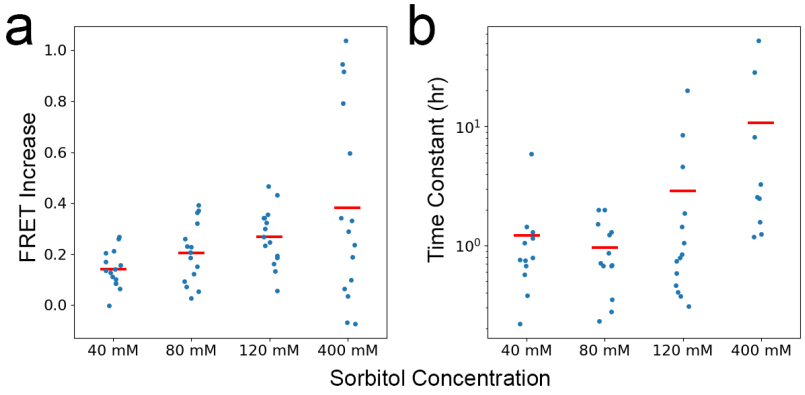

Figure 3.3: Response of HEK293 cells transfected with VPC in response to osmotic stress. Hypertonic stress was induced with the addition of sorbitol at varying concentrations at t = 1 hr. **a**) FRET increase of individual FOVs as a function of sorbitol concentration. **b**) Time constant of VPC relaxation following dynamic activation.

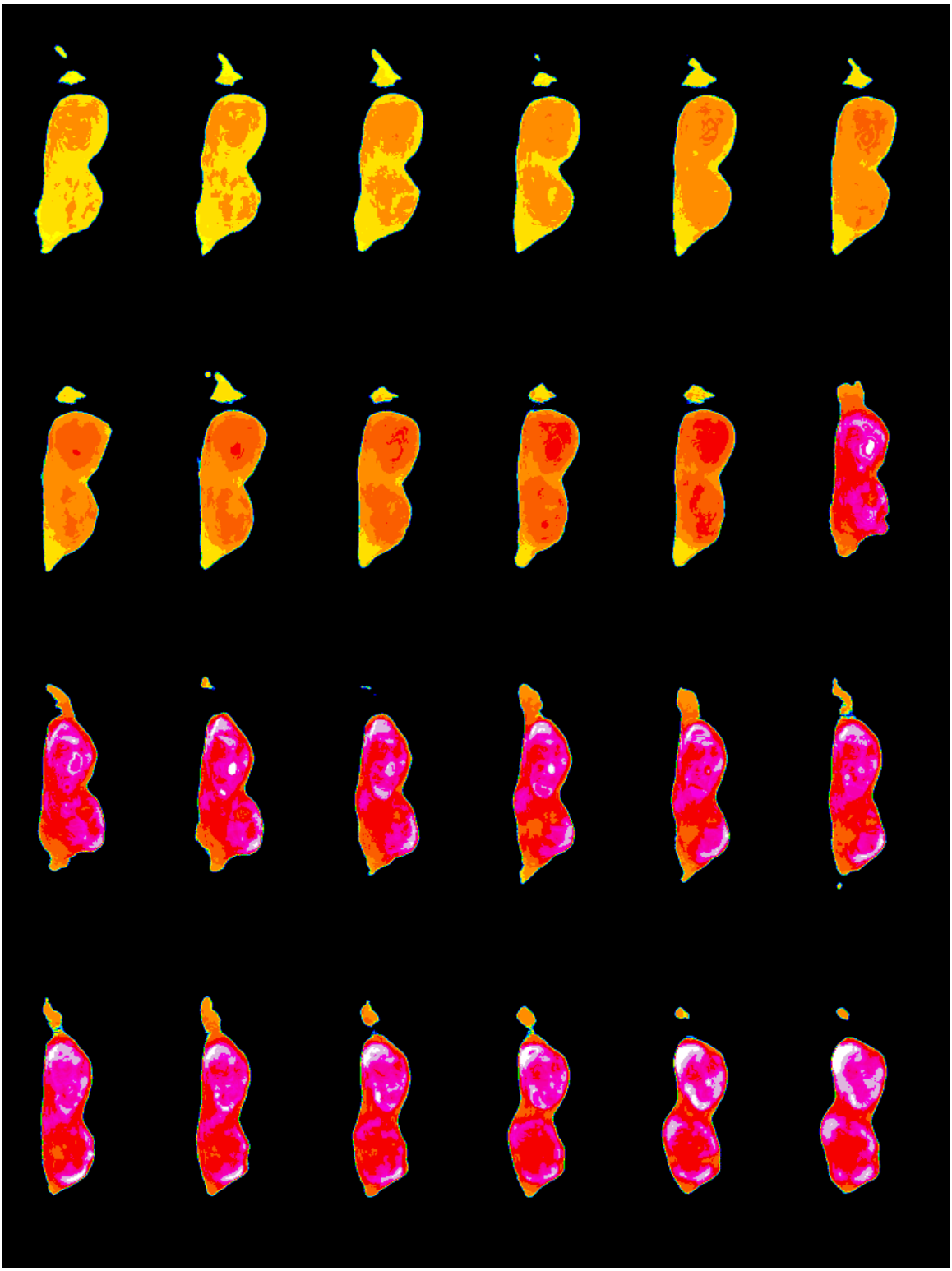

Figure 3.4: Timelapse of HEK293T cells transfected with VPC following induction by 120 mM Sorbitol. Induction occurs between image 11 and 12.

3.2 Fourier Analysis

We start with the FRET signal from the cells as a continuous function of time, $x(t)$. Our measurements are taken at discrete times with sampling period T (typically $T =$ 10 min), corresponding to a sampling frequency $f_s = \frac{1}{T}$. This yields a discrete sequence of N FRET measurements $x_n = \{x_0, x_1, ..., x_{N-1}\}$. We obtain the frequency spectrum of the oscillations, X_f, as the discrete Fourier transform (DFT) of x_n,

$$X_f = \sum_{n=0}^{N-1} x_n e^{-i2\pi fnT}, \tag{3.2}$$

which can be interpreted as the Fourier transform of the continuous FRET signal, $g(f)$,

$$g(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft}\, dt, \tag{3.3}$$

with only a finite discrete sampling of the signal, such that

$$\begin{aligned} X_f &= \int_{-\infty}^{\infty} x_n e^{-i2\pi ft}\, dt \\ &= \int_{-\infty}^{\infty} x(t) \sum_{n=0}^{N-1} \delta(t - nT) e^{-i2\pi ft}\, dt \\ &= \sum_{n=0}^{N-1} \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} \delta(t - nT)\, dt \\ &= \sum_{n=0}^{N-1} x(nT) e^{-i2\pi fnT} \\ &\equiv \sum_{n=0}^{N-1} x_n e^{-i2\pi fnT}. \end{aligned} \tag{3.4}$$

X_f is therefore a discrete spectrum composed of a sample of N evenly spaced frequencies from $f = 0$ to $f = \frac{(N-1)}{NT} = \frac{(N-1)f_s}{N}$, *i.e.*,

$$\{f\} = \left\{0,\ \frac{f_s}{N},\ \frac{2f_s}{N}, \ldots,\ \frac{(N-1)f_s}{N}\right\}. \tag{3.5}$$

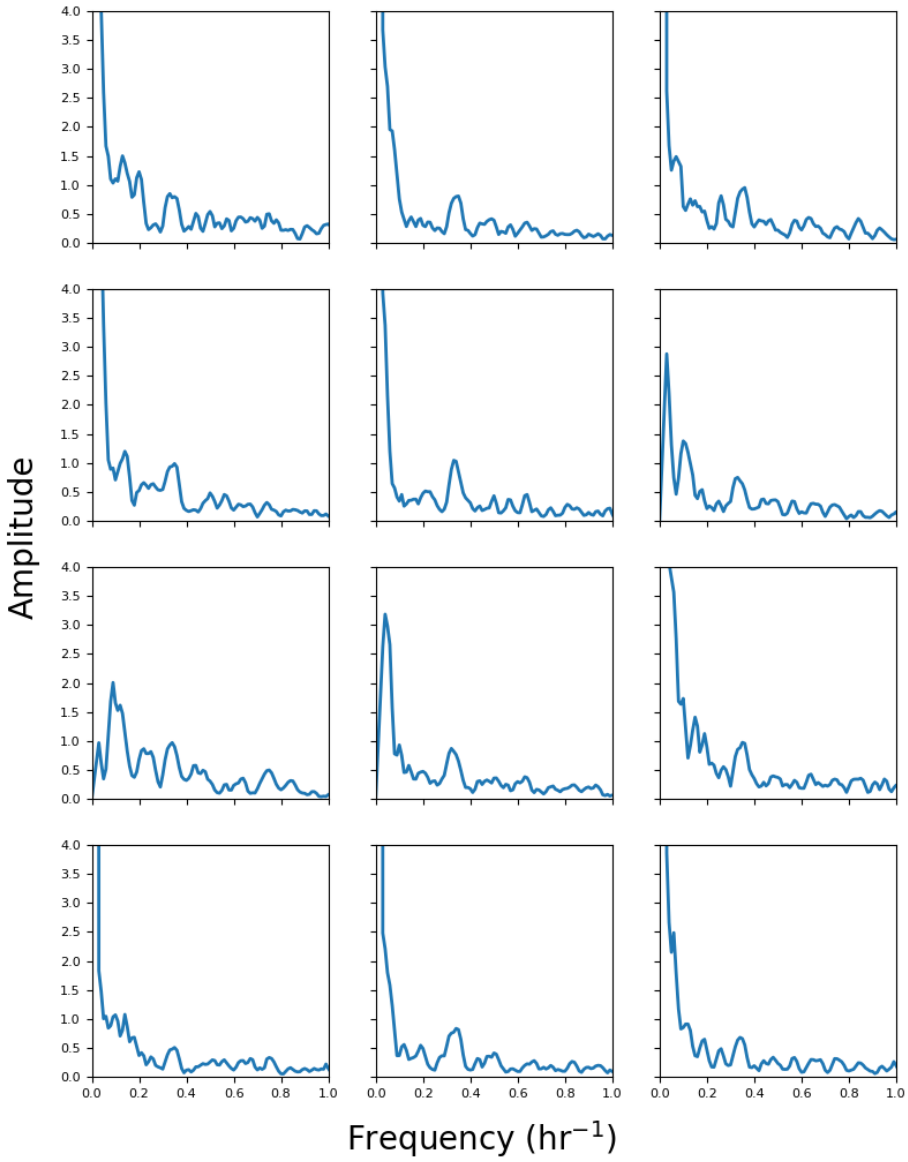

Figure 3.5: Fourier transforms of PerKy293 cells following induction by 300 ng IL-1β on a per-FOV basis. The major periodicity at 0.36 hr^{-1} is broadly represented among the individual FOVs.

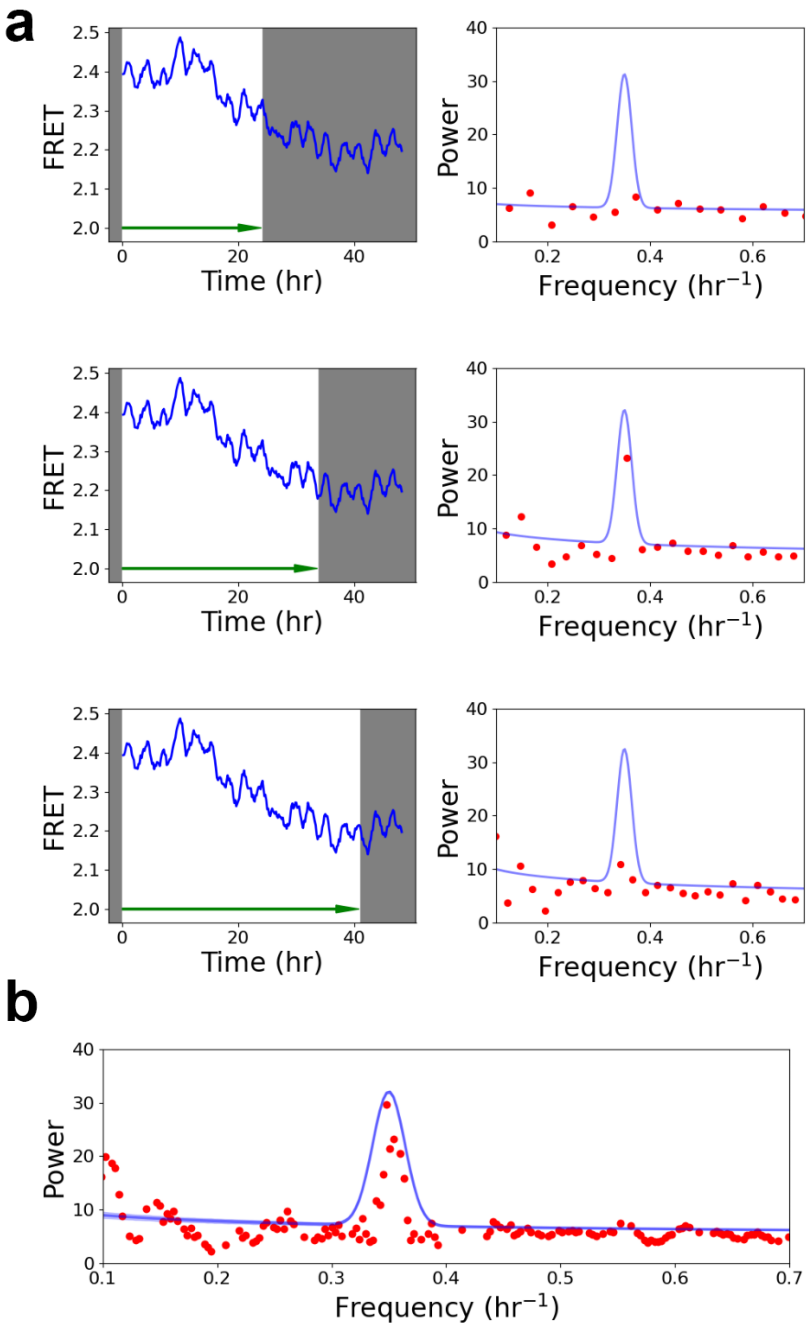

Figure 3.6: Window size is important for a Fourier Transform. If frequency bins are not chosen effectively, periodicities may be missed entirely. **a)** Time domain data on the left is windowed at varying sizes (white area). Output into the Fourier domain on the right (red dots) can differ drastically based based on window choice. Purple represents the Fourier transform as a continuous function. Data is a simulation of Brownian noise with an embedded oscillator with f=0.36 hr^{-1}. **b)** All Fourier transforms combined.

The maximum frequency with unique information will not exceed the Nyquist frequency $f_{Nyquist} = f_s/2$, and hence we take X_f as the subset of the DFT output containing $N/2$ elements with

$$\{f\} = \left\{0, \frac{f_s}{N}, \frac{2f_s}{N}, ..., \frac{f_s}{2}\right\} \tag{3.6}$$

A complication is that, due to the discrete nature of X_f and therefore its limited resolution, if the sampled frequencies do not correspond to the frequencies at which peaks occur in the continuous spectrum, these peaks will not be apparent in the discrete spectrum (Fig. 3.6a). To remedy this issue and to increase the resolution of the spectrum, we implement an expanding window DFT to generate a discrete spectrum with higher frequency resolution, g_f. We first generate a set of discrete spectra with windows of increasing size j, $X_f^{(j)}$, from some minimum sized window $j = j_{min}$ to a maximum $j = N$,

$$X_f^{(j)} = \sum_{t=0}^{j} x_t e^{-i2\pi ft}, \quad j = j_{min}, ..., N, \tag{3.7}$$

each of which yields output with a different set of discrete frequencies. We then collect these together as g_f,

$$g_f = \bigcup_{j=j_{min}}^{N} X_f^{(j)}. \tag{3.8}$$

An example is illustrated in Fig.3.6b. In the following and in the main text, we treat g_f as an approximation of $g(f)$ to keep notation simple.

3.2.1 Cross Power Spectral Density

Given two signals x(t) and y(t), the cross-correlation between the two is written as

$$(x*y)(\tau) \triangleq \int_{-\infty}^{\infty} \overline{x}(t)y(t+\tau)\,dt, \tag{3.9}$$

where $\overline{x}(t)$ denotes the complex conjugate of x(t) and τ denotes a time lag between the two functions. Fourier transforming the cross-correlation over the time lag gives the Cross Power Spectral Density (CPSD)

$$S_{xy}(f) = \int_{-\infty}^{\infty} (x*y)(\tau)e^{-i2\pi f\tau}\,d\tau, \tag{3.10}$$

whereby

$$S_{xy}(f) = S_{yx}(f). \tag{3.11}$$

If the two signals are identical, then the CPSD reduces to being the Power Spectral Density (PSD)[54], which is the squared modulus of the Fourier transform of the autocorrelation function. This metric characterizes commonalities and the phase between two stationary signals, potentially elucidating behaviors in the frequency domain that otherwise would have been missed.

Being a complex number, the CPSD can be written as

$$S_{xy}(f) = C_{xy}(f) + iQ_{xy}(f), \tag{3.12}$$

where C_{xy}(f) is called the cospectrum or the "in-phase signal" and Q_{xy}(f) is known as the quadrature spectrum or "out-of-phase signal"[55]. Spectral coherence takes the squared modulus of the CPSD and normalizes it to each signal's individual PSD by

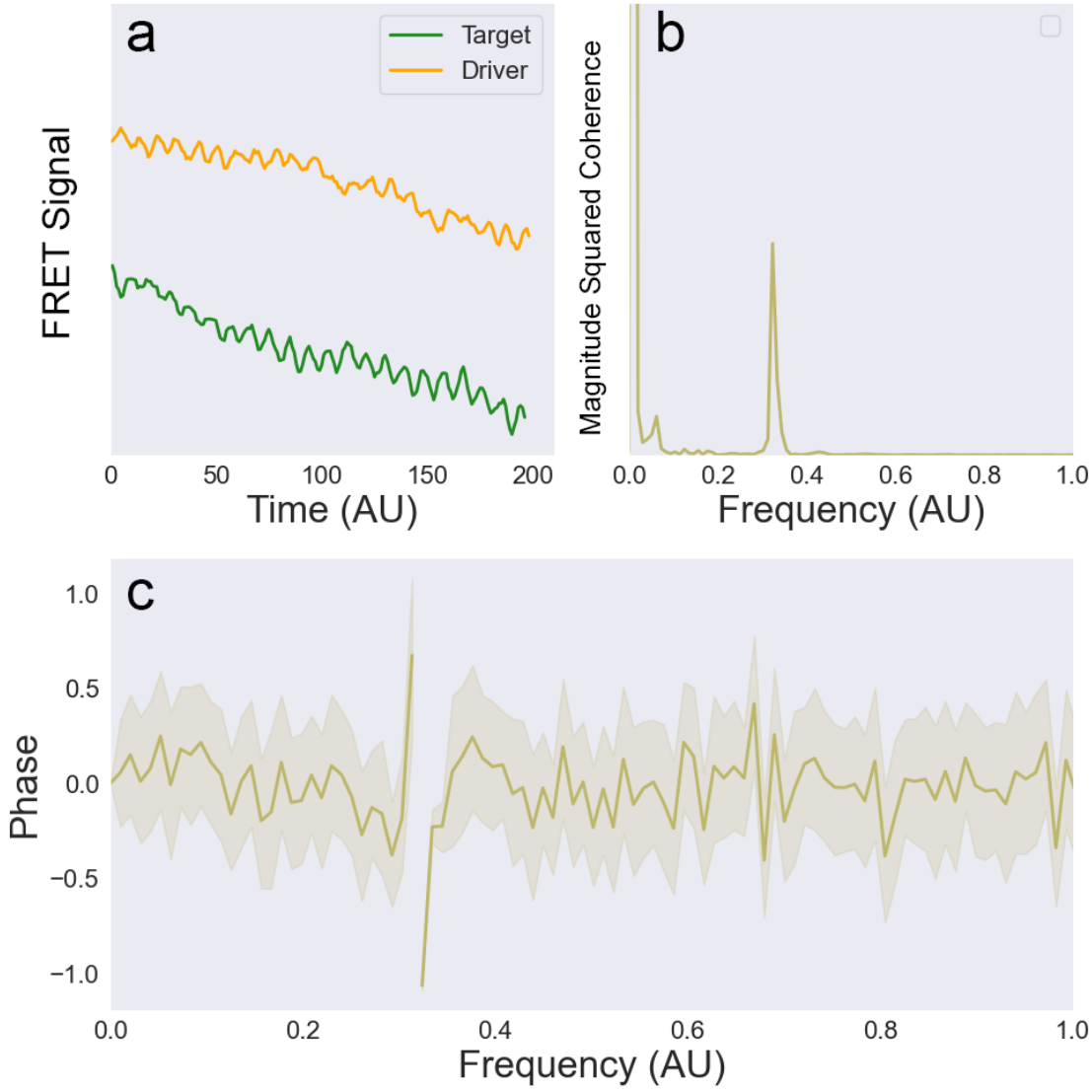

Figure 3.7: Simulated forced damped oscillation dynamics for N = 25 pairs. Driver oscillates at 0.36 AU with added Brownian noise. **a)** Time traces of a driver and its target. **b)** Magnitude squared coherence spectrum shows a prominent peak at the driving frequency. **c)** Phase spectrum shows a discontinuity at the driving frequency.

$$Coh^2 = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)}, \tag{3.13}$$

thereby providing a scalar estimate for the similarity of two signals.

3.2.2 Short Time Fourier Transform

Signals that evolve in time can be visualized in the frequency domain by spectrograms, which are heatmap plots that display frequency and time on the x-y axes and Fourier power with color intensity. One process of producing such a plot is through Short Time Fourier Transform (STFT), where a long time signal is divided into segments of equal length and spacing and then transformed in steps. Mathematically this would be represented as

$$\mathbf{STFT}\{x(t)\}(\tau, \omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-i\omega t}dt \tag{3.14}$$

where $\omega(t-\tau)$ is a window function[56]. For shorter signals there is a trade off between time and frequency accuracy, where narrowing the scanning window improves time resolution at the expense of fewer frequency bins, given by the Heisenberg-Gabor limit

$$\sigma_f \cdot \sigma_t \geq \frac{\pi}{2}. \tag{3.15}$$

As such, a time resolution-oriented approach can quantify the lifetime of a system's oscillatory response.

The Fourier transform treats a time series as repeating infinitely in both directions, thereby considering it a periodic function with an interval equal to the duration of the data. Variable boundary conditions on time series data can result in the emergence of artifacts,

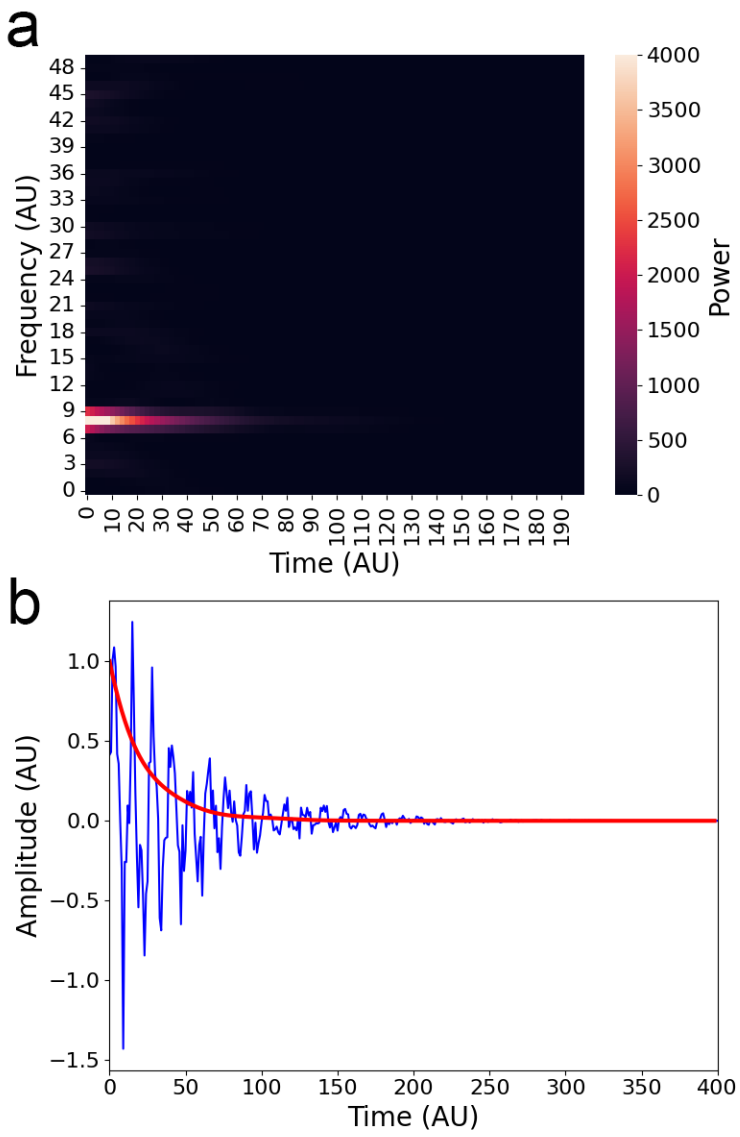

Figure 3.8: Short Time Fourier Transform of a noisy sine function that decays by $e^{-\lambda t}$ shows time evolution. **a)** Spectrogram of the sine function shows a power decay along the Time axis. **b)** Trace of the decaying sine function in the time domain (blue) and the power decay recovered from the spectrogram in **a)** (red).

particularly in STFT. Applying a window function to the time series sets the amplitude at the boundary to 0, smoothing the time evolution. The Blackman Window function[57]

$$w[n] = 0.42 - 0.5\cos\left(\frac{2\pi n}{N}\right) + 0.08\cos\left(\frac{4\pi n}{N}\right) \tag{3.16}$$

was employed for our STFT data. Figure 3.9 shows the process by which windowing a time series affects the resultant Fourier transform. Application of these window functions to our Fourier transforms as in Figure **??** did not qualitatively affect the result, and so we did not employ the Blackman window for our stationary signal analysis.

Applying STFT to FRET data produced in Chapter 2 shows that oscillatory stress response is non stationary (Figure 3.10). Having multiple frequencies that activate and deactivate at different times would indicate a dynamic stress response, but at this time we do not have any additional information to investigate further.

3.3 Calculating Statistical Significance

3.3.1 Brownian Noise Processes

Random variations that are intrinsic to an observable quantity and produce a $1/f^2$ Fourier spectrum are collectively known as Brownian noise[58]. This differs from white noise, where the fluctuations occur independently of each other and are not reliant on a random-walk process. In the frequency domain, the variance of the spectrum decreases with increasing frequency.

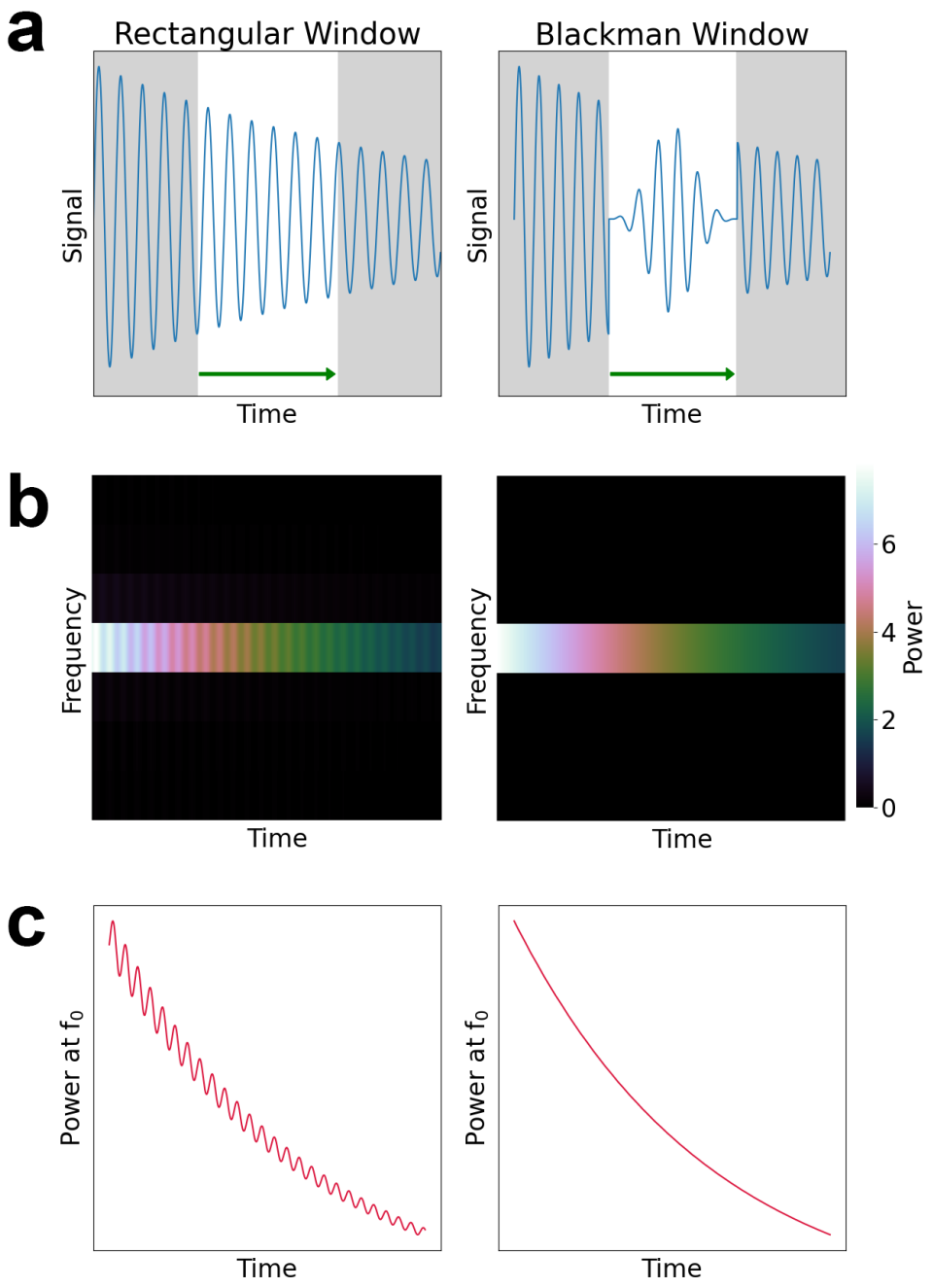

Figure 3.9: STFT is prone to generating artifacts if not windowed properly. **a)** A sine function modified by an exponential decay undergoes STFT with a sliding window that is rectangular (left) or Blackman (right). **b)** Spectrogram of the STFT shows one band at the sine wave frequency f_0. The spectrogram of the time series windowed with a recangular function displays a rippling pattern that is not present in the time series. **c)** Fourier power vs. time of the spectrograms at the sine wave frequency. The rectangular window series displays artifactual oscillations while the Blackman window series is smooth.

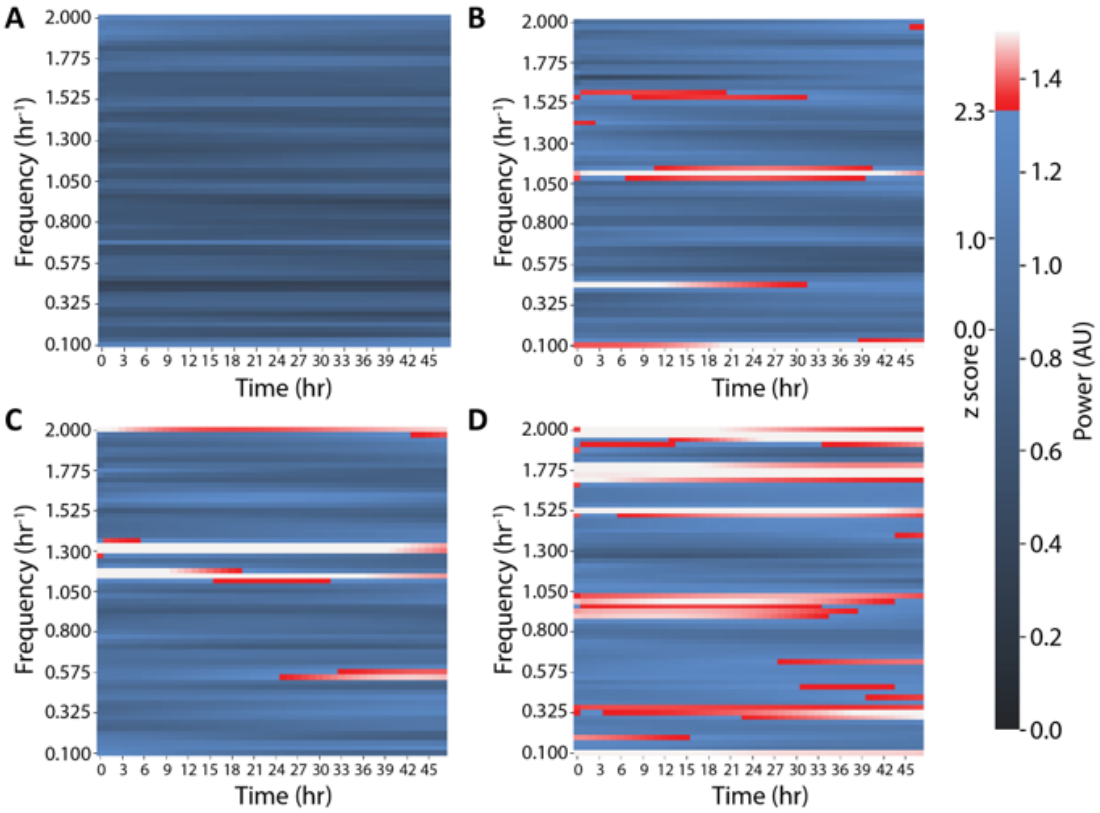

Figure 3.10: STFT spectrograms illustrating how the frequency response of p38 changes over time when subjected to various stimuli. Note that by adding resolution along the time axis, resolution along the frequency axis is reduced, and hence these spectrograms are not directly comparable to the time independent Fourier spectra presented elsewhere, which are effectively (but not mathematically identical to) averages over the time axis of these spectrograms. (**A**) Unstimulated VPC293 control. (**B**) 2 nM SARS-CoV-2 Spike. (**C**) Hyperosmotic shock induced by addition of 120 mM sorbitol. (**D**) 300 ng/mL IL-1β.

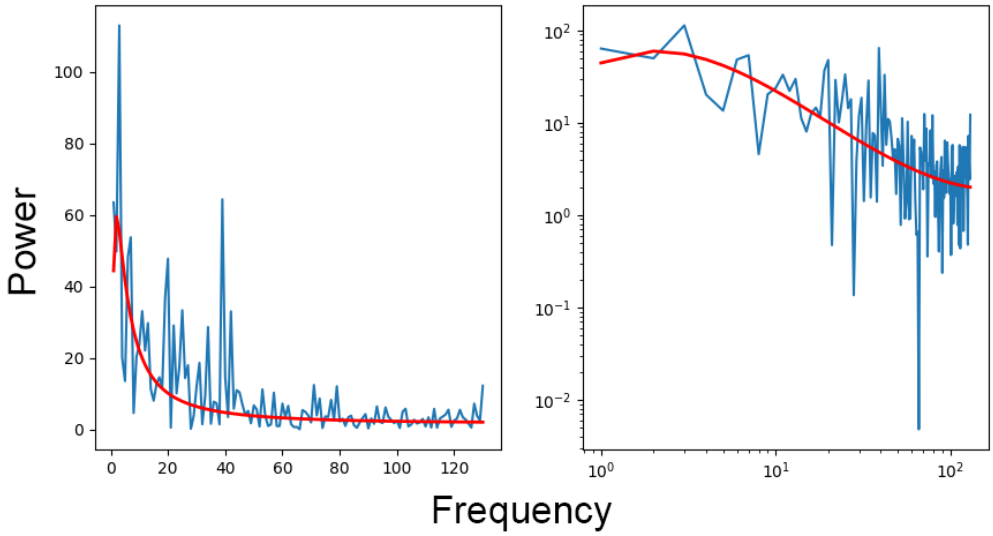

Figure 3.11: Fourier power spectrum derived from a simulated Brownian process. Red indicates a third order polynomial fitted in log space.

Because variance is not constant in the frequency domain, statistical significance should only be derived after the spectrum has been transformed into log space (Figure 3.11). A characteristic of these spectra is that they are capable of spontaneously producing statistically significant peaks where a periodic function would otherwise be absent. These anomalies can be mitigated through ensemble averaging over several unique spectra (Figure 3.12). Statistics in ensemble-averaged Fourier spectra are approximately normal and therefore p values can be derived from the standard deviation of a control experiment.

3.3.2 Harmonic Analysis

Stimulus spectra as in Figure 2.7 display statistically over-represented frequencies that have been thresholded based on significance. The narrowly-defined locations along the

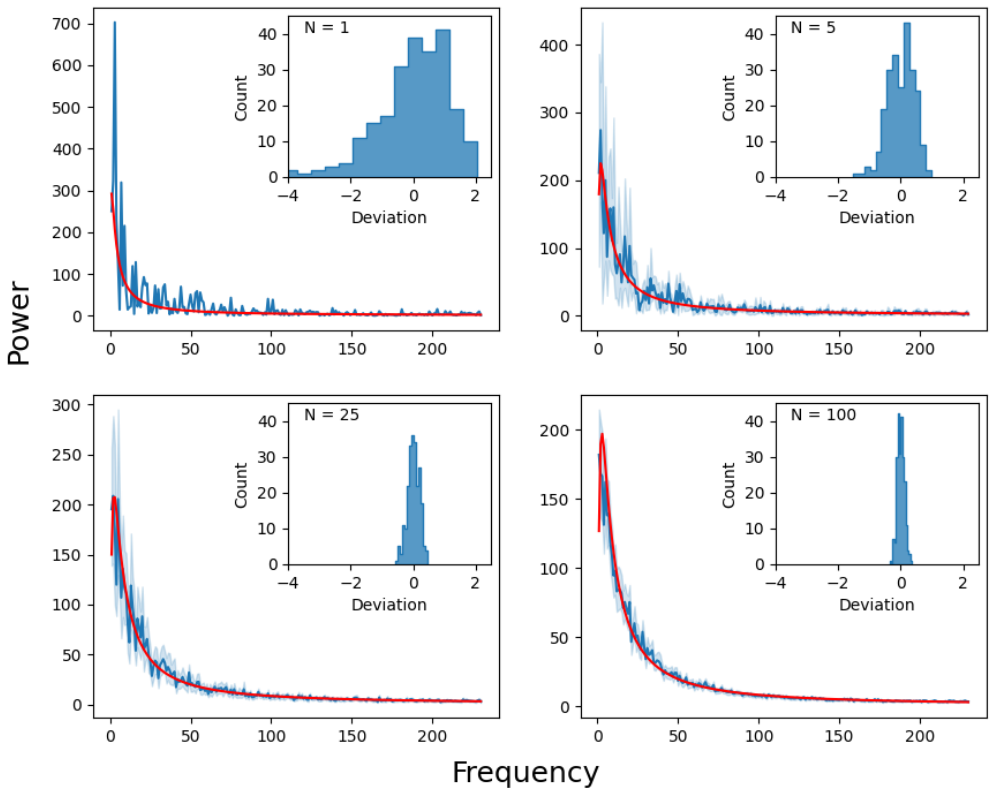

Figure 3.12: Fourier power spectra of simulated Brownian noise processes. Each spectrum is an ensemble average of N independent simulations with light blue shaded region indicating a 95% confidence interval and red curves indicating a third order polynomial fit in log space. Inset: Deviations of points along the ensemble curve in log space from the fitted curve. For N = 1 the shape matches a chi-square distribution and with increasing N converges to a normal distribution.

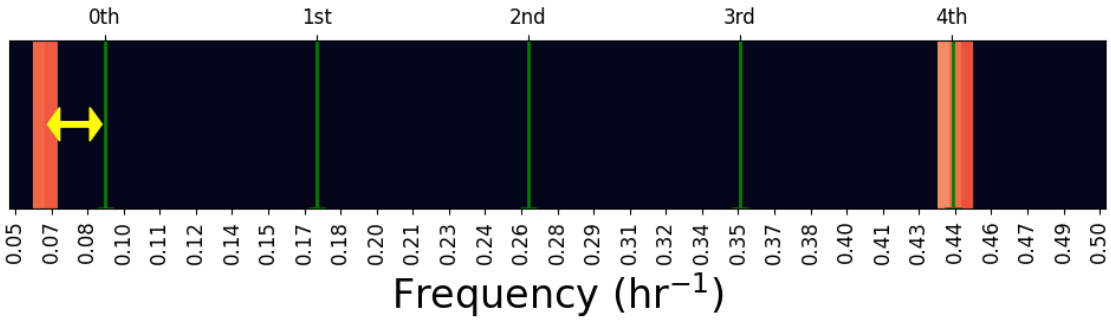

Figure 3.13: Stimulus spectrum with harmonic "tines" applied. Red indicates statistically over-represented frequencies. Green indicates tines evenly spaced by distance f_0. Yellow arrow indicates the distance of a frequency band to the nearest tine.

frequency axis make these plots ideal for employing analysis to determine both fundamental frequency and likelihood of the spectra being harmonically organized. Were one to add a series of delta function "tines" that are evenly spaced by fundamental frequency f_0, the distance of each band on a stimulus spectrum can be measured (Figure 3.13).

For any set of tines, a frequency band can be located no more than $f_0/2$ from the closest tine. For frequency bands that are anharmomincally (randomly) distributed, the average distance of any given frequency band to the closest tine will be $f_0/4$. Therefore, average distances smaller than $f_0/4$ can be considered unlikely for an anharmonic system. Best-fit fundamental frequency can be determined for a system by testing a series of harmonic "combs" of varying f_0 and minimizing the average distance-to-tine (Figure 3.14).

Bates distributions characterize the mean of n statistically independent uniformly distributed random variables[59]. On the interval $0 < x < 0.5\ f_0$, the distribution of n independent random variables is given by

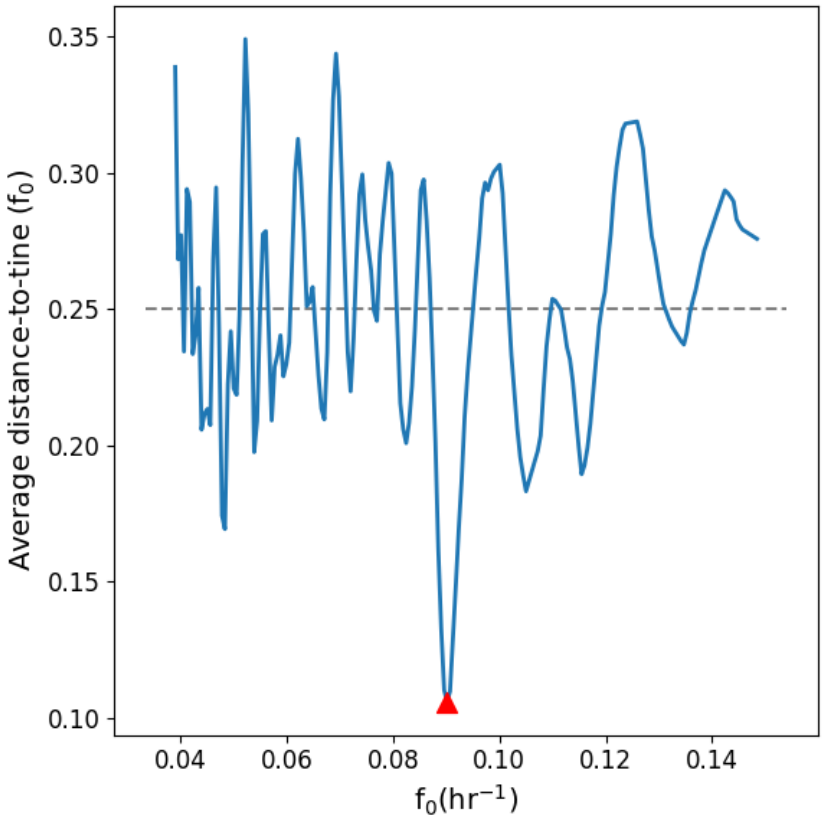

Figure 3.14: Fundamental frequency analysis for aggregated stimulus spectra. A global minimum at 0.09 hr^{-1} indicates that it is the best fit. Gray dotted line indicates the average distance-to-tine for a randomly distributed system. A distance-to-tine of 0 would indicate a perfectly harmonic system.

$$f_X(x;n) = \frac{n}{(n-1)!}\sum_{k=0}^{n}\binom{n}{k}(2nx-k)^{n-1}\mathrm{sgn}(2nx-k). \tag{3.17}$$

where $\binom{n}{k}$ is the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{3.18}$$

and

$$\mathrm{sgn}(2nx-k) = \begin{cases} -1 & 2nx < k \\ 0 & 2nx = k \\ 1 & 2nx > k \end{cases}$$

For a system with n statistically over-represented frequency bands, significance can be derived by comparing to a similar random system with a Bates distribution (Figure 3.15). From there one can determine the likelihood of n frequency bands having an average distance d from the best-fit harmonic frequency bins.

3.4 Applying Machine Learning to Cell Segmentation and Tracking

Cell segmentation is among the most valuable tools to understanding biological dynamics. Epithelial cells such as the HEK293 immortalized line create closely packed structures[60], and a lack of organelle labeling can make it very difficult to discern unique cells. Moreover, these cells have a degree of mobility when not at full confluence, being able

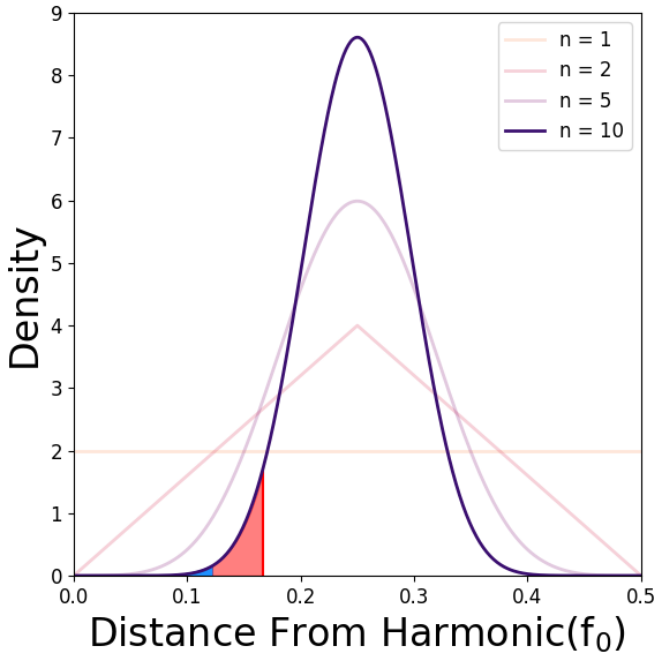

Figure 3.15: Bates distributions of average distance-to-tine for n frequency bands. The n = 1 case represents a uniform distribution and increasing n converges to a normal distribution. Red shaded region denotes the statistical significance of the distances of n = 10 frequency bands as measured from their centers of mass, while blue denotes distances as measured from the closest edge of each band.

crawl in the x-y direction and even along z. This proves disruptive for confocal inverted microscopy of monolayers, with cells often disappearing from the field of view.

Staining techniques are widely employed to abrogate the challenges posed by cell segmentation. DAPI is a blue fluorescent stain that binds to adenine-thymine regions in DNA[61]. However, it is typically employed in fixed-cell imaging due to its specificity and observed mutagenicity. A similarly functioning stain is Hoechst 33342, which does not adversely affect cell viability and therefore is more often employed in live cell imaging[62]. There also exist a number of stains that bind with high affinity to the cell membrane, including DiO, DiI, DiD, DiR and PKH dyes[63].

The advent of Deep Learning has allowed for computers trained on large data sets to emulate the capacity of the human brain to identify unique objects with high confidence. You Only Look Once (YOLO) is an object detection algorithm that employs a convolutional neural network to create bounding boxes around predicted objects[64]. YOLO further is able to track objects through time, eliminating the need to employ mathematical algorithms such as Nearest Neighbor.

Meta's Segment Anything model is a tool for generating exact masks of an object in an image[65]. Simultaneously an advantage and disadvantage of the model is the fact that it has already been trained on billions of images, negating the need for supervised learning. Without a prompt, the system will detect any and all objects that it recognizes in an image. Bounding boxes can be provided to narrow the model's region of interest.

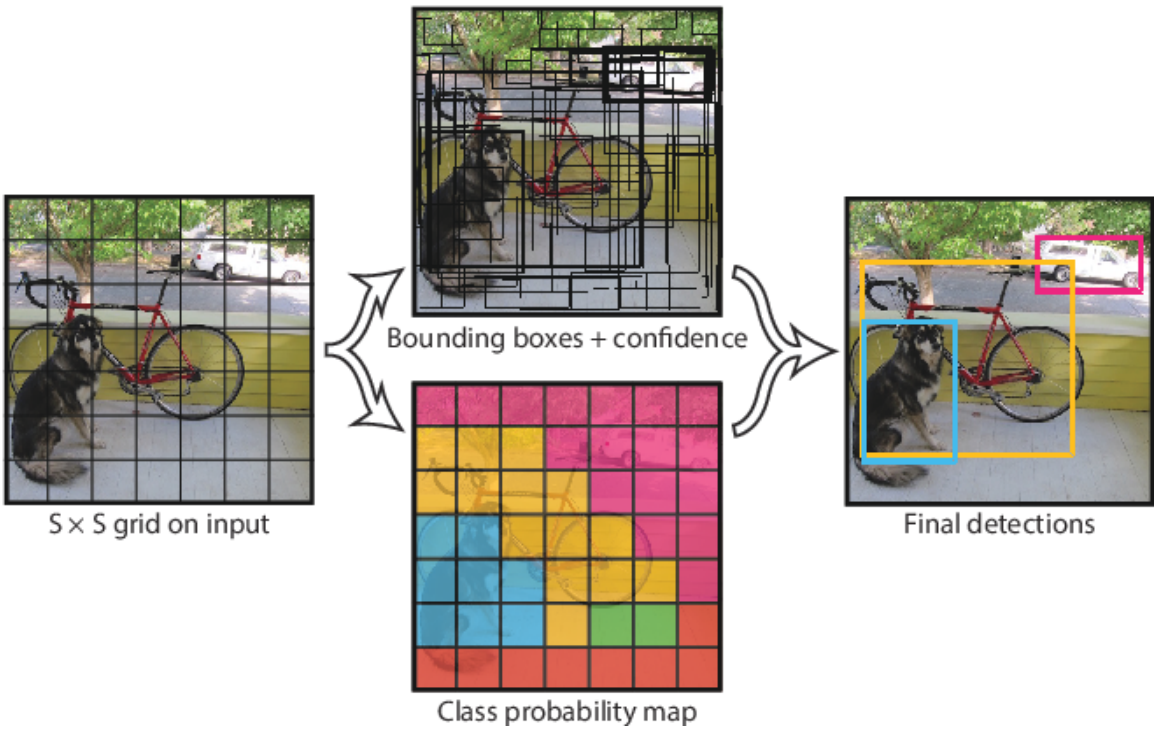

Figure 3.16: The YOLO algorithm divides the image into a grid and creates predictive bounding boxes[64]

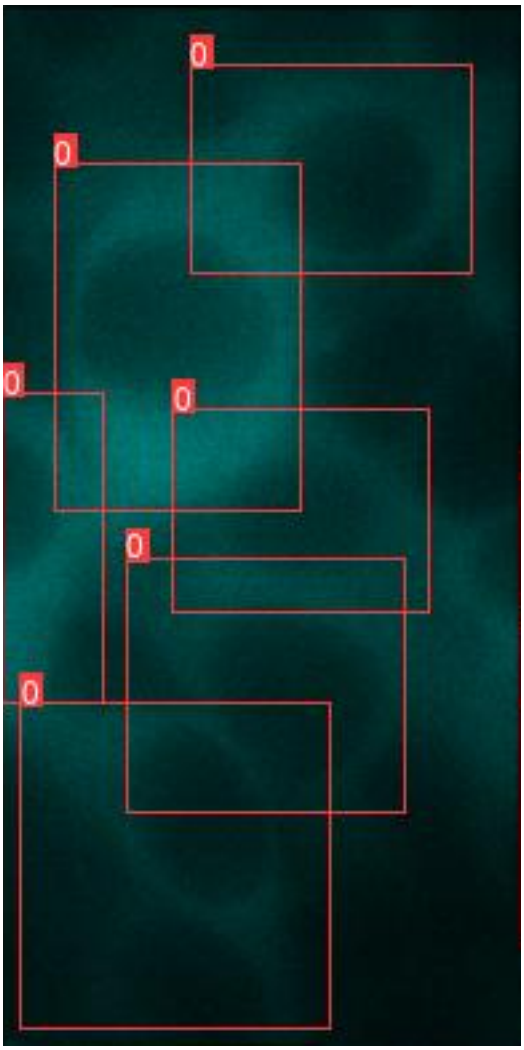

Figure 3.17: PerKy293 cells are detected by YOLO.

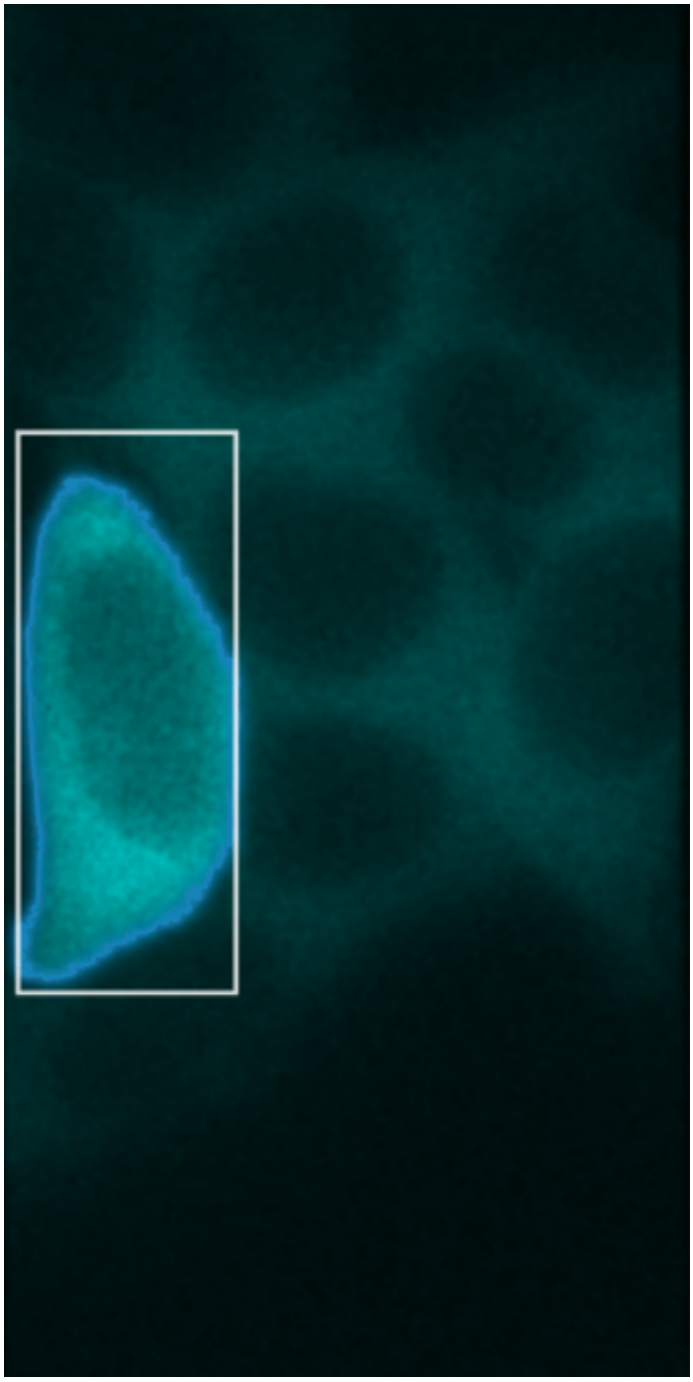

Figure 3.18: PerKy293 cells are masked by Segment Anything. A bounding box provides a prompt to restrict the model's region of interest.

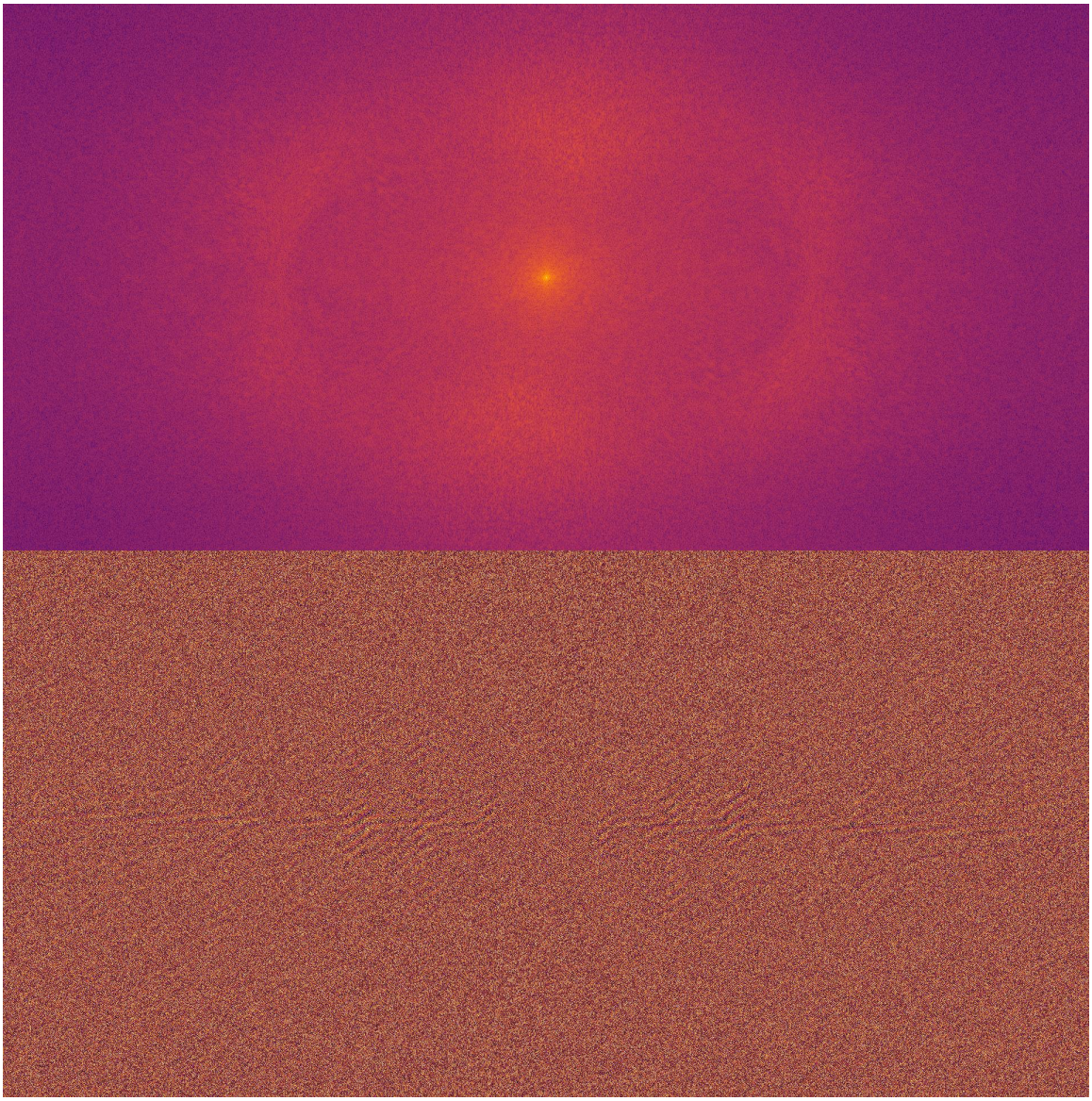

Figure 3.19: I'm sure this has a purpose.

Chapter 4

TnpB

4.1 Introduction

The contents of this chapter were originally published as an article in the Journal of Visualized Experiments in 2023 under the title "Real-Time Quantification of the Effects of IS200/IS605 Family-Associated TnpB on Transposon Activity". Equal contributions were given by myself and Femila Manoj, with Thomas E. Kuhlman serving as the corresponding author. My role in this project was in microscope operation and subsequent data analysis, while Femila Manoj's role was in sample preparation.

Transposable elements (TEs) are genetic elements that mobilize within their host genomes by catalyzed copying or excision followed by genomic reintegration. TEs exist in all domains of life, and transposition restructures the host genome, mutating coding and control regions[66]. This generates mutation and diversity that plays an important role in evolution[67],[68], development[69],[70], and several human diseases[71], including cancer[72].

Using novel genetic constructs that couple aspects of transpositional activity to fluorescent reporters, previous work described the development of an experimental system based on the bacterial TE IS608, a representative of the widespread IS200/IS605 family of TEs, that allows for the real-time visualization of transposition in individual live cells[73] (Figure 4.1). A diagram illustrating the TE system is shown in Figure 4.1A. The TE is composed of the transposase coding sequence, TnpA, flanked by Left End (LE) and Right End (RE) imperfect palindromic repeats (IPs), which are the recognition and excision sites for TnpA. TnpA is expressed using the promoter PLTetO1, which is repressed by tet repressor and is inducible with anhydrotetracycline (aTc)[74]. The TE splits the -10 and -35 sequences of a constitutive PlacIQ1 promoter[75] for the blue reporter mCerulean3[14]. As shown in Figure 4.1C, when TnpA production is induced, the TE can be excised, leading to the reconstitution of the promoter. The resulting cell expresses mCerulean3 and fluoresces blue. The N-terminus of TnpA is fused to the yellow reporter Venus[13], allowing measurement of TnpA levels by yellow fluorescence.

IS608 and other members of the IS200/IS605 family of transposons also typically encode a second gene of thus far unknown function, TnpB[76]. TnpB proteins are an extremely abundant but poorly characterized family of nucleases that are encoded by many bacterial and archaeal TEs[77, 78], which often contain only tnpB[79]. Furthermore, recent studies have renewed interest in TnpB by finding that TnpB functions as a CRISPR/Cas-like programmable RNA-guided endonuclease that, under different conditions, will make either dsDNA or ssDNA breaks[3, 80]. However, it remains unclear what role TnpB may play in regulating transposition. To perform real-time visualization of the effects of TnpB

on IS608 transposition, a version of the transposon, including the coding region of TnpB with an N-terminal fusion to the red fluorescent protein mCherry, was created.

Complementing more detailed bulk-level studies performed by the Kuhlman lab[81], it is shown here how real-time imaging of transposon activity can quantitatively reveal how TnpB or any other accessory proteins impacts transpositional dynamics. By fusing TnpB to mCherry, individual transpositional events are identified by blue fluorescence and correlated with expression levels of TnpA (yellow fluorescence) and TnpB (red fluorescence).

4.2 Results

4.2.1 Cells Expressing TnpB Undergo Higher Excision Rates

This method of visualizing transposon activity in live cells by fluorescence microscopy, while having lower throughput than bulk fluorescence measurements, allows direct visualization of transposon activity in individual live cells. Transposon excision events result in the reconstitution of the promoter for mCerulean3 (Figure 4.1), allowing identification of cells undergoing transposon activity by bright blue fluorescence (Figure 4.2).

It is found that cells expressing the accessory protein TnpB (Figure 4.3, orange) experience 4–5 times higher levels of transposon activity compared to those that do not (Figure 4.3, blue), consistent with the more detailed bulk-level studies[81]. This is particularly notable as the inclusion of the coding sequence of mCherry-tnpB increases the length of the transposon by 2,000 bp, while previous studies have found that IS608 transposon excision is an exponentially decreasing function of transposon length[82].

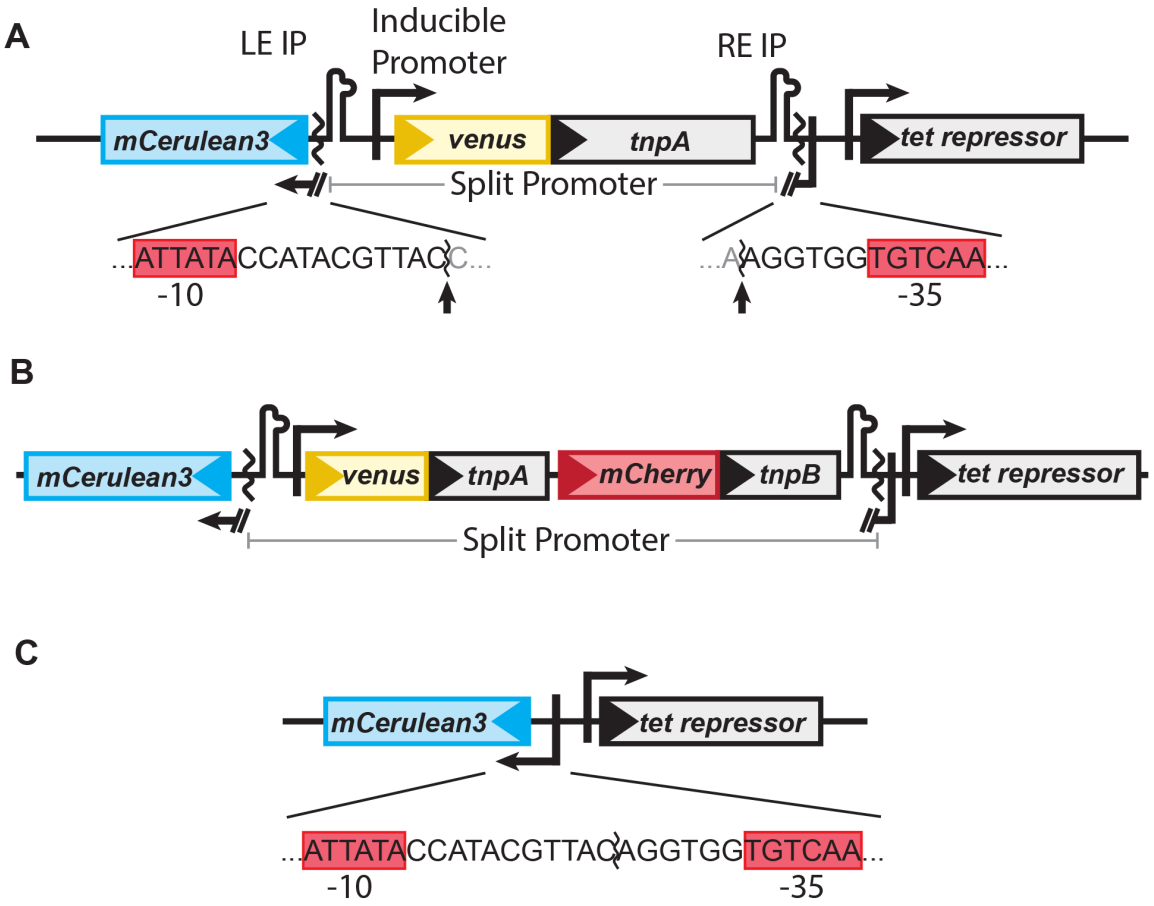

Figure 4.1: (A) The mCerulean3 promoter is disrupted by the TE, the ends of which are flanked by the left end and right end imperfect palindromic sequences (LE IP and RE IP). The transposase, tnpA (gray), is expressed from PLtetO1, which is regulated by tet repressor (gray) and is inducible with anhydrotetracycline (aTc). The sequences of the Promoter/TE junction and promoter -10 and -35 sequences (red boxes) are shown below, and TnpA cleavage sites are indicated by arrows. (B) The TnpB+ construct, where mCherry-tnpB has been transcriptionally fused to venus-tnpA such that both are transcribed as a polycistronic mRNA, mimicking the natural configuration of IS608. (C) Upon excision, the mCerulean3 promoter is repaired, and the cell fluoresces blue. The sequence of the reconstituted promoter is shown below the diagram.

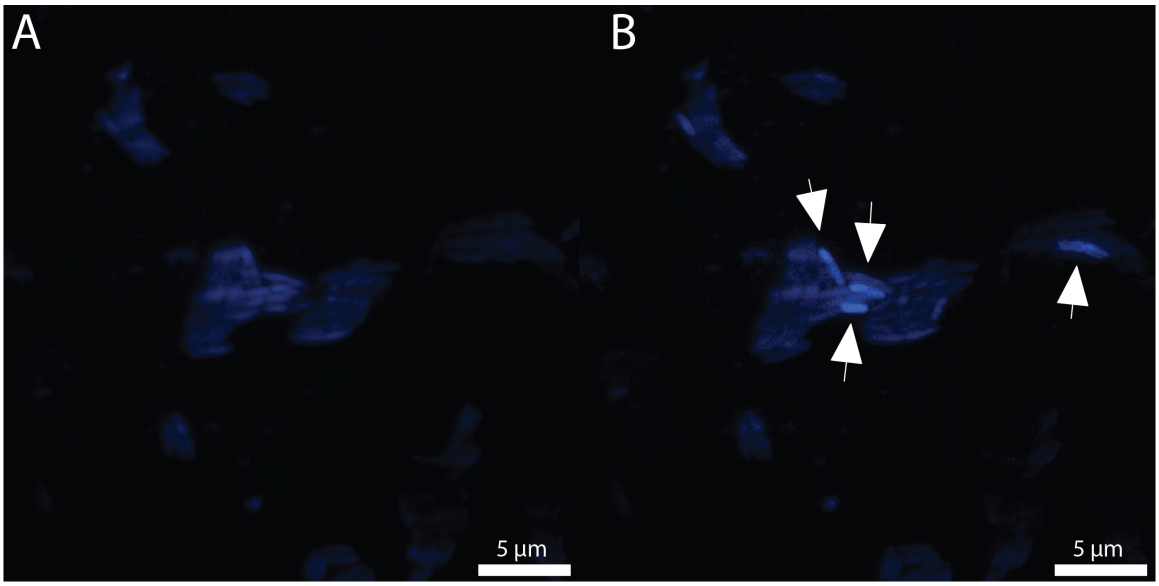

Figure 4.2: Example field of view of TnpB+ cells (A) immediately before and (B) after detecting transposon excision events by blue fluorescence. White arrows indicate excision events. The time difference between the two frames is 20 min. Scale bar = 5 µm.

4.2.2 Excising Cells Express TnpB at Higher Levels

An advantage of real-time imaging is that once identified, cells undergoing transpositional events can be further tracked and analyzed to determine other characteristic parameters, such as growth rate, to determine the distribution of fitness effects or the expression level of accessory proteins to determine their impact on transpositional activity. For example, in TnpB+ cells, cells undergoing transposon excision events have higher expression levels of mCherry-TnpB than the general population (Figure 4.4A). Moreover, for cells undergoing excision events (Figure 4.4B, dark yellow), TnpB+ cells (Figure 4.4B, bottom) express only marginally higher levels of Venus-TnpA transposase than TnpB- cells (Figure 4.4B, top) [TnpB- 158.3 ± 68.2 AU, TnpB+: 193 ± 79.9 AU], which is higher than the yellow fluorescence of the general population (Figure 4.4B, light yellow). Taken

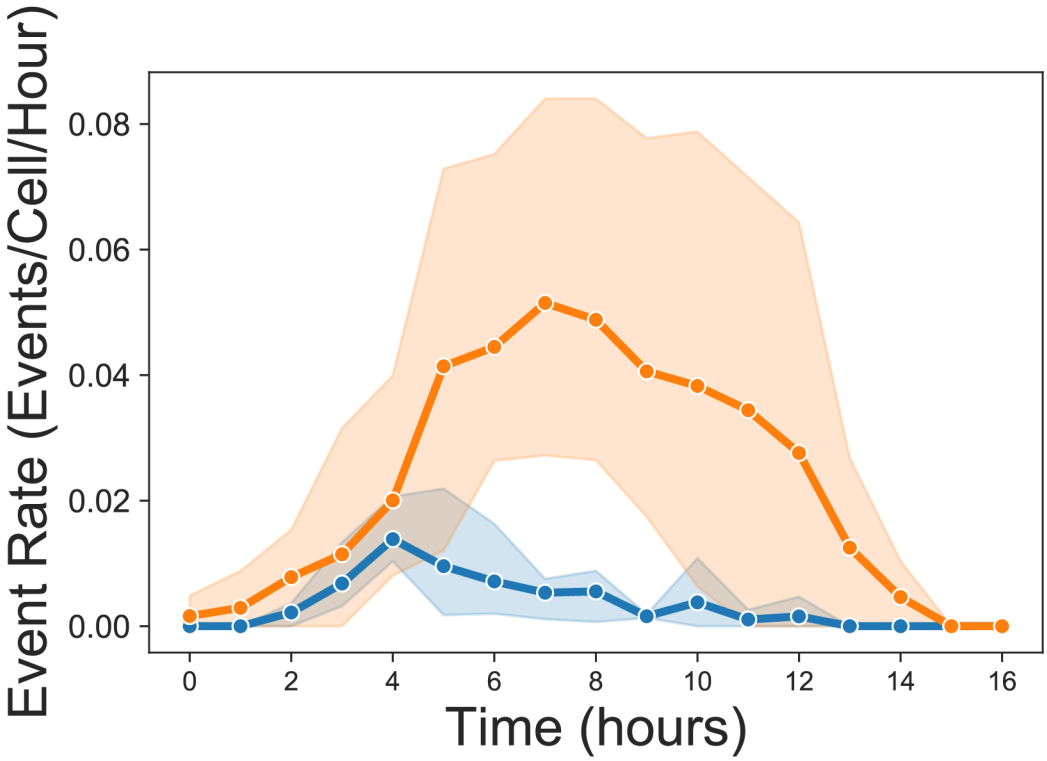

Figure 4.3: Excision rate for TnpB+ cells (orange) and TnpB- (blue) cells. The mean rate from three replicates is shown as points with shaded regions with a 95% confidence interval. Data are aligned so that cells begin excising at t = 0. The maximum measured rate for TnpB+ cells was 5.1 ± 2.4 x 10-2 events per cell per hour, while for TnpB- was 1.4 ± 0.48 x 10-2 events per cell per hour. The average rate over the whole interval shown was 2.6 ± 1.8 x 10-2 events per cell per hour for TnpB+ cells and 5.3 ± 2.9 x 10-3 events per cell per hour for TnpB-.

together, these data suggest that TnpB protein is responsible for the observed higher levels of transpositional activity.

4.3 Discussion

The unique method presented here for real-time imaging of transposable element activity in live cells is a sensitive assay that can directly detect transposition in live cells

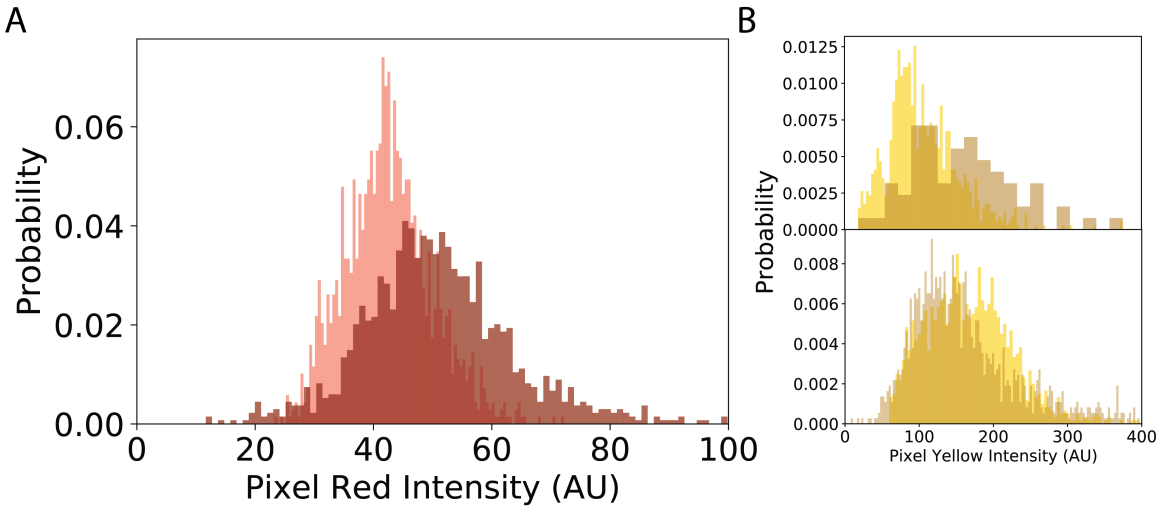

Figure 4.4: Each frame was divided into 64 equal blocks, and fluorescence was measured for cells excising within the block and for all cells contained within the block regardless of excision activity; Probability, as plotted on the y-axis, is measured as the number of pixels of the indicated intensity in each cell type divided by the total number of pixels. The block size for each frame was set to 445 x 445 pixels. (A) Cells that undergo excision events (dark red) express more TnpB than the general population (light red). The average red fluorescence for excising cells was 51.3 ± 15.4 AU (dark red), while that for all cells was 42.5 ± 7.4 AU (light red). (B) Venus-TnpA transposase levels are similar in TnpB- (top) and TnpB+ (bottom) cells. Data sets are normalized so that the mean yellow fluorescences of the total cell population (light yellow) are equal for TnpB+/- at 105.7 AU. Cells that have been identified as undergoing transposon excision events (dark yellow) exhibit higher yellow fluorescence than the general population, with similar distributions for TnpB- (top) and TnpB+ (bottom). Mean yellow fluorescences of the excising populations are TnpB-: 158.3 ± 68.2 AU, TnpB+: 193 ± 79.9 AU.

and in real-time and correlate this activity with the expression of accessory proteins. While the throughput is lower than can be accomplished by bulk methods, this method achieves detailed measurements of TE activity and protein expression in individual living cells.

A variety of tools and techniques can be employed to grow cells directly on the microscope for real-time imaging. The method used here of cell growth on agarose pads has the advantage of being fast, cheap, and easy to perform. A possible disadvantage, depending on the cellular growth state of interest, is that resources available to support cell growth in the agarose pad are limited, and hence cells will naturally exhaust these resources and stop growing after a relatively short period of time (12–24 h). Consequently, care must be taken to prepare the cells in steady state growth and inoculate the pad at a low enough density to give ample time for measurement. Microfluidics can be employed to maintain cells in steady state exponential growth for extended periods of time[83], although these methods require additional expertise, equipment, and setup to be effective.

Complementing more detailed work from the Kuhlman lab[81], it is illustrated here that the IS200/IS605 TE family-associated protein TnpB increases the rate of IS608 excision by up to five-fold, and that increased excision is directly correlated with higher expression levels of TnpB. These methods are one example of improved assay techniques that may help shed light on transposon activity and its impacts on mutational and evolutionary dynamics.

Chapter 5

The DIYncubator

5.1 Introduction

The contents of this chapter are currently in submission to be published as an article by a scientific journal and are available on Biorxiv under the title "A Low-Cost Stage-Top Incubation Device For Human Cell Imaging Using Rapid Prototyping Methods". I am the primary author of this work, with my role being device design, software development, programming, and construction. Secondary authors include Melissa Gomez and Pratyasha Mishra who performed human cell culture preparation, Quintin Meyers who assisted in device construction, and Thomas E. Kuhlman who is acting as the corresponding author for journal submission.

Human and mammalian cell health *in vitro* is maintained partly through use of media supplemented with a bicarbonate buffering system. Environmental CO_2 at concentrations typically ranging from 5-7% dissolves into the medium and allows for regulation of internal pH usually between 7.2 and 7.4 [84]. Maintenance of these optimal growth condi-

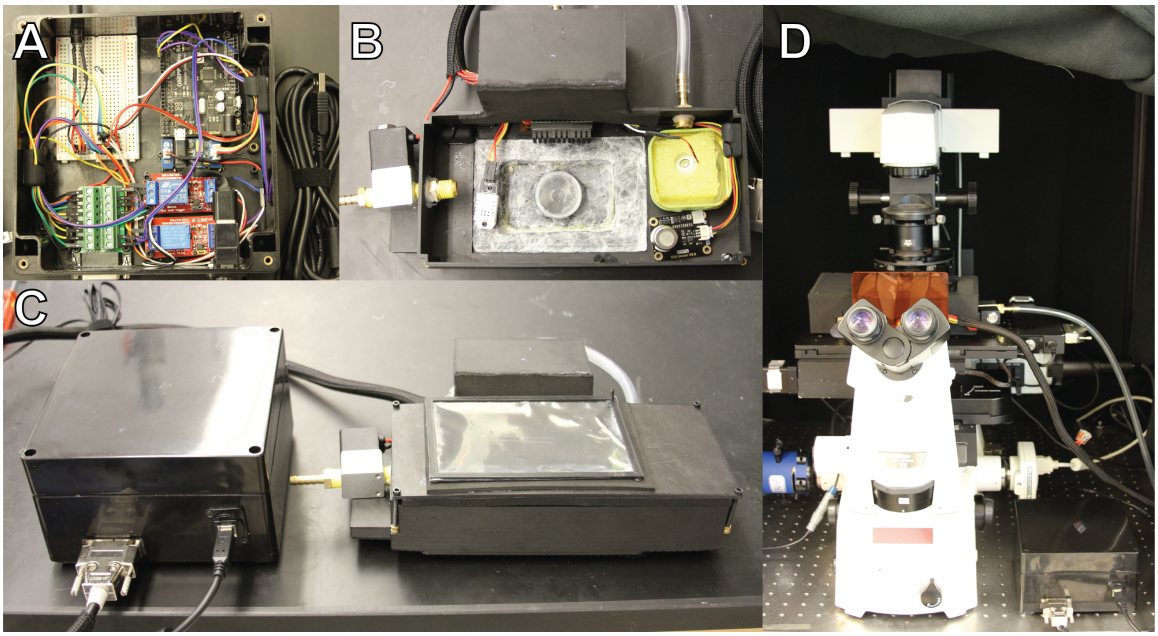

Figure 5.1: (A) The electronics box, containing the Arduino, relays driving the CO_2 solenoid and heating unit, atomizer control board, a small breadboard for connections, and a female DB15 port. (B) Photograph of the bottom half of the DIYncubator chassis with electric components including heating unit and fan, CO_2 solenoid, AM2302 humidity and temperature sensor, MG-811 CO_2 sensor, and ultrasonic atomizer with water reservoir. Note that the bottom of the device and the adapter for holding the sample (here a 29 mm circular dish) is coated with Parafilm to seal the adapter into the chassis and limit movement of the sample. (C) Assembled electronics box and DIYncubator. (D) The DIYncubator deployed, with insulating neoprene cover, on our Nikon TiE inverted microscope.

tions in the lab is typically accomplished via incubators that actively pump CO_2 into the interior. Generally, laboratory incubators are not conducive to facilitating the use of a microscope, making it infeasible to perform frequent microscopy of samples without perturbing the cells unless an environmental chamber for the microscope is purchased or assembled, typically at great financial or labor cost.

The geometry of an inverted microscope, where the objective lens sits underneath the sample, is advantageous for adherent cell cultures, where individual cells settle to the bottom of wells with the assistance of gravity [85]. This arrangement also makes it feasible

to use an incubation device sitting on the top surface of the stage to maintain ideal environmental conditions for imaging live human cell cultures over long periods with minimal perturbations. Imaging of the sample with an inverted microscope means that any culturing device must have an open window on the bottom face.

Commercially available stage-top incubation devices create a sealed atmospheric environment for the sample that can be quickly accessed and regulated with a series of sensors and controllers [86]. Typically, caution is not taken to make the device airtight, and instead the system opts for a constant flow of mixed air-CO_2 atmosphere in and out of the device, with humidity control to reduce evaporation from the sample. Purchasing of commercially available devices can cost between approximately $12,000 to $40,000 in 2023, depending upon the features and manufacturer of the device. Such high cost is a significant barrier of entry for many laboratories. Lower cost custom solutions for mammalian cell incubation reported in the literature are generally suitable only for custom built microscopes and stage arrangements and are typically not "plug and play" stage top devices [87, 88]. Here, we describe a DIY stage-top incubation device, we call the "DIYncubator", that can be constructed simply for approximately two orders of magnitude lower cost than commercial systems (∼$250 USD) and is compatible with existing commercial or custom microscopes and stages.

5.2 Methods

5.2.1 DIYncubator Design and Assembly

The chassis of the DIYncubator (Fig. 5.1) was designed in Autodesk Inventor 2022, formatted for 3D printing with Ultimaker Cura software, and printed on an Ultimaker S7 using black acrylonitrile butadiene styrene (ABS) plastic. Machine screws and threaded screw inserts were used to attach the lid to the device. The contact region between the two major halves of the chassis was lined with neoprene to make the device approximately airtight, limiting the amount of gas required to maintain a constant atmosphere. It is not necessary to make the device completely airtight, but reduction of leakage helps reduce the amount of CO_2 necessary and therefore the cost of operation. Attachment and sealing of other permanent components of the device was achieved with polyurethane or cyanoacrylate adhesive and sealed with silicone caulk typically used for kitchen and bath projects. The shape of the DIYncubator was designed to attach to a Prior ProScan III stage on a Nikon Ti-E inverted microscope. Minor design alterations may be necessary when using a different stage design, but are straightforward to accomplish using 3D modeling software.

The top half of the chassis was designed with a large window to provide quick access to the sample. A covering for the window was constructed from a 3D printed ABS black plastic frame, thin transparent plastic recovered from recycled packaging, and refrigerator magnets for easy detachment from the chassis. The borders of this window were also lined with neoprene to aid in maintaining an airtight environment. The bottom half of the chassis contains a large opening into which a variety of adapters for holding differently sized and shaped samples can be placed. We then coat both the bottom of the chassis and the adapter

in Parafilm, as seen in Fig. 5.1B, which serves both to seal the adapter while in place and also limit movement of the sample when placed on the adapter. Finally, the heating system is placed within a housing at the back of the chassis, and is then covered with a layer of neoprene. To aid in insulation and temperature control of the system, we also fashioned a rough neoprene cover for the entire unit out of scraps remaining from sealing the other components, seen in Fig. 5.1D.

5.2.2 Electronics, Environmental Measurement, and Control

The control system of the DIYncubator was achieved with an Arduino Uno SBC. CO_2 and temperature/humidity readings were taken with a MG-811 CO_2 sensor and an AM2302 temperature and humidity sensor respectively. Heat for the system was generated with two flexible polyimide heating film units covered with heatsinks and distributed with a small fan. During operation, the CO_2 and temperature/humidity sensors were placed as closely as possible to the sample. Gas intake was regulated by a solenoid valve connected to a pure CO_2 gas tank. Humidity control is provided by an NGW-1pc ultrasonic water atomization unit. As a water reservoir for the atomizer, we used an empty mint tin from Trader Joe's that holds 45 - 50 mL of water.

The Arduino was programmed to implement a very simple "bang-bang" control scheme for all parameters, where when the measured value of the parameter fell below/above a set point, the corresponding control device was turned on/off with 100% power. If more precision is required, more complex (and expensive) PID control schemes employing MOSFETs could be implemented. The device was interfaced with a desktop laboratory computer

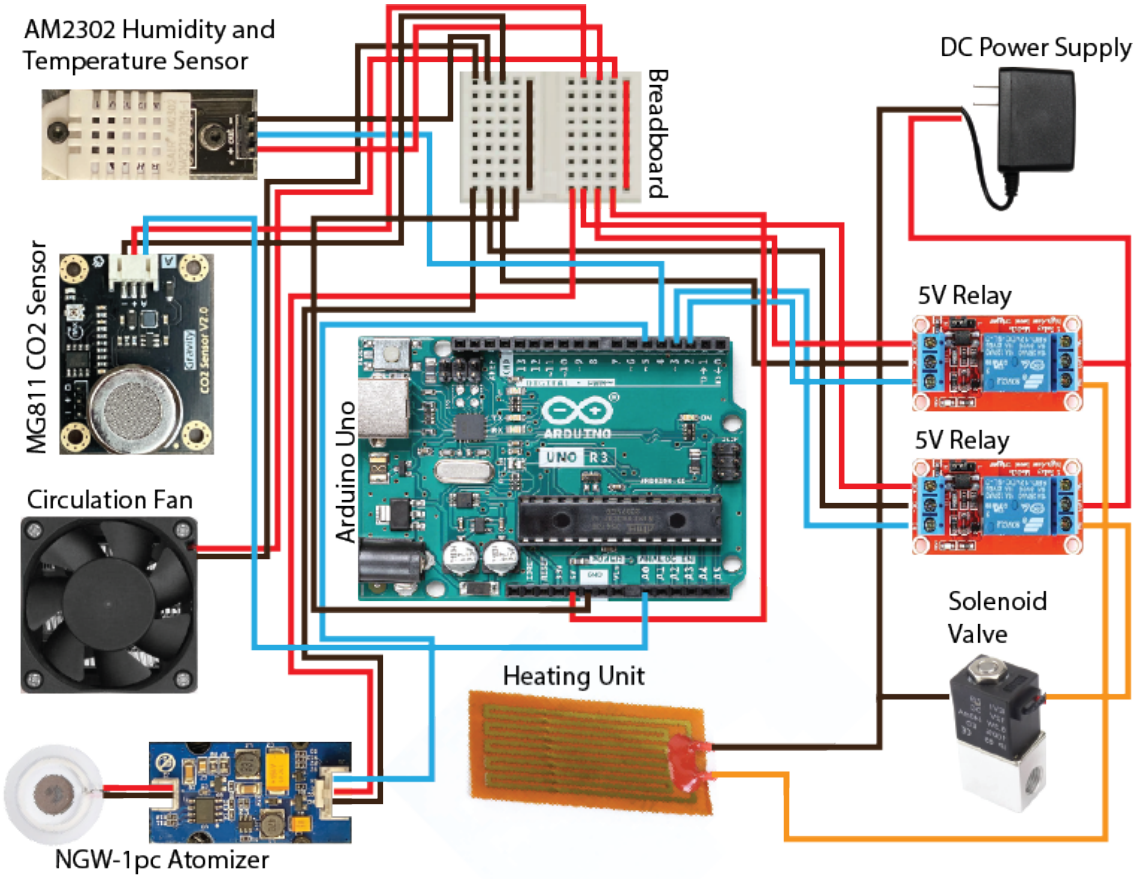

Figure 5.2: Circuit diagram of the device. Red and orange wires: positive connections (before and after passing through relays); black wires: negative connections; blue wires: signal connections

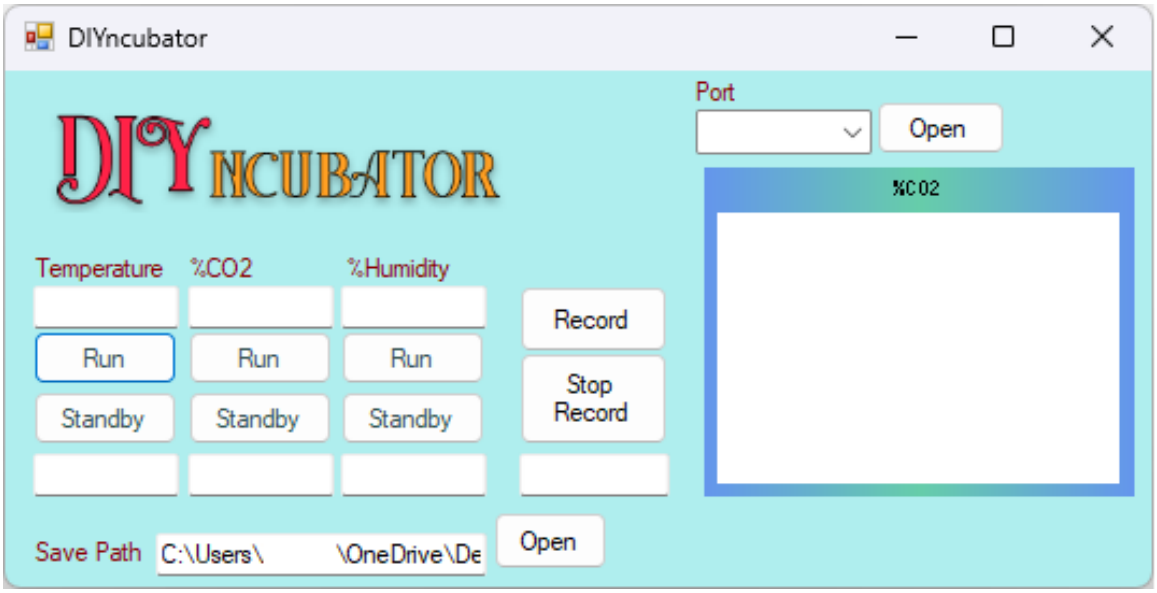

Figure 5.3: Computer control interface for the DIYncubator developed in Visual Studios 2019. Custom software design allows for the implementation of control inputs and measurements that suit the user, in addition to aesthetic freedoms.

via USB connection, and an executable program for user control of the system was developed in Microsoft Visual Studios 2019 for Windows operating systems. The executable program is not absolutely necessary for the device and could be circumvented using the Arduino with an integrated LED display and buttons for operation.

5.3 Results

5.3.1 The DIYncubator Provides Stable Environmental Control Over Long Periods

The DIYncubator was constructed for a total cost of ~$250. Some additional but not required components, such as a plastic project box to house the electronics, DB15 connectors to simplify connections and deployment, and cable cover and management sleeve

brought the cost to ∼$300. It should be noted that purchasing of materials is up to the user's discretion, and more careful sourcing or building custom circuitry may lead to further reduction in cost. For example, some of the materials used were recycled from previous experiments and shipping packaging. Creativity is encouraged in building such a device, provided that the solution is still functional.

Using this feedback and control system (Fig. 5.2), internal CO_2 levels, temperature, and humidity are stably maintained over long periods (Fig. 5.4), limited primarily by the volume of the water container used by the atomizer to supply humidity, which we have found to maintain humidity for up to ∼20 hours. However, the removable lid of the box provides quick access to easily refill the container with minimal perturbation to the sample, allowing for maintenance of conditions essentially indefinitely.

5.3.2 Human Cell Growth Within the DIYncubator Compares Favorably to Commercial Solutions

To test the suitability of the device for live cell culture and imaging, we grew HEK293T cells in a 29 mm circular dish for 24 hours in the DIYncubator and compared the doubling time to the same cell line grown with identical conditions in a commercially purchased ibidi Silver Line stage-top incubator (Fig. 5.5). The doubling time of HEK293T cells inside the DIYncubator was determined to be 39 ± 6.5 hours, compared to a doubling time of 36 ± 7.5 hours in the ibidi Blue Silver incubator.

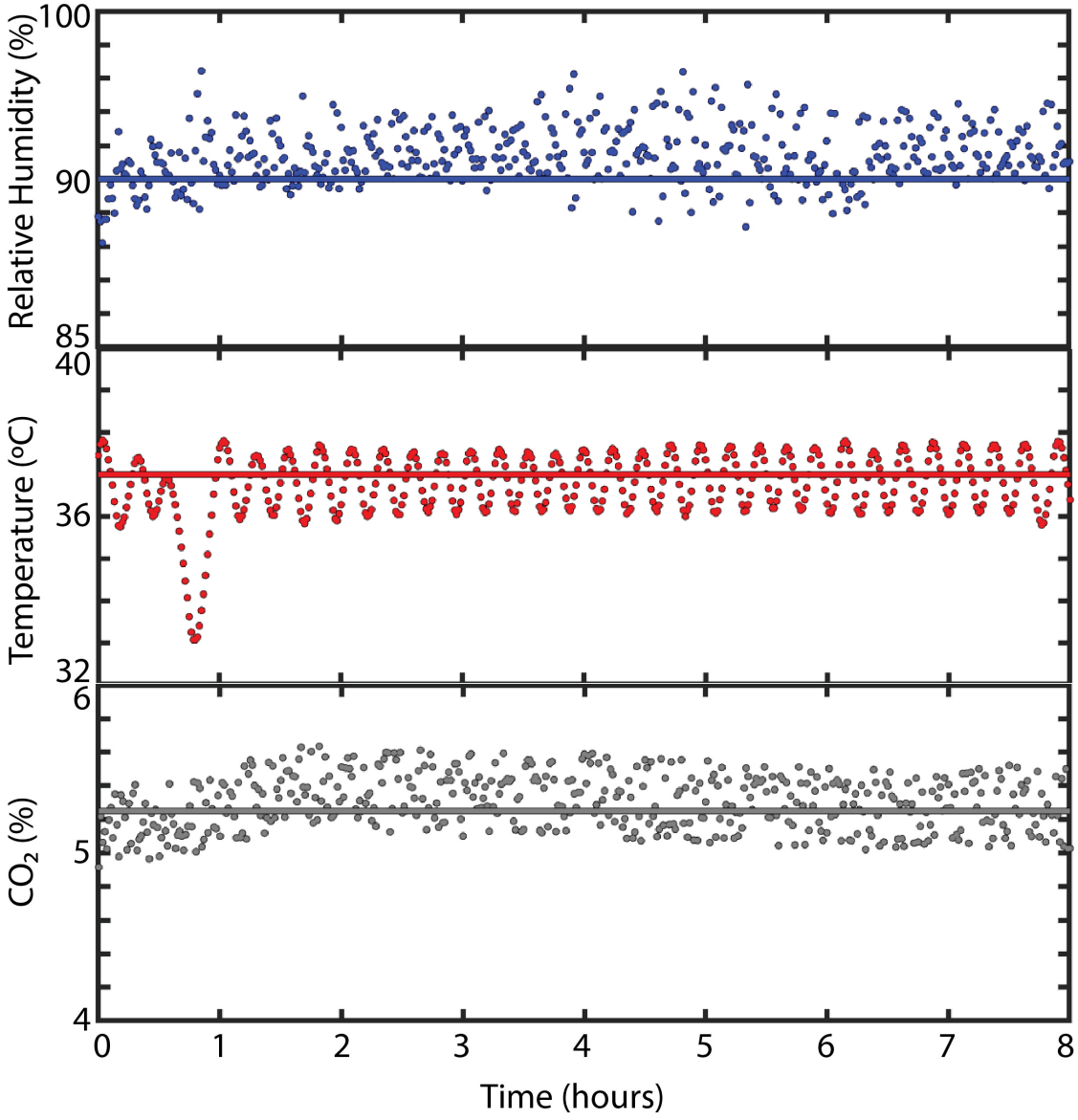

Figure 5.4: Environmental parameters inside the DIYncubator measured at one minute intervals over eight hours. At $t = 0$ the device had been running for 1 hour to come up to temperature, at which point humidity and CO_2 control were engaged. (Top) Percent relative humidity; solid line indicates the set point of 90%, with mean humidity of 90.8 $\pm$ 0.9%. (Middle) Internal temperature, with set point of 37.0 °C, and mean temperature of 36.8 $\pm$ 0.7 °C. (Bottom) Internal CO_2 content, with set point of 5.3%, with equilibrium mean CO_2 content of 5.3 $\pm$ 0.2%

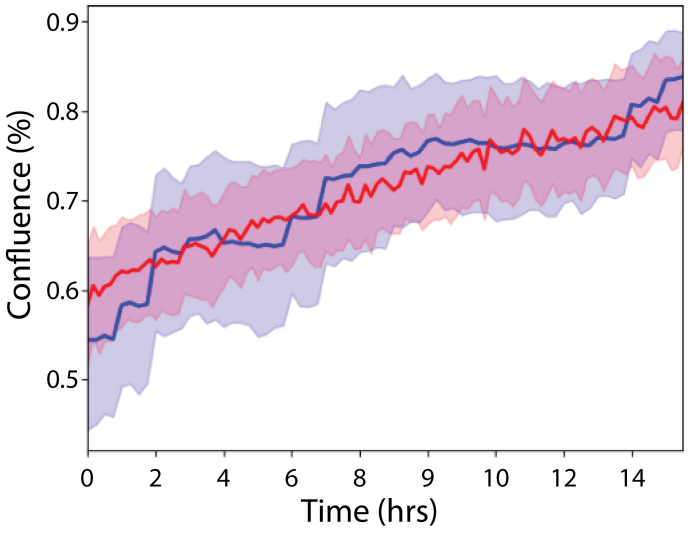

Figure 5.5: Comparison of growth rates between HEK293T cells cultured in an ibidi Silver Line stage-top incubator (blue) vs. the DIYncubator (red). Shaded regions indicate 95% confidence intervals, with solid lines indicating the mean over ∼200 cells.

5.4 Discussion

We have described the design, construction, and implementation of a low cost, easy to build stage top incubation device called the "DIYncubator" for the growth and imaging of human and other mammalian cells. The DIYncubator is able to maintain an environment with controlled temperature, humidity, and CO_2 content over long, multi-day time scales, allowing for fluorescent and brightfield imaging over the long times necessary to, for example, image oscillations in protein expression [10]. This performance compares favorable to commercially available solutions at ∼100x the cost.

However, with the advantage of the low cost comes a variety of disadvantages that should be kept in mind. The DIYncubator, as described here, was fabricated using ABS plastic, which is a low cost material but is relatively flexible and a poor conductor of heat. Additionally, care must be taken with heating approaches as ABS can warp considerably

when exposed to high temperatures. Furthermore, the interior of the DIYcubator can become quite wet over long time periods, as the atomizer sprays moisture into the box and condensation forms on surfaces. As a result, the DIYcubator described here is ideal for fluorescence imaging where illumination is provided through the objective under the sample, but brightfield imaging accomplished by illumination through the removable lid may suffer somewhat as a result of the accumulation of condensation. However, the lid is easily removable and can periodically be wiped clean with minimal perturbation to the sample. Accumulation of condensation and moisture may also play havoc with electronic components over long times. However, after long periods of use we have not had significant failures of electronics, and, in any case, the electronics can easily be replaced many times over before the cost even begins to approach a fraction of that of commercially available systems.

As an alternative, the DIYcubator could be machined out of aluminum, which would yield a number of advantages. First, aluminum is more rigid, and would aid in creating a more airtight environment that would aid in reducing costs of operation by reducing the amount of CO_2 required for a constant environment. Furthermore, the high thermal conductivity of aluminum could be utilized such that, by adhering the heating film directly to the chassis in a variety of locations, the entire body of the box could be used as a heating element to provide a more stable and homogeneous temperature profile across the unit. Finally, heating the chassis of the box in this way, and potentially combined with transparent heating film to make the removable lid, would substantially reduce condensation and therefore provide an environment more amenable to long-term brightfield illumination

and imaging. However, acquiring and machining the requisite materials for this approach would add significantly to the financial and time cost, as well as the skills required, to fabricate the DIYncubator, the suitability of which is left to the discretion of the user.

Chapter 6

SARS-CoV-2 Structural Proteins

6.1 Introduction

The contents of this chapter have been published in the journal Science Advances under the title "Synthesis, Insertion, and Characterization of SARS-CoV-2 Membrane Protein Within Lipid Bilayers". I am a secondary author on this paper, having helped develop the initial methods and optimization procedures for purifying SARS-CoV-2 structural proteins. Yuanzhong Zhang and Aisa Anbir are primary authors, with the former performing AFM and Cryo-EM and the latter performing protein purification. Other secondary authors include Joseph McTiernan, Michael E. Colvin, and Ajay Gopinathan who performed MD and single atom simulations, and Roya Zandi and Siyu Li who performed coarse grained simulations. Further assistance was provided by the lab of Umar Mohideen, and Thomas E. Kuhlman is the corresponding author.

Coronaviruses have caused severe health crises and economic challenges over the centuries. Though the likelihood of another coronavirus type emerging after the eradication

of one has always been high, research on them has often diminished due to limited success in understanding fundamental elements of their functioning and life cycle. This is in sharp contrast to progress on other viruses like HIV and HBV. One of the most significant factors contributing to our limited knowledge of coronaviruses is the difficulty in performing in vitro experiments on viral proteins due to challenges posed by the small size of the proteins and low yield expression methods for them. As an example, while the membrane (M) protein of SARS-CoV-2 is the most abundant protein with a crucial role in the formation of the virus, it is much less well characterized than most other SARS-CoV-2 structural proteins. In this paper, we take important steps toward developing a deeper understanding of M protein structure and function in the physiologically relevant membrane environment by developing an expression technique with orders of magnitude higher yield, inserting the protein into lipid membranes and characterizing the protein-membrane configurations using a combination of AFM, Cryo-EM and MD simulations.

Specifically, using a Small Ubiquitin-related Modifier (SUMO) tag-based expression system, we produced and purified significant quantities of full length, native M protein in Escherichia coli. As far as we are aware, this is the only reported method for producing significant quantities of full length, native M protein, and our method is low cost and produces large quantities of pure native, full length protein with low endotoxin levels. We examined the purity of our M protein product employing western blot and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry and showed that it has low endotoxin activity. Our method produces tens to hundreds of milligrams of protein per liter of

culture, with these high yields enabling many previously inaccessible experiments, including small angle X-ray scattering (SAXS), and Cryo-EM.

In addition, we developed a reliable method for inserting purified M protein dimers into a suspended lipid membrane in a homogeneous orientation, and we used Cryo-EM and AFM to image the M protein dimers in the membrane. We compared our experimental results to all-atom molecular dynamics simulations of the M protein dimers in a lipid membrane to provide structural information about the protein and the surrounding membrane to aid in interpreting the imaging results. While some prior experimental studies [89] revealed two distinct M protein structures, a "short" and a "long" form, other studies only showed the existence of a single M protein structure [90], similar to the short form. Our AFM results, together with our atomistic MD simulations, indicate that M protein dimers within our supported lipid bilayer are entirely in the "short" form identified in earlier structural studies.

Our AFM results also uncovered a thinning of the membrane in the vicinity of the M protein dimer, which we confirmed with MD simulations. Additionally, our MD simulations indicated the propensity of individual M protein dimers to induce local membrane curvature. The line tension resulting from the thinning together with the induced curvature could potentially drive the aggregation viral proteins and the subsequent budding of the virion. To test this idea, we constructed a coarse-grained simulation model of M proteins in a flat membrane and showed that endodomain interactions among M-protein dimers were also critical for budding. Taken together, our work establishes a framework for future studies of viral protein and sheds light on the structure and interactions of M protein dimers

within a lipid membrane yielding insights into important physical processes that drive viral assembly and budding.

6.2 Results

To assess the level of expression of SARS-CoV-2 Membrane (M) protein in E. coli, we constructed a C-terminal fusion of M with the blue fluorescent protein mCerulean3. This fusion protein is borne on the high copy number plasmid pUC57 and expressed from the inducible promoter PlacT7. Upon transformation into the standard expression strain BL21(DE3), we observed extremely low expression regardless of the degree of induction with IPTG (Figure 6.1a). Previous reports have demonstrated that expression of M protein from the original SARS-CoV virus could be significantly enhanced by N-terminal fusion of M with a SUMO tag [91]. Following this same strategy resulted in enhancement of expression by 4-5 times (Figure 6.1b,d). We further optimized yields by quantifying expression of M-mCerulean3 and SUMO-M-mCerulean3 in various strains of E. coli, in particular the Walker strains, C41(DE3) and C43(DE3), which have been specifically selected for high expression of membrane-bound and other difficult proteins. We measured expression in these strains grown in both lysogeny broth (LB) and Terrific Broth (TB), finding that expression of SUMO-M-mCerulean3 in C41(DE3) grown in TB to be 45 – 50 times higher than that of the standard use of BL21(DE3) grown in LB.

Purification of SUMO-M from C41(DE3) grown in TB using standard immobilized metal affinity chromatography (IMAC) resulted in high yields, typically on the order of 50 – 100 mg per liter of culture. Following purification, we refolded the protein by dialysis and

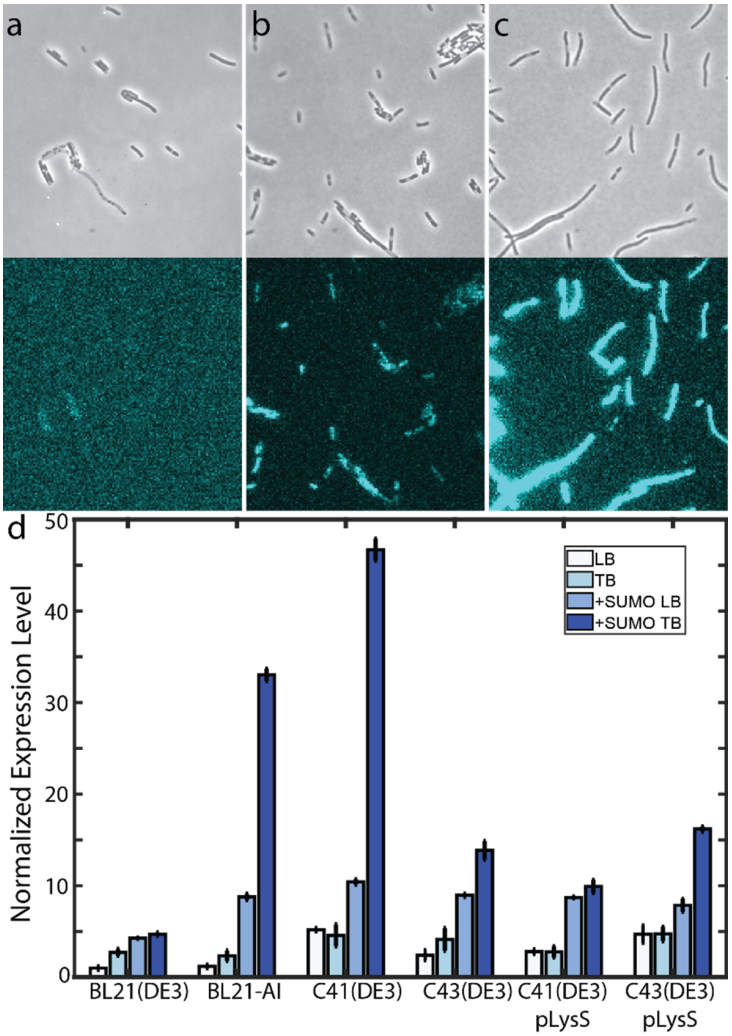

Figure 6.1: Fluorescence microscopy of M-mCerulean3 expression in TB medium, with phase contrast image shown above and mCerulean3 fluorescence channel shown below. **(a)** M-mCerulean3 expressed in BL21(DE3). **(b)** SUMO-M-mCerulean3 expressed in BL21(DE3). **(c)** SUMO-M-mCerulean3 expressed in C41(DE3). All strains were imaged with identical microscope and camera settings, and images are displayed with identical lookup tables (LUTs). **(d)** Quantitative comparison of M-mCerulean3 expression in various E. coli strains and media (white – LB; light blue – TB; SUMO-M-mCerulean3 in LB – medium blue; SUMO-M-mCerulean3 in TB – dark blue. Expression levels are normalized to expression of M-mCerulean3 in BL21(DE3).

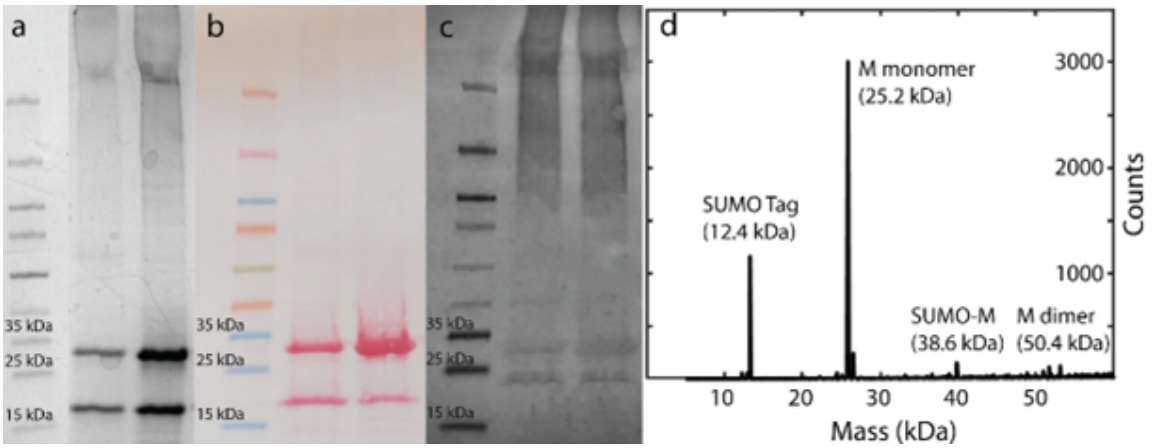

Figure 6.2: **(a)** SDS-PAGE of purified M sample. Lane 1: ladder; Lane 2: After digestion with 0.5 mg SUMO protease per mg M; Lane 3: digestion with 2x amount SUMO protease. **(b)** Ponceau all-protein stain of protein transferred to blot from PAGE gel. Lanes same as **(a)**. **(c)** Western blot of **(a)**. **(d)** MALDI mass spectrum of purified M sample.

removed the 6xHis tag and SUMO tag by digestion with Ulp1 SUMO protease, resulting in pure, full length, native M protein. We quantified the purity by SDS-PAGE, western blot, and matrix-assisted laser desorption/ionization (MALDI) (Figure 6.2) and quantified endotoxin levels by LAL endotoxin assay (0.35 $\pm$ 0.05 EU/mg). SDS-PAGE (Figure 6.2a) and MALDI mass spectrometry (Figure 6.2d) show residual cleaved 6xHis-SUMO tag (12.4 kDa), which was subsequently removed by further IMAC purification. Furthermore, western blot shows (Figure 6.2b,c), in addition to a large band at the predicted M monomer molecular weight of 25.2 kDa, a large streak of high molecular weight multimers and aggregates, as well as a small amount of lower molecular weight products that may be degradation products. However, the majority of detectable products are M monomer and aggregates thereof.

We emphasize that, while E. coli is unable to perform post-translational modifications, such as glycosylation that is predicted to occur post-transcriptionally [92], it is known

for the related SARS-CoV virus that these modifications to M protein are not required for viral assembly and do not significantly impact protein-protein interactions [93, 94, 95, 96].

6.2.1 Cryo-EM of Lipid Bilayer Membrane Reconstituted with M Protein in Large Unilamellar Vesicles (LUVs)

Next, in order to facilitate studies of the M protein in its physiologically relevant context, we developed a procedure for M protein insertion into lipid membranes of large unilamellar vesicles (LUVs) with a phospholipid composition mimicking that of a typical endoplasmid-reticulum-Golgi intermediate compartment (ERGIC) bilayer membrane [described in detail in Methods and in the Supplemental information (Fig. 7.1)]. We then used Cryo-EM and AFM to verify the presence of the inserted M protein and its orientation, as described below. First, Cryo-EM studies of LUVs with and without M protein were done to confirm the presence of M proteins in the bilayer membranes. A representative Cryo-EM image of the M protein embedded in bilayer vesicles is shown in Fig. 6.3a, with the schematic of LUVs with (Fig. 6.3b) and without (Fig. 6.3c) M protein embedded. Figures 3d-f show the typical 2D class average images of Cryo-EM observations from LUVs reconstituted with M protein at a protein-to-lipid mass ratio of 0.015. Two types of 2D class averages were observed in the Cryo-EM images. Figures 3d,e show the M protein as bright spots on the bilayer corresponding to higher electron density of the M protein embedded in the membrane, while in the image in Fig. 6.3f there is no M protein embedded. Most frequently observed spots were 4-5 nm wide, which can be associated with single dimers or possibly some higher order oligomers. We also observed that these spots corresponding to the M protein show a higher intensity on the inner leaflet of the vesicle, indicating they

are likely inserted with the larger and denser C-terminal facing the inside of vesicles. In contrast with regions reconstituted with M proteins, the regions without M proteins show homogeneous intensity from the corresponding uniform electron density throughout the bilayer membrane. In the bottom row as a negative control, blank vesicle samples without M protein were also imaged with Cryo-EM. Figures 3g-i show typical 2D class average images with a homogenous electron density observed for the LUVs without M protein. Lastly, Fig. 6.3j shows a simulation of these Cryo-EM micrographs with M protein calculated from the molecular dynamics simulations of the short form of M protein embedded in a membrane by projecting the atom number density onto a plane perpendicular to the surface of the membrane using a bin size of 0.95 nm x 0.95 nm. Schematics for this MD simulation are shown in Fig. 7.2, with details provided in the methods section. Two different viewing angles are averaged over in creating this simulated image, where lower resolution density maps for both the long and short forms are shown in Fig. 7.3.

6.2.2 AFM Topography of Supported Bilayer (SBL) Reconstituted with Embedded M Protein

As presented above the Cryo-EM confirmed the presence of the M protein in the lipid bilayer. The small molecular weight of 50.2kD for the dimer limits high resolution analysis of the M protein- lipid membrane interaction. The AFM with the height resolution of 0.01nm and a lateral resolution 0.8 nm is ideally suited for understanding the inserted protein morphology, orientation, oligomerization at larger concentration and the protein-lipid bilayer interaction [97, 98]. Figure 4a shows a 1 µm x 1 µm AFM topography of a

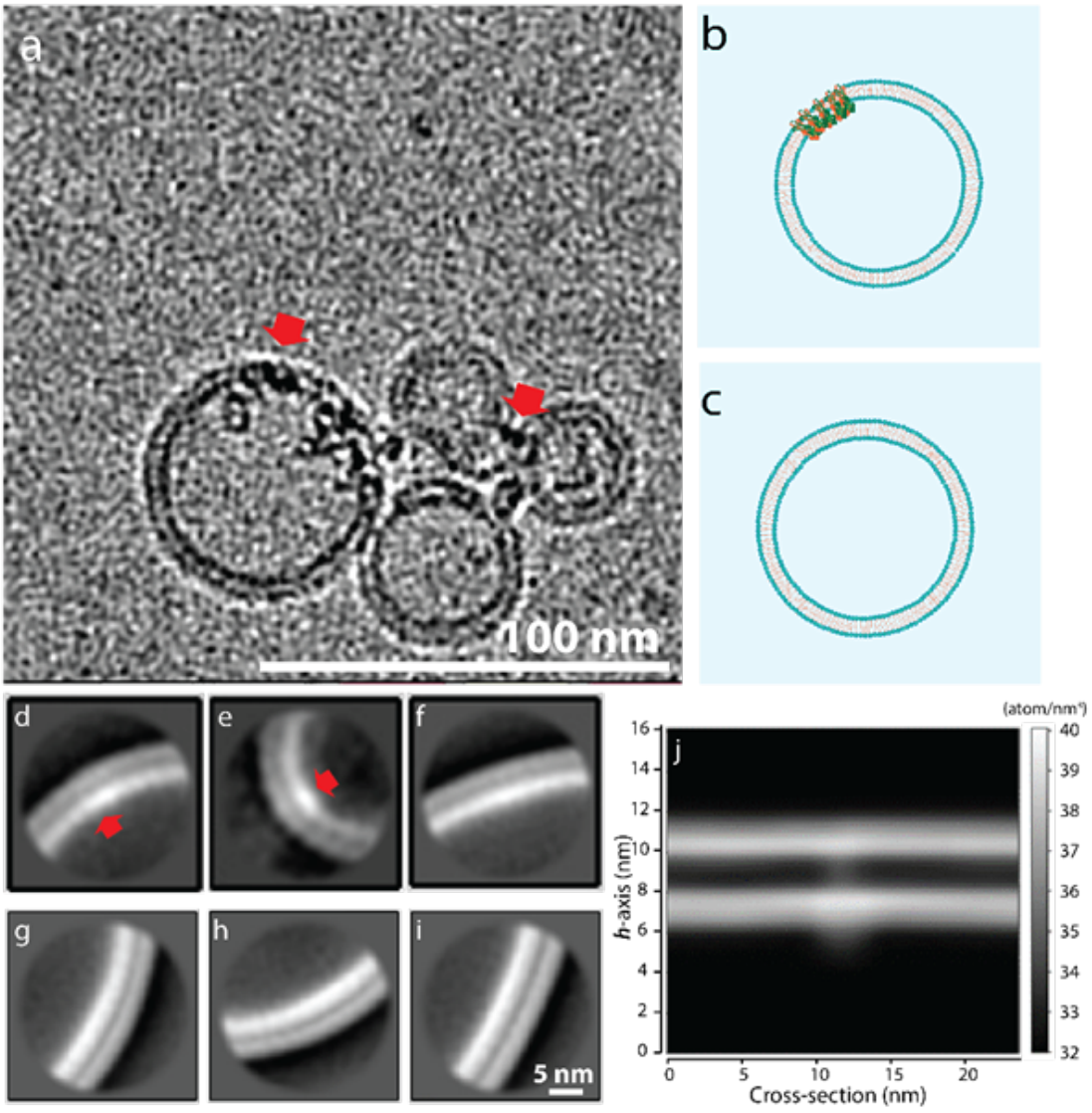

Figure 6.3: (a) Red arrows indicate M proteins inserted in the vesicle membrane. (b) and (c) - schematics of LUVs with and without M reconstitution, respectively. (d) Representative 2D class average image of areas without M protein in reconstituted LUVs. (e) and (f) – two most representative CryoEM 2D class average images from LUV bilayer membranes reconstituted with M protein mass ratio, Rm,M/L = 0.015. Arrows: location of M protein as bright spots. (g-i) 2D class average CryoEM images of blank LUV bilayer membrane. Scale bar = 5.0 nm. (j) Atom number density of the short conformation M protein embedded in a bilayer membrane obtained from all atom MD.

typical endoplasmic reticulum-Golgi intermediate compartment (ERGIC) bilayer membrane generated from LUVs without M protein reconstitution. The height profile along the green dashed line is shown in the lower panel of Fig. 6.4a. The SBL on the mica substrate is smooth with an RMS roughness measured to be 0.3 nm. Figure 4b shows the typical topography of a 1 µm x 1 µm surface of a SBL generated from LUVs reconstituted with M protein at a protein-to-lipid mass ratio of 0.01. In contrast to the smooth surface of the SBL in Fig. 6.4a, scattered single particle protrusions were observed in the AFM image. These can be associated with the embedded protein in the SBL. The line-profile (green dashed line) passing through one of the M proteins shows that the height of the feature above the membrane is 2.4 nm, consistent with the C-terminal height of M protein from literature [89, 90] and simulations presented below. Figures 4c,d show the SBL from LUVs reconstituted at higher protein-to-lipid mass ratios of 0.015 and 0.02 respectively. Figure 4c topography shows a combination of scattered single particles and patches, while patches are mostly observed in Fig. 6.4d. The line profiles show that the patches have the same height as the single particles associated with the M protein. Figure 4e shows the percent of protein-occupied area in the SBL membrane as a function of protein-to-lipid ratio. The area occupied by the protein is calculated from the sum of the single protein and patch areas. The M protein-occupied area increases from 0.3% to 6.2% and 17.1%, as the protein ratio increased from 0.01 to 0.015 and 0.02, respectively.

In addition to examining the planar surface features of the SBL with and without M protein we also characterized the depth profile using AFM indentation studies. Figure 4f shows a typical indentation force profile obtained from three different samples: bare mica

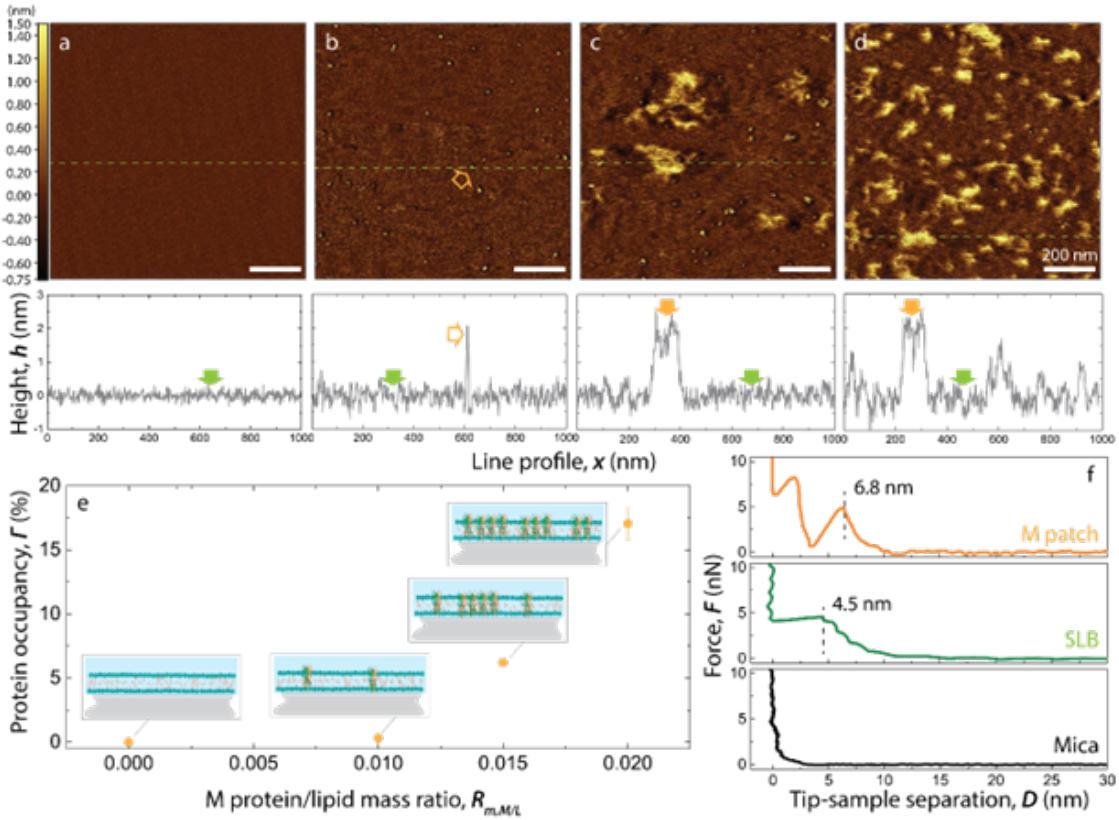

Figure 6.4: (a-d) Height images and representative single line profiles (green dotted lines in images) collected from SBL fabricated from large unilamellar vesicles that were reconstituted with M protein at Rm,M/L = 0, 0.01, 0.015 and 0.02. (e) Area occupation percentage of M protein in SBL as a function of Rm,M/L calculated from A-D. (f) Representative force profiles collected from pristine mica surface (black), SBL (green) and M protein (orange) patches in lipid bilayer, the force profiles were collected during tip engagement.

in the bottom panel; SBL without and with M protein (concentration ratio of Rm,M.L = 0.02) in the middle and top panels respectively. In all three panels of Fig. 6.4f the horizontal axis is the separation distance between the AFM probe tip and the bare mica surface. The bottom panel shows the typical indentation force curve of the bare mica surface in solution with the tip-rigid surface contact at zero followed by a sharp increase in force by the cantilever bending. The middle panel corresponds to the SBL membrane without M protein. Here, the AFM tip comes in contact with the hydration layer at a separation around 7.5 nm with the applied force leading to deformation of the membrane. With increasing applied force, the tip punctures the membrane surface at D = 4.5 $\pm$0.1 nm. This corresponds to the membrane thickness without M protein. With further increase in force the AFM tip penetrates the membrane until it contacts the rigid mica at 0 nm. The top panel shows the indentation force profile obtained from the patch-like structures from the M protein on the SBL. The force showed a discontinuity at 6.1 nm which corresponds to the tip-protein surface interaction. Large forces lead to penetration of the tip into the M protein patch. The discontinuity at 3.2 nm indicates potential disruption of a specific strong interaction region between neighboring M proteins. From the slope of the approach curves in Fig. 6.4f, the Young's modulus can be calculated to be 9.5 MPa on the lipid bilayer and 41.5 MPa on the protein patches respectively. Figure S4 shows the average elastic moduli extracted from nanoindentation force profiles for SBL and the M protein patches. The difference in the Young's modulus indicates that the patches are indeed a different type of material from the lipid bilayer.

6.2.3 Formation of M Protein Aggregate Patches in the Supported Bilayer (SBL)

From Fig. 6.4b we observed that M proteins in SBL start as single M protein assemblies at a M protein to lipid mass concentration Rm,M/L = 0.01. From the size (length, width, and height) they can be identified as M protein dimers. With the increase of M protein concentration in the SBL the dimers aggregate to form patches of M proteins as observed in Figs. 4c and d. As seen in Fig. 6.4e the area of the patches grows nonlinearly with the M protein concentration. The M protein-occupied area increases from 0.3% to 6.2% and 17.1%, as the protein to lipid mass ratio Rm,M/L increased from 0.01 to 0.015 and 0.02. Thus, the M protein incorporation into the membrane is facilitated by the presence of M proteins already in the membrane. Notice that the height of patch-like assemblies is the same as the isolated single M protein dimers observed at the lower M concentration. This indicates that the M proteins assemble in a side-by-side manner inside the bilayer, as demonstrated by schematics shown in Fig. 6.4e; these patches are higher order oligomers.

6.2.4 Morphological Statistics of the Isolated M Protein Particle

We next characterized the properties of single M protein particles embedded in the SBL for protein to lipid mass ratio, Rm,M/L=0.01 (Fig. 6.5). Figure 5a shows a high resolution 250 x 250 nm surface morphology of a typical single isolated M protein at the concentration ratio of Rm,M/L= 0.01. In Fig. 6.5b the characterization parameters the Length (L) and Width (W) are shown. The boundary of the protein particle is identified as an edge higher than the planar membrane height. The planar membrane height is found

by averaging the membrane height far away from the particle (gray region, Fig. 6.5b). The particle thus identified is shown in red in Fig. 6.5b. The Length (L) of the particle is defined as the maximum distance between any two points on the perimeter of the protein particle. The particle Width (W) is defined as the maximum distance between two points on the perimeter, in the direction perpendicular to length. The particle Height (h) is the height of the highest point on the protein surface. For comparison, Fig. 6.5c shows the final frame of an all-atom molecular dynamics simulation for the short conformation M protein embedded in a multicomponent membrane (see Fig. 7.2 for schematics). The analysis of the AFM data is repeated for 214 samples and the histogram of the results for L, W, h and the ratio L/W are shown in Fig. 6.5d-g. Corresponding size classifications from molecular dynamics simulations are shown in Fig. 6.5d-g, averaged over the last 100 ns of the simulation, through the dotted purple lines (see Figs. S5-S7 for more information). Protein length and width were obtained by fitting an ellipse to the projection of the protein above the membrane, while height was based on the difference between the top of the M protein C-terminal and the phospholipid heads adjacent to the protein. The histogram in Fig. 6.5d shows a bimodal distribution which when fit to two normal distributions lead to mean particle lengths, L1 = 4.8 ± 0.1 nm and L2 = 9.4 ± 0.2 nm. Figure 5e shows the width distribution, again seen to be bimodal, which when fit to two normal distributions leads to mean values of W1 = 2.3 ± 0.1 nm and another at W2 = 5.7 ± 0.1 nm. In Fig. 6.5f the aspect ratio 'L/W' again shows two normal distributions, one at L/W1 = 1.9 ± 0.1, the other at L/W2 = 3.1 ± 0.1. Figure 5g shows the histogram obtained from the height

of the M protein particles. The values can be fit to one normal distribution with a mean height h = 2.4 ± 0.1 nm.

6.2.5 M Proteins are Confirmed to be Inserted in the Bilayer Membrane

High resolution Cryo-EM studies have shown that SARS-CoV-2 membrane protein dimers exist in a conformational equilibrium between the long form and the short form, which are 7.2 nm and 8.6 nm in height, respectively [89]. Comparing these heights with Fig. 6.5g, AFM shows the height of exposed protein particles is much shorter (2.4 nm). This is in agreement with the all atom MD simulations, where the height ranges from 2.6 nm to 2.8 nm depending on the protein conformation (short/long, see Fig. 7.5b). The height measured by AFM is more consistent with the short form height from simulations; a more thorough discussion of these height fluctuations can be found in the Supplementary Information (Fig. 7.8). The slight differences between AFM and MD can be attributed to the protein deforming slightly as the AFM tip makes contact. Meanwhile, the lateral dimensions of isolated protein particles observed from AFM in Fig. 6.5d,e are very similar to the literature reported values of 5.0 nm and 5.7 nm for long and short form respectively [89]. A much smaller height and similar lateral dimensions indicate that the M proteins are inserted into the bilayer membranes instead of absorbing on top of the membranes. This is also supported by our Cryo-EM observations in Fig. 6.3d,e. The membrane reconstituted with M protein shows a higher electron density compared with blank membrane. This indicates that a large portion of the protein is buried inside the bilayer membrane.

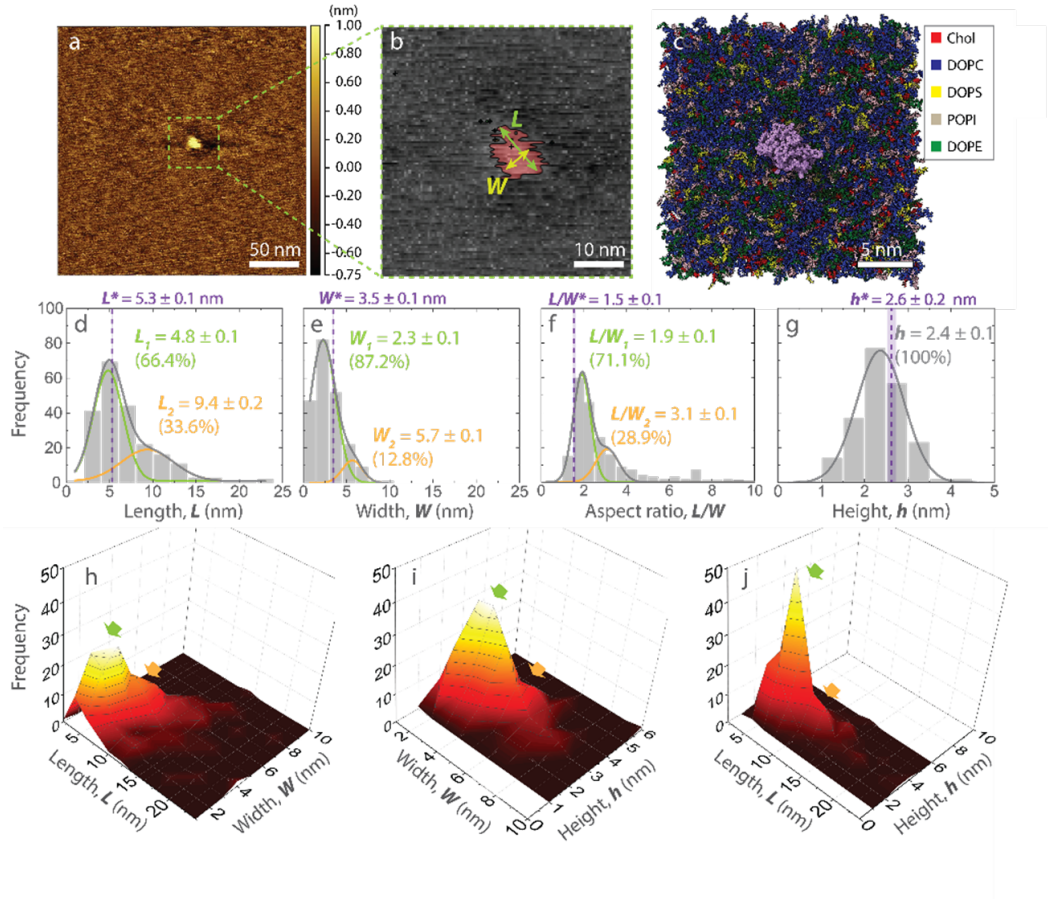

Figure 6.5: (a) Representative AFM height image of individual M protein reconstituted in SBL at Rm,M/L = 0.01. (b) Zoomed-in image of (a), highlighting the protein with red color and lipid bilayer with gray color. The protruded portion of a single M protein particle above lipid bilayer (left) is detected by setting the height of top surface of SBL as a threshold and taking the above portion. (c) Top view of the short conformation M protein (C-terminal in purple) embedded in a multicomponent membrane 1 µs into an all atom MD simulation. Dimensional parameters: length, L, width, W, and height, h, were measured for detected protein particles. (d-g) Histograms of L, W, L/W, and h of protein protruding above lipid membrane. Solid lines are Gaussian fits. Dotted vertical lines mark corresponding values from MD simulation for short-form M protein reconstituted with C-terminal exposed above SBL. (h-j) 3D histograms each showing combined statistical survey of two out of the three dimensions. Green and orange arrows indicate dimer and higher order of oligomers.

6.2.6 Orientation of Reconstituted M Protein has C Terminal Facing the Vesicle Interior

In Fig. 6.3d,e we observed that the bright spots corresponding to reconstituted proteins show a higher electron density in the inner leaflets. This indicates that the C terminus, which is larger than the N terminus and contains a higher density of electron-rich "heavy" atoms, is inserted facing the inside of the vesicles. This is further confirmed in AFM measurements and also matches the simulated Cryo-EM image from MD in Fig. 6.3j. This stems from averaging the protein over different angles, since the N-terminus of the protein on the outer leaflet ranges from small to large arc length along the membrane. Alternatively, the C-terminal domain of the protein along the inner leaflet is largely independent of the viewing angle and will have greater arc length than the N-terminal domain (see Fig. 7.3 for viewing angle differences). In Fig. 6.5g, the height of reconstituted M protein exposed above bilayer membrane show a single distribution with an average height of 2.4 nm, which is consistent with C terminal size observed in high resolution Cryo-EM images and our simulation predictions, indicating that all M proteins in the SBLs observed in AFM are oriented with their C terminals facing the AFM probe. From these results we can deduce that the M proteins were reconstituted in LUVs with their C-terminal encapsulated inside the vesicles. This is because, during the formation of the SBL on mica, the LUVs adsorb to the mica surface with their outer leaflets. Subsequently the LUVs rupture, exposing their inner leaflets facing upwards and accessible to the AFM probe.

6.2.7 Reconstituted M Protein Shows Dimensions and Membrane Thickness Consistent with the Short Form

In Figs. 5h-j the three-dimensional histograms analyze the relationships between protein length (L) and width (W), width and height (h), as well as length and height, respectively. In Fig. 6.5h from the L and W, we can observe two distinct populations with their individual lengths and widths. From Fig. 6.5i plot of the W and h we can observe that both W populations have the same height above the membrane. From Fig. 6.5j the same is observed for the two height populations. Combining the above information with results shown in the previous two-dimensional histograms (Figs. 5d-g), the most frequently observed M protein population in the SBL have a length, width and height of 4.8 $\pm$ 0.1 nm, 2.3 $\pm$ 0.1 nm and 2.4 $\pm$ 0.1 nm, respectively. These dimensions show good agreement with the short form of M protein dimer both from published Cryo-EM studies of M protein in detergent micelles [89] and our simulation predictions (averages shown with the purple lines in Figs. 5d-g, see Fig. 7.5-7.7 for the time evolution of these quantities). The slight discrepancy between the short conformation MD dimensions and AFM results likely stem from the original MD protein structures not including every residue and deformations in the protein from the AFM tip. The second population observed has a much larger size with a length, width and height of 9.4 $\pm$ 0.2 nm, 5.7 $\pm$ 0.1 nm, and 2.4 $\pm$ 0.1 nm, respectively. These dimensions correspond to that of a multimer of M protein dimers.

6.2.8 Reduction of Membrane Thickness around M Protein

Figure 6a shows a typical isolated M protein particle in a magnified view, with color scale corresponding to the membrane thickness shown to the right. We observed that the membrane thickness in proximity to the single M protein particle is reduced. The height of the membrane around each protein particle is measured along two perpendicular axes, shown in Fig. 6.6a as dashed green lines. A similar view from above of the M protein short conformation obtained through all atom MD simulations can be seen in Fig. 6.6b, where red regions represent a thicker membrane and blue thinner (see Fig. 7.9 for unshifted radial plot). The circular cross section of the protein shown in gold arises due to the protein rotation in the membrane during the simulation shown in Fig. 7.10. Figure 6c shows an exaggerated diagram of the membrane thinning near the protein and a corresponding side view of the thickness after 1 µs of molecular dynamics simulation is displayed in Fig. 6.6d. The change in membrane height from the edge of the protein (0 nm) along the dashed lines for 25 such proteins is shown in Fig. 6.6e. Additionally, the red points represent shifted membrane thickness values from Fig. 6.6b, based on the distance from the edge of the protein. The mean change in height with distance from the protein edge is shown by the solid green line. We observed that the bilayer membrane is compressed until the height returns to the normal thickness 12 nm from the edge of the protein. The maximum membrane compression at the edge of the protein has a mean value of 0.5 nm. It is to be emphasized that this AFM-determined thinning profile matches the MD short form thinning profile, as opposed to the long form (long form shown in Fig. 7.9) and agrees with

observations from Neuman et al. for SARS-CoV [99], where the short conformation is found in thinner regions of membrane as opposed to the long form.

6.2.9 Membrane Compression, M Protein Aggregation, and Spontaneous Membrane Curvature

The membrane thickness reduction around M protein observed in Fig. 6.6 is likely due to the height mismatch between the hydrophobic transmembrane domain and the thickness of the lipid bilayer. From the literature [100], and confirmed with our all-atom MD (Fig. 7.5d), the transmembrane region of the M protein dimer is around 4 nm. The experimentally measured thickness of the lipid bilayer is 4.5 nm as shown in Fig. 6.4f. Thus, the protein hydrophobic region is 0.5 nm smaller than the membrane thickness. This hydrophobic mismatch leads to compression of the bilayer in the immediate proximity of the M protein that we show in Fig. 6.6c-e. The thinning of the membrane in the vicinity of the protein results in an energetic cost due to bend and tilt deformations in the lipid layers, which gives rise to an effective line tension [101]. Using our measurements of the mechanical moduli of the membrane, informed by estimates from literature [102] and the magnitude of the thinning, we can estimate the effective line tension to be at least about 0.35 pN (see Chapter 7). This value is consistent with estimates and measurements of line tension for lipid domains with comparable height mismatches [101, 103], which control the growth and size of lipid ordered domains [103, 104], suggesting that the line tension can facilitate the aggregation of M into patches. Our coarse-grained simulation (see next section) shows that this patch formation leads to the spontaneous curvature of the membrane. Additionally, all

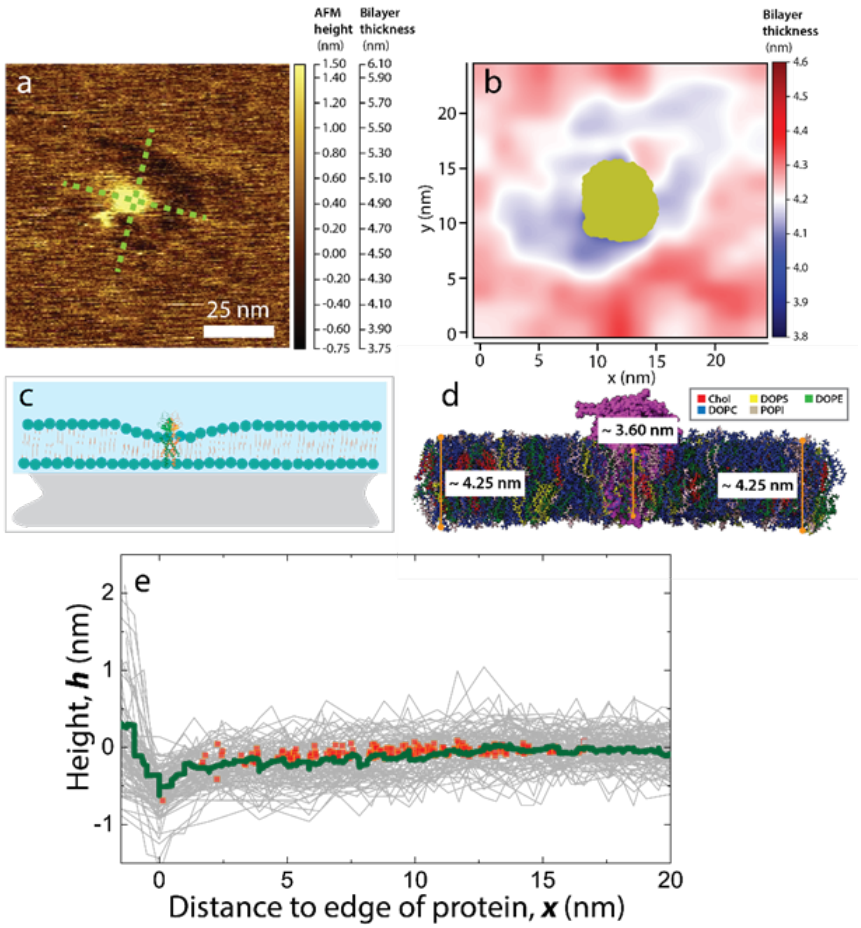

Figure 6.6: (a) Example of individual M protein particles reconstituted in SBL. Height profiles were taken along the green dotted lines starting from the center of the protein. (b) View from above of membrane thickness in nm from 500 ns – 1 µs of an all-atom MD simulation for the short conformation M protein embedded in a multicomponent membrane. (c) Schematics of membrane thinning around M protein particles. (d) Side view of the MD simulation along the y-axis, with the protein in purple, where average thickness is shown dependent on distance from the protein. (e) Solid green line shows the experimentally measured averaged height profile of membrane around individual M protein particles. Inset shows around 100 height profiles obtained from 25 individual protein particles. Here x = 0 nm is defined by the boundary between the M protein particle and lipid membrane. The red points represent vertically shifted values of membrane thickness obtained through MD shown in (b).

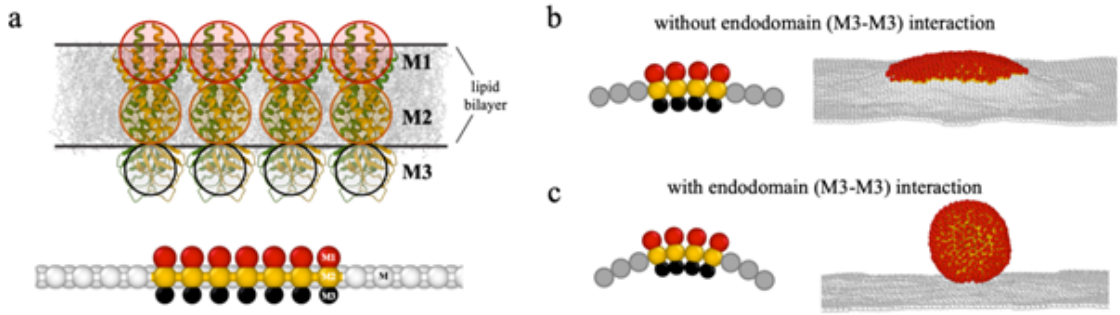

Figure 6.7:

atom MD simulations of both M protein conformations show the propensity to individually induce membrane curvature, as seen in Fig. 7.11. However, the attractive interaction between the C-terminal endodomains is essential for the membrane budding (Fig. 6.7). Interestingly, our coarse-grained ERGIC simulation reveals that N proteins and RNA can effectively introduce membrane curvature as well, providing insight into the mechanism underlying virus budding and encapsulation (Fig. 7.12).

6.2.10 Coarse-grained Molecular Dynamics Simulation of M Protein-induced Budding

In general, our understanding of virus assembly is restricted by the limitations in the spatial and temporal resolution of experimental images, which hinder our ability to observe intermediate structures and states [105, 106, 107, 108]. To address this challenge, coarse-grained simulations have been extensively utilized, successfully improving our understanding of virus assembly processes [109, 110, 111, 112, 113, 114, 115]. Here, we also perform a series of coarse-grained simulations by modeling M proteins and a flat lipid bi-

layer to illustrate how the inserted M proteins can give rise to virus budding, as shown in Figure 7a. Each M protein is built from three hard particles (M1, M2, and M3), where M1 and M2 correspond to the transmembrane domain, and M3 denotes the endodomain. The membrane is modeled as a triangular lattice with hard particles (M) occupying the vertices, on account of the lipids' incompressibility. Since most of experiments and all-atom simulations in this paper are performed on the flat membrane, we also employ a flat membrane to investigate the role of membrane protein on curvature. The simulation results are shown in Figures 7b,c, where M proteins interact with each other through Lennard-Jones potentials. Figure 7b shows that transmembrane domain interaction (M2-M2) cannot give rise to M protein-induced membrane curvature leading to the membrane budding even if the budding particle has lower energy. This is mainly due to the presence of the energy barrier required to bend the membrane. Consistent with the results for a curved ERGIC membrane [116], we find that an attractive endodomain interaction is necessary under all the conditions examined to overcome the bending energy penalties. Figure 7c shows that as we increase the M protein endodomain interaction (M3-M3), the membrane with the embedded M proteins can bud and form a spherical shell. We note that N proteins and RNA can also effectively provide the M protein endodomain interactions and help M proteins bud [116], see Fig. 7.12. The budding prerequisites could indicate the mechanism of how coronavirus can encapsidate N proteins and RNA complex into an infective particle [116].

6.3 Discussion

This paper reports a new expression and purification protocol for SARS-CoV-2 M protein using SUMO tags, which enables us and other members of the scientific community to carry out many new experiments to better understand viral protein function and the life cycle of SARS-CoV-2. We have optimized the method to obtain a very high yield of 50 – 100 mg of protein per liter of culture. Furthermore, we developed a method to reliably insert M protein into phospholipid bilayers in vitro, and our Cryo-EM images confirm that expressed M protein dimers are inserted homogeneously oriented into large unilamellar vesicles. Interestingly, we can precisely control the M protein concentration by adjusting the protein-to-lipid mass ratio used to create the vesicles. Production of large unilamellar vesicles with adjustable concentrations of inserted M protein constitute a new experimental platform for studying the behavior and binding properties of vesicles with realistic densities of M protein.

To create a comprehensive picture of the structure and orientation of M protein in the membrane and the proteins' effect on the surrounding membrane, we deposited the M protein-containing vesicle as a supported membrane on a mica surface that allows high resolution AFM imaging. This reveals the supported bilayer retains a mixture of homogeneously oriented M protein dimers and multimers. This protocol results in the M protein C-terminus being exposed on the surface of the suspended bilayer, allowing measurements of M protein binding to other SARS-CoV-2 structural proteins and viral RNA.

In addition to AFM imaging, we performed atomistic molecular dynamics simulations of the M protein in the suspended membrane. Our AFM and MD results indicate that the M protein is oriented perpendicular to the plane of the membrane, with the top of the C-terminus extending 2.4 nm above the surface of the membrane, while the N-terminus is virtually flush with the membrane surface on the lower leaflet. The membrane-inserted dimer has an elliptical cross section with major and minor diameters of 4.8 and 2.3 nm, respectively. Additionally, a second population of embedded protein is found with the same height, but diameters of 9.4 and 5.7 nm, corresponding to aggregates of dimers. This orientation of the M protein with the C-terminal endodomain extended above the membrane facilitates its proposed role as an anchoring site for other structural proteins.

A particularly interesting result from our study is the structure of the membrane in the immediate vicinity of the M protein dimer. Both the AFM images and the atomistic MD show a region of decreased membrane thickness surrounding the protein, on the order of 0.5 nm which gradually returns to the normal thickness over a distance of 12 nm. This thinning is likely due to the mismatch in the width of the hydrophobic band in the transmembrane region of the M protein dimer (4 nm) and the hydrophobic lipid region of the membrane (4.5 nm), which induces a line tension in the membrane estimated to be 0.35 pN. This line tension can drive aggregation of M protein dimers, even in the absence of specific inter-dimer attraction. Our membrane is suspended on a rigid surface, so we cannot directly observe via AFM whether the M protein dimers cause membrane bending, but our atomistic MD simulations show that the dimer induces a persistent fold in the membrane (which may manifest as a concave indentation for a simulation of a larger membrane patch).

To connect the molecular scale M-protein properties to the mesoscale formation of a coronavirus from all its constituent parts, we constructed a coarse-grained model of a membrane containing 14% M protein dimers. We then carried out MD simulations on this model to understand the combined effects of the endodomains, line tension and spontaneous curvature generation in viral budding . Our simulations revealed that budding and spherical virion formation occur only if M proteins possess an intrinsic curvature and interact with each other via the endodomains. The experimental and simulation techniques developed in this study to date open the door to rapid progress towards fully elucidating the function of M protein. A full understanding of the role of M proteins in driving the budding and release of the newly formed SARS-CoV-2 virion from the cell can pave the path for achieving alternate ways of disrupting their formation.

6.4 Materials and Methods

6.4.1 Strains and Media

For protein expression and purification experiments we employed the commonly used protein expression strains BL21(DE3) (ThermoFisher Scientific) [117, 118], BL21-AI (ThermoFisher Scientific), C41(DE3) [119], C43(DE3) [119], C41(DE3) pLysS [119, 120], and C43(DE3) pLysS [119, 120]. Cells were grown in Lysogeny Broth - Miller (LB-Miller) [10g tryptone (Bacto, ThermoFisher Scientific), 5g yeast extract (Bacto, ThermoFisher Scientific), 10g NaCl (Millipore Sigma), and ultrapure H2O (Milli-Q IQ 7000, Millipore Sigma) to one liter medium [121, 122]] or Modified Terrific Broth (MTB) [20g tryptone, 24g

yeast extract, 4mL glycerol (Fisher Scientific), 0.017 M KH2PO4 (Millipore Sigma), 0.072 M K2HPO4 (Millipore Sigma), and ultrapure H2O to one liter medium [123]].

6.4.2 Plasmid Construction

Plasmids encoding and expressing M with an N-terminal 6xHis tag and with and without SUMO tag fusions were designed in VectorNTI software (ThermoFisher Scientific) and synthesized de novo by GENEWIZ/Azenta Gene Synthesis service. Constructs were synthesized to be expressed from a T7lac promoter [124] with a consensus Shine Dalgarno ribosomal binding site and were cloned into plasmid pUC57-kan-GW. A fluorescently tagged variant, M-mCerulean3 [14] was designed and produced similarly with C-terminal fluorescent fusion.

6.4.3 Cell Growth and Protein Expression and Purification

Cells were grown identically for all experiments. First, a seed culture was started by inoculation of 25 mL medium containing 25 µg/ml kanamycin in a baffled 125 ml Erlenmeyer flask from freezer stock and allowed to grow overnight at 37°C in a C76 shaking water bath (New Brunswick Scientific). This culture was used to inoculate 1 L of medium containing 25 µg/ml kanamycin in a 2 L baffled Erlenmeyer flask and grown at 37°C in a C76 shaking water bath until the optical density at 600 nm (OD600, SmartSpec Plus spectrophotometer, Bio-Rad) reached $0.5 - 0.6$. At this point, the flask was removed from the water bath and transferred to a shaker (Dura-Shaker, VWR) at room temperature. After cooling to room temp, the appropriate concentration of inducer to effect full induction was

added [0.4% w/v L-arabinose for BL21-AI or 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for all others], and the culture allowed to grow overnight to saturation.

6.4.4 Quantification of Protein Expression and Microscopy

To quantify expression of M protein in various E. coli strains and media, cells transformed with the indicated plasmids expressing variants of M-mCerulean3 were grown as described above. 2 mL of culture was then collected and centrifuged (10,000g Eppendorf 5424) and the supernatant medium aspirated away. Cells were then washed with phosphate buffered saline (PBS) three times, and finally resuspended in fresh PBS. Cells were diluted with PBS such that OD600 < 1.0 and distributed to the wells of a 48 well plate (Corning Costar). The plate was placed in a CLARIOstar microplate reader (BMG Labtech) and the OD600 and mCerulean3 fluorescence measured (Excitation at 430/20 nm, emission at 480/20 nm). Expression levels reported in Figure 1 are calculated as mCerulean3 fluorescence divided by OD600. Each measurement was performed with three technical replicates over three experimental/biological replicates.

Fluorescence microscopy was performed on a Nikon Ti-E inverted fluorescence microscope with an Andor DU-897 Ultra camera. For imaging of M-mCerulean3 fluorescence, cells were imaged with a 100x TIRF objective and illuminated with highly inclined 457 nm laser light from a Milles Griot argon laser fitted on a Nikon LU-4A laser launch and using a Chroma filter cube with Z457/10x excitation and ET485/30m emission filters.

6.4.5 Protein Purification

After growth to saturation as described above, the 1 L culture was spun down for 30 min at 4,300g in an Eppendorf 5910Ri refrigerated centrifuge maintained at 4°C and the resulting supernatant discarded. Cells were resuspended in 10 mL lysis buffer (50 mM NaH2PO4, 300 mM NaCl, 10 mM imidazole, pH = 8.0) with 1 mg/ml lysozyme (Millipore Sigma) and 50 U/ml benzonase nuclease (Millipore Sigma) and incubated at 4°C for 30 minutes on a nutating shaker. The cells were flash frozen in liquid nitrogen, followed by incubation in a room temperature water bath until thawed. This slurry was then centrifuged at 10,000 g at 4°C for 30 minutes. The supernatant clarified cell extract (soluble fraction) was collected and stored at -80°C until purification. The pelleted cell debris was then washed three times in inclusion body wash buffer [PBS with 25% w/v sucrose, 5 mM EDTA, and 1% Triton X-100] followed by centrifugation at 20,000 g for 30 minutes at 4°C. After the final wash, inclusion bodies were resuspended in denaturing lysis buffer [8 M Urea (Millipore Sigma), 10 mM Tris (Fisher Scientific), 100 mM NaH2PO4 (Millipore Sigma), 50 mM 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS, Millipore Sigma), 0.3% N-lauroyl sarcosine (Millipore Sigma), 1 mM dithiothreitol (DTT, Millipore Sigma), pH = 11.0] with nutation at room temperature for 30 min. This insoluble fraction was then stored at -80°C until purification.

For purification, samples were processed using a Bio-Rad NGC Quest 10 Plus fast protein liquid chromatography (FPLC) apparatus with a 5 mL EconoFit Nuvia IMAC affinity column (Bio-Rad). To purify protein from the soluble fraction, the column was first equilibrated with Wash Buffer A (50 mM NaH2PO4, 300 mM NaCl, and 20 mM imidazole).

The sample was loaded onto the column and washed with five column volumes of Wash Buffer A. The column was then washed with a linearly increasing mixture of Wash Buffer A mixed with Wash Buffer B (50 mM NaH2PO4, 300 mM NaCl, and 250 mM imidazole) from 3 – 100% Wash Buffer B composition over ten column volumes. Finally, the column was washed with five column volumes of Wash Buffer B. Throughout, the composition of the eluate was monitored by measuring the absorption at 280 nm, and those samples containing the protein of interest were collected and pooled for further processing. To purify protein from the insoluble fraction the procedure was identical using Denaturing Wash Buffer A (50 mM NaH2PO4, 300 mM NaCl, 50 mM CAPS, 0.3% N-lauroyl sarcosine, and 20 mM imidazole, pH = 11.0) and Denaturing Wash Buffer B (50 mM NaH2PO4, 300 mM NaCl, 50 mM CAPS, 0.3% N-lauroyl sarcosine, and 300 mM imidazole, pH = 11.0). All purified SUMO-fused proteins eluted as a single peak.

6.4.6 Protein Folding and Dialysis

To refold denatured proteins purified from the insoluble fraction, samples were dialyzed in regenerated cellulose dialysis membrane (3.5 kDa MWCO, Spectra/Por) over two days at 4°C in 10x volume of refolding buffer (20 mM Tris and 10% glycerol, pH = 8.0) with periodic replacement of buffer at least four times over the course of dialysis.

6.4.7 Cleavage of SUMO tags

To remove the SUMO tag, purified protein (after refolding, if necessary) was mixed with an appropriate volume of 10x SUMO protease cleavage buffer [500 mM Tris, 2% igepal NP-40 (Millipore Sigma), 1.5 M NaCl, and 10 mM DTT, pH = 8.0] along with an excess of

Ulp1 6xHis-SUMO protease previously purified from the E. coli soluble fraction as described above. Digests were incubated in a 30°C water bath overnight.

6.4.8 Removal of SUMO and SUMO Protease for Final Purification

To remove the cleaved 6xHis-SUMO tags and SUMO protease, protein digests were mixed with 1 mL HisPur Ni-NTA resin (ThermoFisher Scientific) and placed on a rotating mixer maintained at 4°C overnight. After incubation, the samples were applied to an empty Poly-Prep chromatography column (Bio-Rad) and the purified proteins collected as the flowthrough. If higher concentrations were required for a given application, these samples were concentrated using a Pierce PES protein concentrator with appropriate molecular weight cutoff (MWCO, ThermoFisher Scientific).

6.4.9 Quantification of Yields and Endotoxin Levels

During purification, absorbance at 280 nm was measured by the NGC Quest 10 Plus optical detector and yields calculated using extinction coefficients estimated using the ExPASy ProtParam online tool based on protein sequence [125]. Estimated concentrations were verified using RC DC protein assay (Bio-Rad). Briefly, 5 µl of protein samples and standards (Protein Standards I and II, Bio-Rad) were added to 25 µl reagent A' and mixed in the wells of a 96 well plate (Corning Costar). To this, 200 µl reagent B was added to each well, mixed, and allowed to incubate at room temperature. After 15 minutes, the absorbance at 750 nm was measured using a CLARIOstar microplate reader, and the concentrations of the samples determined by comparison to linear regression of 2x dilutions of the protein standards.

Endotoxin levels of purified samples were determined using a ToxinSensor chromogenic LAL endotoxin assay kit (GenScript) according to the manufacturer's instructions. 16.67 µl of samples, endotoxin standards, and water as a negative control were added to the wells of a 96 well plate (Corning Costar). To this, 16.67 µl of LAL was added to each well and mixed. The plate was incubated at 37°C for 10 minutes. 16.67 µl reconstituted chromogenic substrate was added, mixed, and incubated at 37°C for six minutes. Then 83.3 µl color stabilizer 1 was added and mixed, followed by 83.3 µl color stabilizer 2, and finally 83.3 µl color stabilizer 3. The absorbance of each sample was measured at 545 nm in a CLARIOstar microplate reader, and the endotoxin levels of the samples were determined by comparison to linear regression using 2x dilutions of the endotoxin standards.

6.4.10 SDS PAGE and Western Blots

To perform SDS polyacrylamide gel electrophoresis (SDS-PAGE), 100 ng of each purified protein was mixed with an appropriate volume of 2x Laemmli sample buffer (Bio-Rad) with 5 mM β-mercaptoethanol (Millipore Sigma). Samples were heated at 95°C for 5 minutes to denature and reduce the proteins and were then separated by electrophoresis through 4-15% Mini-PROTEAN TGX Stain-Free gel (Bio-Rad) with Tris/Glycine/SDS running buffer. Gels were subsequently stained with Oriole fluorescent gel stain (Bio-Rad) and visualized on a UV transilluminator 2000 (Bio-Rad).

For Western blot, the total protein concentration was determined using the bicinchoninic acid (BCA) protein assay kit (Pierce/Thermo Fisher Scientific; Rockford, IL). 25µg of protein samples were mixed with 4x LDS sample buffer (Life Technologies, CA, USA) and 10x reducing agent (Invitrogen; Carlsbad, CA) and heated for 5min at 100°C. Sam-

ples were loaded in a 4-12% SDS-PAGE gel (NUPAGETM; Invitrogen) for electrophoretic separation and subsequently electro-transferred to polyvinyldifluoride (PVDF) membrane (Invitrogen, USA). Membranes were blocked with 5% bovine serum albumin (BSA) in TBST (tris buffered saline with Tween-20) for 1h at room temperature and afterward incubated overnight at 4°C in primary antibodies. Membranes were then washed with TBST (3 times for 15 min) and incubated with anti-rabbit-HRP (111-036-045; Jackson ImmunoResearch; 1:5000) secondary antibody. Membranes were then imaged with Bio-Rad ChemiDoc system (Bio-Rad, Hercules, CA) and analyzed using ImageJ software (NIH). Following primary antibodies were used SARS-CoV-2 membrane protein antibody (ProSci; 9165; 1:1000), SARS-CoV-2 envelope protein antibody (ProSci; 9169; 1:1000), SARS-CoV-2 Nucleocapsid protein (RayBiotech; QHD43423; 1:1000) and SARS-CoV-2 Spike protein antibody (Novus biologicals LLC; NB100-56578; 1:1000).

6.4.11 Protein Reconstitution Into Vesicles

Monodisperse LUVs were extruded using a vesicle extruder. Lipids with composition mimicking that of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) were dissolved in chloroform with solid concentration of 5 mg/mL. The molar ratio of different lipids (Avanti Polar Lipids, Inc, Alabaster, AL) are 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC): 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE): 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol (POPI): 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS): Cholesterol = 0.45: 0.2 : 0.13: 0.07: 0.151. The chloroform solution was dried in a glass vial with gentle N2 gas stream, then vacuumed overnight at -30 in Hg at room temperature. The dried lipid mixture was hydrated with

biological relevant buffer (150 mM NaCl, 20 mM HEPES, pH = 7.2) with 30 s vortex, prior to ten freeze-thaw cycles in dry ice and 37°C. After the final thawing step, the aqueous solution was passed 11 times through a polycarbonate membrane with 100 nm pores (Nuclepore Track-Etch membrane, Whatman, Chicago, IL). The size and zeta potential of extruded membrane were measured with dynamic and phase analysis light scattering (DLS and PALS) techniques using a 90 Plus PALS light scattering machine (Brookhaven instruments, Holtsville, NY).

To reconstitute the M proteins into LUV, the concentrated stock solution of n-dodecyl-β-D-maltoside (DDM Avanti Polar Lipids, Inc, Alabaster, AL) were added to the 5 mg/mL of freshly extruded LUVs solution to reach a final concentration of 80 mM. M-protein stabilized by Triton X-100 (stock solution: 1 wt.% Triton X-100 per every 2 mg/mL M protein) was next added to the LUVs solution at a mass ratio of M/lipid = 1/100, 1/67 and 1/50 after 30 min of incubation. The solution was allowed another 30 min of incubation before addition of 40 mg of wet BioBeads (Bio-Rad, Hercules, CA) per mL of LUVs solution to slowly remove the detergent over 3 hours. The second and third doses of BioBeads at 40 mg/mL were also added at 3 hour intervals. A final dose of Biobeads was added at a ratio of 380 mg/mL and incubated for another 3 hours. The M-reconstituted LUVs were separated from the solution via centrifugation at 13 krpm for 30 min in a microcentrifuge tube.

6.4.12 Preparation of Supported Bilayer Samples

To prepare supported lipid bilayer (SBL) on mica for AFM imaging, 75 μL of LUVs solution collected from the bottom of the microcentrifuge tube was added onto freshly

cleaved pristine mica and incubated for 20 min to 1 hour for different coverage. The sample was then gently rinsed with 5 mL of buffer. Caution was taken not to expose the SBL sample directly to air. The samples were kept submerged in aqueous buffer until imaging in an AFM fluid cell. Imaging was performed within 1 hour of the sample preparation.

6.4.13 AFM Imaging, Force Profile Collection and Analysis

AFM height images were obtained inside a fluid cell in tapping mode using a MSNL cantilever (Bruker, Camarillo, CA). Images were taken in 512x512 line resolution over areas ranging from 250 nm x 250 nm to 1 μm x 1 μm. The effective tip size was calibrated with 10 nm gold nanoparticles (Ted Pella, Redding, CA) using our previously reported protocol2. Membrane structure was probed with nanoindentation to obtain force profiles. The force profiles were collected in force volume mode with a 16 x 16 sampling resolution over the designated area of interest of 500 nm x 500 nm. To probe the continuum elastic property of the membrane and protein patches, an MLCT probe (Bruker, CA) with a larger calibrated tip radius of 23 nm was used. The area of interest was identified from prior tapping mode imaging. The spring constant of the larger tip radius cantilevers was calibrated to be 0.15 N/m using its thermal oscillation spectrum.

AFM image analysis were performed in Mountains SPIP software (Digital Surf, France). Protein particles were identified by setting the height of SBL top surface as the threshold and only identifying out protrusions above the membrane threshold. The software-identified protrusions were then manually examined to exclude noise and false identification of the protein particles. The false identifications by the software can be ruled out as they

are 1-2 pixel wide noise peaks which are much smaller than the proteins which are greater than 40 pixels depending on whether they are isolated protein dimers or large patches.

6.4.14 CryoEM Data Collection and Analysis

Cryo-samples were prepared in a Vitribot Mark IV (Thermo Fisher Scientific). For each sample, Quantifoil grid R2/1 was glow-discharged before a 3.5 μL of aliquot was applied and blotted for 4 s at 95% humidity at room temperature. The samples were then plunged into liquid ethane and preserved in liquid nitrogen. The Cryo-EM micrographs were obtained using a Talos Arctica (FEI) operating at 200 kV equipped with a Falcon 4i Direct Electron Detector (Thermo Fisher Scientific). The imaging was performed at 150,000x magnification with a resolution of 0.95 Å/pixel and a defocus range of -1.5 to -2.5 μm. About 150-250 images were collected for each sample using automated data acquisition using EPU software (Thermo Fisher Scientific). Images were motion corrected within EPU. The data set was processed using Relion software. Approximately 500 areas of interest were manually picked to generate the 2D class average images.

6.4.15 All Atom Molecular Dynamics Simulation

All atom molecular dynamics (MD) simulations were performed using the CHARMM36m force field with the MD package GROMACS, version 2022.3 [126, 127]. The CHARMM-GUI input generator was used to set up the simulated systems with periodic boundary conditions and supplied the six steps used for equilibration [128, 129, 130, 131, 97, 132, 133, 134, 135]. After equilibration, each system was simulated for 1 microsecond with a timestep of 2 femtoseconds in the NPT ensemble. System temperature was maintained at 303.15 K using

the Nose-Hoover thermostat [136, 137], with the pressure maintained semi-isotropically at 1 bar in the x-y dimensions and separately in the z-dimension using the Parrinello-Rahman barostat [138, 139]. The coordinates for each simulation were saved once every 50 thousand timesteps, or every 0.1 ns, for a total of 10 thousand frames.

Three different membrane systems were simulated, one for each form of the M protein dimer ("short" vs "long"), and one as a control only consisting of a membrane. For the simulations involving an M protein dimer, the protein was inserted into the membrane with an orientation and depth that matched other studies [89, 90]. Figure S2 shows schematics for these membrane simulations. In the case of the short form, only residues 9-204 were considered (PDB: 7vgs), while residues 9-206 were included in the long form structure (PDB: 7vgr) [89]. The membrane was composed of Chol 15%; DOPC 45%; DOPE 20%; DOPS 7%; POPI 13% in both leaflets, with the solvent consisting of NaCl at a concentration of 0.15 M and TIP3P water. Each simulation, depending on the inclusion of a protein, consisted of a 25 nm x 25 nm membrane in the x-y plane with at least 5 nm of solvent above and below the protruding protein, yielding a total unit cell thickness of 18 nm in the z dimension.

For runs involving an M protein, the final trajectory was reoriented frame-by-frame such that the protein was centered in the box for all frames. While each trajectory was fitted to eliminate protein translation across the membrane, this was not the case in the direction normal to the membrane and for the rotation of the protein. To analyze the processed trajectory, the python library MDAnalysis was used [140, 141]. Images of these systems were generated using ChimeraX [142].

Additionally, all atom simulations for each conformation of the M protein dimer in solvent were performed to compare protein stability with the membrane simulations. These systems involved a box size of approximately 15(16) nm x 15(16) nm x 15(16) nm for the short(long) conformation. The same process and conditions as the membrane simulations were used, with slight deviations in the equilibration steps due to the lack of a membrane.

6.4.16 Coarse-grained Simulation

The coarse-grained simulation is performed using HOOMD-blue package [143] and BondFlip plugin [116]. For the membrane, we initially assign a 100a×84a triangular lattice plane and place M proteins in the middle disk area with a radius of 15a. The dynamics of the membrane is modeled through Langevin integrator with dt=0.001s in room temperature. In every 100 steps, we apply Monte-Carlo bond-flip to change the local connectivity of the membrane, where each bond has stretching and bending energy with spring constant $20k_B$ T/a^2 and bending rigidity $20k_B$ T. We note that each simulation is run for 10,000s, and a height constraint (h(t)=0) is applied for the region with radial distance r>35a through the simulation time. The simulation is performed on NVIDIA GeForce RTX 3090, and the visualization is performed with OVITO software (58).

Chapter 7

SARS-CoV-2 Structural Proteins - Supplemental Information

7.1 Determination of Detergent Conditions for M Protein Insertion

The proper function of transmembrane M protein requires insertion into the bilayer membrane with the correct orientation. In this study, the protein insertion utilized a detergent assisted procedure and was optimized for M protein in terms of the type and concentration of detergent used. Fig. 7.1 shows the key results obtained during the optimization process. The as-extruded lipid vesicles were measured to exhibit an intensity-averaged hydrodynamic size of 123 nm in DLS (Fig. 7.1a). As shown in Fig. 7.1b, as detergents are gradually introduced into the vesicle solution, the lipid vesicles undergo a three-stage evolution, which has been well established in previous studies [144, 145]. In the

first stage, the detergent molecules insert themselves into the lipid vesicle bilayer structures without dissolving the vesicles. In this stage, with more detergent addition, the detergent concentration in the bilayer membrane increases gradually without decreasing the size and number of vesicles. At the saturation concentration, Csat, the vesicle becomes saturated with detergent. Increasing the detergent concentration beyond saturation leads to the second stage where further detergent addition results in the partial dissociation of the original vesicles. Here, the hydrodynamic size of vesicles significantly decreases with detergent addition, as they are being depleted of lipids. The dissociated lipid molecules form mixed micelles with the detergents. At the end of the second stage, all the vesicles are dissociated completely at the vesicle dissolution concentration, Csol. In this third stage, only mixed micelles exist. Further addition of detergents only increases the detergent composition in the mixed micelles.

In this study, two non-ionic detergents, TX-100 and DDM were tried. Unlike their ionic counterparts, these detergents are mild and do not disrupt protein folding [144]. Figs. 7.1c,d show the hydrodynamic size change with detergent concentration for TX-100 and DDM, where the three stages described in Fig. 7.1b were observed for most cases. TX-100 was found to saturate and dissolve vesicles at much lower detergent concentrations compared with DDM. Fig. 7.1e,f show the saturation and dissolution concentrations as a function of the total lipid concentration in solution for TX-100 and DDM. In both cases, saturation and dissolution concentrations were found to increase with increased lipid concentration. For the concentrations studied TX-100 shows a very gradual linear increase while DDM the saturation concentration increases rapidly and then reaches a plateau at higher lipid

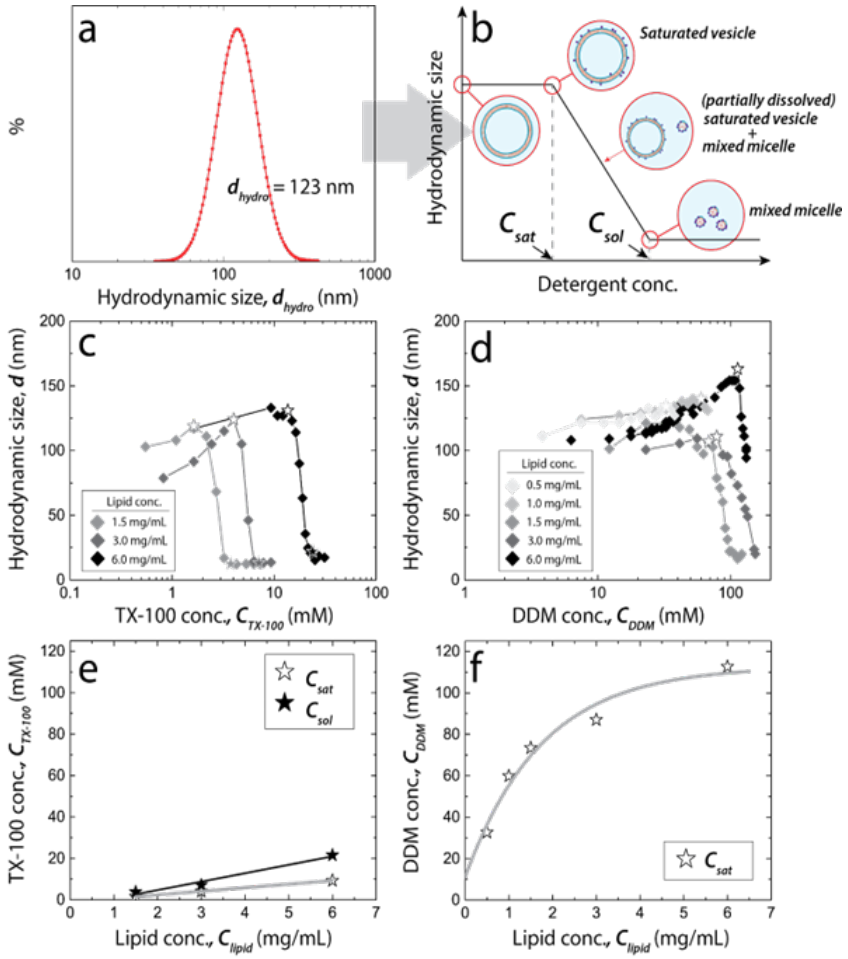

Figure 7.1: **(a)** Intensity distribution vs. hydrodynamic size measured with DLS for as-extruded lipid vesicles in biological buffer. **(b)** Schematics of the three stages of vesicle dissolution with increasing detergent concentration. **(c)** and **(d)** are hydrodynamic size of vesicles measured as functions of TX-100 and DDM detergent concentrations, respectively. The empty and filled stars represent the detergent-saturation concentration, Csat and vesicle dissolution concentration, Csol. **(e)** and **(f)** are C_{sat} and C_{sol} observed at different lipid concentrations for TX-100 and DDM, respectively.

concentrations. Thus, DDM provides much better control and higher tolerance and therefore for the M protein insertion into the lipid bilayers. The DDM concentration used for M protein insertion was 80% of Csat. A similar detergent (octylglucoside) was reported to demonstrate optimum transmembrane protein function and unidirectional insertion in this concentration regime [146]. The unidirectional insertion has been verified by the CryoEM and AFM results as reported below.

After M protein incubation, the detergent removal rate was also optimized. The detergent was removed via hydrophobic adsorption, where hydrophobic polystyrene resin [147] (SM-2) bio beads at a 40 mg/mL were added into the solution. This lower bead concentration in contrast to the usual 80 mg/mL lead to a slower detergent removal rate, which facilitated a higher efficiency in the protein insertion.

7.2 Determination of Elastic Moduli from AFM Measurements

Figure 7.4 shows the average elastic moduli extracted from nanoindentation force profiles for SBL and the M protein patches. SBL exhibits a Young's modulus of 9.5 MPa, which agrees with literature values [148]. The M protein patches demonstrated a much higher modulus value of 41.5 MPa, indicating the patches are indeed a different type of material from the lipid bilayer.

AFM nanoindentation was performed in force volume mode using a MLCT probe (Bruker, CA). The spring constant was calibrated using a thermal tune, k = 0.15 N/m. Tip radius was calibrated with 10 nm gold colloids (Ted Pella, CA), R = 23 nm. The force

profiles were collected over a 1 μm x 1 μm area at a 16 x 16 resolution. To extract the Young's modulus of SBL and the protein patches, a linearized Hertzian contact model was utilized. According to Hertzian model, the relationship between measured force and sample indentation can be expressed as follow:

$$F = \frac{4}{3}\frac{ER^{1/2}}{(1-\sigma^2)}\delta^{3/2}. \tag{7.1}$$

Here, F is the force measured from AFM (Hook's law, F = cantilever deflection x calibrated spring constant). R is the calibrated tip radius of AFM probe, σ is the Poisson's ratio. Here, both σSBL and σMprotein are taken to be 0.5, as most biological materials such as lipids and proteins are considered incompressible. δ is the indentation depth. The acquisition of indentation depth is relatively complicated considering it requires the identification of exact point of tip-sample contact. However, the tip-sample separation, D, can be easily obtained as raw data and the relationship between D and δ is simply:

$$D = c - \delta. \tag{7.2}$$

Here, c is a constant. To simplify the analysis, the Hertzian equation was linearized [149] as follows:

$$F^{2/3} = C - D * \left[\frac{4}{3}\frac{ER^{1/2}}{(1-\sigma^2)}\right]^{2/3} \tag{7.3}$$

Here, C is another constant. Therefore, by obtaining the slope of the linear relationship between F2/3 and D, we can extract the elastic modulus. Here, to rule out the

substrate effect and make sure the E is extracted from the linear elastic region, only the initial 20% of force curve was used to be analyzed by the Hertzian model.

7.3 Atomistic Molecular Dynamics (MD) Simulations of M-protein Stability

Two interconvertible forms of the M protein homodimer have been reported for SARS-CoV and SARS-CoV-2 [89, 99]. To investigate the relative stability of these forms we performed MD simulations of both the long and short forms based on Cryo-EM structures [89]. We ran these simulations in two different environments, embedded within a membrane bilayer (as shown in Fig. 7.2) and in an environment of just water and counter ions. The latter "solvent" simulations will help determine if the relative stability of the long and short M protein dimer conformations is a result of interactions with the membrane or are intrinsic to the structures themselves. While other M protein dimer simulations using experimentally determined structures have found these conformations to be stable in the membrane, the periodic box size for those simulations was either very small or the long form was not considered [89, 90]. For the long conformation, a small box allows the protein to interact with itself, possibly stabilizing it if it does need interactions from other proteins for stability in a planar membrane (as opposed to a micelle from Zhang et al. [89]

From these simulations, the long form in solvent changes most drastically from its initial configuration. The changes in the structural conformation of the short form in solvent and membrane are within the level typical of thermal fluctuations (0.35 nm) as shown in

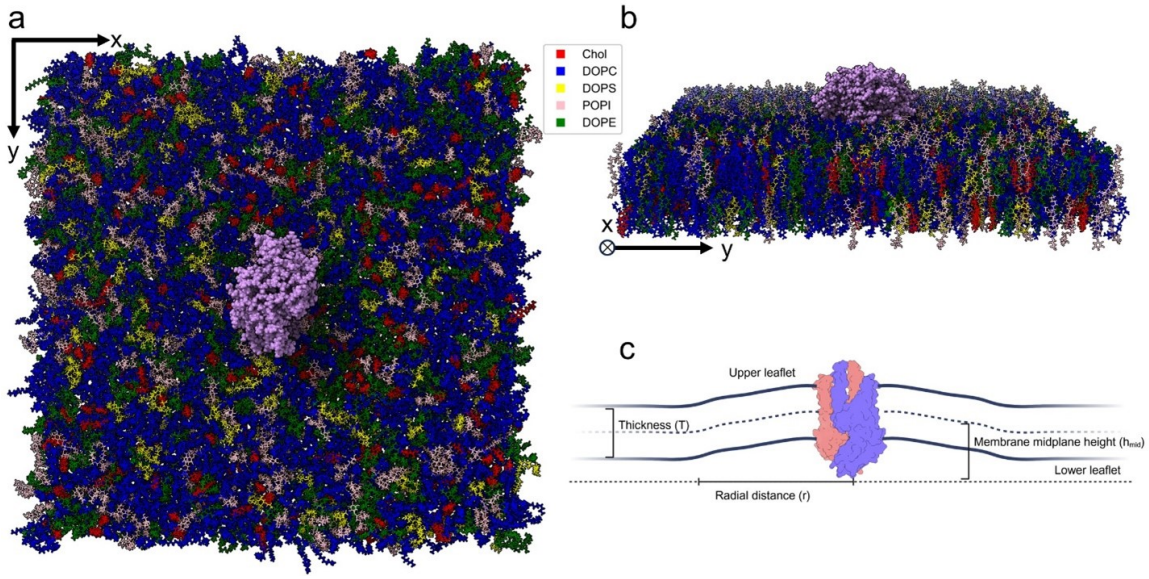

Figure 7.2: **(a)** A view from above looking down on the lower leaflet, with the elliptical shape of the C-terminal for the short form shown in purple. **(b)** A view from the side. **(c)** Diagram of a long form M protein dimer embedded in a membrane, where the differing protein colors refers to the monomers within the dimer (created with BioRender.com).

Fig. 7.8a. In contrast, the solvent-phase long form RMSD is nearly double that observed for the short form solvent-phase simulation after 1 microsecond of simulation (0.7 vs 0.35 nm). In the membrane, the long form structure has a final RMSD that is approximately 30% larger than the short form in membrane, where impact on protein characteristics can be seen in Fig. 7.5-7.7. We see a similar result for the radius of gyration (R_g), where the long form in solvent's R_g decreases over the course of the simulation to match the R_g observed for the short form in both the membrane and solvent. The R_g also decreases in the long form membrane simulation, however, it does not drop to a level comparable to the solvent simulation.

These results suggest that the long form in solvent is not at a minimum energy configuration initially, unlike the short form, which undergoes little change throughout the

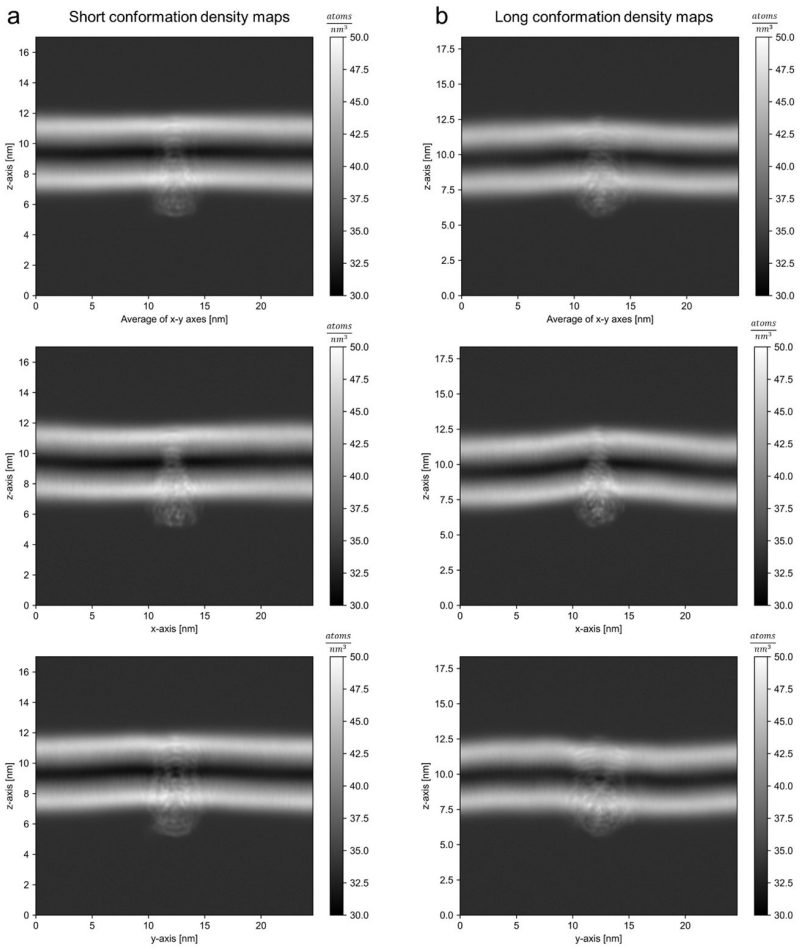

Figure 7.3: Generated using the gmx densmap command from GROMACS, **(a)** and **(b)** represent atom number densities for the short and long forms respectively. The whole simulation temporally and spatially is considered, where every atom other than hydrogen is accounted for, with a bin size of 0.095 nm. Each row represents a different viewing angle, with the top an average of the below two.

simulation. The N-terminal region of the protein curls up without the membrane present to prevent it for the long form. On the other hand, when embedded in the membrane, the alpha helices are split apart by neighboring lipids. Of particular importance is the W31 residue mentioned in Zhang et al. [89], which rotates away from its duplicate due to interactions with nearby DOPC. In comparison to the simulations Zhang et al. ran, these were not performed with the WYF parameter [150]. However, since the nearby DOPC lipids interact with W31 throughout the simulation, including the WYF parameter would only increase this interaction and most likely would not prevent the alpha helices from splitting apart [151]. This supports the suggestion that the long form of the M protein dimer needs both the membrane and possibly other proteins to stabilize its structure [99].

7.4 M Protein's Impact on Membrane Thickness Through MD Simulations

According to these MD simulations, the short and long conformations alter the thickness of the membrane adjacent to the protein differently. Average values of the thickness, calculated as the difference in height between the phosphate groups in the lower and upper leaflets over the last 500 ns of the simulation, can be seen in Fig. 7.9. At each point in time every phospholipid head was placed into a bin (the membrane was split into 15 x 15 bins), where the average upper leaflet height and lower leaflet height for each bin were subtracted from each other to determine the thickness. This process was performed using MDAnalysis.

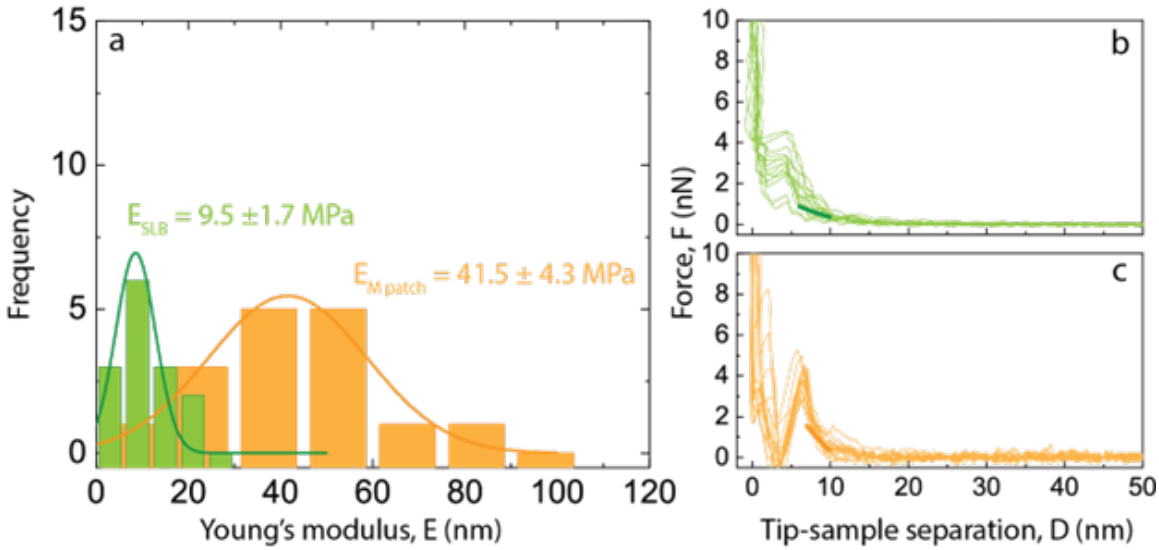

Figure 7.4: **(a)** Histogram of Young's modulus obtained from analyzing the AFM nanoindentation force profiles using Hertzian contact model. **(b)** and **(c)** are the force profiles obtained from SBL and protein patches, respectively. The thick solid lines represent the force profile calculated using the mean elastic modulus values from **(a)**.

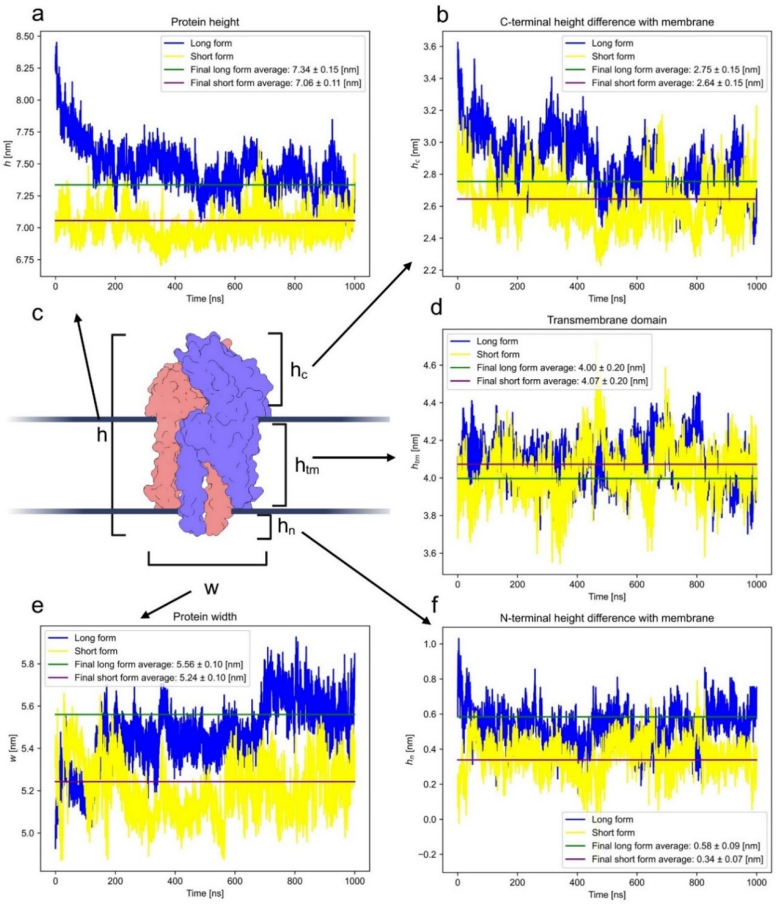

Figure 7.5: These calculations were performed with MDAnalysis. **(a)** Protein height, determined by finding the difference between an average of the five highest and lowest C_α atoms throughout the simulation. **(b)** The difference in height between an average of the five furthest away C-terminal C_α atoms and the phospholipid heads within 2 nm of the protein. **(c)** Schematic showing the different characteristic quantities (created with BioRender.com). **(d)** Transmembrane domain height, determined by subtracting **(b)** and **(f)** from **(a)**. **(e)** Width of the protein, the maximum distance between C_α atoms of like residues. **(f)** The height difference between the N-terminal and the membrane, calculated similarly to **(b)**. Averages were taken over last 100 ns, where the corresponding error is standard deviation.

Based on the line of best fit and the corresponding 95% confidence intervals, the long and short form induce statistically significant changes in membrane thickness that depend on the radial distance from the protein. In contrast, the membrane-only simulation thickness appears to be independent of the distance from the center, hovering around ∼4.20 nm. Thus, the membrane-only system experiences thickness fluctuations, while simulations with M protein exhibit persistent thickness dependencies radially around the embedded protein dimer.

Combining these results with the radially binned averages represented with the black line of each bottom plot, the long form of the protein is surrounded by a region of thicker membrane, while the short form is surrounded by a region of thinner membrane. This agrees with Neuman et al. [99], and their study of M proteins during the formation of viral particles, in which they found long form in thicker regions and short in thinner. These changes are possibly induced by the different shape and/or different membrane-facing surface areas between the two conformations. Alternatively, the different M protein forms may simply stabilize different membrane thicknesses arising due to membrane fluctuations.

7.5 Protein Rotation in MD Simulations

As seen in the superposition of protein cross sections throughout the simulation in Fig. 7.6b, 7.9, and 7.11, the short M protein dimer conformation appears to have a more circular cross section than the long form. Our analysis shows this is a result of the short form's increased rotation in comparison to the long form. This rotation can be measured through the calculation of the protein's inertia tensor, and subsequent computation of the

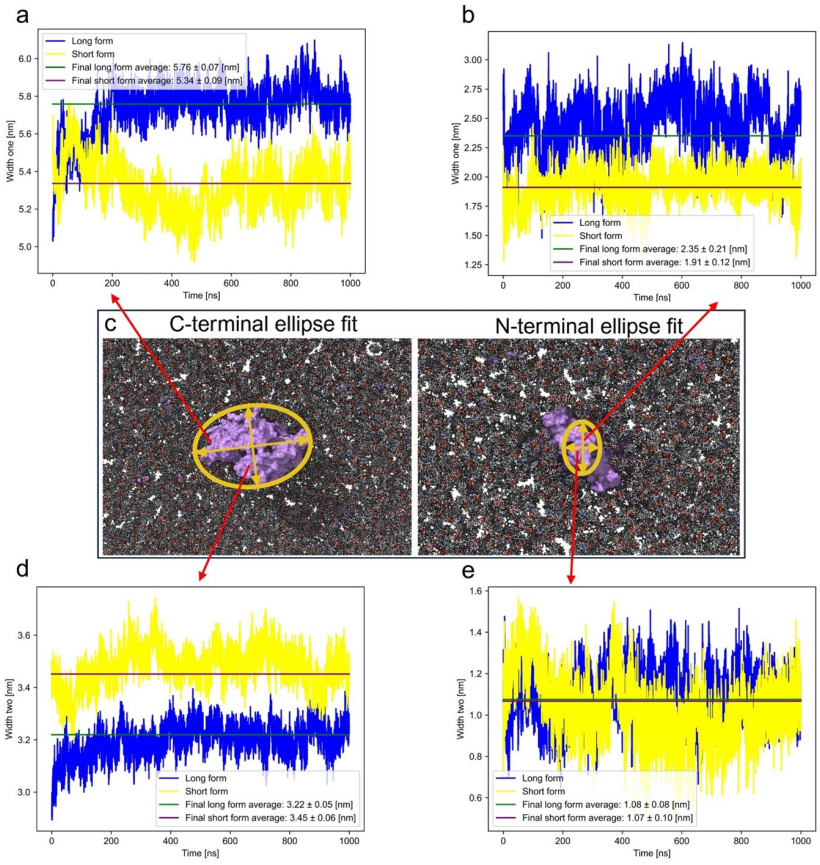

Figure 7.6: **(a)-(b)** and **(d)-(e)** were obtained by calculating the inertia tensor of every protein atom above the height of the corresponding leaflet, diagonalizing it, and then assuming a uniform protein atom density in the shape of an ellipse. **(a)-(b)** correspond to the long axis of the C-terminal and N-terminal respectively, while **(d)-(e)** is the short axis of the C-terminal and N-terminal. **(c)** A schematic of the principal axes of the protein, with the ellipse fit for each terminal shown for the M protein (created with BioRender.com). Averages were taken over last 100 ns, where the corresponding error is standard deviation.

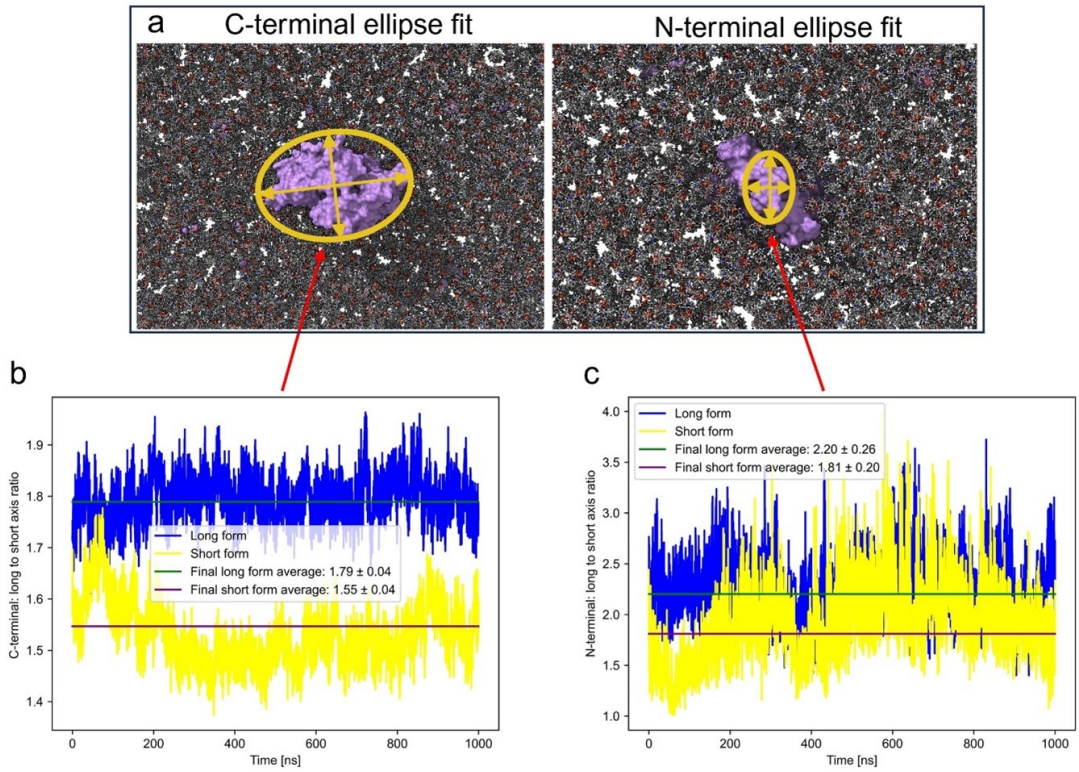

Figure 7.7: **(a)** Repeating the schematic shown in Fig. 7.6. **(b)** Ratios of the C-terminal fit, Fig. 7.6a and 7.6d, and **(c)** ratio of the N-terminal fit, Fig. 7.6b and 7.6e. Averages were taken over last 100 ns, where the corresponding error is standard deviation.

dot product between the x-axis and an inertia eigenvector. Performing this analysis on both the short and long form leads to the results shown in Fig. 7.10.

As seen in Fig. 7.10b-c, the short conformation rotates over a wider angular range than the long conformation. The range of rotation is twice as large for the short form than the long form with several peaks in the frequency plot. For the case of the long form, the protein long axis is aligned with the trough of the membrane, which could be an explanation for the bend's formation: that different axes for the protein induce different levels of curvature. Furthermore, the bending aligned with respect to the long axis of the protein can imply a possible process for the formation of higher order oligomers as other

M protein dimers align in this trough. However, it is unknown if this is the preferred curvature profile of M protein or if it is impacted by the small membrane patch size and periodic boundary conditions. Imaging studies of the M protein for SARS-CoV have shown line-like formations across the virion [99][152].

7.6 Impact of M Protein on Membrane Curvature in MD Simulations

To decouple the height of the membrane from the thickness, membrane midplane height is analyzed over the last 500 ns of the simulation. The midplane is defined as the average surface between the lower and upper leaflet shown in Fig. 7.2c. As seen in Fig. 7.11, while the membrane-only simulation has slight height variation, it is clearly not as substantial as either the short or long form runs. Thus, it can be assumed that these represent membrane fluctuations that persist on the scale of 500 ns.

From the line of best fit, with the given 95% confidence intervals for each simulation, the long conformation of the M protein dimer seems to have the greatest effect on membrane deformation. Additionally, since the confidence interval does not include the horizontal (slope = 0), this radial dependence of average membrane midplane height appears significant.

The shape of the membrane curvature can be determined from the radial average. While the deformation of the membrane-only simulation matches that of height fluctuations, both membranes containing either the long or short M protein conformations seem to adopt

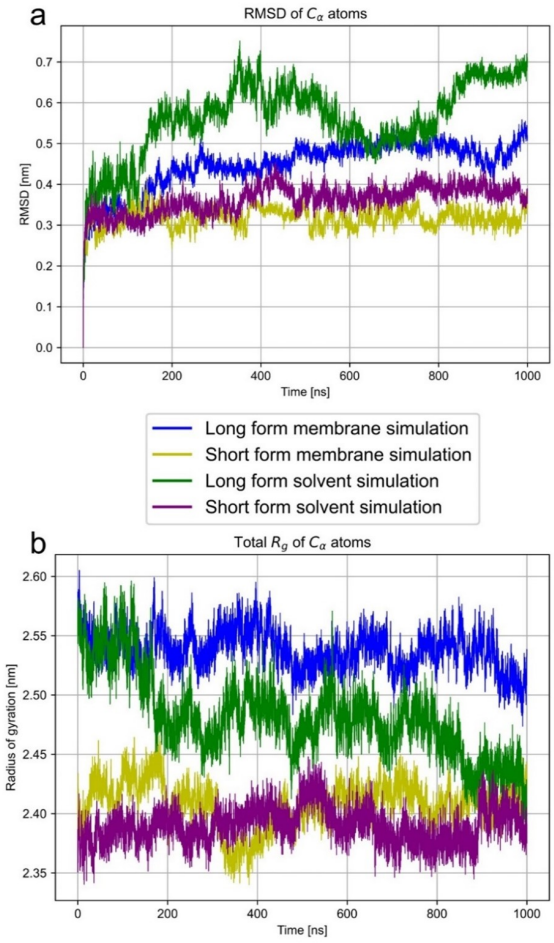

Figure 7.8: The long/short (blue/yellow) form membrane simulations and long/short (green/purple) form solvent simulations. **(a)** Root mean square distance (RMSD) represents the difference of each C_α atom at the corresponding time with its original position squared, summed over every C_α atom, and then divided by the number of C_α atoms and taken to the half power. **(b)** Total radius of gyration R_g, defined as the radial distance to a point in which the moment of inertia would match the actual inertia if the body was condensed there, representing how compact the protein is. RMSD and R_g were calculated using the GROMACS commands gmx rms and gmx gyrate respectively.

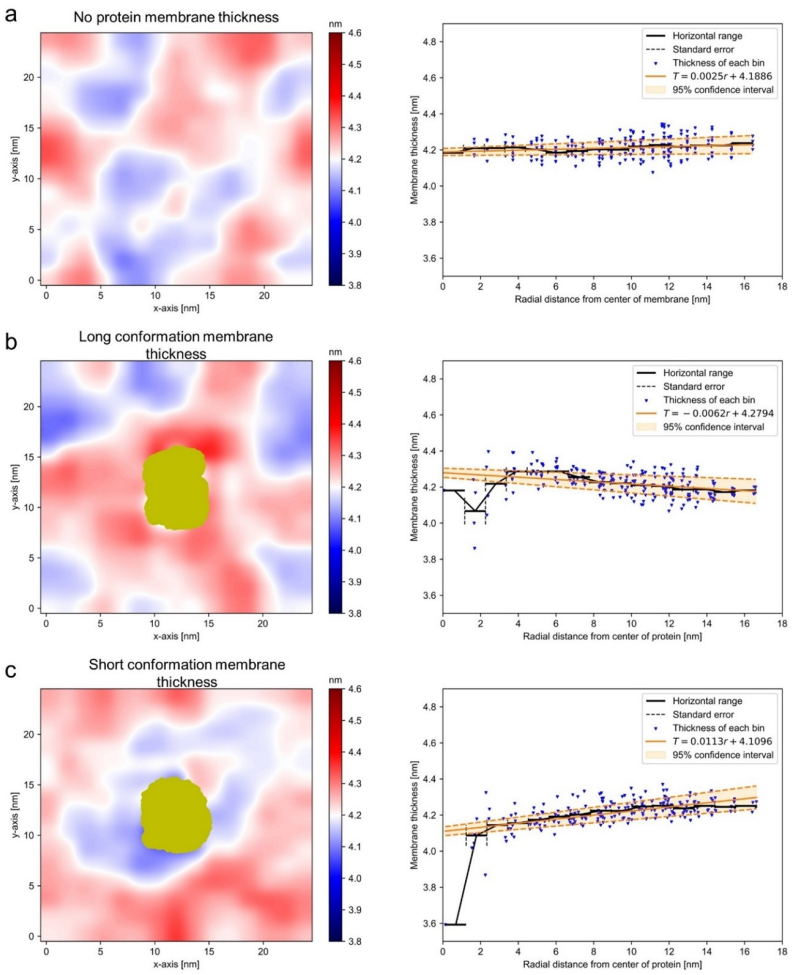

Figure 7.9: **(a)** – **(c)** Visual from above in nm, with a radial plot from the center of the box or center of the protein shown to the right. Each blue dot represents the average thickness at a given bin, orange defines the 95% confidence interval for a linear line of best fit, and black is a radial average of average thickness. Every atom in the protein at every frame averaged over is shown in gold. **(c)** Shows the non-shifted radial plot used in Fig. 7.6e.

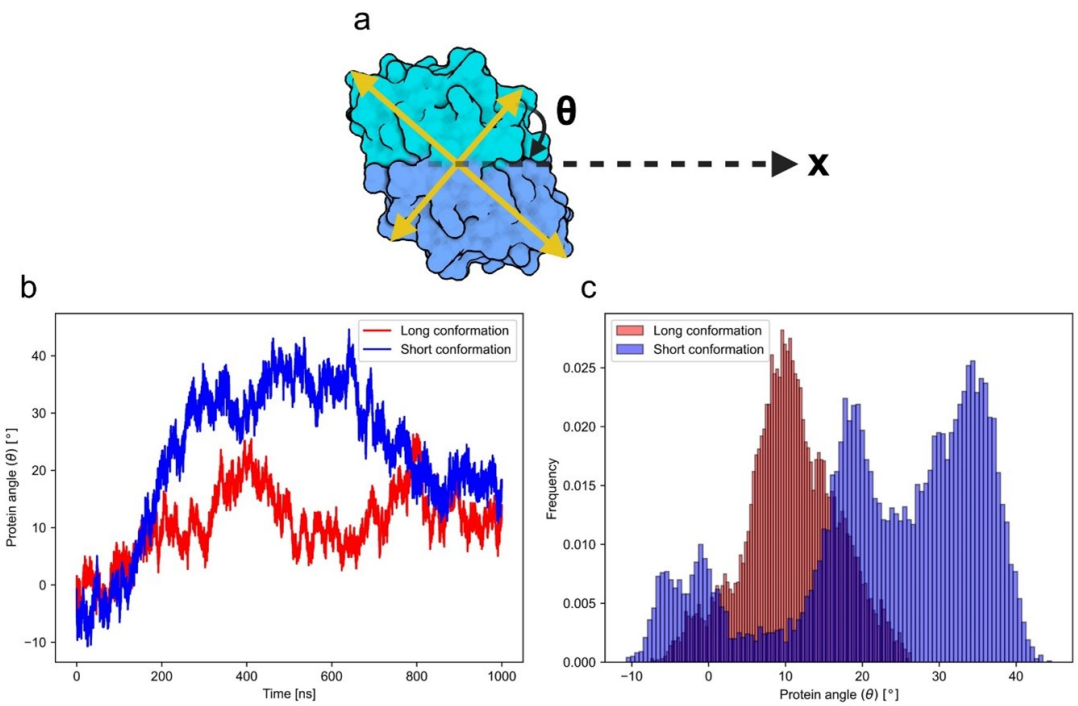

Figure 7.10: **(a)** Diagram displaying rotation in the x-y plane, where θ defines the angle between the short axis of the protein and the x-axis. The yellow lines represent projections of the long and short axis of the protein (created with BioRender.com). **(b)** Time evolution of the long/short (red/blue) form orientation (θ) with respect to the x-axis. **(c)** Shows the frequency of angles throughout the simulation for both short and long forms.

persistent curvature, in the form of a valley-like fold in the membrane enclosing the protein dimer. This shows that the M protein induces curvature in the membrane. Due to the small size of the lipid membrane patch simulated and the periodic boundary conditions imposed, this is only a qualitative confirmation of the membrane curvature induced by the protein. As discussed in Neuman et al. [99] for SARS-CoV, the long conformation appears to induce more curvature than the short conformation, agreeing with our observed membrane deformation.

7.7 Coarse-grained Molecular Dynamics Simulation of M Protein Effects on Membrane Curvature

The role of M Proteins in inducing membrane curvature and budding in ERGIC was previously explored [116]. ERGIC was modeled as a sphere built from a triangular lattice as shown in Fig. 7.12, which shows how the M-M interaction can result in the budding of the region of the ERGIC membrane where the M proteins are embedded. The ERGIC is shown as a gray vesicle in Fig. 7.12 and the three regions of M protein are depicted with three hard spheres (M1 in red, M2 in orange, and M3 in black), where M1 and M2 correspond to the transmembrane domain, and M3 denotes the endodomain. Fig. 7.12a shows that in the absence of endodomain interaction, the membrane does not curve. However, the attractive interaction between endodomains results in the membrane curvature and budding as illustrated in Fig. 7.12b. However, in the absence of interaction between endodomains, the attractive interaction between N proteins (blue and yellow beads) and M proteins, see Fig. 7.12c, or between RNA and M proteins (Fig. 7.12d) can induce curvature and budding.

Since most of the experiments and all-atom simulations in this paper are performed on the flat membrane, for the purpose of this paper, we simulated the role of M proteins embedded in a flat membrane, see the main text. Previously, it was found that the curvature of ERGIC helps with overcoming the energy barriers and leads to budding. In this paper, we found that M proteins can curve even the flat membranes and induce budding if the endodomains interact with each other.

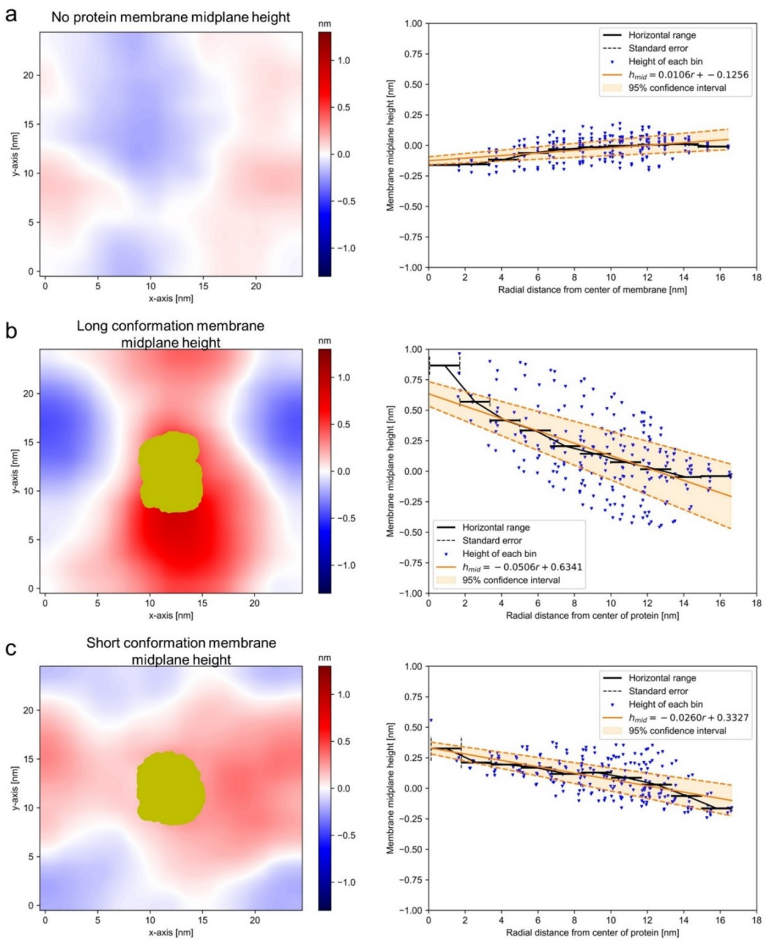

Figure 7.11: **(a)** – **(c)** Visual from above in nm, with radial plots to the right showing the average heights of each bin (blue) in comparison to either the radial distance from the center of the membrane or the radial distance from the center of the protein. Every atom in the protein at every frame averaged over is shown in gold.

7.8 Estimation of Line Tension from Membrane Thinning

For a given height mismatch, δ, between a lipid ordered domain of monolayer thickness hr and the surrounding lipid disordered membrane of thickness hs, the line tension generated due to the elastic cost of bend and tilt deformation is given by [101]

$$\gamma = \left(\frac{\delta}{h}\right)^2 \frac{\sqrt{B_s K_s}\sqrt{B_r K_r}}{\sqrt{B_s K_s} + \sqrt{B_r K_r}}, \tag{7.4}$$

where B_s, r and K_s, r are the bending and tilt modulus of the soft and rigid phases respectively, and h=(h_r+h_s)/2. Since the stiffness of the M protein region (41.5 MPa) is much higher than that of the lipid layer (9.5 MPa), we can take $B_r >> B_s$ and $K_r >> K_s$, simplifying our expression to

$$\gamma = \left(\frac{\delta}{h}\right)^2 \sqrt{B_S K_S}. \tag{7.5}$$

We can estimate the bending modulus B_s from the measured Young's modulus (E) via [153]

$$B_s = \frac{E h_s^3}{12(1-\sigma^2)}. \tag{7.6}$$

With $\sigma = 0.5$ (incompressible) the bending modulus is approximately 3 k_BT and we can take K_s= 3 k_BT/nm^2, which is consistent with lower values reported from simulations and experiments [102]. We can use the line tension expression for a monolayer assuming the height mismatch is evenly distributed among the two leaflets. Using these moduli along with δ = 0.25 nm and h = 2.125 nm, the simplified expression for the line tension gives γ

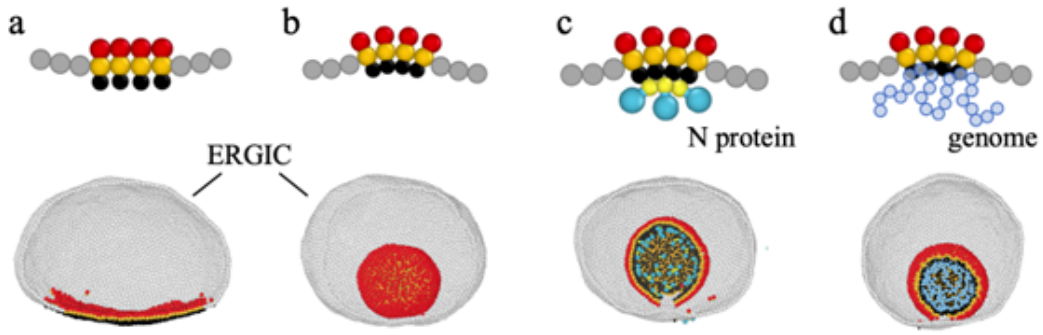

Figure 7.12: **(a)** M proteins embedded in the bottom of ERGIC have no endodomain interactions, and no budding is observed in this scenario. **(b)** M proteins have endodomain interactions, proteins bud into ERGIC and form a spherical shell in 4000s. **(c)** M proteins have no endodomain interactions, a budding process happens when N proteins are added. Note that N proteins interact with M proteins endodomain and effectively introduce endodomain interactions. **(d)** M proteins have no endodomain interactions, a budding happens when a short RNA genome (Length = 400a) is added. The M proteins interact with the genome through screened electrostatic interactions, which effectively cause endodomain interactions.

= 0.35 pN (accounting for both leaflets). If we assume the deformation is entirely confined to one leaflet ($\delta = 0.5$ nm), the line tension increases to $\gamma = 0.66$ pN.

Bibliography

[1] Huan-Xiang Zhou. Q&A: What is biophysics? *BMC Biology*, 9(1):13, December 2011.

[2] Inverted and upright microscopy at a glance. `https://ibidi.com/content/212-inverted-and-upright-microscopy`. Accessed: 2023-11-14.

[3] Han Altae-Tran, Soumya Kannan, F. Esra Demircioglu, Rachel Oshiro, Suchita P. Nety, Luke J. McKay, Mensur Dlakić, William P. Inskeep, Kira S. Makarova, Rhiannon K. Macrae, Eugene V. Koonin, and Feng Zhang. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science*, 374(6563):57–65, October 2021.

[4] Femila Manoj, Laura W. Tai, Katelyn Sun Mi Wang, and Thomas E. Kuhlman. Targeted insertion of large genetic payloads using cas directed LINE-1 reverse transcriptase. *Scientific Reports*, 11(1):23625, December 2021.

[5] Begoña Canovas and Angel R. Nebreda. Diversity and versatility of p38 kinase signalling in health and disease. *Nature Reviews Molecular Cell Biology*, 22(5):346–366, May 2021.

[6] Jeremy C. Burton, William Antoniades, Jennifer Okalova, Morgan M. Roos, and Neil J. Grimsey. Atypical p38 Signaling, Activation, and Implications for Disease. *International Journal of Molecular Sciences*, 22(8):4183, April 2021.

[7] Adrián Martínez-Limón, Manel Joaquin, María Caballero, Francesc Posas, and Eulàlia De Nadal. The p38 Pathway: From Biology to Cancer Therapy. *International Journal of Molecular Sciences*, 21(6):1913, March 2020.

[8] Jiahuai Han, Jianfeng Wu, and John Silke. An overview of mammalian p38 mitogen-activated protein kinases, central regulators of cell stress and receptor signaling. *F1000Research*, 9:653, June 2020.

[9] Taichiro Tomida, Mutsuhiro Takekawa, and Haruo Saito. Oscillation of p38 activity controls efficient pro-inflammatory gene expression. *Nature Communications*, 6(1):8350, September 2015.

[10] Naama Geva-Zatorsky, Erez Dekel, Eric Batchelor, Galit Lahav, and Uri Alon. Fourier analysis and systems identification of the p53 feedback loop. *Proceedings of the National Academy of Sciences*, 107(30):13550–13555, July 2010.

[11] Li Sun, Guozhe Yang, Mone Zaidi, and Jameel Iqbal. TNF-induced gene expression oscillates in time. *Biochemical and Biophysical Research Communications*, 371(4):900–905, July 2008.

[12] Matthew Fosbrink, Nwe-Nwe Aye-Han, Raymond Cheong, Andre Levchenko, and Jin Zhang. Visualization of JNK activity dynamics with a genetically encoded fluorescent biosensor. *Proceedings of the National Academy of Sciences*, 107(12):5459–5464, March 2010.

[13] Takeharu Nagai, Keiji Ibata, Eun Sun Park, Mie Kubota, Katsuhiko Mikoshiba, and Atsushi Miyawaki. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology*, 20(1):87–90, January 2002.

[14] Michele L. Markwardt, Gert-Jan Kremers, Catherine A. Kraft, Krishanu Ray, Paula J. C. Cranfill, Korey A. Wilson, Richard N. Day, Rebekka M. Wachter, Michael W. Davidson, and Megan A. Rizzo. An Improved Cerulean Fluorescent Protein with Enhanced Brightness and Reduced Reversible Photoswitching. *PLoS ONE*, 6(3):e17896, March 2011.

[15] Xiaowei Gong, Xiaoyan Ming, Peng Deng, and Yong Jiang. Mechanisms regulating the nuclear translocation of p38 MAP kinase. *Journal of Cellular Biochemistry*, 110(6):1420–1429, August 2010.

[16] Haitao Zhang, George G. Chen, Zhiyi Zhang, Sukying Chun, Billy Cheuk Sing Leung, and Paul B. S. Lai. Induction of autophagy in hepatocellular carcinoma cells by SB203580 requires activation of AMPK and DAPK but not p38 MAPK. *Apoptosis*, 17(4):325–334, April 2012.

[17] Yingli He, Hua She, Ting Zhang, Haidong Xu, Lihong Cheng, Manuel Yepes, Yingren Zhao, and Zixu Mao. p38 MAPK inhibits autophagy and promotes microglial inflammatory responses by phosphorylating ULK1. *Journal of Cell Biology*, 217(1):315–328, January 2018.

[18] Chaogeng Zhu, Guiyun He, Qinqin Yin, Lin Zeng, Xiangli Ye, Yongzhong Shi, and Wei Xu. Molecular biology of the SARs-CoV-2 spike protein: A review of current knowledge. *Journal of Medical Virology*, 93(10):5729–5741, October 2021.

[19] Hyun Kim, Yeongjin Hong, Keigo Shibayama, Yasuhiko Suzuki, Nobutaka Wakamiya, and Youn Uck Kim. Functional analysis of the receptor binding domain of SARS coronavirus S1 region and its monoclonal antibody. *Genes & Genomics*, 36(3):387–397, June 2014.

[20] Anthony M Kyriakopoulos, Greg Nigh, Peter A McCullough, and Stephanie Seneff. Mitogen Activated Protein Kinase (MAPK) Activation, p53, and Autophagy Inhibition Characterize the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Spike Protein Induced Neurotoxicity. *Cureus*, December 2022.

[21] Joël Raingeaud, Alan J. Whitmarsh, Tamera Barrett, Benoit Dérijard, and Roger J. Davis. MKK3- and MKK6-Regulated Gene Expression Is Mediated by the p38 Mitogen-Activated Protein Kinase Signal Transduction Pathway. *Molecular and Cellular Biology*, 16(3):1247–1255, March 1996.

[22] M. Takekawa. A human homolog of the yeast Ssk2/Ssk22 MAP kinase kinase kinases, MTK1, mediates stress-induced activation of the p38 and JNK pathways. *The EMBO Journal*, 16(16):4973–4982, August 1997.

[23] Syu-ichi Hirai, Masaru Katoh, Masaaki Terada, John M. Kyriakis, Leonard I. Zon, Ajay Rana, Joseph Avruch, and Shigeo Ohno. MST/MLK2, a Member of the Mixed Lineage Kinase Family, Directly Phosphorylates and Activates SEK1, an Activator of c-Jun N-terminal Kinase/Stress-activated Protein Kinase. *Journal of Biological Chemistry*, 272(24):15167–15173, June 1997.

[24] Tetsuo Moriguchi, Noriyo Kuroyanagi, Kyoko Yamaguchi, Yukiko Gotoh, Kenji Irie, Takahisa Kano, Kyoko Shirakabe, Yoshinao Muro, Hiroshi Shibuya, Kunihiro Matsumoto, Eisuke Nishida, and Masatoshi Hagiwara. A Novel Kinase Cascade Mediated by Mitogen-activated Protein Kinase Kinase 6 and MKK3. *Journal of Biological Chemistry*, 271(23):13675–13679, June 1996.

[25] Hidenori Ichijo, Eisuke Nishida, Kenji Irie, Peter Ten Dijke, Masao Saitoh, Tetsuo Moriguchi, Minoru Takagi, Kunihiro Matsumoto, Kohei Miyazono, and Yukiko Gotoh. Induction of Apoptosis by ASK1, a Mammalian MAPKKK That Activates SAPK/JNK and p38 Signaling Pathways. *Science*, 275(5296):90–94, January 1997.

[26] Hong Sun, Catherine H. Charles, Lester F. Lau, and Nicholas K. Tonks. MKP-1 (3CH134), an immediate early gene product, is a dual specificity phosphatase that dephosphorylates MAP kinase in vivo. *Cell*, 75(3):487–493, November 1993.

[27] Caroline Nunes-Xavier, Carlos Roma-Mateo, Pablo Rios, Celine Tarrega, Rocio Cejudo-Marin, Lydia Tabernero, and Rafael Pulido. Dual-Specificity MAP Kinase Phosphatases as Targets of Cancer Treatment. *Anti-Cancer Agents in Medicinal Chemistry*, 11(1):109–132, January 2011.

[28] Sharol Su Lei Cho, Jian Han, Sharmy J. James, Chin Wen Png, Madhushanee Weerasooriya, Sylvie Alonso, and Yongliang Zhang. Dual-Specificity Phosphatase 12 Targets p38 MAP Kinase to Regulate Macrophage Response to Intracellular Bacterial Infection. *Frontiers in Immunology*, 8:1259, October 2017.

[29] KEGG PATHWAY: MAPK signaling pathway - Reference pathway.

[30] Ubaldo Gioia, Sara Tavella, Pamela Martínez-Orellana, Giada Cicio, Andrea Colliva, Marta Ceccon, Matteo Cabrini, Ana C. Henriques, Valeria Fumagalli, Alessia Paldino, Ettore Presot, Sreejith Rajasekharan, Nicola Iacomino, Federica Pisati, Valentina Matti, Sara Sepe, Matilde I. Conte, Sara Barozzi, Zeno Lavagnino, Tea Carletti, Maria Concetta Volpe, Paola Cavalcante, Matteo Iannacone, Chiara Rampazzo, Rossana Bussani, Claudio Tripodo, Serena Zacchigna, Alessandro Marcello, and Fabrizio d'Adda Di Fagagna. SARS-CoV-2 infection induces DNA damage, through CHK1 degradation and impaired 53BP1 recruitment, and cellular senescence. *Nature Cell Biology*, 25(4):550–564, April 2023.

[31] Silvia Rossi, Caterina Motta, Valeria Studer, Giulia Macchiarulo, Elisabetta Volpe, Francesca Barbieri, Gabriella Ruocco, Fabio Buttari, Annamaria Finardi, Raffaele Mancino, Sagit Weiss, Luca Battistini, Gianvito Martino, Roberto Furlan, Jelena Drulovic, and Diego Centonze. Interleukin-1 causes excitotoxic neurodegeneration and multiple sclerosis disease progression by activating the apoptotic protein p53. *Molecular Neurodegeneration*, 9(1):56, December 2014.

[32] Han Shan, Yaqi Bian, Zhaoma Shu, Linxia Zhang, Jialei Zhu, Jianhua Ding, Ming Lu, Ming Xiao, and Gang Hu. Fluoxetine protects against IL-1-induced neuronal apoptosis via downregulation of p53. *Neuropharmacology*, 107:68–78, August 2016.

[33] Magdalena C. Liebl and Thomas G. Hofmann. Cell Fate Regulation upon DNA Damage: p53 Serine 46 Kinases Pave the Cell Death Road. *BioEssays*, 41(12):1900127, December 2019.

[34] Philipp Thomas, Arthur V. Straube, Jens Timmer, Christian Fleck, and Ramon Grima. Signatures of nonlinearity in single cell noise-induced oscillations. *Journal of Theoretical Biology*, 335:222–234, October 2013.

[35] Katrina M. Waters, Brian S. Cummings, Harish Shankaran, Natalie E. Scholpa, and Thomas J. Weber. ERK Oscillation-Dependent Gene Expression Patterns and Deregulation by Stress Response. *Chemical Research in Toxicology*, 27(9):1496–1503, September 2014.

[36] Myong-Hee Sung, Luigi Salvatore, Rossana De Lorenzi, Anindya Indrawan, Manolis Pasparakis, Gordon L. Hager, Marco E. Bianchi, and Alessandra Agresti. Sustained Oscillations of NF-B Produce Distinct Genome Scanning and Gene Expression Profiles. *PLoS ONE*, 4(9):e7163, September 2009.

[37] Li Sun, Guozhe Yang, Mone Zaidi, and Jameel Iqbal. TNF-induced oscillations in combinatorial transcription factor binding. *Biochemical and Biophysical Research Communications*, 371(4):912–916, July 2008.

[38] Jared C. Bronski and Lan Wang. Partially Phase-Locked Solutions to the Kuramoto Model. *Journal of Statistical Physics*, 183(3):46, June 2021.

[39] Dmitri Bratsun, Dmitri Volfson, Lev S. Tsimring, and Jeff Hasty. Delay-induced stochastic oscillations in gene regulation. *Proceedings of the National Academy of Sciences*, 102(41):14593–14598, October 2005.

[40] Yi Qun Xiao, Ken Malcolm, G. Scott Worthen, Shyra Gardai, William P. Schiemann, Valerie A. Fadok, Donna L. Bratton, and Peter M. Henson. Cross-talk between ERK and p38 MAPK Mediates Selective Suppression of Pro-inflammatory Cytokines by Transforming Growth Factor-. *Journal of Biological Chemistry*, 277(17):14884–14893, April 2002.

[41] Raquel M. Melero-Fernandez De Mera, Li-Li Li, Arkadiusz Popinigis, Katryna Cisek, Minna Tuittila, Leena Yadav, Andrius Serva, and Michael J. Courtney. A simple optogenetic MAPK inhibitor design reveals resonance between transcription-regulating circuitry and temporally-encoded inputs. *Nature Communications*, 8(1):15017, May 2017.

[42] Wei Lin, Gary C.H. Mo, Sohum Mehta, and Jin Zhang. DrFLINC Contextualizes Super-resolution Activity Imaging. *Journal of the American Chemical Society*, 143(37):14951–14955, September 2021.

[43] Kathryn E. Medders, Natalia E. Sejbuk, Ricky Maung, Maya K. Desai, and Marcus Kaul. Activation of p38 MAPK Is Required in Monocytic and Neuronal Cells for HIV Glycoprotein 120-Induced Neurotoxicity. *The Journal of Immunology*, 185(8):4883–4895, October 2010.

[44] Evan Xu, Yan Xie, and Ziyad Al-Aly. Long-term neurologic outcomes of COVID-19. *Nature Medicine*, 28(11):2406–2415, November 2022.

[45] 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, July 2012.

[46] Anne E Carpenter, Thouis R Jones, Michael R Lamprecht, Colin Clarke, In Kang, Ola Friman, David A Guertin, Joo Chang, Robert A Lindquist, Jason Moffat, Polina Golland, and David M Sabatini. *Genome Biology*, 7(10):R100, 2006.

[47] Philip Cohen. The origins of protein phosphorylation. *Nature Cell Biology*, 4(5):E127–E130, May 2002.

[48] Erik A. Rodriguez, Robert E. Campbell, John Y. Lin, Michael Z. Lin, Atsushi Miyawaki, Amy E. Palmer, Xiaokun Shu, Jin Zhang, and Roger Y. Tsien. The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends in Biochemical Sciences*, 42(2):111–129, February 2017.

[49] James B. Pawley, editor. *Handbook Of Biological Confocal Microscopy*. Springer US, Boston, MA, 2006.

[50] Pavithra Sundaresan and Richard W Farndale. p38 mitogen-activated protein kinase dephosphorylation is regulated by protein phosphatase 2A in human platelets activated by collagen. *FEBS Letters*, 528(1-3):139–144, September 2002.

[51] Xue Lu, Chun Li, Yong-Kun Wang, Kun Jiang, and Xiao-Dong Gai. Sorbitol induces apoptosis of human colorectal cancer cells via p38 MAPK signal transduction. *Oncology Letters*, 7(6):1992–1996, June 2014.

[52] Isabelle Corre, François Paris, and Jacques Huot. The p38 pathway, a major pleiotropic cascade that transduces stress and metastatic signals in endothelial cells. *Oncotarget*, 8(33):55684–55714, August 2017.

[53] Peter Juo, Calvin J. Kuo, Susan E. Reynolds, Richard F. Konz, Joël Raingeaud, Roger J. Davis, Hans-Peter Biemann, and John Blenis. Fas Activation of the p38 Mitogen-Activated Protein Kinase Signalling Pathway Requires ICE/CED-3 Family Proteases. *Molecular and Cellular Biology*, 17(1):24–35, January 1997.

[54] Spectral density. `https://en.wikipedia.org/wiki/Spectral_density#Cross_power_spectral_density`. Accessed: 2023-11-14.

[55] Chris Bretherton. Atm s 552 lecture 14: Cross-spectral analysis, Winter 2015.

[56] R. Lyons. dsp tips & tricks - the sliding DFT. *IEEE Signal Processing Magazine*, 20(2):74–80, March 2003.

[57] Blackman window. `https://mathworld.wolfram.com/BlackmanFunction.html`. Accessed: 2023-11-14.

[58] S. Vaughan. A simple test for periodic signals in red noise. *Astronomy & Astrophysics*, 431(1):391–403, February 2005.

[59] Norman L Johnson, Samuel Kotz, and N Balakrishnan. *Continuous Univariate Distributions, Volume 1*. Wiley Series in Probability and Statistics. John Wiley & Sons, Nashville, TN, 2 edition, October 1994.

[60] Tomas Ganz. Epithelia: Not just physical barriers. *Proceedings of the National Academy of Sciences*, 99(6):3357–3358, March 2002.

[61] Brad Chazotte. Labeling Nuclear DNA Using DAPI. *Cold Spring Harbor Protocols*, 2011(1):pdb.prot5556, January 2011.

[62] S A Latt, G Stetten, L A Juergens, H F Willard, and C D Scher. Recent developments in the detection of deoxyribonucleic acid synthesis by 33258 Hoechst fluorescence. *Journal of Histochemistry & Cytochemistry*, 23(7):493–505, July 1975.

[63] Sarah Hassdenteufel and Maya Schuldiner. Show your true color: Mammalian cell surface staining for tracking cellular identity in multiplexing and beyond. *Current Opinion in Chemical Biology*, 66:102102, February 2022.

[64] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. 2015. Publisher: arXiv Version Number: 5.

[65] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment Anything. 2023. Publisher: arXiv Version Number: 1.

[66] Michael Cowley and Rebecca J. Oakey. Transposable Elements Re-Wire and Fine-Tune the Transcriptome. *PLoS Genetics*, 9(1):e1003234, January 2013.

[67] Dominique Schneider and Richard E Lenski. Dynamics of insertion sequence elements during experimental evolution of bacteria. *Research in Microbiology*, 155(5):319–327, June 2004.

[68] Lin Chao, Christopher Vargas, Brian B. Spear, and Edward C. Cox. Transposable elements as mutator genes in evolution. *Nature*, 303(5918):633–635, June 1983.

[69] Nicole G. Coufal, José L. Garcia-Perez, Grace E. Peng, Gene W. Yeo, Yangling Mu, Michael T. Lovci, Maria Morell, K. Sue O'Shea, John V. Moran, and Fred H. Gage. L1 retrotransposition in human neural progenitor cells. *Nature*, 460(7259):1127–1131, August 2009.

[70] Hiroki Kano, Irene Godoy, Christine Courtney, Melissa R. Vetter, George L. Gerton, Eric M. Ostertag, and Haig H. Kazazian. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes & Development*, 23(11):1303–1312, June 2009.

[71] Victoria P Belancio, Prescott L Deininger, and Astrid M Roy-Engel. LINE dancing in the human genome: transposable elements and disease. *Genome Medicine*, 1(10):97, 2009.

[72] John L Goodier. Retrotransposition in tumors and brains. *Mobile DNA*, 5(1):11, 2014.

[73] Neil H. Kim, Gloria Lee, Nicholas A. Sherer, K. Michael Martini, Nigel Goldenfeld, and Thomas E. Kuhlman. Real-time transposable element activity in individual live cells. *Proceedings of the National Academy of Sciences*, 113(26):7278–7283, June 2016.

[74] R Lutz. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*, 25(6):1203–1210, March 1997.

[75] Michèle P. Calos and Jeffrey H. Miller. The DNA sequence change resulting from the I Q1mutation, which greatly increases promoter strength. *Molecular and General Genetics MGG*, 183(3):559–560, November 1981.

[76] Dangeruta Kersulyte, Billie Velapatiño, Giedrius Dailide, Asish K. Mukhopadhyay, Yoshiyuki Ito, Lizbeth Cahuayme, Alan J. Parkinson, Robert H. Gilman, and Douglas E. Berg. Transposable Element IS *Hp608* of *Helicobacter pylori* : Nonrandom Geographic Distribution, Functional Organization, and Insertion Specificity. *Journal of Bacteriology*, 184(4):992–1002, February 2002.

[77] Sergey Shmakov, Aaron Smargon, David Scott, David Cox, Neena Pyzocha, Winston Yan, Omar O. Abudayyeh, Jonathan S. Gootenberg, Kira S. Makarova, Yuri I. Wolf, Konstantin Severinov, Feng Zhang, and Eugene V. Koonin. Diversity and evolution

of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology*, 15(3):169–182, March 2017.

[78] Weidong Bao and Jerzy Jurka. Homologues of bacterial TnpB_is605 are widespread in diverse eukaryotic transposable elements. *Mobile DNA*, 4(1):12, 2013.

[79] Patricia Siguier, Edith Gourbeyre, and Mick Chandler. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiology Reviews*, 38(5):865–891, September 2014.

[80] Tautvydas Karvelis, Gytis Druteika, Greta Bigelyte, Karolina Budre, Rimante Zedaveinyte, Arunas Silanskas, Darius Kazlauskas, Česlovas Venclovas, and Virginijus Siksnys. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature*, 599(7886):692–696, November 2021.

[81] Davneet Kaur and Thomas E. Kuhlman. IS200/IS605 Family-Associated TnpB Increases Transposon Activity and Retention. preprint, Synthetic Biology, October 2022.

[82] Bao Ton-Hoang, Cécile Pasternak, Patricia Siguier, Catherine Guynet, Alison Burgess Hickman, Fred Dyda, Suzanne Sommer, and Michael Chandler. Single-Stranded DNA Transposition Is Coupled to Host Replication. *Cell*, 142(3):398–408, August 2010.

[83] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust Growth of Escherichia coli. *Current Biology*, 20(12):1099–1103, June 2010.

[84] Charis-P. Segeritz and Ludovic Vallier. Cell Culture. In *Basic Science Methods for Clinical Researchers*, pages 151–172. Elsevier, 2017.

[85] Kurt Thorn. A quick guide to light microscopy in cell biology. *Molecular Biology of the Cell*, 27(2):219–222, January 2016.

[86] ibidi GmbH, Lochhamer Schlag 11, 82166, Gräfelfing, Germany. *Heating System Slide/Dish – Silver Line.*

[87] Vincent M. Rossi, Katherine C. Davidson, and Lauren E. Moore. Arduino-based, low-cost imaging incubator for extended live cell imaging. *Applied Optics*, 61(17):5282, June 2022.

[88] Fehmi Can Ay, Nesim Bilici, Rahmetullah Varol, Atasangu Yilmaz, Ufuk Gorkem Kirabali, Abdurrahim Yilmaz, and Huseyin Uvet. Study on the Concept and Development of a Mobile Incubator. 2022. Publisher: arXiv Version Number: 1.

[89] Zhikuan Zhang, Norimichi Nomura, Yukiko Muramoto, Toru Ekimoto, Tomoko Uemura, Kehong Liu, Moeko Yui, Nozomu Kono, Junken Aoki, Mitsunori Ikeguchi, Takeshi Noda, So Iwata, Umeharu Ohto, and Toshiyuki Shimizu. Structure of SARS-CoV-2 membrane protein essential for virus assembly. *Nature Communications*, 13(1):4399, August 2022.

[90] Kimberly A Dolan, Mandira Dutta, David M Kern, Abhay Kotecha, Gregory A Voth, and Stephen G Brohawn. Structure of SARS-CoV-2 M protein in lipid nanodiscs. *eLife*, 11:e81702, October 2022.

[91] Xun Zuo, Shuisen Li, John Hall, Michael R. Mattern, Hiep Tran, Joshua Shoo, Robin Tan, Susan R. Weiss, and Tauseef R. Butt. Enhanced Expression and Purification of Membrane Proteins by SUMO Fusion in Escherichia coli. *Journal of Structural and Functional Genomics*, 6(2-3):103–111, September 2005.

[92] Yanqiu Gong, Suideng Qin, Lunzhi Dai, and Zhixin Tian. The glycosylation in SARS-CoV-2 and its receptor ACE2. *Signal Transduction and Targeted Therapy*, 6(1):396, November 2021.

[93] M. Oostra, C. A. M. De Haan, R. J. De Groot, and P. J. M. Rottier. Glycosylation of the Severe Acute Respiratory Syndrome Coronavirus Triple-Spanning Membrane Proteins 3a and M. *Journal of Virology*, 80(5):2326–2336, March 2006.

[94] Cornelis A. M. De Haan, Lili Kuo, Paul S. Masters, Harry Vennema, and Peter J. M. Rottier. Coronavirus Particle Assembly: Primary Structure Requirements of the Membrane Protein. *Journal of Virology*, 72(8):6838–6850, August 1998.

[95] Cornelis A. M. De Haan, M. Smeets, F. Vernooij, H. Vennema, and P. J. M. Rottier. Mapping of the Coronavirus Membrane Protein Domains Involved in Interaction with the Spike Protein. *Journal of Virology*, 73(9):7441–7452, September 1999.

[96] H Laude, J Gelfi, L Lavenant, and B Charley. Single amino acid changes in the viral glycoprotein M affect induction of alpha interferon by the coronavirus transmissible gastroenteritis virus. *Journal of Virology*, 66(2):743–749, February 1992.

[97] Emilia L. Wu, Xi Cheng, Sunhwan Jo, Huan Rui, Kevin C. Song, Eder M. Dávila-Contreras, Yifei Qi, Jumin Lee, Viviana Monje-Galvan, Richard M. Venable, Jeffery B. Klauda, and Wonpil Im. CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *Journal of Computational Chemistry*, 35(27):1997–2004, October 2014.

[98] Shaolong Chen, Jun Xu, Mingyue Liu, A. L. N. Rao, Roya Zandi, Sarjeet S. Gill, and Umar Mohideen. Investigation of HIV-1 Gag binding with RNAs and lipids using Atomic Force Microscopy. *PLOS ONE*, 15(2):e0228036, February 2020.

[99] Benjamin W. Neuman, Gabriella Kiss, Andreas H. Kunding, David Bhella, M. Fazil Baksh, Stephen Connelly, Ben Droese, Joseph P. Klaus, Shinji Makino, Stanley G. Sawicki, Stuart G. Siddell, Dimitrios G. Stamou, Ian A. Wilson, Peter Kuhn, and Michael J. Buchmeier. A structural analysis of M protein in coronavirus assembly and morphology. *Journal of Structural Biology*, 174(1):11–22, April 2011.

[100] Rumana Mahtarin, Shafiqul Islam, Md. Jahirul Islam, M Obayed Ullah, Md Ackas Ali, and Mohammad A. Halim. Structure and dynamics of membrane protein in SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, 40(10):4725–4738, July 2022.

[101] Peter I. Kuzmin, Sergey A. Akimov, Yuri A. Chizmadzhev, Joshua Zimmerberg, and Fredric S. Cohen. Line Tension and Interaction Energies of Membrane Rafts Calculated from Lipid Splay and Tilt. *Biophysical Journal*, 88(2):1120–1133, February 2005.

[102] M. Doktorova, D. Harries, and G. Khelashvili. Determination of bending rigidity and tilt modulus of lipid membranes from real-space fluctuation analysis of molecular dynamics simulations. *Physical Chemistry Chemical Physics*, 19(25):16806–16818, 2017.

[103] Ana J. García-Sáez, Salvatore Chiantia, and Petra Schwille. Effect of Line Tension on the Lateral Organization of Lipid Membranes. *Journal of Biological Chemistry*, 282(46):33537–33544, November 2007.

[104] Rebecca D. Usery, Thais A. Enoki, Sanjula P. Wickramasinghe, Michael D. Weiner, Wen-Chyan Tsai, Mary B. Kim, Shu Wang, Thomas L. Torng, David G. Ackerman, Frederick A. Heberle, John Katsaras, and Gerald W. Feigenson. Line Tension Controls Liquid-Disordered + Liquid-Ordered Domain Size Transition in Lipid Bilayers. *Biophysical Journal*, 112(7):1431–1443, April 2017.

[105] Rees F. Garmann, Aaron M. Goldfain, and Vinothan N. Manoharan. Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proceedings of the National Academy of Sciences*, 116(45):22485–22490, November 2019.

[106] Guillaume Tresset, Jingzhi Chen, Maelenn Chevreuil, Naïma Nhiri, Eric Jacquet, and Yves Lansac. Two-Dimensional Phase Transition of Viral Capsid Gives Insights into Subunit Interactions. *Physical Review Applied*, 7(1):014005, January 2017.

[107] Roya Zandi, Bogdan Dragnea, Alex Travesset, and Rudolf Podgornik. On virus growth and form. *Physics Reports*, 847:1–102, March 2020.

[108] Rees F. Garmann, Mauricio Comas-Garcia, Ajaykumar Gopal, Charles M. Knobler, and William M. Gelbart. The Assembly Pathway of an Icosahedral Single-Stranded RNA Virus Depends on the Strength of Inter-Subunit Attractions. *Journal of Molecular Biology*, 426(5):1050–1060, March 2014.

[109] William M. Gelbart and Charles M. Knobler. Pressurized Viruses. *Science*, 323(5922):1682–1683, March 2009.

[110] Michael F. Hagan. Modeling Viral Capsid Assembly. In Stuart A. Rice and Aaron R. Dinner, editors, *Advances in Chemical Physics*, pages 1–68. Wiley, 1 edition, April 2014.

[111] Jason D Perlmutter, Cong Qiao, and Michael F Hagan. Viral genome structures are optimal for capsid assembly. *eLife*, 2:e00632, June 2013.

[112] Siyu Li, Gonca Erdemci-Tandogan, Jef Wagner, Paul Van Der Schoot, and Roya Zandi. Impact of a nonuniform charge distribution on virus assembly. *Physical Review E*, 96(2):022401, August 2017.

[113] Eric C. Dykeman, Peter G. Stockley, and Reidun Twarock. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proceedings of the National Academy of Sciences*, 111(14):5361–5366, April 2014.

[114] Puja Adhikari, Neng Li, Matthew Shin, Nicole F. Steinmetz, Reidun Twarock, Rudolf Podgornik, and Wai-Yim Ching. Intra- and intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: implication for ACE2 receptor binding. *Physical Chemistry Chemical Physics*, 22(33):18272–18283, 2020.

[115] Sanaz Panahandeh, Siyu Li, Bogdan Dragnea, and Roya Zandi. Virus Assembly Pathways Inside a Host Cell. *ACS Nano*, 16(1):317–327, January 2022.

[116] Siyu Li and Roya Zandi. Biophysical Modeling of SARS-CoV-2 Assembly: Genome Condensation and Budding. *Viruses*, 14(10):2089, September 2022.

[117] F.William Studier and Barbara A. Moffatt. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189(1):113–130, May 1986.

[118] J Grodberg and J J Dunn. ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *Journal of Bacteriology*, 170(3):1245–1253, March 1988.

[119] Bruno Miroux and John E. Walker. Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology*, 260(3):289–298, July 1996.

[120] F.William Studier. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology*, 219(1):37–44, May 1991.

[121] G. Bertani. STUDIES ON LYSOGENESIS I: The Mode of Phage Liberation by Lysogenic Escherichia coli. *Journal of Bacteriology*, 62(3):293–300, September 1951.

[122] K. Brooks Low. *Experiments in Molecular Genetics.* Jeffrey H. Miller. *The Quarterly Review of Biology*, 49(2):151–151, June 1974.

[123] K. D. Tartoff and C. A. Hobbs. Improved media for growing plasmids and cosmid clones. *Bethesda Research Laboratories Focus*, 9(12), 1987.

[124] John W. Dubendorf and F.William Studier. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *Journal of Molecular Biology*, 219(1):45–59, May 1991.

[125] Elisabeth Gasteiger, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. Protein Identification and Analysis Tools on the ExPASy Server. In John M. Walker, editor, *The Proteomics Protocols Handbook*, pages 571–607. Humana Press, Totowa, NJ, 2005.

[126] Paul Bauer, Berk Hess, and Erik Lindahl. GROMACS 2022.3 Manual. September 2022. Publisher: Zenodo Version Number: 2022.3.

[127] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L De Groot, Helmut Grubmüller, and Alexander D MacKerell. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1):71–73, January 2017.

[128] Sunhwan Jo, Taehoon Kim, and Wonpil Im. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE*, 2(9):e880, September 2007.

[129] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865, August 2008.

[130] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614, July 2009.

[131] Sunhwan Jo, Joseph B. Lim, Jeffery B. Klauda, and Wonpil Im. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical Journal*, 97(1):50–58, July 2009.

[132] Jumin Lee, Xi Cheng, Jason M. Swails, Min Sun Yeom, Peter K. Eastman, Justin A. Lemkul, Shuai Wei, Joshua Buckner, Jong Cheol Jeong, Yifei Qi, Sunhwan Jo, Vijay S. Pande, David A. Case, Charles L. Brooks, Alexander D. MacKerell, Jeffery B. Klauda, and Wonpil Im. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, 12(1):405–413, January 2016.

[133] Jumin Lee, Dhilon S. Patel, Jonas Ståhle, Sang-Jun Park, Nathan R. Kern, Seonghoon Kim, Joonseong Lee, Xi Cheng, Miguel A. Valvano, Otto Holst, Yuriy A. Knirel, Yifei Qi, Sunhwan Jo, Jeffery B. Klauda, Göran Widmalm, and Wonpil Im. CHARMM-GUI *Membrane Builder* for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *Journal of Chemical Theory and Computation*, 15(1):775–786, January 2019.

[134] Jumin Lee, Manuel Hitzenberger, Manuel Rieger, Nathan R. Kern, Martin Zacharias, and Wonpil Im. CHARMM-GUI supports the Amber force fields. *The Journal of Chemical Physics*, 153(3):035103, July 2020.

[135] Soohyung Park, Yeol Kyo Choi, Seonghoon Kim, Jumin Lee, and Wonpil Im. CHARMM-GUI Membrane Builder for Lipid Nanoparticles with Ionizable Cationic Lipids and PEGylated Lipids. *Journal of Chemical Information and Modeling*, 61(10):5192–5202, October 2021.

[136] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, March 1985.

[137] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, June 1984.

[138] Shuichi Nosé and M.L. Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50(5):1055–1076, December 1983.

[139] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, December 1981.

[140] Richard Gowers, Max Linke, Jonathan Barnoud, Tyler Reddy, Manuel Melo, Sean Seyler, Jan Domański, David Dotson, Sébastien Buchoux, Ian Kenney, and Oliver Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. pages 98–105, Austin, Texas, 2016.

[141] Naveen Michaud-Agrawal, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10):2319–2327, July 2011.

[142] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004.

[143] Joshua A. Anderson, Jens Glaser, and Sharon C. Glotzer. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Computational Materials Science*, 173:109363, February 2020.

[144] Jean-Louis Rigaud and Daniel Lévy. Reconstitution of Membrane Proteins into Liposomes. In *Methods in Enzymology*, volume 372, pages 65–86. Elsevier, 2003.

[145] Marie Therese Paternostre, Michel Roux, and Jean Louis Rigaud. Mechanisms of membrane protein insertion into liposomes during reconstitution procedures involving the use of detergents. 1. Solubilization of large unilamellar liposomes (prepared by reverse-phase evaporation) by Triton X-100, octyl glucoside, and sodium cholate. *Biochemistry*, 27(8):2668–2677, April 1988.

[146] Jean Louis Rigaud, Marie Therese Paternostre, and Aline Bluzat. Mechanisms of membrane protein insertion into liposomes during reconstitution procedures involving the use of detergents. 2. Incorporation of the light-driven proton pump bacteriorhodopsin. *Biochemistry*, 27(8):2677–2688, April 1988.

[147] Theresa M. Allen, Alice Y. Romans, Henri Kercret, and Jere P. Segrest. Detergent removal during membrane reconstitution. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 601:328–342, 1980.

[148] Dominik Drabik, Martynas Gavutis, Ramunas Valiokas, and Arturas Ulčinas. Determination of the Mechanical Properties of Model Lipid Bilayers Using Atomic Force Microscopy Indentation. *Langmuir*, 36(44):13251–13262, November 2020.

[149] Senli Guo and Boris B. Akhremitchev. Packing Density and Structural Heterogeneity of Insulin Amyloid Fibrils Measured by AFM Nanoindentation. *Biomacromolecules*, 7(5):1630–1636, May 2006.

[150] Han Liu, Haohao Fu, Christophe Chipot, Xueguang Shao, and Wensheng Cai. Accuracy of Alternate Nonpolarizable Force Fields for the Determination of Protein–Ligand Binding Affinities Dominated by Cation Interactions. *Journal of Chemical Theory and Computation*, 17(7):3908–3915, July 2021.

[151] Hanif M. Khan, Alexander D. MacKerell, and Nathalie Reuter. Cation- Interactions between Methylated Ammonium Groups and Tryptophan in the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, 15(1):7–12, January 2019.

[152] B.W. Neuman and M.J. Buchmeier. Supramolecular Architecture of the Coronavirus Particle. In *Advances in Virus Research*, volume 96, pages 1–27. Elsevier, 2016.

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

[154] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, July 2012.

[155] David R. Stirling, Madison J. Swain-Bowden, Alice M. Lucas, Anne E. Carpenter, Beth A. Cimini, and Allen Goodman. CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics*, 22(1):433, December 2021.

[156] Ting Liu, Lingyun Zhang, Donghyun Joo, and Shao-Cong Sun. NF-B signaling in inflammation. *Signal Transduction and Targeted Therapy*, 2(1):17023, July 2017.

[157] Nathan C Shaner, Robert E Campbell, Paul A Steinbach, Ben N G Giepmans, Amy E Palmer, and Roger Y Tsien. Improved monomeric red, orange and yellow fluorescent

proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology*, 22(12):1567–1572, December 2004.

[158] Cinzia Dello Russo, Natalia Cappoli, Isabella Coletta, Daniele Mezzogori, Fabiola Paciello, Giacomo Pozzoli, Pierluigi Navarra, and Alessandra Battaglia. The human microglial HMC3 cell line: where do we stand? A systematic literature review. *Journal of Neuroinflammation*, 15(1):259, December 2018.

[159] Houmin Lin, Steven Grant Dixon, Wei Hu, Eric D. Hamlett, Junfei Jin, Adviye Ergul, and Gavin Y. Wang. p38 MAPK Is a Major Regulator of Amyloid Beta-Induced IL-6 Expression in Human Microglia. *Molecular Neurobiology*, 59(9):5284–5298, September 2022.

[160] Hui Li, Feng Wang, Xuqi Guo, and Yugang Jiang. Decreased MEF2A Expression Regulated by Its Enhancer Methylation Inhibits Autophagy and May Play an Important Role in the Progression of Alzheimer's Disease. *Frontiers in Neuroscience*, 15:682247, June 2021.

[161] Adrián Martínez-Limón, Manel Joaquin, María Caballero, Francesc Posas, and Eulàlia De Nadal. The p38 Pathway: From Biology to Cancer Therapy. *International Journal of Molecular Sciences*, 21(6):1913, March 2020.

[162] Harish Shankaran, Danielle L Ippolito, William B Chrisler, Haluk Resat, Nikki Bollinger, Lee K Opresko, and H Steven Wiley. Rapid and sustained nuclear–cytoplasmic ERK oscillations induced by epidermal growth factor. *Molecular Systems Biology*, 5(1):332, January 2009.

[163] John G. Albeck, Gordon B. Mills, and Joan S. Brugge. Frequency-Modulated Pulses of ERK Activity Transmit Quantitative Proliferation Signals. *Molecular Cell*, 49(2):249–261, January 2013.

[164] Jacob Stewart-Ornstein, Ho Wa (Jacky) Cheng, and Galit Lahav. Conservation and Divergence of p53 Oscillation Dynamics across Species. *Cell Systems*, 5(4):410–417.e4, October 2017.

[165] Peter Ruoff and Nobuaki Nishiyama. Frequency switching between oscillatory homeostats and the regulation of p53. *PLOS ONE*, 15(5):e0227786, May 2020.

[166] Eric Batchelor, Caroline S. Mock, Irun Bhan, Alexander Loewer, and Galit Lahav. Recurrent Initiation: A Mechanism for Triggering p53 Pulses in Response to DNA Damage. *Molecular Cell*, 30(3):277–289, May 2008.

[167] Jeremy E. Purvis, Kyle W. Karhohs, Caroline Mock, Eric Batchelor, Alexander Loewer, and Galit Lahav. p53 Dynamics Control Cell Fate. *Science*, 336(6087):1440–1444, June 2012.

[168] Andrew L. Paek, Julia C. Liu, Alexander Loewer, William C. Forrester, and Galit Lahav. Cell-to-Cell Variation in p53 Dynamics Leads to Fractional Killing. *Cell*, 165(3):631–642, April 2016.

[169] Galit Lahav, Nitzan Rosenfeld, Alex Sigal, Naama Geva-Zatorsky, Arnold J Levine, Michael B Elowitz, and Uri Alon. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nature Genetics*, 36(2):147–150, February 2004.

[170] M. K. Rasmussen, L. Iversen, C. Johansen, J. Finnemann, L. S. Olsen, K. Kragballe, and Borbala Gesser. IL-8 and p53 are inversely regulated through JNK, p38 and NF-B p65 in HepG2 cells during an inflammatory response. *Inflammation Research*, 57(7):329–339, July 2008.

[171] Xun Zuo, Shuisen Li, John Hall, Michael R. Mattern, Hiep Tran, Joshua Shoo, Robin Tan, Susan R. Weiss, and Tauseef R. Butt. Enhanced Expression and Purification of Membrane Proteins by SUMO Fusion in Escherichia coli. *Journal of Structural and Functional Genomics*, 6(2-3):103–111, September 2005.

[172] Dong Gui, Sharad Gupta, Jun Xu, Roya Zandi, Sarjeet Gill, I-Chueh Huang, A.L.N. Rao, and Umar Mohideen. A novel minimal in vitro system for analyzing HIV-1 Gag-mediated budding. *Journal of Biological Physics*, 41(2):135–149, March 2015.

[173] Alexander Stukowski. Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering*, 18(1):015012, January 2010.

[174] Dennis Hartmann. Chapter 6b: Spectra, data windows, significance testing, Winter 2017.

[175] K. D. Tartoff and C. A. Hobbs. Improved media for growing plasmids and cosmid clones. *Bethesda Research Laboratories Focus*, 9(12), 1987.