THE UNIVERSITY OF CHICAGO

THE MEMBRANE CONTEXT OF PHOSPHATIDYLSERINE EXPOSURE INFLUENCES ITS RECOGNITION BY THE TIM PROTEINS: A STRUCTURALLY GUIDED STUDY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
AND
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOPHYSICAL SCIENCES

BY
DANIEL HENRY SOLZ KERR

CHICAGO, ILLINOIS
AUGUST 2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

While this thesis work mainly represents seven years of my labor, it also includes the contributions and labor of several other people. Many people were instrumental to this work, without whom the story told here would be a lot less satisfying.

Firstly, I'd like to thank my advisor, Ka Yee Lee, who provided nearly unconditional support for me and my work. I had a few rough spells on my journey, and Ka Yee patiently gave me the time and encouragement to pull myself up by my own bootstraps and find my way. It is a testament to Ka Yee's commitment to scientific inquiry that she let me challenge accepted methods when there was no indication anywhere in the field that these approaches were problematic. Overall, Ka Yee is just an incredibly warm person and it was a pleasure to work for her throughout my graduate career.

In addition to Ka Yee's contribution as a mentor, she contributed in cultivating a lab a supportive and cooperative community. The Lee Lab is the perfect work environment I would ever ask for. I could turn to just about anybody for anything and know I would have their full attention. In particular, I have to thank Greg Tietjen and Zhiliang Gong as constant collaborators for the TIM proteins, even well after both of them left the lab. Greg started this project and taught me everything I knew about it. He left a huge amount of both data and potential lines of inquiry that provided me with a clear, defined set of questions for my thesis. Zhiliang and I pursued these questions vigorously, and we were very much partners in solving the mysteries of the TIM proteins. Other Lee lab members were also instrumental and not just as a source of support. For instance, the x-ray data that serves as the backbone of my thesis work was acquired by several Lee Lab members. I thank Kathleen Cao, Charles Heffern, and Luke Hwang for devoting several hours to obtaining this data. I am also thankful to Michael Henderson who led by example and demonstrated how to efficiently run an Argonne study. No beam time has ever run as smoothly as it did with Michael. I also have to thank a number of undergraduates who have contributed a huge amount to this project. Tiffany Suwatthee regularly assisted Zhiliang and me in obtaining binding data. Simon Kit Sang Chu was instrumental in developing the framework for integrating simulations into x-ray analysis and without his sharp, critical questions of my instructions, we would not have completed this work so quickly. Renee Scarpaci spent just one summer with us but amazingly obtained key data in that short time. And while I did not directly work alongside every Lee Lab member, I am thankful to the constant support I received from Ben Slaw, Peter Chung, Alessandra Leong, and Andrew Molina. Finally, I am thankful to our newest member, Sofi Maltseva, who I know will take this project into even more enriching territory. She has only been working on the project for my last few months here but has already pushed my understanding of the system significantly in that time.

I also thank a number of mentors who supported and encouraged me to keep pushing forward. Ted Steck, in addition to being my committee chair, has been a constant motivator for deeper inquiry. Without our regular meetings, entire avenues of this project would not have been explored. Ideas that started out as tangential curiosities evolved into a complete framework for this project. Ted's combined interest and expertise in the project made this possible, as he asked the right questions to catalyze my development of this work. I also would like to thank Mark Schlossman for his contributions to the x-ray methodology of this project. One of the aforementioned inconsistencies I encountered was the analysis of x-ray data standard for the whole field. As an early graduate student, I panicked when the accepted analysis methodology produced inconsistent results. Naturally I blamed myself and my own understanding but I could not figure out where I had gone wrong. I turned to Mark and, after spending considerable time on the problem, reassured me that my uncertainties were well-founded. This validation was instrumental in my resolution of the problem. I thought his contributions were so invaluable I requested he be added to my committee despite not being associated with the University of Chicago, which he gratefully accepted. Lastly, Adam Hammond and his consistently open door were huge influences in my direction. Particularly early on when I was floundering, figuring out what to do both scientifically and regarding life in general, Adam was always there to discuss these difficulties. Adam's broad interests within and beyond science pushed me forward in many ways and expanded my view of what science could be.

This work was also greatly enhanced by our collaborations with other labs. Absolutely essential to this project were the contributions of Erin Adams and her lab. Adrienne Luoma, Charlie Dulberger, and Sobhan Roy are all amazing people to work with, each of them incredibly patient and generous. None of them studied the TIM proteins but were nonetheless completely committed to assisting us in our studies. The simulation component of this work was similarly completely reliant on the contributions of Emad Tajkhorshid and his former student, Javier Baylon. They developed the simulation methods I used in my thesis work and Javier showed us how to run these simulations ourselves with his work on TIM1. I also thank Michael Stein of the statistics department. I had struggled with the correct statistical analysis of x-ray data for over a year even after consulting multiple textbooks and scientists. In just one meeting with Michael, these statistical difficulties were resolved such that a week later I had a complete, consistent framework.

In addition, I am indebted to those who maintained the equipment necessary for my work. Binhua Lin, Wei Bu, and Mati Meron have made experiments at APS sector 15, ChemMatCARS, so incredibly productive because of their highly competent support. Even if we were stranded out there at 3am with a technical failure, we could easily rely on Binhua or Wei to rescue us. Wei, in particular, has made on-site data analysis possible with his recent python code. This has allowed us to analyze data as we get it making our precious beam time even more useful. I also thank Lorenzo Pesce and Joe Urbanski for running Beagle. We are not a computational group, but these two got me completely up and running with simulations. I am deeply indebted to their efforts and Lorenzo's direct demeanor in getting this working for us. I am similarly indebted to the efforts of Elena Solomaha for her support with the fluorolog. All of these individuals maintain shared equipment, which is a difficult job given the propensity of users to mess things up for everybody else. I hope I have not been one of those users and I appreciate their efforts in ensuring that I am minimally affected by the messes of other users.

The biophysics program has been a similarly great and supportive community. I have been here for several generations of students but the community itself remains a constant. This is in large part due to the efforts of Michele Wittels and Julie Feder, who are the two greatest administrators I have ever relied on. The students themselves also play a large role in fostering a great program and I have to thank pioneers such as David Medovoy, Tobias Falzone, Greg Tietjen, Phil Long, Caitlin Sullivan Trejo, and James Crooks for immediately making the program feel comfortable. I also have to thank my cohort and the year below as well for their friendship, Sean Gibbons, Pete Dahlberg, Guillermina Ramirez San Juan, Bo Osinski, Vu Dinh, Kristen Hitchcox, Herman Gudjonson, Kevin Song, Monika Scholtz, Eugene Leypunskiy, Alan Hutchison, and Ryan Mork. As the last of this group to leave the University of Chicago, I am inspired by all of their accomplishments and incredibly thankful for the community they provided throughout my time here.

I also wish to thank a number of mentors from before graduate school. It is easy to take the path that led to graduate school for granted but several individuals inspired and pushed me to go into science. I firstly thank my high school advisor and physics teacher, Tengiz Bibilashvili. He recognized my potential early on and immediately took me under his wing. Most of my time in school was spent working on physics problems under his direction. He pushed me to participate in a science tournament for two summers. In this experience, I was further mentored by Lewis Forsheit and Dr. Bruce Oldaker. From UCSB, I have many advisors and professors to thank: The college of creative studies physics program put me under the tutelage of Francesc Roig, Sathya Guruswamy, and Everett Lipman, who all provided me with an excellent theoretical and experimental physics background. I am also indebted to many other professors of math and biology including Mihai Putinar, Paul Atzberger, Kathy Foltz, Armin Kuris, and Bob Warner, all of whom expanded my scientific literacy beyond physics.

Throughout the hardest moments of work, my girlfriend Soo Ji has been there for me both emotionally and physically. In addition to her help, Soo Ji has been my partner in life the last two years: We climb together, play tennis together, travel together, eat exorbitantly together, do yoga together and etc... In writing this dissertation I also am thankful she pushed us to get our dog, Maru. Just having unconditional love from the two of them at the deepest moments of stress ameliorated the whole process.

Lastly, I am indebted to my family both personally and scientifically. My parents both pushed me early on to pursue physics and math. I remember encountering such difficulties early on such as failing to have the intuition that the distance between the sun and the earth has to be much, much larger than the distance between the earth and the moon (this was 8th grade just for reference). I think most kids would have been discouraged at such massive blunders of intuition, declaring themselves to be “not good at math,” but because of my parents’ interests and their support, I never had this narrative for myself. I had a mixed history of excelling and falling behind in these subjects and regardless my parents pushed me further. Alongside me, my sister Johanna was similarly reaching high in the sky but unlike me, consistently succeeded. Despite being younger, my sister obtained her doctorate earlier than me, guided by a laser sharp focus on realizing her goals. Her efforts have been a constant source of inspiration and pride. My grandparents have also been constant source of support. Despite arriving in the US with no money after the Holocaust, they were able to accrue significant savings to pay for my college (and more). Because of them I don’t have to deal with debt as many in my generation do. They served an example for this family that I hope I can live up to.

There are many other people who made my life here great who I did not get to mention in my acknowledgements. Thank you all.

CHAPTER 1

INTRODUCTION

1.1 Structural Basis of Phospholipid Recognition by Peripheral Membrane-Binding Proteins

The broad goal of this thesis work is to reveal the structural elements of peripheral membrane-binding proteins, specifically the Transmembrane immunoglobulin and mucin domain (TIM) protein family, that directly confer sensitivity to phospholipid membranes. Peripheral membrane-binding proteins can distinguish their target membranes from those of different cells or even of the same cells in different states (1, 2). This specific recognition is only possible due to the varying membrane compositions associated with each organelle, type of cell, or state of a cell (1, 3, 4). Therefore, peripheral membrane-binding proteins must be structurally “fine-tuned” to the physical and chemical properties of their target membranes.

While the structures of many phospholipid-binding proteins have been determined in solution (5, 6), the membrane-bound structures of these proteins are typically intuited assuming that the membrane induces no conformational change in the protein. Alternatively, these structures are simulated with no direct experimental constraints and are indirectly validated in their comparison with experimental results (7, 8). Unlike structures of protein-protein interactions, which isolate specific molecular interactions, protein-membrane models lack such molecular details. Without this information, it is exceedingly difficult to determine how disease variants of these proteins disrupt membrane-binding or to design proteins with a desired membrane specificity. The field demands a rigorous method for determining the membrane-bound states of peripheral membrane-binding proteins and thus the protein residues which directly interact with the membrane. The largest contributions of this thesis are the development of such a method and its application to the recognition of phosphatidylserine (PS) by three murine members of the TIM protein family, TIM1, TIM3, and TIM4.

In this dissertation, the structures of the TIM proteins are used to correlate differences in their association with PS-containing membranes to the residues that interface with the lipid membrane. These models implicate specific energetic contributions to protein-membrane binding, such as hydrophobic insertion and electrostatic engagement. The relative strength of each of these energetic terms correlates with the selectivity of the TIM proteins for particular membranes. Moreover, the membrane-bound structures identify the amino acid residues that directly implement these interactions. This work provides both the molecular details that underlie the recognition of PS-containing membranes by the TIM proteins and the prescription for obtaining such molecular details for the broader class of peripheral membrane-binding proteins.

1.2 A Structure-Informed Characterization of Protein-Membrane Binding Reveals the Mechanism of Membrane Selectivity

The determination of membrane-bound structures is part of a larger framework to generally characterize the interaction of proteins with phospholipid membranes. This framework is composed of three approaches: (i) Characterization of binding to model membranes determined by equilibrium thermodynamic methods; (ii) Determination of membrane-bound structures using x-ray reflectivity of proteins adsorbed to a lipid monolayer at the air-water interface, which serves a mimic for the outer leaflet of a bilayer; (iii) Molecular dynamics (MD) simulations of these structures on model bilayers. Combining the information obtained from these seemingly disjoint approaches presents a challenge. Thus, a major focus of this thesis is the integration of these approaches in a single paradigm, as outlined in **Figure 1.1**.

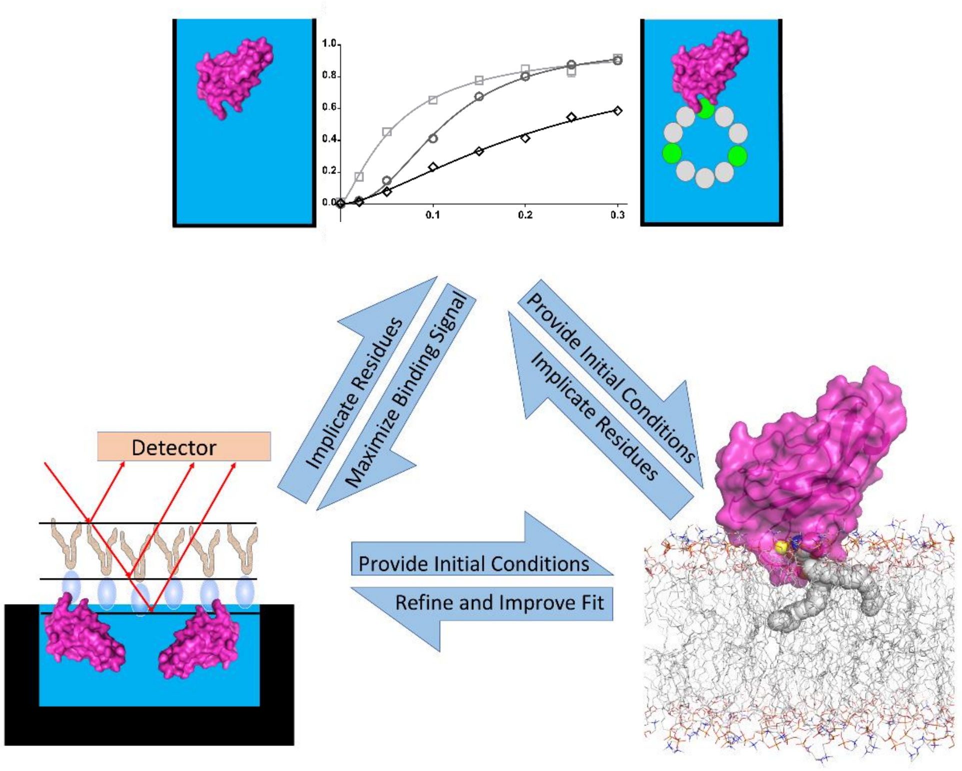


Figure 1.1. Experimental Paradigm for Peripheral Membrane-Binding Proteins. The combination of thermodynamic equilibrium binding, x-ray reflectivity, and MD simulations characterize the specificity of peripheral membrane-binding proteins to various membranes. These techniques each inform the others as well, such as binding experiments providing maximal binding conditions for x-ray reflectivity and MD or x-ray reflectivity implicating the membrane-contact surface that confers the membrane specificity measured in binding experiments. Figure adapted from (9).

In particular, the obtained membrane-bound structures of the TIM proteins inspired a comprehensive mathematical approach that describes the association of peripheral membrane-binding proteins to anionic phospholipids. In the literature, protein-membrane binding is measured by a myriad of approaches and analyzed by Michaelis-Menten and Hill equations despite the multi-ligand interactions occurring in a complex combination of three-dimensional and two-dimensional reactions (10–13). The equations developed in this work explicitly model these complex reactions and explains a wide range of disparate results not just for the TIM proteins, but other anionic lipid-binding proteins such as Annexin V, Protein Kinase C, and the Synaptotagmins (10–12, 14, 15). Additionally, the equations prescribe a set of binding experiments which efficiently determine the parameters that best differentiate the membrane selectivity of various proteins based on their physicochemical interactions. The structure-centric approach thus culminates in offering a standard set of experimental parameters in which to compare this broad class of proteins.

1.3 Organization

The dissertation divides the explication of the TIM protein system and the three approaches presented above in the following chapters. Chapter 2 presents background on the lipid membranes of cells, the recognition of specific membranes by peripheral membrane-binding proteins, and the PS recognition of the TIM proteins in their physiological context. The experimental methods used in this thesis, tryptophan fluorescence and x-ray reflectivity, are discussed in Chapter 3, along with the procedures used to produce the TIM proteins and components of these experiments. Chapter 4 is entirely devoted to the analytical tools developed to analyze x-ray reflectivity and MD simulations in parallel. Chapter 5 describes the determination of the membrane-bound structure of TIM1 using the methods presented in Chapter 4. The determination of the membrane-bound structure of TIM3 is discussed in Chapter 6. The mathematical binding model is developed in Chapter 7, followed by the application of the derived equations to the TIM proteins. The results of Chapters 5-7 are integrated in Chapter 8, followed by a discussion of the physiological implications of this work. The thesis then concludes with an overview of the future directions entailed by this work.

1.4 List of Publications

Gong Z, Kerr D, Hwang HL, Henderson JM, Suwatthee T, Slaw BR, Cao KD, Lin B, Bu W, Lee KYC (2017) Quantitative analysis of total reflection X-ray fluorescence from finely layered structures using XeRay. *Review Scientific of Instruments*; 88: 033112.

Tietjen GT, Baylon JL, Kerr D, Gong Z, Henderson JM, Heffern CTR, Meron M, Lin B, Schlossman ML, Adams EJ, Tajkhorshid E, Lee KYC (2017) Coupling X-Ray Reflectivity and In Silico Binding to Yield Dynamics of Membrane Recognition by Tim1. *Biophysical Journal*; 113: 1505–1519.

Kerr D, Tietjen GT, Gong Z, Tajkhorshid E, Adams EJ, Lee KYC (2018) Sensitivity of peripheral membrane proteins to the membrane context: A case study of phosphatidylserine and the TIM proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes*; 1860: 2126–2133.

Kerr D, Gong Z, Suwatthee T, Tietjen GT, Luoma A, Roy S, Scarpaci R, Hwang HL, Henderson JM, Cao KD, Bu W, Lin B, Steck TL, Adams EJ, Lee KYC (2019) Phosphatidic acid enhances association of Tim3 with phosphatidylserine membranes. *In preparation*.

Kerr D, Gong Z, Suwatthee T, Chu KS, Tietjen GT, Luoma A, Dulberger C, Roy S, Hwang HL, Henderson JM, Cao KD, Heffern CTR, Bu W, Meron M, Lin B, Steck TL, Schlossman ML, Adams EJ, Lee KYC (2020) Determination of membrane-bound Tim3 structure under physiological low affinity conditions. *In preparation*.

Kerr D, Maltseva S, Suwatthee T, Dulberger C, Roy S, Steck TL, Adams EJ, Lee KYC (2020) General binding equations for peripheral membrane-binding proteins. *Under Investigation*.

1.5 References

1. Harayama T, Riezman H (2018) Understanding the diversity of membrane lipid composition. *Nat Rev Mol Cell Biol* 19(5):281–296.
2. Bigay J, Antonny B (2012) Curvature, lipid packing, and electrostatics of membrane organelles: defining cellular territories in determining specificity. *Dev Cell* 23(5):886–895.
3. Van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* 9(2):112–124.
4. Klose C, et al. (2012) Flexibility of a eukaryotic lipidome--insights from yeast lipidomics. *PLoS ONE* 7(4):e35063.
5. Stace CL, Ktistakis NT (2006) Phosphatidic acid- and phosphatidylserine-binding proteins. *Biochim Biophys Acta* 1761(8):913–926.
6. Caffrey M (2015) A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallogr Sect F Struct Biol Commun* 71(Pt 1):3–18.
7. Cho H, Wu M, Bilgin B, Walton SP, Chan C (2012) Latest developments in experimental and computational approaches to characterize protein-lipid interactions. *Proteomics* 12(22):3273–3285.
8. Monje-Galvan V, Klauda JB (2016) Peripheral membrane proteins: Tying the knot between experiment and computation. *Biochim Biophys Acta* 1858(7 Pt B):1584–1593.
9. Kerr D, et al. (2018) Sensitivity of peripheral membrane proteins to the membrane context: A case study of phosphatidylserine and the TIM proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1860(10):2126–2133.
10. Tait JF, Gibson DF, Smith C (2004) Measurement of the affinity and cooperativity of annexin V-membrane binding under conditions of low membrane occupancy. *Anal Biochem* 329(1):112–119.
11. Tran HT, Anderson LH, Knight JD (2019) Membrane-Binding Cooperativity and Coinsertion by C2AB Tandem Domains of Synaptotagmins 1 and 7. *Biophys J* 116(6):1025–1036.
12. Jeppesen B, Smith C, Gibson DF, Tait JF (2008) Entropic and enthalpic contributions to annexin V-membrane binding: a comprehensive quantitative model. *J Biol Chem* 283(10):6126–6135.
13. Tietjen GT, et al. (2014) Molecular mechanism for differential recognition of membrane phosphatidylserine by the immune regulatory receptor Tim4. *Proc Natl Acad Sci U S A* 111(15):E1463–72.

14. Bazzi MD, Nelsestuen GL (1991) Highly sequential binding of protein kinase C and related proteins to membranes. *Biochemistry* 30(32):7970–7977.

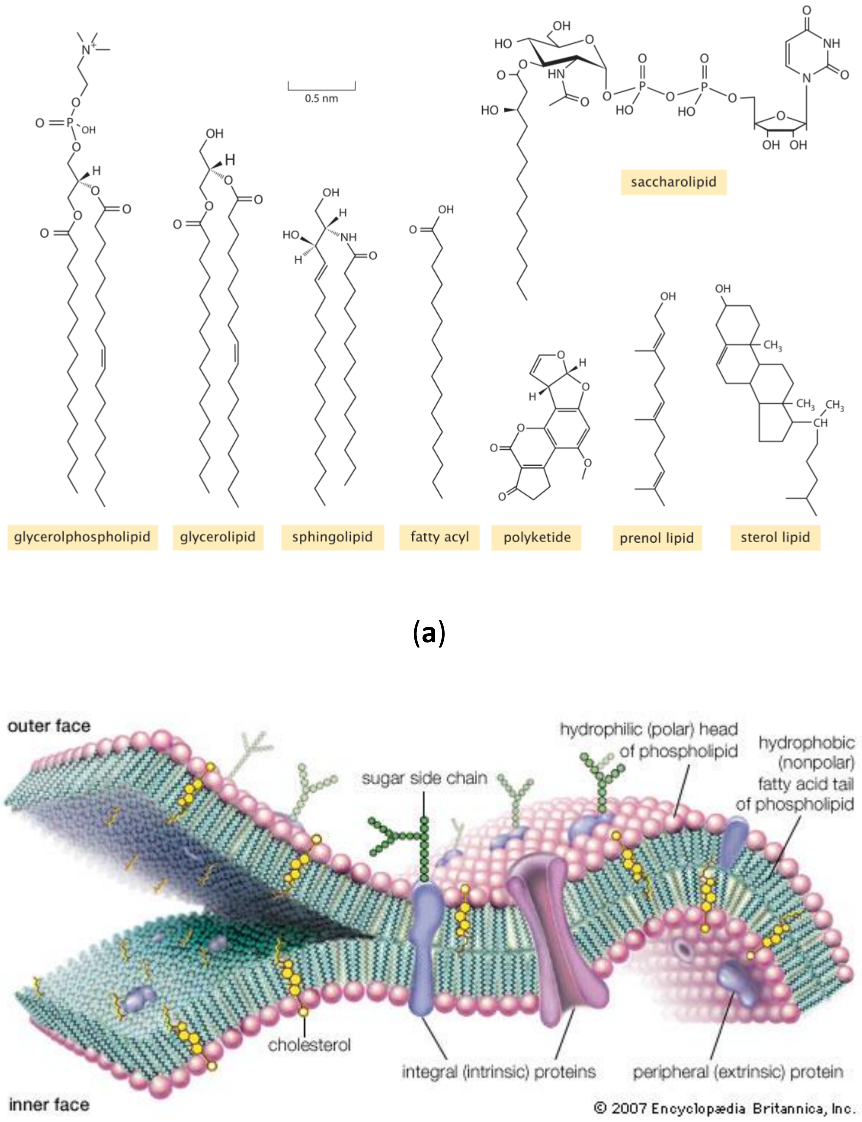
15. Mosior M, Newton AC (1998) Mechanism of the apparent cooperativity in the interaction of protein kinase C with phosphatidylserine. *Biochemistry* 37(49):17271–17279.

CHAPTER 2

BACKGROUND: PERIPHERAL MEMBRANE-BINDING PROTEINS AND THE RECOGNITION OF PHOSPHATIDYLSERINE BY TIM PROTEINS

2.1 The Identity and State of a Cell are Encoded in its Membrane Composition

Lipid membrane define the boundaries of cells or organelles by encapsulating them from their surrounding conditions. These membranes prevent the open exchange of water and other materials between the cell and its exterior, allowing cells and organelles to maintain distinct biochemical environments. The organization of these biological membranes arise as emergent structures determined by the chemical nature of their lipid components. A lipid molecule consists of a hydrophilic head group attached to a hydrophobic tail group of acyl chains. There are several molecular architectures of lipids that comprise cell membranes, a few of which are shown in **Figure 2.1a**. As a result of their amphiphilic structures, lipids in solution self-assemble as to maximally expose the hydrophilic head groups to their fluid environment and bury the hydrophobic tail groups. Biological membranes are organized in a bilayer structure in which two layers of lipids are aligned tail to tail, presenting a hydrophilic interface to the cell and its exterior separated by a hydrophobic barrier (1). However, the function of biological membranes is not restricted to enforcing a boundary, the membrane is also the interface at which cells and organelles communicate with one another. This communication is facilitated by proteins and other molecules embedded in the membrane alongside the lipids. Typically, the membrane is imagined as an inert two-dimensional fluid in which various active proteins and sugars function and interface with the cell exterior. This picture, introduced in 1972 and presented in **Figure 2.1b**, is called the fluid-mosaic model and has served as the *de facto* model of the cell membrane in biology since its conception (2).



(**a**)

(**b**)

Figure 2.1. Cell membranes are composed of lipids. (**a**) The structure of several types of lipids which comprise biological membranes. Reproduced from (3, 4). (**b**) Proposed in 1972 by Singer and Nicholson, the fluid-mosaic model proposes that the lipid membrane serves as the inert backdrop in which active proteins and sugars interface with the cell exterior. © 2007 Encyclopædia Britannica, Inc.

Biological membranes are composed of thousands of different lipid species. Each of these lipids confer emergent properties to the membranes they comprise such as curvature, membrane packing density, and electrostatic charge (5, 6). Contrary to the supposition of the fluid mosaic model that the lipid membrane serves as an inert backdrop, the properties of lipid membranes are specialized to the biochemical and physical requirements of the organelles or cells of which they enclose. Thus, the lipid composition of cell membranes significantly varies among different cell types, organelles, and organisms (7, 8), suggesting that the two-dimensional lipid membrane effectively encodes the identity of a cell or organelle. Changes in the lipid composition, asymmetry, or morphology of cell membranes reflect changes in the cell's internal state (9–11), allowing the membrane to transmit dynamic and contextual information to other cells through the recognition of peripheral membrane-binding proteins.

2.2 Peripheral Membrane-Binding Proteins Recognize Specific Membranes

Protein-lipid interactions underlie several biological phenomena such as the firing of a neuron, catalysis of chemical reactions, and transport of cargo across membranes. As opposed to membrane-embedded proteins which associate with lipids in plane, peripheral membrane-binding proteins are water-soluble and directly bind or adsorb to lipids. Once associated with the membrane, these proteins serve as structural scaffolds, induce changes in local membrane properties, or mediate cellular-scale processes such as phagocytosis, clot formation, and neurotransmitter release. Binding to a lipid membrane can cause downstream gene activation or repression, clustering of surface receptors, or membrane permeabilization. The regulation of these processes is governed, in part, by the specificity of these proteins for particular lipids and membrane conditions (12–14).

Peripheral membrane-binding proteins recognize particular membranes through one or more of the following methods (also illustrated in **Figure 2.2**): The burial of solvent-exposed amino acids, neutralization of exposed surface charges, post-translational modifications such as palmitoylation, and specific lipid coordination. The first three mechanisms depend on bulk membrane properties like fluidity, curvature, or charge, while the last relies on the presence of a specifically recognized lipid ligand (15). Understanding how these mechanisms mediate recognition of specific membranes would lead to numerous applications, for example, drug delivery platforms that target particular cells based on their distinct membranes.

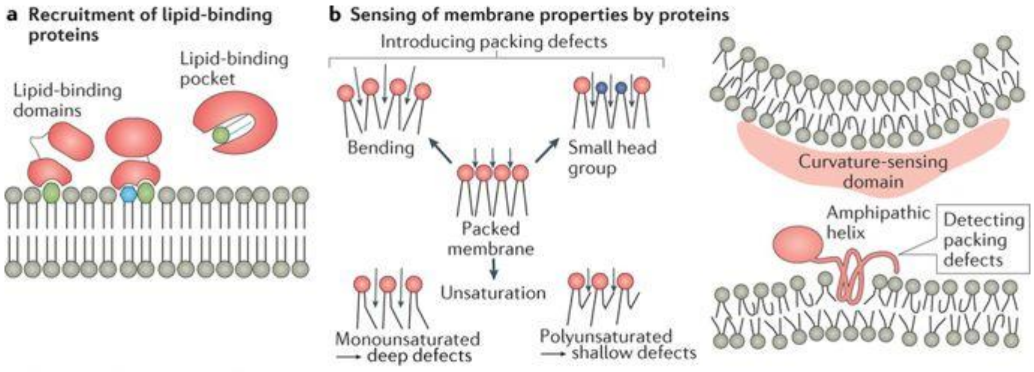


Figure 2.2. Peripheral membrane-binding proteins sense membrane properties. Membrane recognition by proteins can be achieved through multiple means by detection of various physical properties. (**a**) Proteins can recognize specific lipid species and/or interact non-specifically with anionic lipids. (**b**) The binding of a protein can be mediated by lipid packing, intrinsic curvature, or membrane defects. Figure is reproduced from (6).

2.3 Phosphatidylserine is a Signaling Lipid Recognized by a Large Class of Peripheral Membrane-Binding Proteins

Several lipid-binding proteins specifically recognize phosphatidylserine (PS) (16). In a healthy cell, PS is typically confined to the inner leaflet. The scrambling of PS to the outer leaflet of cell membranes serves as a signal for multiple pathways, some of which are depicted in **Figure 2.3** and the most prominent of which is apoptosis (17–23). The clearance of apoptotic cells by PS-recognizing phagocytes prevent damage to surrounding cells and tissues (24). During apoptosis, PS exposure is accompanied by changes in intracellular and local extracellular calcium concentrations, as well as membrane fluidity and the scrambling of additional inner leaflet lipids such as phosphatidylinositides (PI) and phosphatidylethanolamine (PE) (25–28). PS exposure in apoptosis is markedly higher and less transient than non-apoptotic PS signaling (29), implicating PS surface density as an important contextual signal (20, 30). These changes typically do not accompany non-apoptotic PS exposure and constitutive exposure of PS is not sufficient to induce phagocytosis (31), suggesting that the surrounding membrane is a significant component of the signal.

The diversity of PS presentation seemingly requires a complementary diversity of PS-receptors. PS-receptors are specialized for several extracellular and intracellular processes, such as the recognition of apoptotic cells, signal transduction, exocytosis, phosphorylation of phosphoinositide, and coagulation (16). Each of these processes has its own set of PS-receptors and occurs on distinct membranes. It is therefore likely that PS-binding proteins recognize both PS and the membrane context of PS presentation. Indeed, several PS-receptors have been demonstrated to be sensitive to the membrane environment of PS exposure. Annexin V, which inhibits blood coagulation by outcompeting other PS-receptors, binds cooperatively to PS and calcium, indicating a preference for membranes with high PS surface density (30, 32). Protein Kinase C (PKC) family members bind PS cooperatively with PI (33, 34) and the coagulation factor Protein Z binds PS and PE (35).

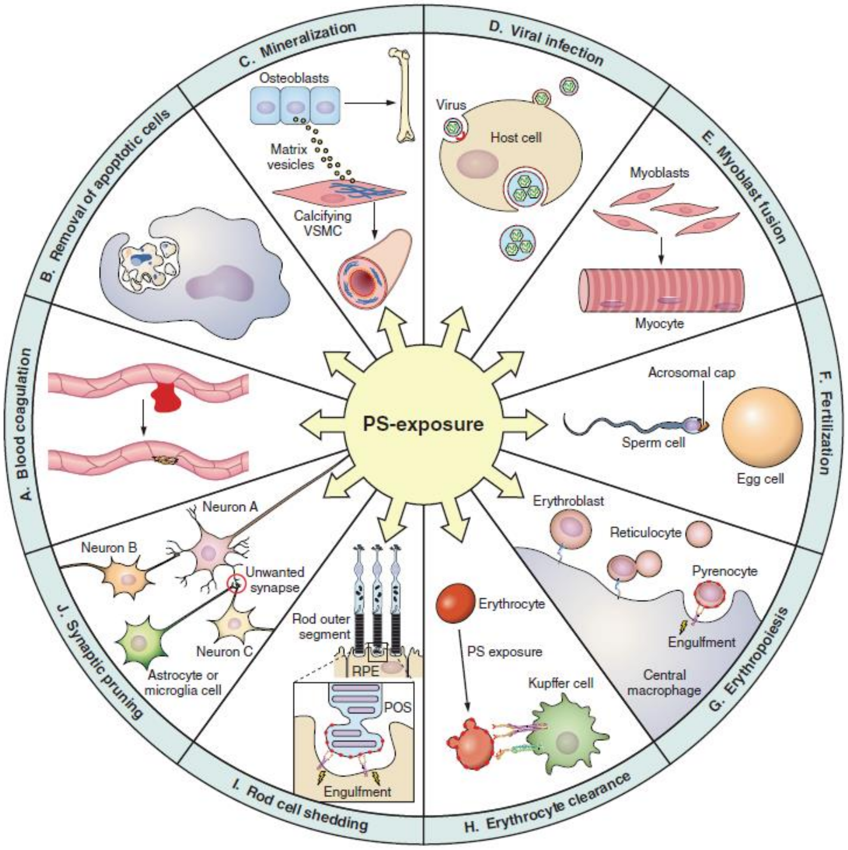


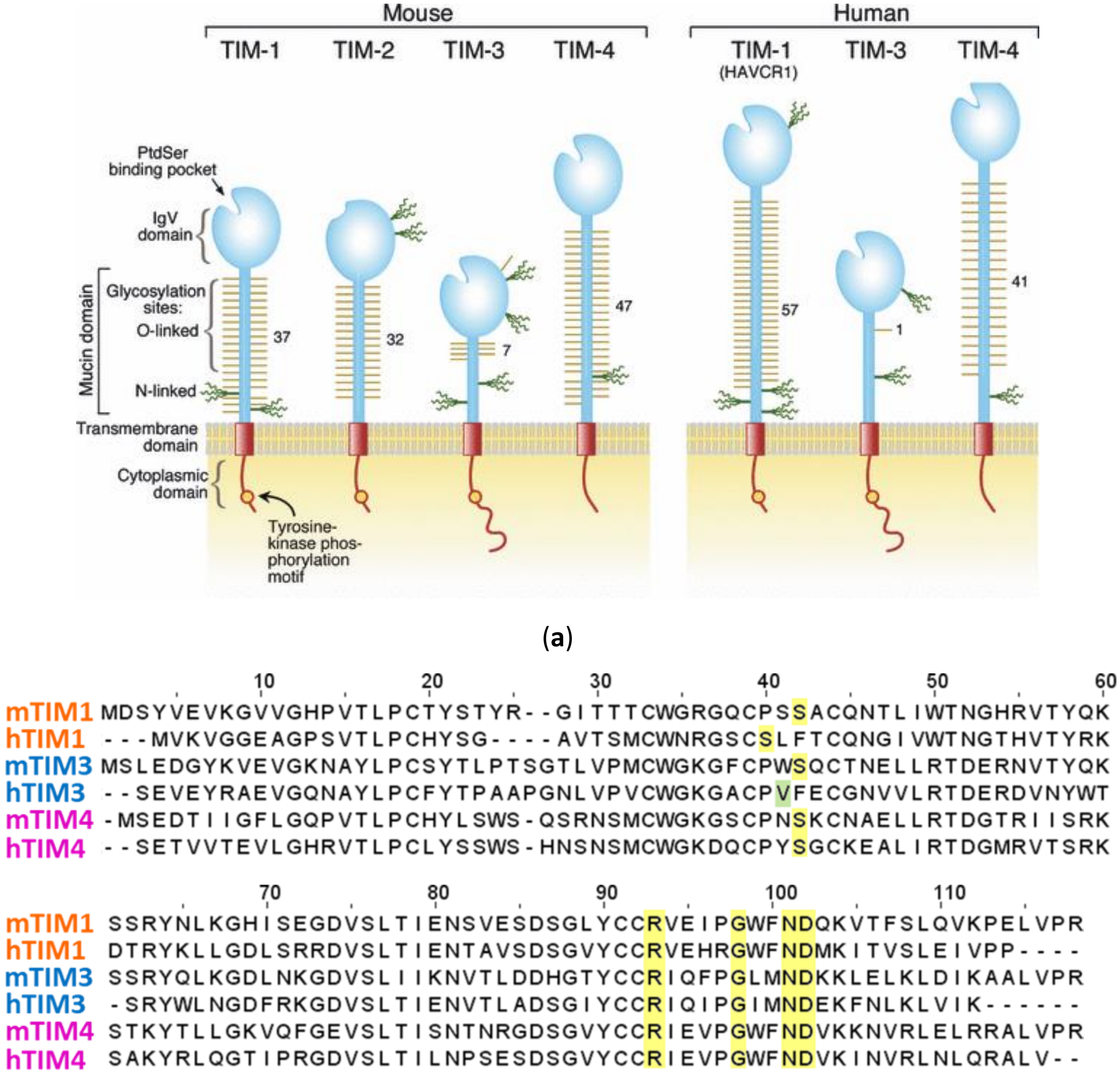
Figure 2.3. Exposure of phosphatidylserine signals multiple biological processes. The scrambling of PS occurs in several physiological functions in order to recruit cells bearing PS-receptors for phagocytosis, general engulfment of cells into tissues, or anchoring to tissues. Reproduced from (18).

2.4 The TIM Proteins Recognize PS in Distinct Physiological Processes

There are three human proteins in the Transmembrane Immunoglobulin and Mucin domain (TIM) family and four murine proteins, of which TIM1, TIM3, and TIM4 are homologous to the human protein family members (36). The TIM proteins take their name from the domains that comprise them: The immunoglobulin variable (IgV), mucin, transmembrane, and cytoplasmic domains as seen in **Figure 2.4a** (36). The TIM proteins were initially identified as the protein products expressed from a gene region associated with a mouse model for asthma (37). TIM1 and TIM3 were initially recognized as surface proteins demarcating distinct T cell populations that had opposing roles in the progression of the disease (38–43). TIM4 was later discovered to be a protein product of the same gene region and was the first TIM protein found to bind PS (44–46). These proteins are expressed on the exterior of the plasma membrane of several cell populations; TIM1 and TIM3 are predominantly expressed on T cells and B cells (40, 42) while TIM4 is mostly expressed on macrophages (44, 45, 47). All of these cells are believed to use apoptotic cells as a signal for mounting an immune response in the form of tolerance induction, inflammation, or phagocytosis.

TIM1, TIM3, and TIM4 specifically recognize a single PS coordinated by exactly one calcium ion in a conserved binding pocket located in the immunoglobulin domain (45, 46, 48–50) defined by the residues highlighted in **Figure 2.4b**. This pocket is preserved in the human and murine homologues. The mucin domain acts as a tether for the IgV domain to the cell membrane and contains several glycosylation and binding sites for additional ligands. The cytoplasmic domain contains a tyrosine phosphorylation motif in all but TIM4, suggesting that activation or clustering of the TIM proteins (possibly upon binding of PS) results in downstream regulation. The Ig and/or mucin domains of TIM1, TIM3, and TIM4 are also expressed as soluble isoforms, or are cleaved from the full-length isoforms (51–54). These soluble isoforms are exploited in disease states to downregulate the host immune response, presumably by outcompeting the PS recognition by the immune system. For example, soluble TIM1 is expressed in lymphoma (55) and soluble TIM3 is enriched in the plasma of untreated HIV patients and correlates with disease progression (51). The PS binding function is thus equally relevant for the full-length and soluble isoforms of the TIM proteins.

Given the high conservation of the PS binding pocket in the IgV domains between the various TIM protein members, as well as across species, binding PS is likely a significant function for all TIM proteins. However, the exact roles of the TIM proteins and the functional relevance of their recognition of PS have been difficult to characterize. It is thus natural to question why all cells do not use the same, single TIM protein or even just one PS-receptor. Instead, there are three distinct PS-binding TIM proteins – the same three conserved in mice, rats, and humans.



(a)

```
            10        20        30        40        50        60
mTIM1 MDSYVEVKGVVGHPVTLPCTYSTYR--GITTTCWGRGQCPSSACQNTLIWTNGHRVTYQK
hTIM1 ---MVKVGGEAGPSVTLPCHYSG----AVTSMCWNRGSCSLFTCQNGIVWTNGTHVTYRK
mTIM3 MSLEDGYKVEVGKNAYLPCSYTLPTSGTLVPMCWGKGFCPWSQCTNELLRTDERNVTYQK
hTIM3 --SEVEYRAEVGQNAYLPCFYTPAAPGNLVPVCWGKGACPVFECGNVVLRTDERDVNYWT
mTIM4 -MSEDTIIGFLGQPVTLPCHYLSWS-QSRNSMCWGKGSCPNSKCNAELLRTDGTRIISRK
hTIM4 --SETVVTEVLGHRVTLPCLYSSWS-HNSNSMCWGKDQCPYSGCKEALIRTDGMRVTSRK

              70        80        90       100       110
mTIM1 SSRYNLKGHISEGDVSLTIENSVESDSGLYCCRVEIPGWFNDQKVTFSLQVKPELVPR
hTIM1 DTRYKLLGDLSRRDVSLTIENTAVSDSGVYCCRVEHRGWFNDMKITVSLEIVPP----
mTIM3 SSRYQLKGDLNKGDVSLIIKNVTLDDHGTYCCRIQFPGLMNDKKLELKLDIKAALVPR
hTIM3 -SRYWLNGDFRKGDVSLTIENVTLADSGIYCCRIQIPGIMNDEKFNLKLVIK------
mTIM4 STKYTLLGKVQFGEVSLTISNTNRGDSGVYCCRIEVPGWFNDVKKNVRLELRRALVPR
hTIM4 SAKYRLQGTIPRGDVSLTILNPSESDSGVYCCRIEVPGWFNDVKINVRLNLQRALV--
```

(b)

Figure 2.4. Domain architecture of human and murine TIM proteins. (**a**) The TIM proteins each feature immunoglobulin variable (IgV), mucin stalk, cytoplasmic, and transmembrane anchoring domains. TIM1, TIM3, and TIM4 all bind PS through IgV domains. Reproduced from (36). (**b**) Sequence alignment of murine and human TIM proteins. Conserved pocket residues are highlighted in yellow with conservative replacements highlighted in green.

The need for multiple PS-receptors is best demonstrated by the role PS recognition plays in TIM4 physiology. The role of PS recognition in the physiology of TIM4 is the best understood of the

TIM proteins (56). TIM4 is predominantly expressed on the surface of macrophages and acts as a tether for sensing apoptotic cells. In addition to effecting phagocytosis of apoptotic cells, TIM4 has been shown to bind naïve and activated T cells, some of which display low levels of PS for non-apoptotic signaling (57), subsequently inducing the proliferation of the bound T cells (56, 58, 59). Since TIM4 lacks a tyrosine phosphorylation motif in its cytoplasmic domain, TIM4 does not regulate phagocytosis through genetic regulation (60). Instead TIM4 uses integrin proteins to effect phagocytosis by linking the macrophage's membrane to the actin cytoskeleton (61). It was shown that for microglial cells, macrophages that reside in the central nervous system, TIM4 only anchored apoptotic neurons—through its PS link—to the actin cytoskeleton after engulfment had already initiated. The engulfment was instead initiated by another PS-receptor working in tandem with TIM4, BAI1 (62). This demonstrates that PS recognition is a multi-component process that requires the cooperation of several PS-receptors. Indeed, TIM4 and BAI1-bearing macrophages contain many more PS-receptors that participate in phagocytosis as shown in **Figure 2.5**. Rather than every PS-receptor being tuned to recognize a specific PS-exposing membrane, the distinct capabilities of each PS-receptor might instead correspond to a particular task in the downstream process following PS recognition.

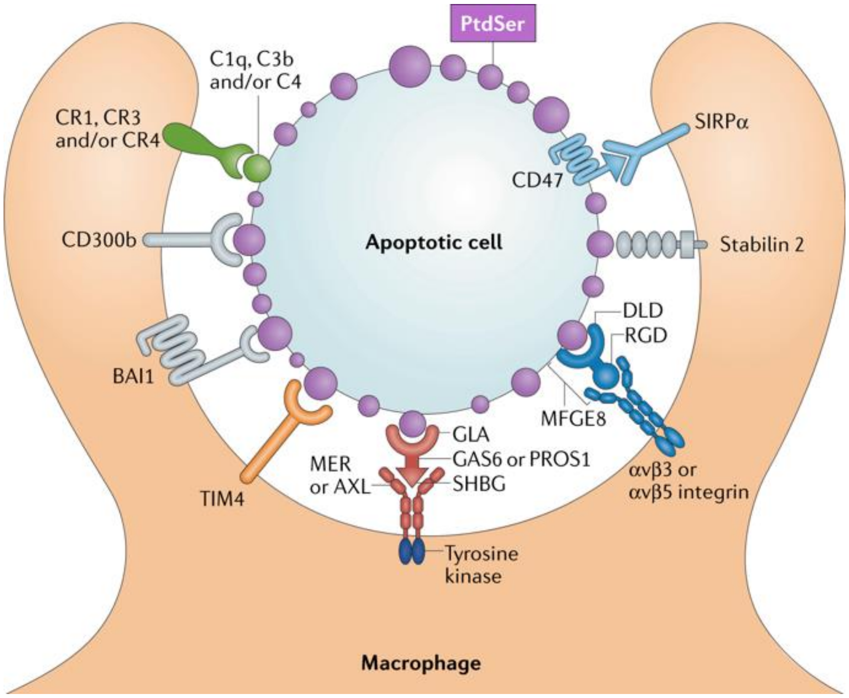


Figure 2.5. TIM4 is one of many PS-receptors in macrophages that mediate phagocytosis. Membrane-bound PS-receptors, such as TIM4, BAI1, Stabilin 2, and CD300b, directly anchor the apoptotic cell to the macrophage. Soluble PS-receptors such as MFGE8, GLA-domain containing proteins, GAS6, and PROS1 bridge apoptotic cells to macrophages through binding surface-bound proteins such as integrins. The individual contributions of each PS-receptor are active areas of research. Reproduced from (63).

In general, the literature for TIM1 and TIM3 does not implicate a direct role for PS binding. While TIM1 and TIM3 have been found on cells that phagocytose, endothelial cells and macrophages, respectively, the cell populations that predominantly express TIM1 or TIM3, T helper cells and B cells, respectively, do not phagocytose. Binding PS is most likely an activation or repression signal for TIM1 and TIM3 expressed on non-phagocytosing cells. TIM1 is enriched in intracellular compartments such as the Golgi complex, early endosome, and lysosomes, and is thought to use PS recognition to traffic itself to the outer leaflet (50, 64). T cells actively scramble PS from the inner to the outer leaflet under a variety of conditions and TIM1 similarly traffics to the outer membrane upon PS exposure (46, 50, 64–66), though the exact mechanism has yet to be characterized. TIM3 is typically studied in context with galectin-9 (41), a suspected protein ligand for TIM3, although no bound structure has been obtained. There are additional suggested protein ligands for TIM3 including CEACAM-1 (67) and HMGB-1 (68). It is unclear how relevant PS binding is in light of all of these other potential interactions for TIM3. However, antibodies that disrupt TIM3's immunological functions *in vitro* in humans and *in vivo* in mice interfere with the binding of PS (and CEACAM-1) but not with its other suggested ligands (69), implying that PS is indeed an important ligand for TIM3.

The diverse physiological roles of the TIM proteins and their PS recognition establish them as a representative subset of the broader class of PS-binding proteins. Furthermore, the structural similarities of the TIM proteins allow for a controlled comparison of the few divergent structural motifs which might confer their differing sensitivities to PS-containing membranes. The combined simplicity and far-reaching physiological consequences of the TIM protein family culminate in a near-ideal model system to study peripheral membrane-binding proteins.

2.5 References

1. Bustin S (2015) Molecular biology of the cell, sixth edition; ISBN: 9780815344643; and molecular biology of the cell, sixth edition, the problems book; ISBN 9780815344537. *ijms* 16(12):28123–28125.
2. Singer SJ, Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. *Science* 175(4023):720–731.
3. Fahy E, et al. (2005) A comprehensive classification system for lipids. *J Lipid Res* 46(5):839–861.
4. Handel A (2017) *Cell Biology by the Numbers* . By Ron Milo and Rob Phillips; illustrated by Nigel Orme. New York: Garland Science (Taylor & Francis Group). $49.95 (paper). xlii + 356 p.; ill.; index. ISBN: 978-0-8153-4537-4. 2016. *Q Rev Biol* 92(4):477–477.
5. Bigay J, Antonny B (2012) Curvature, lipid packing, and electrostatics of membrane organelles: defining cellular territories in determining specificity. *Dev Cell* 23(5):886–895.
6. Harayama T, Riezman H (2018) Understanding the diversity of membrane lipid composition. *Nat Rev Mol Cell Biol* 19(5):281–296.
7. Van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* 9(2):112–124.
8. Dopico AM, Tigyi GJ (2007) A glance at the structural and functional diversity of membrane lipids. *Methods Mol Biol* 400:1–13.
9. Zeniou-Meyer M, et al. (2007) Phospholipase D1 production of phosphatidic acid at the plasma membrane promotes exocytosis of large dense-core granules at a late stage. *J Biol Chem* 282(30):21746–21757.
10. Fadeel B, Xue D (2009) The ins and outs of phospholipid asymmetry in the plasma membrane: roles in health and disease. *Crit Rev Biochem Mol Biol* 44(5):264–277.
11. Zhang Y, Chen X, Gueydan C, Han J (2018) Plasma membrane changes during programmed cell deaths. *Cell Res* 28(1):9–21.
12. Lemmon MA (2008) Membrane recognition by phospholipid-binding domains. *Nat Rev Mol Cell Biol* 9(2):99–111.
13. Stahelin RV, Scott JL, Frick CT (2014) Cellular and molecular interactions of phosphoinositides and peripheral proteins. *Chem Phys Lipids* 182:3–18.
14. Cho W, Stahelin RV (2005) Membrane-protein interactions in cell signaling and membrane trafficking. *Annu Rev Biophys Biomol Struct* 34:119–151.
15. Stillwell W (2016) in *An introduction to biological membranes* (Elsevier), pp 89–110.

16. Stace CL, Ktistakis NT (2006) Phosphatidic acid- and phosphatidylserine-binding proteins. *Biochim Biophys Acta* 1761(8):913–926.

17. Yeung T, et al. (2008) Membrane phosphatidylserine regulates surface charge and protein localization. *Science* 319(5860):210–213.

18. Bevers EM, Williamson PL (2016) Getting to the outer leaflet: physiology of phosphatidylserine exposure at the plasma membrane. *Physiol Rev* 96(2):605–645.

19. Leventis PA, Grinstein S (2010) The distribution and function of phosphatidylserine in cellular membranes. *Annu Rev Biophys* 39:407–427.

20. Birge RB, et al. (2016) Phosphatidylserine is a global immunosuppressive signal in efferocytosis, infectious disease, and cancer. *Cell Death Differ* 23(6):962–978.

21. Connor J, Pak C, SchroitS A Exposure of Phosphatidylserine in the Outer Leaflet of Human Red Blood Cells.

22. Lang KS, et al. (2005) Mechanisms of suicidal erythrocyte death. *Cell Physiol Biochem* 15(5):195–202.

23. Segawa K, Nagata S (2015) An apoptotic “eat me” signal: phosphatidylserine exposure. *Trends Cell Biol* 25(11):639–650.

24. Arandjelovic S, Ravichandran KS (2015) Phagocytosis of apoptotic cells in homeostasis. *Nat Immunol* 16(9):907–917.

25. Hofer AM, Brown EM (2003) Extracellular calcium sensing and signalling. *Nat Rev Mol Cell Biol* 4(7):530–538.

26. Bratton DL, et al. (1997) Appearance of phosphatidylserine on apoptotic cells requires calcium-mediated nonspecific flip-flop and is enhanced by loss of the aminophospholipid translocase. *J Biol Chem* 272(42):26159–26165.

27. Hampton MB, Vanags DM, Pörn-Ares MI, Orrenius S (1996) Involvement of extracellular calcium in phosphatidylserine exposure during apoptosis. *FEBS Lett* 399(3):277–282.

28. Gibbons E, et al. (2013) Molecular details of membrane fluidity changes during apoptosis and relationship to phospholipase A(2) activity. *Biochim Biophys Acta* 1828(2):887–895.

29. Rysavy NM, et al. (2014) Beyond apoptosis: the mechanism and function of phosphatidylserine asymmetry in the membrane of activating mast cells. *Bioarchitecture* 4(4-5):127–137.

30. Appelt U, et al. (2005) Viable, apoptotic and necrotic monocytes expose phosphatidylserine: cooperative binding of the ligand Annexin V to dying but not viable cells and implications for PS-dependent clearance. *Cell Death Differ* 12(2):194–196.

31. Segawa K, Suzuki J, Nagata S (2011) Constitutive exposure of phosphatidylserine on viable cells. *Proc Natl Acad Sci U S A* 108(48):19246–19251.

32. Tait JF, Gibson DF, Smith C (2004) Measurement of the affinity and cooperativity of annexin V-membrane binding under conditions of low membrane occupancy. *Anal Biochem* 329(1):112–119.

33. Mosior M, Newton AC (1998) Mechanism of the apparent cooperativity in the interaction of protein kinase C with phosphatidylserine. *Biochemistry* 37(49):17271–17279.

34. Manna D, et al. (2008) Differential roles of phosphatidylserine, PtdIns(4,5)P2, and PtdIns(3,4,5)P3 in plasma membrane targeting of C2 domains. Molecular dynamics simulation, membrane binding, and cell translocation studies of the PKCalpha C2 domain. *J Biol Chem* 283(38):26047–26058.

35. Sengupta T, Manoj N (2016) Phosphatidylserine and Phosphatidylethanolamine Bind to Protein Z Cooperatively and with Equal Affinity. *PLoS ONE* 11(9):e0161896.

36. Freeman GJ, Casasnovas JM, Umetsu DT, DeKruyff RH (2010) TIM genes: a family of cell surface phosphatidylserine receptors that regulate innate and adaptive immunity. *Immunol Rev* 235(1):172–189.

37. McIntire JJ, et al. (2001) Identification of Tapr (an airway hyperreactivity regulatory locus) and the linked Tim gene family. *Nat Immunol* 2(12):1109–1116.

38. Sabatos CA, et al. (2003) Interaction of Tim-3 and Tim-3 ligand regulates T helper type 1 responses and induction of peripheral tolerance. *Nat Immunol* 4(11):1102–1110.

39. De Souza AJ, Oriss TB, O’malley KJ, Ray A, Kane LP (2005) T cell Ig and mucin 1 (TIM-1) is expressed on in vivo-activated T cells and provides a costimulatory signal for T cell activation. *Proc Natl Acad Sci U S A* 102(47):17113–17118.

40. Umetsu SE, et al. (2005) TIM-1 induces T cell activation and inhibits the development of peripheral tolerance. *Nat Immunol* 6(5):447–454.

41. Zhu C, et al. (2005) The Tim-3 ligand galectin-9 negatively regulates T helper type 1 immunity. *Nat Immunol* 6(12):1245–1252.

42. Monney L, et al. (2002) Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature* 415(6871):536–541.

43. Sánchez-Fueyo A, et al. (2003) Tim-3 inhibits T helper type 1-mediated auto- and alloimmune responses and promotes immunological tolerance. *Nat Immunol* 4(11):1093–1101.

44. Shakhov AN, et al. (2004) SMUCKLER/TIM4 is a distinct member of TIM family expressed by stromal cells of secondary lymphoid tissues and associated with lymphotoxin signaling. *Eur J Immunol* 34(2):494–503.

45. Miyanishi M, et al. (2007) Identification of Tim4 as a phosphatidylserine receptor. *Nature* 450(7168):435–439.

46. Santiago C, et al. (2007) Structures of T cell immunoglobulin mucin protein 4 show a metal-Ion-dependent ligand binding site where phosphatidylserine binds. *Immunity* 27(6):941–951.

47. Wong K, et al. (2010) Phosphatidylserine receptor Tim-4 is essential for the maintenance of the homeostatic state of resident peritoneal macrophages. *Proc Natl Acad Sci U S A* 107(19):8712–8717.

48. DeKruyff RH, et al. (2010) T cell/transmembrane, Ig, and mucin-3 allelic variants differentially recognize phosphatidylserine and mediate phagocytosis of apoptotic cells. *J Immunol* 184(4):1918–1930.

49. Kobayashi N, et al. (2007) TIM-1 and TIM-4 glycoproteins bind phosphatidylserine and mediate uptake of apoptotic cells. *Immunity* 27(6):927–940.

50. Santiago C, et al. (2007) Structures of T Cell immunoglobulin mucin receptors 1 and 2 reveal mechanisms for regulation of immune responses by the TIM receptor family. *Immunity* 26(3):299–310.

51. Clayton KL, et al. (2015) Soluble T cell immunoglobulin mucin domain 3 is shed from CD8+ T cells by the sheddase ADAM10, is increased in plasma during untreated HIV infection, and correlates with HIV disease progression. *J Virol* 89(7):3723–3736.

52. Schweigert O, et al. (2014) Soluble T cell immunoglobulin and mucin domain (TIM)-1 and -4 generated by A Disintegrin And Metalloprotease (ADAM)-10 and -17 bind to phosphatidylserine. *Biochim Biophys Acta* 1843(2):275–287.

53. Geng H, et al. (2006) Soluble form of T cell Ig mucin 3 is an inhibitory molecule in T cell-mediated immune response. *J Immunol* 176(3):1411–1420.

54. Bailly V, et al. (2002) Shedding of kidney injury molecule-1, a putative adhesion protein involved in renal regeneration. *J Biol Chem* 277(42):39739–39748.

55. Kishimoto W, et al. (2016) Expression of Tim-1 in primary CNS lymphoma. *Cancer Med* 5(11):3235–3245.

56. McGrath MM (2018) Diverse roles of TIM4 in immune activation: implications for alloimmunity. *Curr Opin Organ Transplant* 23(1):44–50.

57. Fischer K, et al. (2006) Antigen recognition induces phosphatidylserine exposure on the cell surface of human CD8+ T cells. *Blood* 108(13):4094–4101.

58. Albacker LA, et al. (2010) TIM-4, a receptor for phosphatidylserine, controls adaptive immunity by regulating the removal of antigen-specific T cells. *J Immunol* 185(11):6839–6849.

59. Rodriguez-Manzanet R, et al. (2008) TIM-4 expressed on APCs induces T cell expansion and survival. *J Immunol* 180(7):4706–4713.

60. Park D, Hochreiter-Hufford A, Ravichandran KS (2009) The phosphatidylserine receptor TIM-4 does not mediate direct signaling. *Curr Biol* 19(4):346–351.

61. Flannagan RS, Canton J, Furuya W, Glogauer M, Grinstein S (2014) The phosphatidylserine receptor TIM4 utilizes integrins as coreceptors to effect phagocytosis. *Mol Biol Cell* 25(9):1511–1522.

62. Mazaheri F, et al. (2014) Distinct roles for BAI1 and TIM-4 in the engulfment of dying neurons by microglia. *Nat Commun* 5:4046.

63. Lemke G (2019) How macrophages deal with death. *Nat Rev Immunol*.

64. Balasubramanian S, Kota SK, Kuchroo VK, Humphreys BD, Strom TB (2012) TIM family proteins promote the lysosomal degradation of the nuclear receptor NUR77. *Sci Signal* 5(254):ra90.

65. Angiari S, Constantin G (2014) Regulation of T cell trafficking by the T cell immunoglobulin and mucin domain 1 glycoprotein. *Trends Mol Med* 20(12):675–684.

66. Angiari S, et al. (2014) TIM-1 glycoprotein binds the adhesion receptor P-selectin and mediates T cell trafficking during inflammation and autoimmunity. *Immunity* 40(4):542–553.

67. Huang Y-H, et al. (2015) CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. *Nature* 517(7534):386–390.

68. Chiba S, et al. (2012) Tumor-infiltrating DCs suppress nucleic acid-mediated innate immune responses through interactions between the receptor TIM-3 and the alarmin HMGB1. *Nat Immunol* 13(9):832–842.

69. Sabatos-Peyton CA, et al. (2018) Blockade of Tim-3 binding to phosphatidylserine and CEACAM1 is a shared feature of anti-Tim-3 antibodies that have functional efficacy. *Oncoimmunology* 7(2):e1385690.

CHAPTER 3

METHODOLOGY

3.1 Introduction

A diverse set of experimental and computational techniques were required to correlate structural details of the TIM proteins with their binding of phospholipid membranes. This thesis work relied on three independent approaches that bridged physical chemistry, biochemistry, and computation into a single framework. The membrane-bound orientations of the TIM proteins were obtained from x-ray reflectivity of protein-bound lipid monolayers at an air-water interface. These data were analyzed and corroborated by independent molecular dynamics (MD) simulations that both implicated the membrane contact of specific protein residues and resolved the membrane-bound conformation of the protein. The association of the TIM proteins to PS-containing lipid membranes was characterized by equilibrium binding measurements based on the intrinsic fluorescence of a solvent-exposed tryptophan in each of the TIM proteins. This chapter discusses the foundation and implementation of tryptophan fluorescence, x-ray reflectivity, and MD to the extent required to obtain primary data. The extraction of molecular and thermodynamic parameters is developed in Chapter 4 for x-ray reflectivity and MD, and Chapter 7 for equilibrium binding.

3.2 Tryptophan Fluorescence

3.2.1 Background and mechanism

Tryptophan is one of the twenty naturally occurring amino acids. As one of the most hydrophobic amino acids, it is ubiquitous in protein structures, often acting as part of the hydrophobic core of folded proteins. Alongside phenylalanine and tyrosine, tryptophan is a fluorescent molecule often used to quantify the concentration of protein solutions. Tryptophan has a peak excitation wavelength of 280 nm and an emission spectrum in the range of 300-400 nm (1), which makes UV-vis spectroscopy an ideal measurement for this system. As compared with phenylalanine and tyrosine, tryptophan has a high quantum yield emission and is minimally affected by neighboring peptide bonds (2). Tryptophan has two excitation pathways, one of which has a large permanent dipole sensitive to the electronic environment and the other which is relatively insensitive. (3). For the vast majority of proteins, the former excitation mode is the favored pathway and the other mode is only prominent for certain proteins with excitation peaks near 310 nm (4, 5). Dipole interactions with water, through nonradiative interactions, redshift the emission spectrum of tryptophan and lower its quantum yield (3). These properties confer tryptophan with a high sensitivity to its local environment.

Solvent-exposed tryptophan residues are excellent reporters for the polarity of a given environment and are used to probe the folding or membrane burial of proteins (1). The emission spectrum of tryptophan directly reports on the environment of the protein and can be used to infer if a protein is solvent-exposed. The spectrum is composed of multiple emission peaks, each of which represents an excited vibrational mode of the environment that neutralizes the permanent dipole of the excited tryptophan (3). In a given experiment, the tryptophan residues of a protein are in an ensemble of environments each with different sets of vibrational modes. The resulting spectrum is therefore smooth and broad with no discernible peaks.

Throughout this thesis, spectra will be labeled as “bound” or “unbound” when corresponding to environments that promote complete or no binding. However, a “bound” spectrum corresponds to an ensemble of different membrane-bound environments in which the tryptophan is in many possible orientations. These labels are intended to represent average environments of the membrane-bound or solvent-exposed protein and do not refer to a specific binding mode or conformation of the protein. A key assumption is that the ensemble of “bound” states does not appreciably change with reactant conditions. However, if this assumption is invalid, the analysis can circumvent this difficulty. This will be addressed in the next section.

3.2.2 Experimental Design

Tryptophan fluorescence is measured using a Horiba Fluorolog-3 spectrophotometer with USHIO Xenon short arc lamp and a 1-cm path-length quartz cuvette. The signal is maximized for such an apparatus if binding is measured in bulk and not at a single interface such as that used in ELISA assays. We use large unilamellar vesicles (LUVs), model bilayer systems with an average diameter of ~100 nm, suspended in buffer. We can control the lipid composition of LUVs, allowing us to assess the sensitivity of proteins for various membrane compositions.

The TIM proteins each contain a tryptophan residue which buries upon binding of membranes (6–9). Moreover, this residue is located adjacent to the central PS binding pocket, meaning that an engaged pocket will necessarily place the tryptophan at the membrane interface. When the TIM proteins are free in solution, this tryptophan is exposed to the solvent. Since the hydrophobic core of a lipid bilayer is much more apolar than the solvent, the tryptophan blueshifts and increases its emission upon protein binding. **Figure 3.1** illustrates the experimental design, showing example spectra for membrane-bound and free protein.

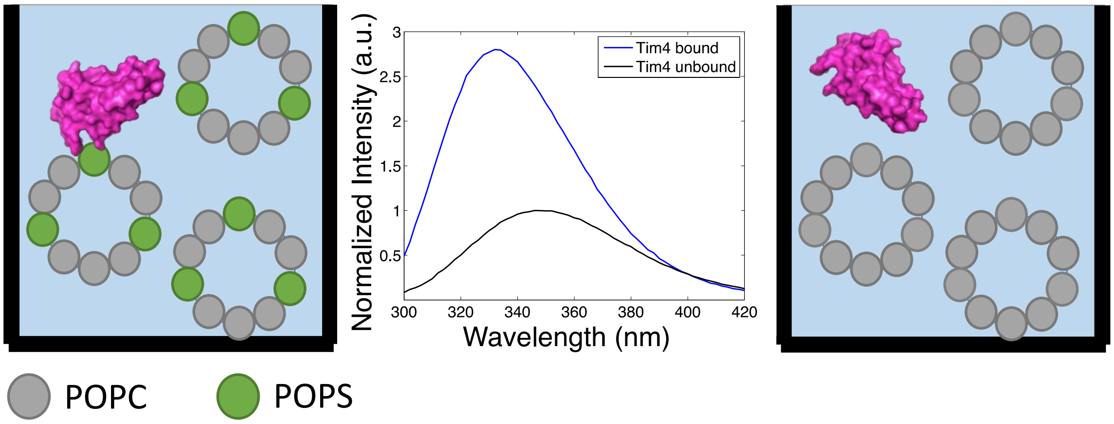


Figure 3.1. Diagram of tryptophan fluorescence corresponding to membrane-bound and free TIM4. The plot in the middle shows the spectra of the membrane-bound (blue) and unbound (black) states normalized to the peak emission of the unbound state. Free in solution or with vesicles containing only POPC (cartoon on the right), TIM4's tryptophan fluoresces with a peak emission at 350 nm. Upon binding vesicles containing POPS (cartoon on the left), the peak blueshifts 18 nm and the intensity increases ~2.8 fold. POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, POPS: 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine, a.u.: Arbitrary units.

While "bound" and "unbound" spectra can be used to qualitatively infer membrane-binding, quantitative estimates of the bound fraction of protein at each reactant concentration are required to determine thermodynamic parameters. Bound fractions are typically inferred from the increase of intensity at a single wavelength (peak emission of the bound spectra for example) upon added reactant (1). However, the measurement at any one wavelength is often noisy and not necessarily directly correlated to the bound fraction. As discussed in the previous section, the shape of the spectrum contains information for the tryptophan's environment, including wavelengths which are not the peak. We therefore analyze the spectra at intermediate binding as a linear combination of the bound and unbound spectra obtained at saturating and zero reactant conditions, respectively (9). This analysis is explicated in the following equation:

$$S(R) = b \cdot S(R_{max}) + (1 - b) \cdot S(0) \tag{3.1}$$

$S(R)$ represents the measured spectra at reactant concentration R, and is generally a vector of intensities at each wavelength. $R_{\max}$ is the reactant concentration of maximal binding with $S(R_{\max})$ being its corresponding spectrum and $S(0)$ being the unbound spectrum. b is the bound fraction of observed protein binding and takes values between zero and one. In general, b relates the degree of binding at all reactant concentrations to that of the final measured point. If $R_{\max}$ is nearly the saturating concentration of reactant, then the spectra at infinite reactant will be negligibly different from that at $R_{\max}$,

$$S(\infty) - S(R_{max}) \approx 0 \tag{3.2}$$

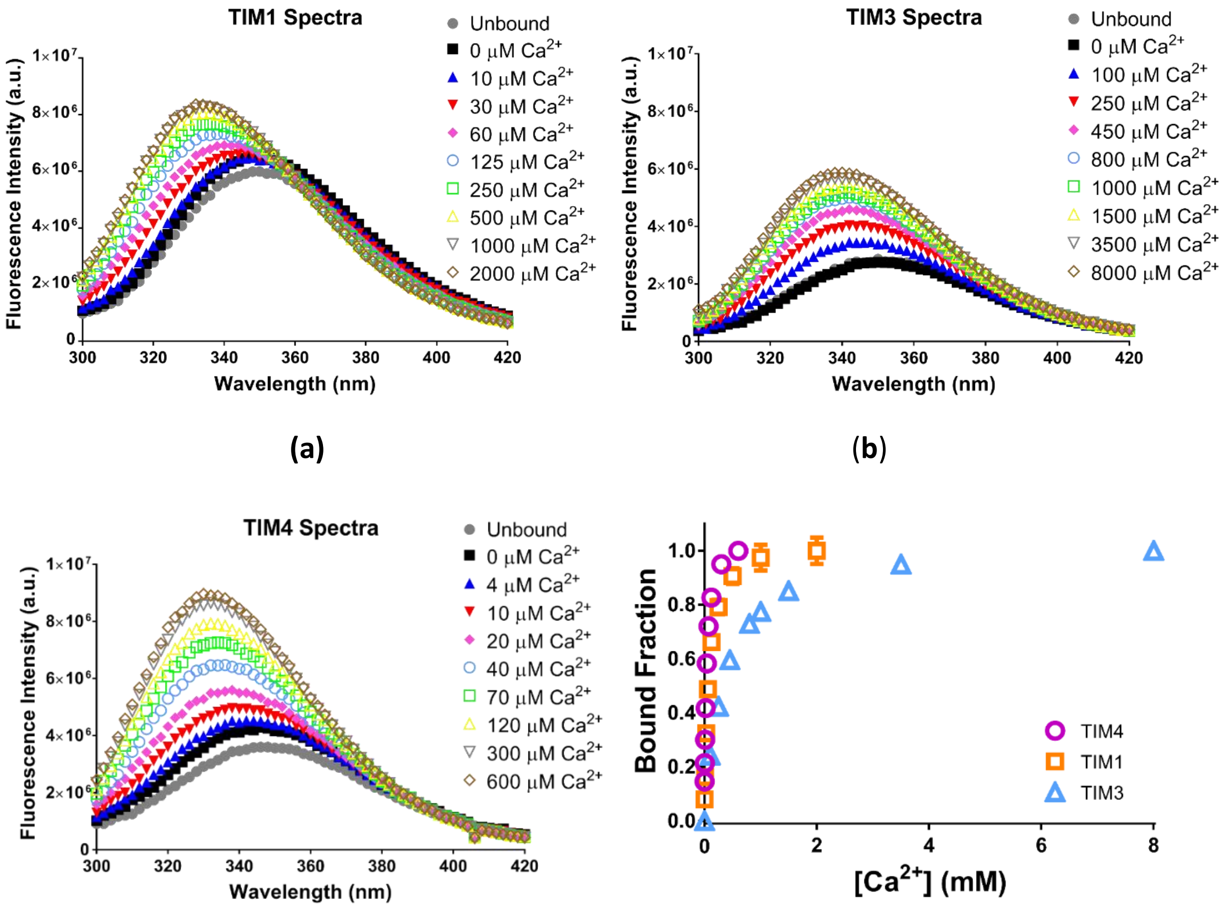
If **Equation 3.2** holds, then b can be interpreted as the bound fraction of total protein. In general **Equation 3.2** does not hold and the bound fraction of total protein at $R_{\max}$ is inferred from analysis using a particular binding model. Unlike the observed protein bound fraction b, the bound fraction of total protein is a model-dependent parameter. Once this value is obtained, the true bound spectrum is predicted from the following equation:

$$S(\infty) = b_{max} \cdot S(R_{max}) - (1 - b_{max}) \cdot S(0) \tag{3.3}$$

Here b_{max} represents the maximal observed protein binding predicted at saturation normalized to the binding at R_{max}. For example, if the binding is 80% saturated at R_{max}, then b_{max} will be equal to 1.25. Thus, b_{max} represents the binding of total protein. **Equation 3.1** can be rewritten as a function of the bound fraction of total protein, $b/\ b_{\mathrm{max}}$, by substituting in **Equation 3.3**:

$$S(R) = \frac{b}{b_{max}} S(\infty) + \left(1 - \frac{b}{b_{max}}\right) S(0) \tag{3.4}$$

Equation 3.4 demonstrates that the bound fraction of observed protein is a scalar multiple of the bound fraction of total protein. Therefore, we can use b in a thermodynamic analysis even without knowing b_{max} provided we treat it as a relative binding fraction. Spectra of the TIM proteins at intermediate reactant concentrations and the inferred bound fractions of observed protein, b, are shown in **Figure 3.2**. From these plots we can compare the rate at which reactant increases binding. However, without a relevant model, we cannot compare the bound fractions of one protein to another; this analysis will be developed in Chapter 7.



(**a**) (**b**)

(**c**) (**d**)

Figure 3.2. Tryptophan spectra of TIM proteins with PS-containing membranes. (**a**) TIM1, (**b**) TIM3, and (**c**) TIM4 tryptophan spectra with 7:3 POPC:POPS LUVs. TIM binding is modulated by the amount of Ca^{2+} in the system in order to achieve intermediate levels of binding. The TIM proteins respond differently to membrane binding; TIM1 and TIM4 blueshift by 18 nm while TIM3 blueshifts by only 10 nm. The ordinate for panels a, b, and c are in the same arbitrary units and thus can be directly compared. (**d**) The inferred bound fractions, *b*, of protein were obtained using the unbound spectrum corresponding to each TIM protein and the bound spectra corresponding to the maximally saturating Ca^{2+} concentrations.

The underlying assumption of this approach is that the protein exists in only two states, a bound ensemble and an unbound ensemble. As stated previously, this assumption is valid if the ensembles of tryptophan environments are insensitive to the reactant concentration. If there are spectra corresponding to distinct bound states, then the same analysis can be applied but with a basis set of three or more distinct spectra, as opposed to the two-state analysis described here. The bound spectrum is highly sensitive to the hydration and lipid packing of a given membrane. Experiments comparing the effect of membrane tension, curvature, or acyl chain unsaturation must be analyzed carefully as not to assume the bound spectra for one condition is equivalent to that of another. The TIM proteins have only one bound spectrum per membrane condition, which is independent of the lipid composition (provided that the acyl chains are the same for all lipids). The largest environmental change for the TIM proteins is likely the exposure to the acyl chains of the lipid and a decreased exposure to water. Reorientation of the tryptophan due to other lipid species is a seemingly minor contribution to the fluorescence.

3.2.3 Materials

All analytical grade reagents and solvents were obtained from Fisher Scientific [Hampton, NH]. All lipids were obtained from Avanti Polar Lipids [Alabaster, AL]. The medium used in most studies was 150 mM NaCl buffered to pH 7.5 with 10 mM HEPES (HBS), also obtained from Fisher Scientific.

3.2.4 Protein Production

The cDNAs of TIM4, TIM1, and TIM3 were a generous gift from Shigekazu Nagata (Osaka University, Osaka). The cloning, expression, and purification were conducted in collaboration with Professor Erin Adams at the University of Chicago. The cDNA corresponding to the murine form of the PS recognition domains (Ig domain) of TIM1 [residues D2- E112; Protein Data Bank code 2OR8 (38)], TIM3 [residues G6-K112; Protein Data Bank code 3KAA (30)], and TIM4 [residues D4-A112; Protein Data Bank code 3BIB (23)] were cloned in frame with a C-terminal 6× His tag into the pAcGP67A vector (BD Biosciences [San Jose, CA]). Recombinant TIM1, TIM3, and TIM4 were all produced in Hi5 cells via baculovirus transfection as described in (10) except that purification was performed in 450 mM NaCl-10 mM Hepes (pH 7.2). Higher NaCl buffer was critical for maintaining protein solubility during the purification process. The purified protein was stable for several months at 4 °C. Separate protein preparations were used for the experiments in this thesis, and consistency among these protein preparations was established with $CaCl_2$ dependence assays using 300 µM vesicles of 70:30 POPC:POPS. The bound and unbound spectra and $[Ca^{2+}]_{½}$ values were consistent among these protein preparations.

3.2.5 Vesicle Preparation

LUVs were produced by extrusion (11). Lipids in the powder form were dissolved in chloroform, mixed at the desired concentration and composition, then dried to form a thin film inside the glass vials under nitrogen flow and constant rotation by hand. The lipids were then further dried under vacuum for 1 h then suspended in HBS and vortexed at a moderate speed for 1 h at 40 °C. The lipid/buffer mixtures were then subjected to five freeze-thaw cycles using a dry ice/ethanol slurry. Finally, the lipids were extruded 17-21 times using 100-nm-diameter filters (Avanti Polar Lipids [Alabaster, AL]). Vesicle diameter and polydispersity were around 130 nm and 10%, respectively, as determined by dynamic light scattering (Zen3600 Malvern Nano Zetasizer). Total lipid concentration was determined by using Molybdenum Blue (12).

3.2.6 Fluorescence Methods and Data Collection

All experiments were conducted using 170 nM protein in HBS with varied $CaCl_2$, and were at 23-25 ˚C. A Horiba Fluorolog-3 spectrophotometer with USHIO Xenon short arc lamp and a 1-cm path-length quartz cuvette were used with excitation at 280 nm and emission collected from 300 to 420 nm. Spectral signals were linear with protein binding. Typically, Ca^{2+} dependence curves were obtained by adding increments of $CaCl_2$ of <0.2% of volume with vigorous stirring. Spectra were collected after two minutes of equilibration until saturation was reached. Results were accepted only if the lamp current was stable within 10^{-3} µA for more than 10 minutes before and throughout the experiment. Experiments were repeated at least three times.

3.3 X-ray Reflectivity

3.3.1 Introduction

Characterization of membrane-bound protein structures presents a formidable challenge since lipids cannot be crystallized as membranes. While lipids cannot form ordered three-dimensional structures, lipids have a preferred orientation in the membrane. For example, at an air-water interface, lipids align with the head groups facing the water and the tail group facing the air. This provides a one-dimensional order that scattering techniques can probe. One such technique, reflectivity, measures the specular reflection of either neutrons or x-rays at an interface. The interfacial structure is non-trivially embedded in this signal and requires a multi-step analysis to extract it. In this section, the physics underlying this technique is discussed, specifically with regards to x-rays. The development here follows the treatment of (13) with added insight from (14–17).

3.3.2 Experimental Design and Geometry

Reflectivity experiments for a lipid monotlayer are typically conducted on a Langmuir trough, such as one depicted in **Figure 3.3**. The trough is filled with a subphase (usually a water-based solvent), and the interface is deposited either at the solid-liquid or air-liquid interface. In x-ray experiments, lipid monolayers are deposited at an air-water interface, as x-rays have low penetration depth. As a model of the lipid bilayer, a monolayer is a valid representation of one of the leaflets of the membrane if the protein of interest does not span more than one leaflet. The lipid monolayer is compressed to bilayer equivalence in packing using a set of interfacial barriers. The surface pressure is monitored by measuring the force exerted by the surface on a Wilhelmy plate. The protein is injected into the subphase below the monolayer. The result is an air-water interface with a monolayer sitting on top and a single layer of bound proteins.

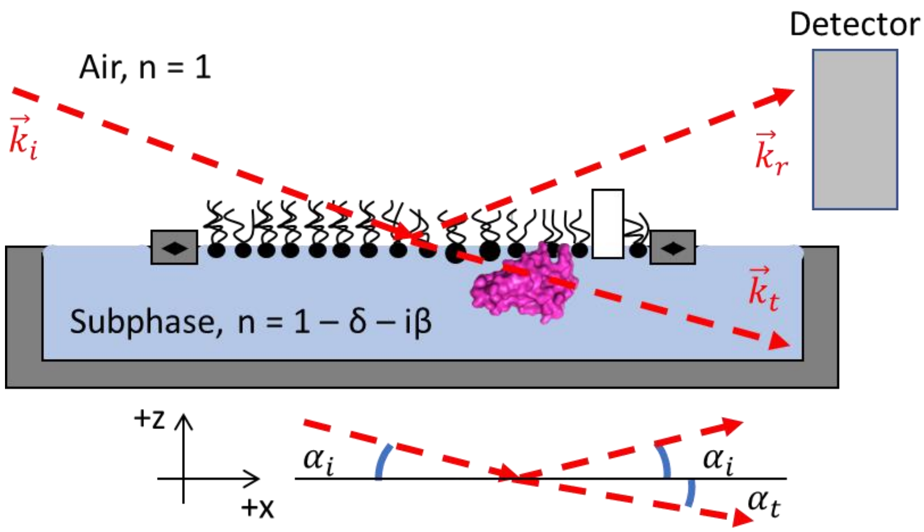


Figure 3.3. Scattering geometry of x-ray reflectivity experiment. Lipids are deposited at the air-water interface and barriers (gray rectangles at the surface) compress the lipids until the desired surface pressure is reached, as measured by a Wilhelmy plate (white rectangle near the rightmost barrier). The z-coordinate is defined along the direction normal to the interface while the x-coordinate is defined in plane. X-rays are incident on an air-water interface at an angle α_i with respect to *x* and reflections at the same angle are measured in a detector.

The reflectivity experiment is analyzed in the following geometry: the air-water interface defines the x-y plane, with the z-coordinate defined as its normal. An x-ray beam approaches the interface with initial wavevector $\vec{k}_i$ at an angle α_i, then is transmitted and reflected with wavevectors $\vec{k}_t$ and $\vec{k}_r$ at angles α_i and α_t, respectively. $z = 0$ is the location of the interface while $z > 0$ corresponds to air with an index of refraction of ~1 and $z < 0$ corresponds to the interfacial structure and subphase, which together have an index of refraction of $n(z) = 1 - \delta(z) - i\beta(z)$. δ represents the dispersion by the medium while β represents the absorption by the medium (in practice, the absorption is often negligible and will be omitted from this analysis). δ is approximately given by the following expression,

$$\delta(z) \cong \frac{2\pi r_e}{\left|\vec{k}_i\right|^2}\langle\rho(z)\rangle \qquad (3.5)$$

Here r_e is the classical electron radius and $\langle\rho(z)\rangle$ is the average electron density at position z. **Equation 3.5** demonstrates that the index of refraction is linearly dependent on the electronic structure of the interface. The amount of light that gets transmitted or reflected at the interface depends on the index of refraction and therefore is determined by the electron density of the interface.

3.3.3 Theoretical Development

When light encounters an interface between two media with different electron densities, a portion of the light will be transmitted while the rest will be reflected. As an example, we consider normal incidence at a single interface between two regions with constant electron densities ρ_i and ρ_f, illustrated in **Figure 3.4**.

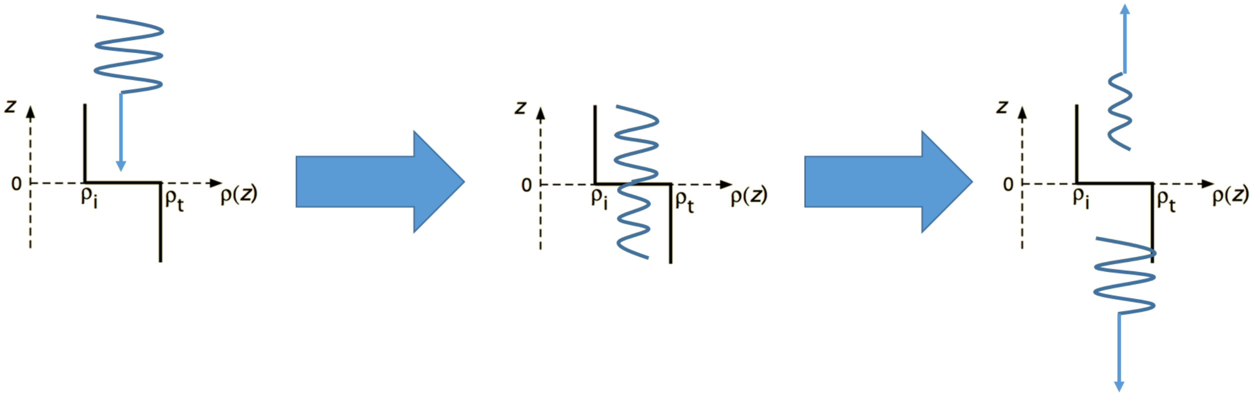


Figure 3.4. Fresnel reflection and transmission at normal incident to a single interface. The left panel depicts an incident electromagnetic wave propagating towards an interface along the normal. The middle panel depicts the formation of the transmitted and reflected waves at the interface with the latter interfering with the incident wave. The right panel shows the resulting transmitted and reflected waves.

Mathematically, this interaction can be represented as a set of wave equations,

$$\tilde{E}(z) = \begin{cases} \tilde{E}_i e^{k_{i,z}z-\omega t} + \tilde{E}_r e^{-k_{i,z}z-\omega t} & z > 0 \\ \tilde{E}_t e^{k_{t,z}z-\omega t} & z \leq 0 \end{cases} \tag{3.5}$$

$\tilde{E}$ represents the total electric field amplitude while the subscripts *i*, *t*, and *r* denote incident, transmitted, and reflected waves respectively. k_z is the *z* component of the wave vector and is inversely proportional to the projection of the wavelength along the *z* coordinate and ω is the angular frequency. Maxwell's equations constrain the wavevectors and amplitudes in this equation, which allows us to calculate the reflection and transmission coefficients, *r* and *t*,

$$r = \frac{\tilde{E}_r}{\tilde{E}_i} = \frac{k_{i,z} - k_{t,z}}{k_{i,z} + k_{t,z}} \tag{3.6}$$

$$t = \frac{\tilde{E}_t}{\tilde{E}_i} = \frac{2k_{i,z}}{k_{i,z} + k_{t,z}} \tag{3.7}$$

Equations 3.6 and **3.7** are Fresnel's equations for the normal component at a single, sharp interface between two regions of constant electron density. Moreover, these equations describe waves with *s*-polarization, i.e. their electric fields are perpendicular to the *x*-*z* plane depicted in **Figure 3.5**. These coefficients, *r* and *t*, are equal to the ratio of the transmitted or reflected wave amplitudes and the incident wave amplitude, respectively. This approach for a single, sharp interface can be generalized to a medium with varying electron density with little difficulty, thanks to the contribution of Parratt (18). **Figure 3.5** demonstrates how reflectivity from an object with varying, average electron density, $\langle\rho(z)\rangle$, can be represented as a series reflection events from multiple interfaces.

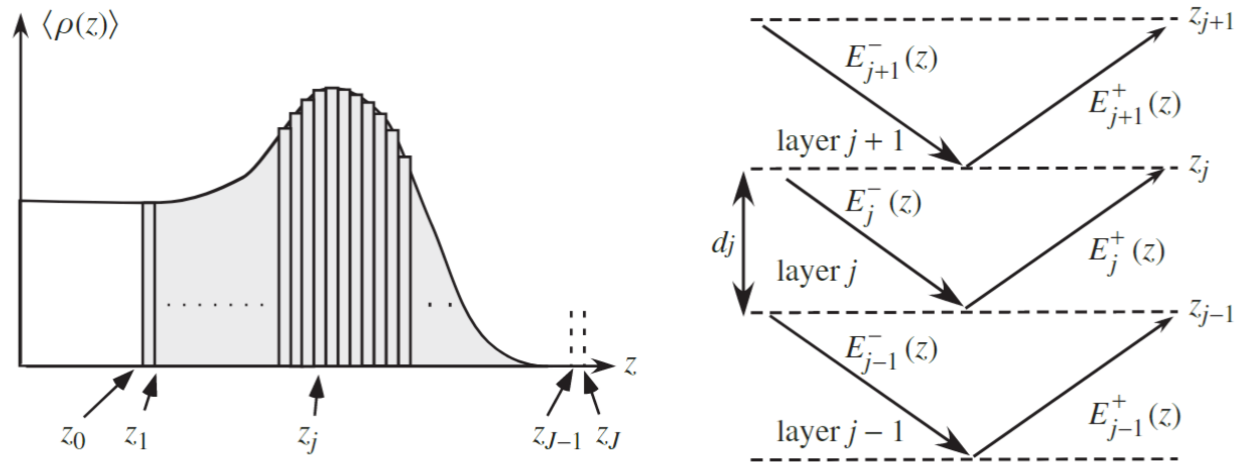


Figure 3.5. The non-constant electron density of a medium represented as a series of interfaces. Each interface acts as an origin for a Fresnel reflection and transmission event. From left to right, the bulk is up until z_0, the interfacial structure until z_J, and air beyond z_J. Figure adapted from (13).

Each interface is described by its own set of Fresnel equations given by,

$$r_{j,j-1} = \frac{k_{j,z} - k_{j-1,z}}{k_{j,z} + k_{j-1,z}} \tag{3.8}$$

$$t_{j,j-1} = \frac{2k_{j,z}}{k_{j,z} + k_{j-1,z}} = 1 + r_{j,j-1} \tag{3.9}$$

These coefficients give the amplitude of the transmitted and reflected waves. Waves that have been transmitted, reflected, or any combination thereof multiple times, are represented as products of these coefficients. A matrix representation of reflectivity can then be formulated as follows: the incident and reflected waves at each interface can be represented as a column vector,

$$\tilde{E}_j(z_j) = \begin{bmatrix} \tilde{E}_{i,j}(z_j) \\ \tilde{E}_{r,j}(z_j) \end{bmatrix} = \begin{bmatrix} Ae^{-ik_{j,z}z} \\ Be^{ik_{j,z}z} \end{bmatrix} \tag{3.10}$$

A and *B* are the amplitudes for the incident and reflected waves respectively. z_j is the coordinate of the j^{th} interface. As these waves propagate between interfaces, the phase of these waves increases. This is represented by a propagation matrix,

$$\tilde{E}_j\left(z_{j-1}\right) = \begin{bmatrix} e^{ik_{j,z}\left(z_j - z_{j-1}\right)} & 0 \\ 0 & e^{-ik_{j,z}\left(z_j - z_{j-1}\right)} \end{bmatrix} \begin{bmatrix} \tilde{E}_{i,j}\left(z_j\right) \\ \tilde{E}_{r,j}\left(z_j\right) \end{bmatrix} = P_j \tilde{E}_j\left(z_j\right) \tag{3.11}$$

This matrix propagates both the incident wave further down and the reflected wave further up along *z*. $\tilde{E}_j\left(z_{j-1}\right)$ represents the incident and reflected waves above the interface at z_{j-1}. The interfacial scattering events can similarly be represented in matrix form to obtain the incident and reflected waves below this interface,

$$\tilde{E}_{j-1}\left(z_{j-1}\right) = \frac{1}{1 + r_{j-1,j}} \begin{bmatrix} 1 & r_{j-1,j} \\ r_{j-1,j} & 1 \end{bmatrix} \begin{bmatrix} \tilde{E}_{i,j}\left(z_{j-1}\right) \\ \tilde{E}_{r,j}\left(z_{j-1}\right) \end{bmatrix} = I_{j-1}^{j} \tilde{E}_j\left(z_{j-1}\right) \tag{3.12}$$

Here $\tilde{E}_{j-1}\left(z_{j-1}\right)$ serves as the transmitted wave, for which the transmission coefficient has been divided to the right-hand side of the equation. Unlike the geometry implied in **Equations 3.8** and **3.9**, the scattering event in **Equation 3.12** is considered to be incident from below the interface and the index j increases with increasing *z*, hence the reflection coefficients are $r_{j-1,j} = -r_{j,j-1}$. The reflectivity from the top-most interface can be expressed as a product of matrices representing propagation and scattering from the lowest interface:

$$\begin{bmatrix} \tilde{E}_{i,0}(0) \\ \tilde{E}_{r,0}(0) \end{bmatrix} = \left(I_0^1 \prod_{j=1}^{J} P_j I_{j-1}^{j} \right) \begin{bmatrix} \tilde{E}_{i,J}\left(z_J\right) \\ \tilde{E}_{r,J}\left(z_J\right) \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \tilde{E}_{i,J}\left(z_J\right) \\ \tilde{E}_{r,J}\left(z_J\right) \end{bmatrix} \tag{3.13}$$

The index *J* corresponds to the top-most interface, meaning $\tilde{E}_{r,J}\left(z_J\right)$ is the reflected wave the experiment measures and $\tilde{E}_{i,J}\left(z_J\right)$ is the incident x-ray beam. Since there is no propagating wave from below the bulk, i.e. $\tilde{E}_{r,0}(0) = 0$, we can read off the bottom row of the matrix to express the reflected wave in terms of the incident wave:

$$M_{21}\tilde{E}_{i,J}\left(z_J\right) + M_{22}\tilde{E}_{r,J}\left(z_J\right) = 0 \tag{3.14}$$

$$R = \left|\frac{\tilde{E}_{r,J}\left(z_J\right)}{\tilde{E}_{i,J}\left(z_J\right)}\right|^2 = \left|\frac{M_{21}}{M_{22}}\right|^2 \tag{3.15}$$

The total reflectivity is the product of multiple transmission and reflection events. If these interfaces are spaced in multiples of the wavelength (as projected along the *z* axis) then the reflections exhibit constructive interference. Varying the incident angle changes the projection of the wave vector onto the membrane normal, effectively changing the wavelength. As the wavelength changes, the spacing of the interfaces changes the constructive and deconstructive components of the interference pattern. In this way, reflectivity encodes the interfacial structure as a set of interfaces specifying their spacing and their change in electron density. However, since we only measure the intensity of the reflected wave, we lose all of the information encoded in the phase. Due to the propagation matrices, the phase contains the coordinates of each of the interfaces. The remaining information encodes the degree of constructive interference as a function of the spacing of interfaces. The next section explains how this information can be used to constrain model structures.

3.3.4 Analysis

Since the reflectivity signal cannot be directly, analytically manipulated to recover the structure, the reflectivity must instead be used as a constraint to distinguish proposed structural models.

This is accomplished by fitting simulated reflectivity experiments to the data and selecting the model which is most consistent with the data. Such an approach requires a direct calculation of the reflectivity given an electron density. **Equations 3.13** and **3.15** provides the algorithm that produces the predicted reflectivity from a given electron density. However, these equations are written in a general form and several approximations must be made in order to tractably relate them to the electron density. The key assumptions are listed as follows: (i) The angle of incident is small, allowing the use of the small angle approximation; (ii) The incoming x-rays are *s*-polarized (though p-polarization is not significantly different for small angles); (iii) The x-ray energy is far above the binding energy of atomically-bound electrons. With these considerations, **Equation 3.8** is re-expressed in terms of experimental parameters:

$$r_{j-1,j} = \frac{\sqrt{{Q_z}^2 - {Q_{c,j-1}}^2} - \sqrt{{Q_z}^2 - {Q_{c,j}}^2}}{\sqrt{{Q_z}^2 - {Q_{c,j-1}}^2} + \sqrt{{Q_z}^2 - {Q_{c,j}}^2}} \tag{3.16}$$

Q_z is the momentum transfer vector and describes the momentum transferred to the system from the reflection. It is related to the wavevector by the following equation:

$$Q_z = k_{r,z} - k_{i,z} = \frac{4\pi}{\lambda} \sin \alpha_i \tag{3.17}$$

$Q_{c,j}$ is the critical momentum transfer vector corresponding to the critical angle beyond which there is total internal reflection. This quantity is directly related to the electron density,

$$Q_{c,j} = \frac{4\pi}{\lambda} \sin \alpha_{c,j} \cong 4\sqrt{\pi r_e \langle \rho(z_j) \rangle} \tag{3.18}$$

By substituting **Equations 3.16** and **3.18** into **3.13** and **3.15**, any model electron density can be used to calculate the reflectivity.

The problem is now reduced to defining the appropriate electron density function for a lipid monolayer. “Slab” models are often used to construct the electron density; they treat the interface as a series of slabs that each have constant electron density. For example, most lipid monolayers are described by a two-slab model, one slab which corresponds to the lipid headgroups and one that corresponds to the lipid tailgroups. These models can include effects due to thermal smearing of the interfaces in the form of a parameter called the roughness, σ. **Equation 3.19** presents a general *N*-slab model,

$$\langle\rho(z)\rangle = \sum_{n=0}^{N} \frac{(\rho_n - \rho_{n-1})}{2} \operatorname{erf}\left(\frac{z_n - z_{n-1}}{\sqrt{2\sigma^2}}\right) \tag{3.19}$$

This equation is a convolution of a series of step functions with a Gaussian, resulting in the error function, *erf*(*x*). This convolution can be interpreted as a Gaussian, thermal smearing of the location of the interface between electron densities ρ_n and ρ_{n-1}. Fitting the simulated reflectivity of **Equation 3.19** yields estimates for its $2N + 1$ parameters and thus the structure of the interface. However, the value of the roughness can be calculated a priori and held constant in the fitting procedure. The modeling of the membrane-bound protein is a type of slab model as well, the details of which are covered in Chapter 4.

3.3.5 Sample Preparation and Procedure

X-ray reflectivity experiments were performed at ChemMatCARS sector 15ID at the Advanced Photon Source of Argonne National Laboratory. Experiments were performed on two custom troughs, one with a volume of 30 mL and another with 70 mL. The former lacked barriers to compress films and thus required deposition of lipids to the final surface pressure of 25 mN/m. The second custom trough had a movable barrier and a control loop was used to maintain the surface pressure of the lipid monolayer in the range of 25-35 mN/m. This surface pressure range was chosen so that the area per lipid corresponded to that of a bilayer, namely ~65 Å^2 per lipid. The buffer contained 10 mM HEPES with 150 mM NaCl and 1-2 mM $CaCl_2$ pH 7.4. One of the two troughs was placed in an airtight helium filled chamber, with oxygen kept below 1% to minimize oxidation. After 30 minutes of relaxation on the 30 mL trough or after compression of the 70 mL trough, reflectivity scans on the lipid-only system were performed for each film prior to injecting protein. Protein was injected into the 70 mL subphase (final concentration ~500 nM) and following the injection, an equal amount of buffer was removed from the subphase to restore the initial liquid level. A stir bar was placed in the trough before lipid deposition to accelerate protein equilibration.

After two hours of equilibration, stirring was stopped and x-ray reflectivity scans were taken within a Q_z range of 0.018-0.55 Å^{-1} or 0.016-0.65 Å^{-1}. The wavelength of the x-ray source was approximately 1.24 Å. Due to the limited range of the detector, the x-ray intensity had to be attenuated to get sufficient statistics for every Q_z value. This was accomplished with a set of titanium absorbers that were used to diminish the beam at low Q_z values. The varying attenuation at different Q_z values introduced discontinuities in the signal which were later patched together by normalization. These patches significantly overlapped in the Q_z values scanned for different attenuation levels as to better assist the normalization procedure. These overlap regions were selected in regions of low curvature of the reflectivity signal. The processing of the raw data and the patching of overlap regions were accomplished using either IDL or python code developed by Mati Meron or Wei Bu of ChemMatCARS. At least two curves were gathered on each film. Phospholipids containing stearoyl tails were used in these x-ray experiments rather than the palmitoyl tails used in binding experiments because longer saturated acyl chains reduce radiation damage.

3.4 Molecular Dynamics Simulations

3.4.1 Introduction

The previous section described how membrane-bound protein structures can be obtained using x-ray reflectivity. However, such models are static averages of the complex structural dynamics occurring at the membrane interface. It is important to validate the reflectivity-derived models and investigate whether the placement of specific residues on the protein near a membrane actually correlates with an interaction. MD simulations fulfill this need as they produce atomic trajectories which evolve deterministically according to the laws of physics. The interactions can be directly inferred from the association of these residues with specific lipids in the membrane.

As will be addressed in Chapter 4, MD simulations are also used to obtain improved protein structures for fitting to x-ray reflectivity data.

In previous work with TIM4, MD was used to validate membrane-bound protein structure derived from x-ray reflectivity (9). The simulation was initiated with TIM4 embedded in the membrane according to the parameters of the x-ray reflectivity-derived structure. This approach confirmed the association of specific residues and showed that the derived membrane-bound state was stably bound, at least as a local minimum. However, the approach did not independently validate the bound orientation as that was provided as an initial condition.

Membrane association with the protein starting in the bulk could, in theory, be simulated, however the lipid reorganization that must occur to facilitate protein insertion can take much longer than computers are able to reasonably simulate. Several methods have been developed to overcome this issue, such as coarse-grained simulations which abandon all-atom representations for simpler models with fewer degrees of freedom. There are several inequivalent coarse-grained approaches that have been applied to peripheral membrane-binding proteins, but all have the drawback that the reduced representations need to be parameterized by all-atom simulations of the system (19). This reduction is therefore not necessarily a reduction in computational resources. We instead opted for a different method that was computationally low-cost but could simulate membrane binding within feasible simulation times.

3.4.2 Highly Mobile Membrane Mimetic Molecular Dynamics

We collaborated with the laboratory of Professor Emad Tajkhorshid (UIUC) and utilized their recently developed all-atom model called Highly Mobile Membrane Mimetic (HMMM) (20). This method accelerates membrane-binding by selectively speeding up the dynamics of the lipid in the membrane and not that of the protein or buffer. The idea behind this approach is that the rate-limiting step is the formation of a space in the membrane which can accommodate the protein. In typical all atom simulations, the protein will collide with the membrane multiple times but not at the right point to facilitate insertion. Most of the simulation is thus spent waiting for a collision in which there is space for the protein to penetrate into the membrane. By speeding up the lipid dynamics, a collision of the protein with the membrane allows the protein to sample far more configurations before "bouncing away." With this increased sampling, the protein is far more likely to insert.

The lipid dynamics are determined largely by the interactions of the tail groups. Through steric clashes of neighboring tails, lipid dynamics are relatively slow. Ample simulation time is required to observe the significant membrane reorganization that accommodates protein binding. The HMMM model therefore shortens the tails of the lipids and instead replaces them with the hydrophobic solvent, 1,1-dichloroethane (dClE). The shortened lipids are constrained to stay in the membrane and maintain their total area. In this model, depicted in **Figure 3.6**, the lipids float freely on the solvent, which captures the chemical properties of the membrane hydrophobic core but without acting as an impedance to lipid mobility. When the protein approaches the membrane, the increased mobility of the lipids allow the protein to better displace the membrane. Provided that the shortened lipids are long enough that the protein does not penetrate into the dClE layer, the obtained, equilibrated bound state serves as an initial condition for a full-tail lipid membrane simulation. This method thus requires no orientation information from x-ray reflectivity to be initiated.

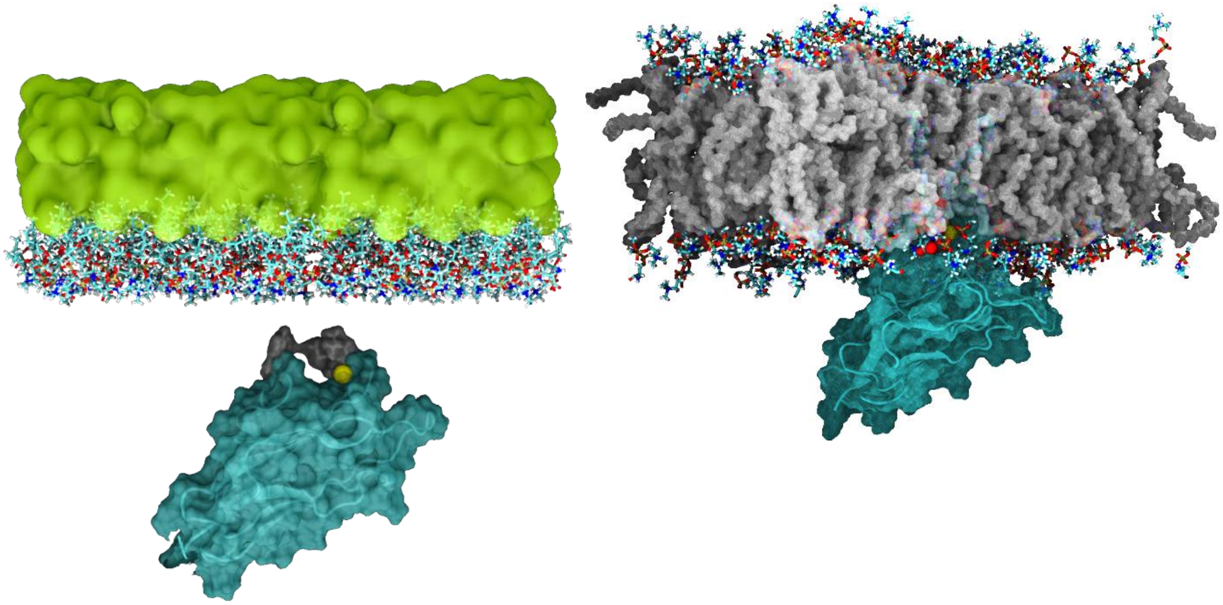


Figure 3.6. HMMM model of TIM3 membrane-binding. Left panel: TIM3 starts in the bulk 10-20 Å from the membrane edge. The lipid tails are replaced by dClE, represented in green. Right panel: After the protein equilibrates with the membrane, the tails are added back in for a standard all-atom MD simulation.

3.4.3 Simulation Protocols for TIM1 and TIM3

Simulations were performed using NAMD2 (21), with the CHARMM27 force field with cMAP corrections for TIM1, the CHARMM36 or CHARMM36M force field for TIM3 (22), and CHARMM36 force-field parameters for lipids (23). The TIP3P model was used for water (24). All the HMMM simulations were performed in an NP_nAT ensemble at 1.0 atm and 310 K, and with a time step of

2 fs. The full-membrane POPC/POPS systems (extended-tail simulations) were simulated as an NPT ensemble with similar parameters, except for the preliminary system derived with CHARMM-GUI, which was simulated in the NP_nAT ensemble with constant area A = 4605 $Å^2$ for TIM1 and A = 6495 $Å^2$ for TIM3 to maintain the area per lipid close to experimental value (∼65 Å/lipid). Constant pressure was maintained using the Nosé-Hoover Langevin piston method (25, 26) and constant temperature was maintained by Langevin dynamics with a damping coefficient of 0.5 ps^{-1} applied to all atoms. Nonbonded interactions were cut off after 12 Å with a smoothing function applied after 10 Å. The particle mesh Ewald method (27) was used for long-range electrostatic calculations with a grid density of >1 $Å^{-3}$.

3.4.4 Docked Membrane Simulations of TIM1 and TIM3

To initially obtain a protein model for fitting to x-ray data, we first performed a short simulation of Tim1 in an aqueous environment with a Ca^{2+} ion positioned into the putative PS pocket based on homology to Tim4. The crystal structures of TIM1 (7) (PDB: 2OR8) and TIM3 (8) were (PDB: 3KAA) obtained from the RCSB PDB (http://www.rcsb.org). The position of the missing Ca^{2+} ion and its coordinating residues in TIM1 were modeled after the ion binding site of TIM4, which shares the same residues in the ion binding site (6). TIM1-Ca^{2+} structure was placed in a box with dimensions of 80 × 80 × 90 $Å^3$ and containing ∼20,000 water molecules, generated with the SOLVATE plugin of VMD (28), and neutralized with 150 mM NaCl ions using the AUTOIONIZE plugin. The system was then energy minimized for 10,000 steps and simulated for 100 ps with the Cα atoms of the protein harmonically restrained (k = 1 kcal mol^{-1} $Å^{-2}$ or 1.8 kT $Å^{-2}$ at 310 K), followed by a 10-ns run with no constraints. After equilibrating the TIM1-Ca^{2+} structure, we next built a PS lipid into the central PS pocket, again based on homology to TIM4.

The CHARMM-GUI Membrane Builder (29) (available at http://www.charmm-gui.org/?doc=input/membrane) was used to construct an atomistic lipid bilayer around the TIM proteins with the best-fit TIM1 orientation and the suspected best-fit TIM3 orientation. The simulation systems included ~140 lipids with the PS residing entirely in the protein-facing leaflet. A 15-20 Å water cushion was included beyond both the proteins and membrane structure to prevent interactions between periodic cells. The system was equilibrated following the default simulation protocol method produced by the CHARMM-GUI Membrane Builder with 1 ns of production performed as the final step of equilibration, with 200 ns of unrestrained simulation.

3.4.5 Preparation of HMMM Membranes

The HMMM membranes for TIM1 simulations were constructed by placing two leaflets of short-tailed lipids at the interface of water and dClE. The lipids used in the membrane patch were constructed starting from POPC and POPS molecules as templates, by shortening lipid tails to only five carbons. To match the experimental lipid concentration, a lipid ratio of 7:3 PC:PS was employed in the simulations. The HMMM membrane was assembled by constructing a dClE box, with dimensions of $80 \times 80 \times 10$ Å^3 containing 740 molecules of the organic solvent, and by placing 160 short lipids on its large faces, 80 lipids on each side. The resulting bilayer was then solvated using the SOLVATE plugin of VMD, yielding a system with ∼36,000 atoms. The solvated membrane mimetic systems were energy minimized for 10,000 steps and simulated for 2 ns without any constraints on the lipids, using an NP_nAT ensemble with constant area, and with a target pressure and temperature of 1.0 atm and 310 K, respectively. A constant area of 6400 $Å^2$ (80×80 $Å^2$) was employed, yielding an area per lipid of $\sim$80 $Å^2$. This increased area per lipid was chosen to facilitate the insertion of the protein in the timescale of the simulations. Upon binding the protein, the area per lipid reduces to ~ 65 $Å^2$. The resulting membrane was employed in the subsequent membrane-binding simulations.

The HMMM membranes for TIM3 were constructed using the HMMM membrane builder in the CHARMM GUI (30). The protein-facing leaflet contained 7:3 POPC:POPS (90 lipids total) and the other leaflet contained only POPC (86 lipids total). All lipids were shortened to 6 (instead of 5 for TIM1) carbons and the lipid area was scaled by 1.1 compared to a membrane of full-length lipids (to facilitate protein insertion). In total, ten unique membranes were generated for each TIM3 simulation condition.

3.4.6 HMMM Membrane-Binding Simulations of TIM1 and TIM3

The resulting structure of TIM1 with a bound Ca^{2+} ion obtained from the simulation in aqueous environment described previously was used as the starting point for the membrane-binding simulations with the HMMM model. The crystal structure of TIM3 was used as its initial structure, with unresolved residues added onto the N-terminal and C-terminal tails. These structures of TIM1 and TIM3 were added to their respective 7:3 PC:PS HMMM membranes. The protein was initially placed above the membrane surface of the cis leaflet, ∼10-15 Å away from the membrane, with two different initial orientations for both TIM1 and TIM3. An additional layer (20 Å for TIM1 and 30 Å for TIM3) of water was added using the SOLVATE plugin to the resulting structures to ensure sufficient hydration. To neutralize the systems, 150 mM NaCl was added using the AUTOIONIZE plugin of VMD. The initial systems consisted of a box with dimensions of $80 \times 80 \times 120$ Å^3 containing ∼69,000 atoms for TIM1 and ~84,000 atoms for TIM3.

The systems were energy minimized for 10,000 steps and simulated for 100 ps with the Cα atoms of the protein harmonically restrained (k = 1 kcal mol^{-1} Å^{-2} or 1.8 kT Å^{-2} at 310 K), followed by a production run of 50 ns for TIM1, and 220 ns for TIM3. Each initial orientation of TIM1 was tested in seven independent simulations, yielding a total of 14 simulations where membrane binding of TIM1 was assessed. Each initial orientation of TIM3 was tested in five independent simulations for a total of ten independent simulations. A harmonic constraint along the z axis, with a force constant k = 0.01 kcal mol^{-1} Å^{-2} (or 0.018 kT Å^{-2} at 310 K), was applied to the carbonyl atoms of the PC and PS lipids to restrain the position of the short-tailed lipids to better mimic the atomic distributions of a full lipid bilayer and to prevent their occasional diffusion into the aqueous phase.

3.4.7 Full-Length Lipid Membrane Simulations of TIM1 and TIM3

To further examine the stability of the resulting membrane-bound TIM1 and TIM3, simulations employing a full POPC/POPS lipid bilayer were also performed. For TIM1, the last frame of the

HMMM systems, in which a PS headgroup was bound to the putative PS binding pocket of TIM1 (four systems), was employed as the starting structures for further simulations. For TIM3, several PS-bound pocket conformations were observed even within single trajectories. We were uncertain if these states were transitory in nature or fluctuations belonging to the ensemble of the equilibrated bound state. These states were categorized by the coordination of specific residues and atoms of the bound PS defined within a 3.5 Å interaction distance. The protein and PS structures that belonged to these states were averaged together. A representative frame that best matched the average was chosen as the initial state of a full-length lipid membrane simulation. As a result, a single HMMM trajectory could produce up to three initial conditions for full-length lipid membrane simulations. The short-tailed lipids of HMMM membranes were grown into full lipids after removing the dClE molecules by adding the missing carbons of the lipid tails while preserving the positions of the lipid atoms already present in the HMMM model (headgroups and initial few carbons of the tails). To further optimize the process, the positions of the newly added atoms in the lipid tails were adopted from the coordinates of randomly selected lipid molecules from a separately equilibrated POPC/POPS membrane. During these steps, the original contacts established between the lipids and proteins during the HMMM membrane-binding simulations were preserved. These systems were then minimized and equilibrated for 100 ps while constraining (k = 1 kcal mol^{-1} Å^{-2} or 1.8 kT Å^{-2} at 310 K) the heavy atoms of the protein and the short-tailed lipids (already present before adding the new atoms) to allow for the newly added atoms to relax. Following this step, TIM1 systems were equilibrated without constraints for 100 ns, with the last 50 ns used for structural analysis. TIM3 systems were equilibrated without constraints for 220 ns and equilibrium was defined by a stabilization of the Euler angles of the protein and the RMSD of the protein and the lipids it contacted.

3.5 References

1. Kraft CA, Garrido JL, Leiva-Vega L, Romero G (2009) Quantitative analysis of protein-lipid interactions using tryptophan fluorescence. *Sci Signal* 2(99):pl4.
2. Cowgill RW (1963) Fluorescence and the structure of proteins. ii. fluorescence of peptides containing tryptophan or tyrosine. *Biochim Biophys Acta* 75:272–273.
3. Callis PR (1997) in *Flourescence Spectroscopy*, Methods in Enzymology. (Elsevier), pp 113–150.
4. Vivian JT, Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys J* 80(5):2093–2109.
5. Broos J, et al. (2007) The emitting state of tryptophan in proteins with highly blue-shifted fluorescence. *Angew Chem Int Ed Engl* 46(27):5137–5139.
6. Santiago C, et al. (2007) Structures of T cell immunoglobulin mucin protein 4 show a metal-Ion-dependent ligand binding site where phosphatidylserine binds. *Immunity* 27(6):941–951.
7. Santiago C, et al. (2007) Structures of T Cell immunoglobulin mucin receptors 1 and 2 reveal mechanisms for regulation of immune responses by the TIM receptor family. *Immunity* 26(3):299–310.
8. DeKruyff RH, et al. (2010) T cell/transmembrane, Ig, and mucin-3 allelic variants differentially recognize phosphatidylserine and mediate phagocytosis of apoptotic cells. *J Immunol* 184(4):1918–1930.
9. Tietjen GT, et al. (2014) Molecular mechanism for differential recognition of membrane phosphatidylserine by the immune regulatory receptor Tim4. *Proc Natl Acad Sci U S A* 111(15):E1463–72.
10. Scharf L, et al. (2010) The 2.5 Å structure of CD1c in complex with a mycobacterial lipid reveals an open groove ideally suited for diverse antigen presentation. *Immunity* 33(6):853–862.
11. MacDonald RC, et al. (1991) Small-volume extrusion apparatus for preparation of large, unilamellar vesicles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1061(2):297–303.
12. Morrison WR (1964) A fast, simple and reliable method for the microdetermination of phosphorus in biological materials. *Anal Biochem* 7(2):218–224.
13. Pershan PS, Schlossman M (2012) in *Liquid Surfaces and Interfaces: Synchrotron X-ray Methods* (Cambridge University Press, Cambridge), pp 88–172.

14. Hamley IW, Pedersen JS (1994) Analysis of neutron and X-ray reflectivity data. I. Theory. *J Appl Crystallogr* 27(1):29–35.

15. Pedersen JS, Hamley IW (1994) Analysis of neutron and X-ray reflectivity data. II. Constrained least-squares methods. *J Appl Crystallogr* 27(1):36–49.

16. Als-Nielsen J, et al. (1994) Principles and applications of grazing incidence X-ray and neutron scattering from ordered molecular monolayers at the air-water interface. *Physics Reports* 246(5):251–313.

17. Attwood D (1999) *Soft X-rays and Extreme Ultraviolet Radiation: Principles and Applications* (Cambridge University Press, Cambridge).

18. Parratt LG (1954) Surface Studies of Solids by Total Reflection of X-Rays. *Phys Rev* 95(2):359–369.

19. Monje-Galvan V, Klauda JB (2016) Peripheral membrane proteins: Tying the knot between experiment and computation. *Biochim Biophys Acta* 1858(7 Pt B):1584–1593.

20. Ohkubo YZ, Pogorelov TV, Arcario MJ, Christensen GA, Tajkhorshid E (2012) Accelerating membrane insertion of peripheral proteins with a novel membrane mimetic model. *Biophys J* 102(9):2130–2139.

21. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.

22. Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–1415.

23. Vanommeslaeghe K, et al. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31(4):671–690.

24. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926.

25. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* 103(11):4613–4621.

26. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* 101(5):4177.

27. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089.

28. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33–8, 27.

29. Wu EL, et al. (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* 35(27):1997–2004.

30. Qi Y, et al. (2015) CHARMM-GUI HMMM Builder for Membrane Simulations with the Highly Mobile Membrane-Mimetic Model. *Biophys J* 109(10):2012–2022.

CHAPTER 4

RESOLUTION OF MEMBRANE-BOUND STRUCTURES THROUGH INTEGRATION OF X-RAY REFLECTIVITY AND MOLECULAR DYNAMICS

4.1 Introduction

“Protein structure determines function,” as a dogma has motivated the search for protein structures for more than half of a century. As a fundamental cornerstone of biochemistry, structural techniques have resulted in ~150,000 solved protein structures on the RCSB Protein Databank at the time of this writing (1). The determination of biologically native structures of proteins is a difficult task. Instead, most protein structures are characterized in non-functional contexts. Indeed, ~134,000 of the solved protein structures in the protein databank were determined using x-ray crystallography, which requires the proteins to be in highly packed conditions at a far higher density than in most cellular conditions.

Despite these limitations, crystallization can still provide valuable information, especially for proteins that bind small ligands. The interaction surface is typically a small part of the protein complex that can be isolated from neighboring proteins in the lattice. The structure of the binding site is less likely to be influenced by the high packing conditions. Crystallography provides an unparalleled amount of molecular detail underlying the coordination of protein and their ligands. If the ligand of a protein is a lipid membrane, then co-crystallization is not an option. Co-crystallization of soluble, short chained lipids or membrane-mimetics are possible, however, structural techniques which, at best, only yield the protein structure complexed with a single lipid cannot identify the residues responsible for conferring sensitivity to the membrane context.

Techniques such as lipid cubic phase crystallography and cryo-electron microscopy (cryo-EM) of 2D crystals, Nuclear Magnetic Resonance (NMR), and Electron Paramagnetic Resonance (EPR) have been utilized to determine bound-structures of peripheral membrane-binding proteins to short-chained lipids or mimetic bilayers such as bicelles, micelles, or lipid nanodiscs (2–6). These mimetic membranes can significantly differ from bilayer or monolayer membranes and may induce significantly different structures of bound proteins. Fourier-Transform Infrared Spectroscopy (FTIR) approaches have been used with monolayer and bilayer membranes to obtain membrane-bound protein structures (7) and while these approaches sometimes identify the membrane burial of specific residues, they do not easily yield the overall surface burial of the protein (2, 8). The insertion of the protein is a highly informative structural parameter that correlates with the interaction strength between the protein and the membrane; the stronger the interaction, the more energy that is available for the protein to displace lipids in the membrane. Structural techniques which do not probe this parameter provide limited information for protein-membrane interactions.

Reflectivity techniques that utilize x-rays or neutrons have been used to probe the orientation of proteins bound to lipid monolayers (9–19). Reflections are scattering events caused by changes in electron density, or scattering length density, normal to an interface. Reflectivity is often used to characterize systems with homogeneous or periodic in-plane structure. However, globular proteins do not form a homogeneous, adsorbed layer; instead, they patchily cover the lipid monolayer and are randomly oriented in-plane. The one-dimensional electron density corresponding to such a system is an average over regions of the monolayer with randomly oriented proteins and regions lacking proteins altogether. Additionally, reflectivity-derived protein structures cannot be directly "solved" but must be inferred from theoretical models of the interfacial structure. These issues present a significant barrier in the interpretation of such techniques applied to protein-membrane systems.

These difficulties have motivated the development of an x-ray reflectivity fitting model that assumes a rigid configuration of the bound protein. The model requires an initial protein structure obtained from another structural technique, such as crystallography or NMR, to represent the bound configuration. This approach was successfully applied to PKCα (9, 10, 20) and then later to TIM4 (21), resolving membrane-bound structures for both proteins. However, if the protein structure is unknown, reflectivity can only estimate the average length of the adsorbed protein layer normal to the membrane. This information can be used to distinguish proposed binding models on the basis of only the length of the bound state. If the membrane-bound state consists of multiple binding configurations, this approach is highly limited in its ability to extract orientation parameters, but it is possible with structural models that capture these dynamics.

While the reflectivity model based on the crystal structure of TIM4 performed well in elucidating the bound orientation of the protein at the membrane surface, several difficulties arose when applying the same method to TIM1 and TIM3. If the crystal structure is not representative of the membrane-bound structure, this technique can result in poorly or incorrectly resolved membrane-bound orientations. Employing simulations to resolve more relevant, functionally contextual structures could potentially mitigate this issue. Typically, studies seek correspondence of membrane-bound structures yielded by MD with structures independently obtained from reflectivity or other structural techniques such as NMR (22–26). We find that the greatest synergy of MD and reflectivity is achieved when the MD structure serves as the model for analysis of x-ray reflectivity. This approach has been developed for neutron reflectivity (27) but the modeling is significantly different from that of x-ray reflectivity.

In this chapter, the details of the integration of x-ray reflectivity and MD are directly explicated. The analysis of x-ray reflectivity data is divided into two parts: Section 4.2 provides an overview of the fitting technique used in previous publications (9, 10, 21) and introduces refinements to the model. Improved statistical and numerical methods for the implementation of the model are described, resulting in an efficient algorithm to compute electron density profiles from three-dimensional protein structures. Section 4.3 describes the integration of MD with x-ray reflectivity. Several candidate approaches are compared, with one arising as the clear preferred method. Section 4.4 provides a meditation on the nature of this model and the information it contains.

4.2 A Model for the Electron Density of a Membrane-Protein Complex

4.2.1 The Slab Model for Lipid Monolayers

As described in Chapter 3, the analysis of reflectivity data reduces down to the modeling of the interfacial electron density. The electron density of lipid monolayers is well-characterized by slab models,

$$\langle \rho(z) \rangle = \sum_{n=1}^{N} \frac{(\rho_n - \rho_{n-1})}{2} \operatorname{erf}\left(\frac{z - z_{n-1}}{\sqrt{2\sigma^2}}\right) \tag{4.1}$$

Here, ρ_n is the electron density of the n$^{\text{th}}$ slab, σ is the interfacial roughness of the membrane due to thermodynamics, and z_n is the position of the n$^{\text{th}}$ interface between slabs. The coordinate *z* is aligned such that *z* = 0 is the air interface while *z* < 0 descends into the interfacial structure and subsequently the subphase. For lipid monolayers, this model posits only two slabs for the membrane, with two additional, constrained slabs for the air above and the buffer below. **Equation 4.1** is thus parameterized for lipid monolayers as,

$$\begin{aligned}\langle \rho(z) \rangle = {} & \frac{(\rho_{air} + \rho_{buff})}{2} + \frac{(\rho_{air} - \rho_{tail})}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2\sigma^2}}\right) + \frac{(\rho_{tail} - \rho_{head})}{2} \operatorname{erf}\left(\frac{z + L_{tail}}{\sqrt{2\sigma^2}}\right) \\ & + \frac{(\rho_{head} - \rho_{buff})}{2} \operatorname{erf}\left(\frac{z + L_{tail} + L_{head}}{\sqrt{2\sigma^2}}\right)\end{aligned} \tag{4.2}$$

The electron densities and interface locations have been labeled by “air,” “buff,” “head,” and “tail,” reflecting the structural features the slabs are intended to capture. The interface locations are expressed as a sum of lengths, L, of the head and tail group regions of the lipid monolayer. If the interface between the tail groups and air is given by z = 0, the interface between the head groups and the buffer is the sum of the lengths of the tail and head groups, $L_{tail} + L_{head}$. These parameters are illustrated in **Figure 4.1**.

When protein is bound to the lipid monolayer, the protein electron density is thermally smeared by the capillary waves of the membrane as well. It is therefore convenient to model the electron density of the combined protein-membrane system before modeling the contribution of the capillary waves. We return our attention to the lipid monolayer alone and separate the effect of the capillary waves. This separation is expressed in **Equation 4.2** as a convolution of a series of hard interfaces with a Gaussian, given by **Equations 4.3** and **4.4**,

$$\langle \rho(z) \rangle = \int_{-\infty}^{\infty} \langle \rho(z') \rangle_L \cdot \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left(\frac{-(z-z')^2}{2\sigma^2}\right) dz' \tag{4.3}$$

$$\begin{aligned}\langle \rho(z) \rangle_L = {} & \rho_{air}[1 - \mathrm{H}(z)] + \rho_{tail}[\mathrm{H}(z) - \mathrm{H}(z + L_{tail})] \\ & + \rho_{head}[\mathrm{H}(z + L_{tail}) - \mathrm{H}(z + L_{tail} + L_{head})] + \rho_{buff}\mathrm{H}(z + L_{tail} + L_{head})\end{aligned} \tag{4.4}$$

Equation 4.4 represents the hard interfaces between slabs as the Heaviside step function, $\mathrm{H}(z)$, defined by **Equation 4.5**,

$$\mathrm{H}(z) = \begin{cases} 0 & z < 0 \\ 1 & z \geq 0 \end{cases} \tag{4.5}$$

Equations 4.3-4.5 have intuitive interpretations: The electron density of the head and tail groups of a lipid monolayer are assumed to be constant, illustrated by the orange curve in **Figure 4.1**. Since a reflectivity measurement samples a large area of the lipid monolayer for a long time, it also samples the intrinsic, thermodynamic fluctuations of the monolayer. These fluctuations are approximately Gaussian distributed, with a roughness parameter determined by the probing wavelength of x-ray and the chemical identity of the interface (28). Therefore, the intrinsic electron density of the membrane is successfully separated from its dynamics.

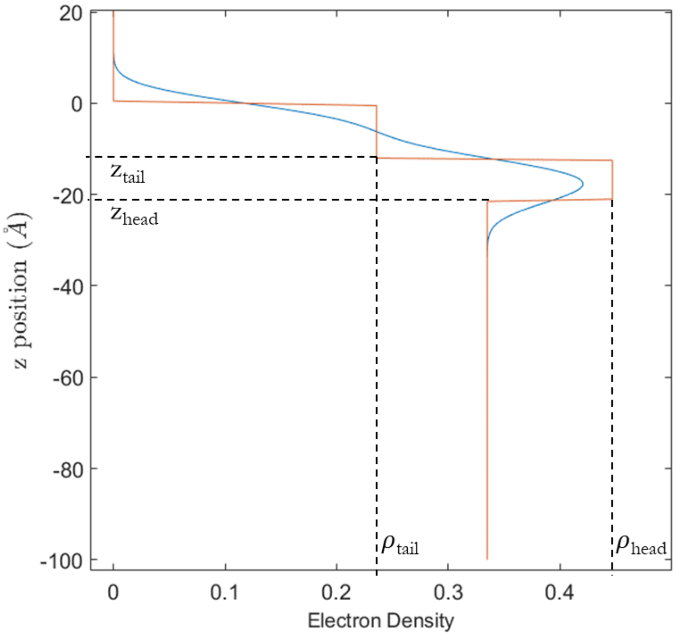


Figure 4.1. The Two-Slab Model for the Electron Density of a Lipid Monolayer. The orange curve represents the "intrinsic" electron density of the lipid monolayer with no thermal contributions. The head and tail groups have constant electron densities ρ_{head} and ρ_{tail} with lengths L_{head} and L_{tail}, respectively. The blue curve is the electron density of the monolayer with thermodynamic smearing described by the roughness, σ. The blue curve can be thought of as a sum of Gaussians with standard deviation, σ, centered at every point of z, with each Gaussian's amplitude given by the corresponding value in the orange curve.

4.2.2 Model of the Electron Density of a Protein-Monolayer System

While the slab model for lipid monolayers is a tractable model with only five parameters, a similar model for a protein would require many more parameters. As a consequence, the modeled electron density might not uniquely implicate a protein structure. Instead, we use a combined modeling approach in which only the electron density of the lipid monolayer is parameterized as a slab model while that of the protein is directly calculated from a crystal structure. The method used in this thesis is a modification of that used in references (10, 21). The treatment here expands on these references by grounding the model in a more mathematical formulation.

We are interested in determining the total electron density of the protein-membrane system. We choose to represent the system as a combination of protein-containing and membrane-only components of the monolayer,

$$\langle \rho(z) \rangle_I = C\langle \rho(z) \rangle_{P+L} + (1 - C)\langle \rho(z) \rangle_L \tag{4.6}$$

The subscript “*P* + *L*” denotes the average electron density of the protein-bound lipid regions of the monolayer, the subscript “*L*” denotes electron density of the lipid-only regions (given by **Equation 4.4**), and the subscript “*I*” reminds us that **Equation 4.6** describes sharp interfaces that are unconvoluted with thermodynamic smearing. The coverage parameter, *C*, denotes the fractional area of the membrane bound by the protein and is a free parameter in fitting reflectivity data. Therefore, the goal is to determine an expression for the complete electron density, particularly that of the protein, to evaluate the following (**Equation 4.3** rewritten to include protein),

$$\langle \rho(z) \rangle = \int_{-\infty}^{\infty} \langle \rho(z') \rangle_I \cdot \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left(\frac{-(z - z')^2}{2\sigma^2}\right) dz' \tag{4.7}$$

We start by analyzing the electron density of the protein alone. This is easily obtained from the atomic structure of the protein. Resolved structures of proteins are freely available from the protein data bank (1) in an accessible format which lists the three-dimensional coordinates of all atoms. We therefore use these coordinates for the *N* atoms, and represent them as the set $\{\vec{a}_i\}_{\, i \in \{1,\ldots,N\}}$. These atomic coordinates are rotated and translated during the fitting procedure to determine the best-fit orientation and membrane insertion depth. The parameters describing these transformations are implemented as a change of coordinates in a matrix formalism:

$$\vec{a}_i{}' = R_{\theta,\phi}\vec{a}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \vec{a}_i \tag{4.8}$$

$$\vec{a}_i{}'' = \vec{a}_i{}' + \left(\left[d_p - L_{head} - L_{tail}\right] - \left[\max_i\left(a'_{i,z} + r_i\right)\right]\right)\hat{z} \tag{4.9}$$

$R_{\theta,\phi}$ is the rotation matrix for two Euler angles and d_p is the insertion depth of the protein. These formulae implicitly assume that the protein's coordinates are expressed in the center of mass frame. The insertion of the protein is defined such that the edge of the topmost atom is placed at the coordinate $\left[d_p - L_{head} - L_{tail}\right]$. For example, $d_p > L_{head}$ implies that the top of the protein penetrates into the tail group slab. These parameters are illustrated in **Figure 4.2**. **Equations 4.8** and **4.9** combine to yield **Equation 4.10** for the fully transformed protein coordinates,

$$\vec{a}_i{}'' = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \cos\theta\sin\phi & \cos\theta\cos\phi & -\sin\theta \\ \sin\theta\sin\phi & \sin\theta\cos\phi & \cos\theta \end{bmatrix} \vec{a}_i$$

$$+ \begin{bmatrix} 0 \\ 0 \\ \left[d_p - L_{head} - L_{tail}\right] - \max_i\left(\sin\theta\sin\phi\, a_{i,x} + \sin\theta\cos\phi\, a_{i,y} + \cos\theta\, a_{i,z} + r_i\right) \end{bmatrix} \quad (4.10)$$

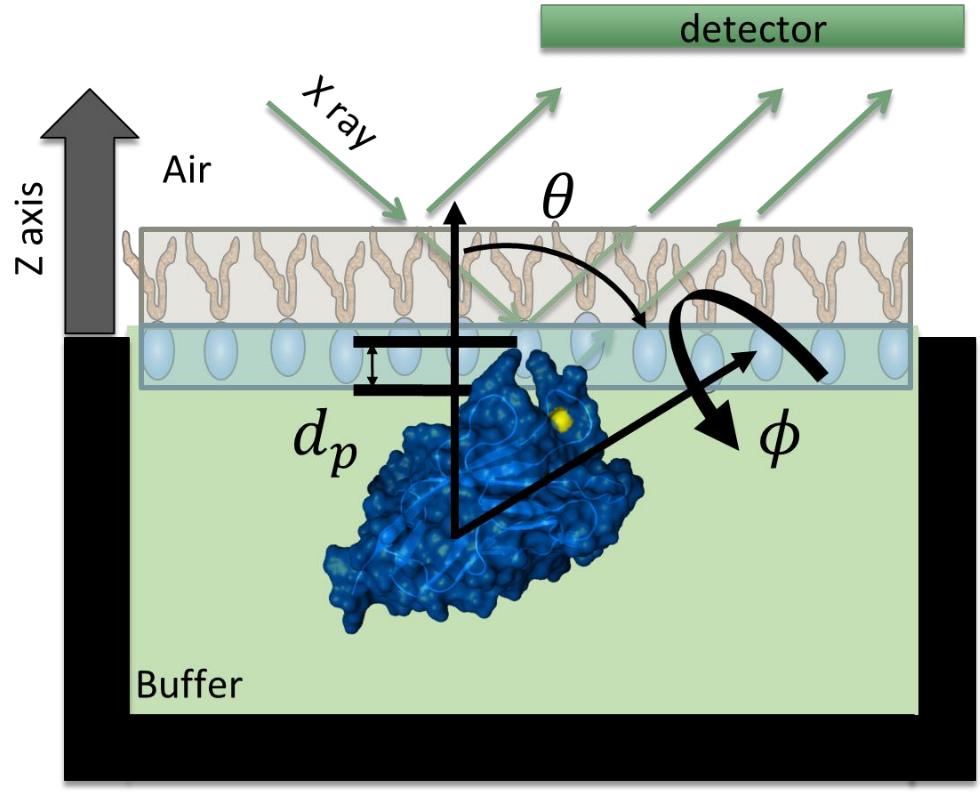


Figure 4.2. Model of Membrane-Bound Protein for the Calculation of Electron Density. The protein is rotated about its own axis by an angle ϕ, then rotated with respect to the membrane normal by an angle, θ. The protein is then inserted into the membrane such that the topmost atom is inserted an amount d_p past the head-buffer interface.

There are many possible choices to represent the electron density of the protein, the simplest of which is to assume that each atom's electrons are uniformly distributed within in a van der Waal's sphere with radii, $\{r_i\}_{i\in\{1,\ldots,N\}}$. We also have a corresponding set of number of electrons per atom, $\{n_i\}_{i\in\{1,\ldots,N\}}$, which we use to calculate the electron density,

$$\rho_i(\vec{r}) = \frac{n_i}{\frac{4}{3}\pi {r_i}^3}[1 - \mathrm{H}(|\vec{r} - \vec{a}_i''|^2 - {r_i}^2)] \tag{4.11}$$

The Heaviside step function ensures that inside the sphere, the electron density is a constant and outside the sphere, the electron density is zero. Alternatively, the electron density can be represented with a Gaussian:

$$\rho_i(\vec{r}) = \frac{n_i}{(2\pi)^{3/2}{r_i}^3}\exp\left(\frac{|\vec{r} - \vec{a}_i''|^2}{2{r_i}^2}\right) \tag{4.12}$$

In our experience, **Equations 4.11** and **4.12** are negligibly different for the fitting of reflectivity data. We thus proceed using the simpler **Equation 4.11**; although, the following steps are just as applicable to **Equation 4.12**. **Equation 4.11** gives the value of the electron density at a single point, $\vec{r}$; however, we require the electron density of the plane of *x*-*y* points located at *z*. By integrating **Equation 4.11** over a set area, and dividing by that set area, we obtain the average planar electron density of a slice located at *z*,

$$\langle\rho_i(z)\rangle = \frac{3n_i}{4\pi {r_i}^3}\left[1 - \mathrm{H}\left(\left|z - a''_{i,z}\right|^2 - {r_i}^2\right)\right] \cdot \left[\frac{\pi\left({r_i}^2 - \left(z - a''_{i,z}\right)^2\right)}{\pi {r_i}^2}\right] \tag{4.13}$$

Here, $a_{i,z}$ is the z-component of $\vec{a}_i$ also equal to $\vec{a}_i \cdot \hat{z}$. The second bracketed term is the area of a circular slice of a sphere at a position, *z*, divided by the projected area of the whole sphere. **Equation 4.13** is written more succinctly as,

$$\langle\rho_i(z)\rangle = \frac{3n_i\left[{r_i}^2 - \left(z - a''_{i,z}\right)^2\right]}{4\pi {r_i}^5}\left[1 - \mathrm{H}\left(\left|z - a''_{i,z}\right|^2 - {r_i}^2\right)\right] \tag{4.14}$$

Note that **Equation 4.14** only applies to each atom individually. However, we are interested in the electron density of the whole protein, in which multiple atoms might overlap. Summing **Equation 4.14** would double-count the regions with overlapping atoms. We therefore perform this integration of $\rho_i(\vec{r})$ only after summing over all the atoms:

$$\langle\rho(z)\rangle_P = \frac{\iint_S \sum_i^N \rho_i(\vec{r})\, ds}{\iint_S ds} \tag{4.15}$$

Here, *S* is an arbitrary area that encapsulates the entire protein and the index "*P*" denotes the average electron density of the protein. In the works cited (10, 21), *S* is chosen as,

$$S_{rect} = \left(\max_i\left(a''_{i,x} + r_i\right) - \min_i\left(a''_{i,x} - r_i\right)\right) \times \left(\max_i\left(a''_{i,y} + r_i\right) - \min_i\left(a''_{i,y} - r_i\right)\right) \tag{4.16}$$

That is, the area is a rectangle enclosing the protein with sides given by the longest *x* and *y* projected lengths of the protein. This rectangle gives a clear interpretation to **Equations 4.6** and **4.7**: The coverage parameter represents the fraction of rectangles with area *S* under the x-ray beam footprint which contain protein. However, the coverage parameter then does not directly relate to the number of proteins, but rather abstractly to the number of protein-containing rectangles comprising the lipid monolayer. These rectangles cannot be rotated or otherwise manipulated to facilitate closer protein packing as shown in **Figure 4.3**. If an orientation of the protein is oblong, this rectangle may prohibit protein coverages that require rectangles to overlap even if the contained proteins themselves do not overlap.

The first refinement introduced in this thesis work was to redefine the area. With a more encapsulating area, the coverage parameter has a clearer link to the number of surface-adsorbed proteins. Additionally, the model has a greater flexibility in dealing with cases of large protein coverage. This modification is expressed more clearly in terms of indicator functions, fundamental functions in real analysis and probability theory for defining areas and volumes,

$$\mathbf{1}_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases} \tag{4.17}$$

Here, *A* is an arbitrary set though typically a subset of the real numbers. The indicator function returns the value 1 only if the argument *x* belongs to the set *A*. As an aside, the Heaviside step function, $\mathrm{H}(x)$, is an indicator function in which *A* is defined as the set of numbers greater than or equal to 0. We can then define the area *S* as the following integral,

$$S = \iint_S ds = \iint_{-\infty}^{\infty} \mathbf{1}_A(x, y)\, dx dy \tag{4.18}$$

The set *A* therefore defines encapsulating area of the protein. The indicator function allows us to define this area in terms of set theoretic operations, giving us significantly more expressive power to define an area than the dimensions of a polygon. We choose the minimal possible area which encapsulates the protein, namely the projection of the protein onto the *x*-*y* plane:

$$A = \bigcup_{i=1}^{N} \left\{ \left| \vec{r} \cdot \hat{x} - a''_{i,x} \right|^2 + \left| \vec{r} \cdot \hat{y} - a''_{i,y} \right|^2 \leq {r_i}^2 \right\} \tag{4.19}$$

Equation 4.19 is the set of all the *x*-*y* points that belong to the projected *x*-*y* areas of every atom in the protein. The set union operation ensures that no coordinate is double counted in the area, avoiding the problem of overlapping atoms. **Figure 4.3** illustrates the enclosed area described by **Equations 4.18** and **4.19** in comparison with the approach described by **Equation 4.16**.

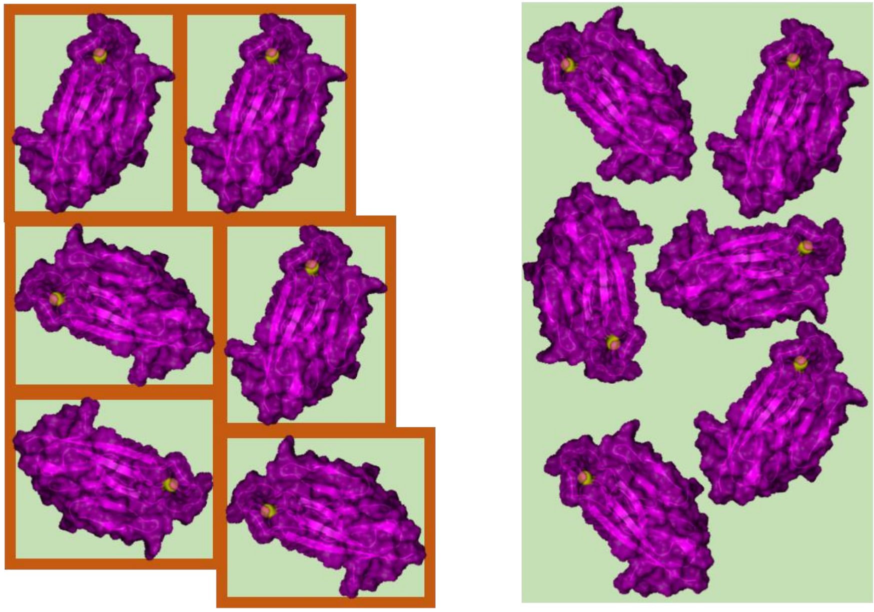


Figure 4.3. Comparison of Protein Area Calculation Methods. Left panel: The area enclosing the protein is calculated as a rectangle according to **Equation 4.16**. The dimensions of the box are given by the longest *x* and *y* dimensions of the projected protein area, the latter of which is highlighted in magenta. As a result of this rectangular structure, the membrane can only be modeled as a series of boxes. Any protein packing that requires an intersection of boxes will result in a coverage parameter, *C*, greater than one. Right panel: The area enclosing the protein is given by the projected area of the protein itself, illustrated in magenta. **Equations 4.18** and **4.19** provide the necessary instructions for computing this shape. As a result, a membrane can be modeled as any arbitrary, in-plane tessellation of the protein-containing and membrane-only regions. The protein imaged here is TIM4.

Taking **Equations 4.18** and **4.19** to define the area *S*, the integrals in **Equation 4.15** are well defined and written as,

$$\langle\rho(z)\rangle_P = \frac{\iint_{-\infty}^{\infty} \sum_i^N \rho_i(\vec{r})\, \mathbf{1}_A(x,y)dxdy}{\iint_{-\infty}^{\infty} \mathbf{1}_A(x,y)dxdy} \tag{4.20}$$

This equation has no explicit dependence on the coordinate *z* in its current form. In order to evaluate this integral to obtain an explicit function of *z*, we rewrite **Equation 4.11** using indicator functions,

$$\rho_i(\vec{r}) = \frac{3n_i}{4\pi {r_i}^3}\mathbf{1}_{B_i(z)}(x,y) \tag{4.21}$$

$$B_i(z) = \left\{\left(\left|\vec{r}\cdot\hat{x} - a''_{i,x}\right|^2 + \left|\vec{r}\cdot\hat{y} - a''_{i,y}\right|^2 \leq \left({r_i}^2 - \left[z - a''_{i,z}\right]^2\right)\right)\right.$$

$$\left.\wedge\left({r_i}^2 \geq \left|z - a''_{i,z}\right|^2\right)\right\} \tag{4.22}$$

The set, $B_i(z)$, represents the set of coordinates occupied by atom *i* that also intersect the *x*-*y* plane at coordinate *z*. Note that the following condition holds,

$$\forall z\ (B_i(z) \subseteq A) \tag{4.23}$$

That is, for every slice at *z*, the area occupied by an atom at that slice, $B_i(z)$, is a subset of *A*. This is essentially a tautology since *A* was defined as the projected area of the protein (i.e. every atom) to the *x*-*y* plane,

$$A = \bigcup_z \left\{\bigcup_{i=1}^{N} B_i(z)\right\} \tag{4.24}$$

Therefore, the product of indicator functions is,

$$\mathbf{1}_A(x,y)\mathbf{1}_{B_i(z)}(x,y) = \mathbf{1}_{B_i(z)}(x,y) \tag{4.25}$$

And finally, we can write **Equation 4.20** as,

$$\langle\rho(z)\rangle_P = \frac{\iint_{-\infty}^{\infty} \sum_i^N \frac{3n_i}{4\pi {r_i}^3} \mathbf{1}_{B_i(z)}(x,y)\, dxdy}{\iint_{-\infty}^{\infty} \mathbf{1}_A(x,y) dxdy} \tag{4.26}$$

While this equation entirely accounts for the protein density, there are values of *x*, *y*, and *z* where the indicator function returns 0. For those coordinates (provided they do not reside in the protein interior), the electron density should not be zero, but rather should be that of the buffer or lipid membrane. This is easily addressed by assigning the area of a given slice that is occupied by the protein with the electron density of the two-slab model. Therefore, $\langle\rho(z)\rangle_{P+L}$ is given by the sum of these areas of lipid electron density and **Equation 4.26**. We can now express **Equation 4.6** in terms of the lipid and protein electron density,

$$\langle\rho(z)\rangle_I = C\langle\rho(z)\rangle_P + \left(1 - C\frac{\iint_{-\infty}^{\infty} \mathbf{1}_{\bigcup_{i=1}^N B_i(z)}(x,y) dxdy}{\iint_{-\infty}^{\infty} \mathbf{1}_A(x,y) dxdy}\right)\langle\rho(z)\rangle_L \tag{4.27}$$

Here, $\mathbf{1}_{\bigcup_{i=1}^N B_i(z)}$ is the indicator function for the set that contains the intersection of all atoms with the slice at coordinate *z*. If this area is exactly equal to the projected area of the protein, then there is no area to fill in with lipid or buffer electron density and the parenthetical terms is given by $1 - C$. Alternatively, if the given slice does not intersect with any atom, then the parenthetical term is equal to 1 and the entire area is assigned the electron density of the two-slab model. In the previous works that utilized this approach (10, 21), the correction only applied to atoms of the protein in the buffer. In the membrane, the protein electron density was instead added on top of the lipid electron density. The equation describing this approach (which also used a rectangular area) is given by,

$$\langle\rho(z)\rangle_S = C\langle\rho(z)\rangle_P + \Bigg(1 - C + C\cdot\mathrm{H}(z + L_{head} + L_{tail})\langle\rho(z)\rangle_L$$

$$C\cdot[1 - \mathrm{H}(z + L_{head} + L_{tail})]\frac{\iint_{S_{rect}}\left[S_{rect} - \mathbf{1}_{\bigcup_{i=1}^{N} B_i(z)}(x,y)\right]dxdy}{S_{rect}}\Bigg)\langle\rho(z)\rangle_L \qquad (4.28)$$

The step functions ensure that the correction is zero above the head group-buffer interface. We suspected that the use of **Equation 4.28** biases the estimate for d_p to be more withdrawn. If the protein were placed in a substantially inserted position, the lipid would have to be locally displaced near the protein. It is unclear what the local electron density of the monolayer would actually be in the presence of the protein, but nonetheless the lipid density in this region would have to be lower. **Equation 4.27** implicitly models this "replacement" and avoids the complications introduced by "addition." The difference of these two approaches is illustrated in **Figure 4.4**.

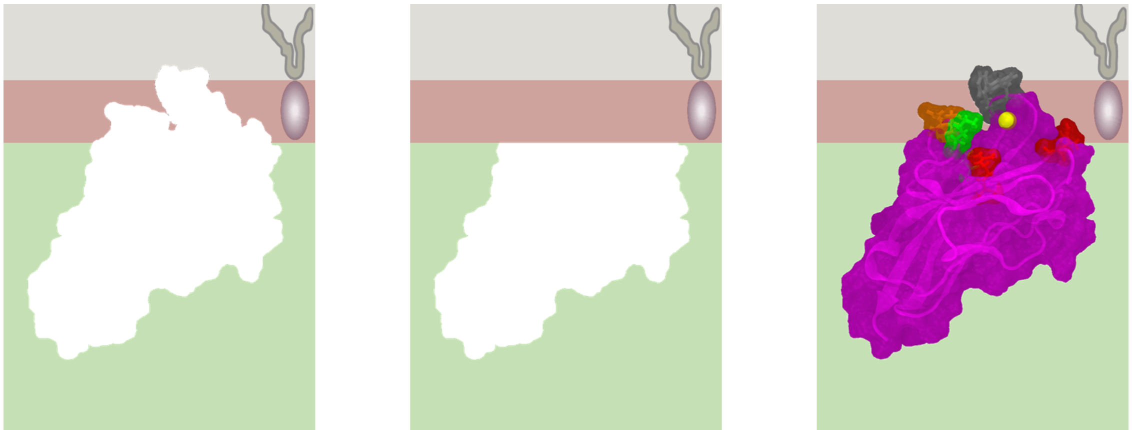


Figure 4.4. Corrections to the Empty Volumes in the Protein Electron Density. Left panel: The electron density "replacement" method described by **Equation 4.27**. The protein electron density replaces the buffer and lipid electron densities. Center panel: The electron density "addition" method described by **Equation 4.28**. The protein electron density replaces the buffer but not the lipid electron density. Right panel: The combined protein-membrane electron density obtained by adding the protein electron density to the left or center panels. The protein visualized here is TIM4.

Combining **Equations 4.26** and **4.27**, we obtain the final equation for the combined system's electron density,

$$\langle\rho(z)\rangle_I = C\left(\frac{\iint_{-\infty}^{\infty}\sum_i^N \frac{3n_i}{4\pi {r_i}^3}\mathbf{1}_{B_i(z)}(x,y)\,dxdy}{\iint_{-\infty}^{\infty}\mathbf{1}_A(x,y)dxdy}\right)$$

$$+\left(1 - C\frac{\iint_{-\infty}^{\infty}\mathbf{1}_{\cup_{i=1}^N B_i(z)}(x,y)dxdy}{\iint_{-\infty}^{\infty}\mathbf{1}_A(x,y)dxdy}\right)\langle\rho(z)\rangle_L \tag{4.29}$$

The dependence on the protein orientation parameters are implicit in the sets A and $B(z)$, making **Equation 4.29** highly non-analytic. This equation must instead be calculated numerically. The choice to represent the electron density in terms of indicator functions allows us to easily identify a procedure for calculating these integrals. The *x*, *y*, and *z* coordinates are discretized into a grid with dimensions Δx, Δy, and Δz. The indicator functions act as Boolean operators which determine, which atoms contribute electron density to a given grid space and how much. The integrals are thus evaluated on a lattice, transforming them into approximating sums,

$$S = \iint_{-\infty}^{\infty}\mathbf{1}_A(x,y)dxdy \rightarrow S = \sum_y\sum_x \mathbf{1}_A(x,y)\Delta x\Delta y \tag{4.30}$$

The space is cut up into a series of grids each with volume, $\Delta x\Delta y\Delta z$. The set A is similarly discretized as,

$$A = \bigcup_{i=1}^{N}\left\{\left|\left(\vec{r}\cdot\hat{x} \pm \frac{\Delta x}{2}\right) - a''_{i,x}\right|^2 + \left|\left(\vec{r}\cdot\hat{y} \pm \frac{\Delta y}{2}\right) - a''_{i,y}\right|^2 \le {r_i}^2\right\} \tag{4.31}$$

Equation 4.31 states that an atom is considered to be in a grid space with dimensions Δx, Δy if at least one corner of that grid space intersects the van der Waal's sphere of the atom. **Figure 4.5** illustrates the action of the discretized indicator functions described by **Equation 4.31**. Similarly, the set $B_i(z)$ is discretized as,

$$\begin{aligned} B_i(z_j) = \Bigg(\left|\left(\vec{r}_j \cdot \hat{x} \pm \frac{\Delta x}{2}\right) - a''_{i,x}\right|^2 + \left|\left(\vec{r}_j \cdot \hat{y} \pm \frac{\Delta y}{2}\right) - a''_{i,y}\right|^2 \\ \le \left(r_i^2 - \left[\left(z_j \pm \frac{\Delta z}{2}\right) - a''_{i,z}\right]^2\right)\Bigg) \wedge \left(r_i^2 \ge \left|\left(z_j \pm \frac{\Delta z}{2}\right) - a''_{i,z}\right|^2\right) \end{aligned} \tag{4.32}$$

The coordinate *z* is discretized as a set of slices, *z_j*, which mark the centers of thick sliced regions with constant electron density. $B_i(z_j)$ represents a thick slice at z_j with its top interface given by $z_j + \frac{\Delta z}{2}$ and its bottom interface, $z_j - \frac{\Delta z}{2}$. An atom is considered to be in this slice if it penetrates either one or both of the interfaces. In practice, we have found that $\Delta x = \Delta y = \Delta z = 0.5$ Å is computationally efficient for the size of the TIM proteins and any mesh more finely divided negligibly changes the fit result. The final computational scheme is given by the discretization of **Equation 4.29**,

$$\begin{aligned} \langle \rho(z_j) \rangle_S = C \left(\frac{\sum_y \sum_x \sum_i^N \frac{3n_i}{4\pi r_i^3} \mathbf{1}_{B_i(z_j)}(x,y)\, \Delta x \Delta y}{\sum_y \sum_x \mathbf{1}_A(x,y) \Delta x \Delta y} \right) \\ + \left(1 - C \frac{\sum_y \sum_x \mathbf{1}_{\cup_{i=1}^N B_i(z_j)}(x,y) \Delta x \Delta y}{\sum_y \sum_x \mathbf{1}_A(x,y) \Delta x \Delta y} \right) \langle \rho(z_j) \rangle_L \end{aligned} \tag{4.33}$$

The convolution integral, **Equation 4.7**, is evaluated analytically. Since the argument of **Equation 4.33** is a discrete variable z_j, the equation can be made a function of a continuous argument if it is instead written as a sum of Heaviside step functions,

$$\langle \rho(z)\rangle_S = \sum_j \Bigg\{ \Bigg(\frac{C}{S}\sum_y\sum_x\sum_i^N \frac{3n_i}{4\pi r_i{}^3}\mathbf{1}_{B_i(z_j)}(x,y)\Delta x\Delta y$$

$$+\left[1-\frac{C}{S}\sum_y\sum_x \mathbf{1}_{\cup_{i=1}^N B_i(z_j)}(x,y)\Delta x\Delta y\right]\langle\rho(z_j)\rangle_L\Bigg)\left[\mathrm{H}(z-z_j)-\mathrm{H}(z-z_{j+1})\right]\Bigg\} \tag{4.34}$$

Thus **Equation 4.**34 is convoluted with the Gaussian in **Equation 4.**7,

$$\langle \rho(z)\rangle = \int_{\min_j(z_j)-7\sigma}^{\max_j(z_j)+7\sigma} \langle\rho(z')\rangle_S \cdot \left(\sqrt{2\pi}\sigma\right)^{-1}\exp\left(\frac{-(z-z')^2}{2\sigma^2}\right)dz' \tag{4.35}$$

We calculated this convolution up to 7σ away from both the topmost and bottommost interfaces. This value was chosen as the reflectivity needs to be computed over a region with sufficient padding of constant electron density in order to avoid boundary effects and this value was more than large enough while still being computationally inexpensive. The evaluation of **Equation 4.35** results in a sum of error functions akin to **Equation 4.2**,

$$\langle \rho(z)\rangle = \frac{1}{2}\sum_j \operatorname{erf}\left(\frac{z-z_{j-1}}{\sqrt{2\sigma^2}}\right)\Bigg\{\frac{C}{S}\left[\sum_y\sum_x\sum_i^N \frac{3n_i}{4\pi r_i{}^3}\left(\mathbf{1}_{B_i(z_j)}(x,y)-\mathbf{1}_{B_i(z_{j-1})}(x,y)\right)\Delta x\Delta y\right]$$

$$+\left(\langle\rho(z_j)\rangle_L-\langle\rho(z_{j-1})\rangle_L\right)$$

$$-\frac{C}{S}\left[\sum_y\sum_x\left(\mathbf{1}_{\cup_{i=1}^N B_i(z_j)}(x,y)\langle\rho(z_j)\rangle_L-\mathbf{1}_{\cup_{i=1}^N B_i(z_{j-1})}(x,y)\langle\rho(z_{j-1})\rangle_L\right)\Delta x\Delta y\right]\Bigg\} \tag{4.36}$$

Equation 4.36 serves as a complete algorithm for calculating the protein-membrane electron density. The continuous argument z is evaluated on a mesh which we typically set to 0.25 Å as smaller values do not result in sufficiently different reflectivity curves (not to be confused with the mesh that was used to evaluate the set of electron density interfaces of the protein). **Equation 4.10** is used to generate the protein orientation while **Equations 4.31**, and **4.32** are used to determine the criteria for including a grid space in the volume of an atom. Therefore, the indicator functions conveniently, directly translate to Boolean statements in any coding language.

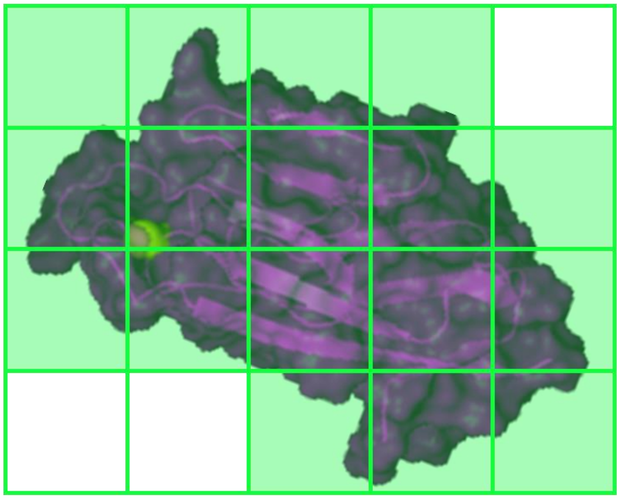


Figure 4.5. Illustration of the Discrete Calculation of Electron Density of TIM4. The electron density is calculated on a grid, with wide spacing to clarify the action of the algorithm. The indicator functions serve as Boolean operators that determine if specific grid spaces are occupied by the protein. If any part of the protein is contained in a grid space, then the indicator function assigns the grid space the electron density of the protein. The indicator function assigns empty grid spaces the electron density of either the lipid or buffer. These assignments are represented by the coloring of the grid spaces where if the indicator function returns “1,” the space is green, and if the indicator function returns “0,” the space is white.

4.2.3 Statistical Evaluation of Best-Fit Parameters

Equations 4.10, **4.31**, **4.32**, and **4.36** contain a total of eight free parameters; the protein parameters θ, ϕ, d_p, and *C*, and the lipid parameters, ρ_{head}, ρ_{tail}, L_{head}, and L_{tail}. The remaining parameters, ρ_{air}, ρ_{buff}, and σ are calculated *a priori* and the values (for SOPC/SOPS films on water or HBS buffer), are 0, 0.334 Å^{-3} and 3.4 Å, respectively (10, 21). **Equation 4.36** is then used to calculate a simulated reflectivity curve (see Chapter 3), which is fit to the data via non-linear least squares. Eight is a significant number of parameters in a fitting procedure, considering that the number of individual points in a reflectivity curve can vary from 60 to 100, and some of which potentially contain redundant information. While the best-fit values of the parameters are straightforward to obtain with a variety of non-linear least squares methods, the evaluation of the error, and therefore the confidence, of these obtained values is non-trivial and model-specific. Adding to the complexity, this model is a highly non-linear function in terms of these parameters, presenting a large obstacle to calculating the error in any theoretical way. Therefore, the statistical evaluation of this model requires special attention.

We start with a brief overview of the statistical problem. Recall that the goodness of fit, χ^2, is given by,

$$\chi^2 = \sum_{m=1}^{M} \frac{\left(R(Q_{z,m}) - R_T(Q_{z,m}; \theta, \phi, d_p, C, \rho_{head}, \rho_{tail}, L_{head}, L_{tail})\right)^2}{{s_m}^2} \tag{4.37}$$

where *R* is the value of the reflectivity at the measured momentum transfer value $Q_{z,m}$, s_m is the measurement error and R_T is the theoretical reflectivity curve calculated using the electron density of **Equation 4.36**. The best-fit values for the parameters are obtained by taking a gradient of **Equation 4.37** and solving for the values in which this gradient is zero. Since the underlying model is non-analytic, this gradient is also non-analytic, and thus its calculation is numerically approximated. There are many choices available for calculating this gradient in an optimization procedure, we use the trust-region-reflective approach implemented in MATLAB (29, 30). Regardless of the chosen method, the error of the parameter values is often outputted alongside the best-fit parameters and are often published as the error. However, this output does not reflect the confidence of the best-fit values. The intended purpose of these outputted errors is to provide an estimation for a more likely best-fit value than the currently determined one if the optimization procedure were to be continued. These methods estimate the error based on local, analytic approximations to the function, typically second-order Taylor expansions. These approximations are only valid in a small neighborhood of parameter values near the best-fit values. The quadratic Taylor expansion diverges from the original function even at relatively small distances from the expansion point, meaning the estimated error is unlikely to be correct. Taken together, these errors should not be reported as uncertainties.

In the previous work that utilized the protein-membrane model (10, 21), the errors were calculated using an approach based on the analysis presented in Bevington (31). In this approach, the obstacles described above are circumvented by manually comparing the χ^2 value at various points in the (eight dimensional) parameter space. This procedure is onerous for all eight parameters, but feasible for two or three of the parameters. Since the protein orientation parameters are the most important, we scan through values of the two Euler angles, θ and ϕ, and perform non-linear least-squares on the remaining six parameters. The result is a table of $\chi^2(\theta,\phi)$ values. By the definition of χ^2, the likelihood of the value of the parameters is given by:

$$L \propto \exp\left(\frac{-\chi^2(\theta,\phi)}{2}\right) \tag{4.38}$$

An increase in the value of χ^2 of one from its minimum value corresponds to a standard deviation of the likelihood function. The set of θ and ϕ values that resulted in such an increase was used to estimate the standard deviation of the parameters for PKC-α (10). However, the likelihood function gives the probability of a particular set of values of the parameters for a specific measurement outcome. A single measurement of reflectivity could be randomly skewed by experimental error, misrepresenting the “true” values of the parameters. Therefore, the error that we desire is the error of the “true” set of values. This is instead determined by the formula (32),

$$\chi^2(\theta,\phi) \le \min_{\theta,\phi} \chi^2(\theta,\phi) + \Delta\chi^2{}_{\upsilon=2,p} \tag{4.39}$$

$\Delta\chi^2{}_{\upsilon=2,p}$ is a value determined at a desired level of confidence (the *p*-value) by the χ^2 distribution with $\upsilon = 2$ degrees of freedom. The degrees of freedom are determined by the variation, $\Delta\chi^2$, from the best-fit value, induced by varying the values of θ and ϕ. Since two parameters generate this variation, $\Delta\chi^2$, its distribution has two degrees of freedom. The set of values for θ and ϕ which satisfy **Equation 4.39** are within 1 – *p* confidence of being the true parameter value. A similar approach was employed for TIM4 (21). An illustration of this analysis is depicted in **Figure 4.6** as a χ^2-map, in which values of θ and ϕ are color-coded by their confidence level.

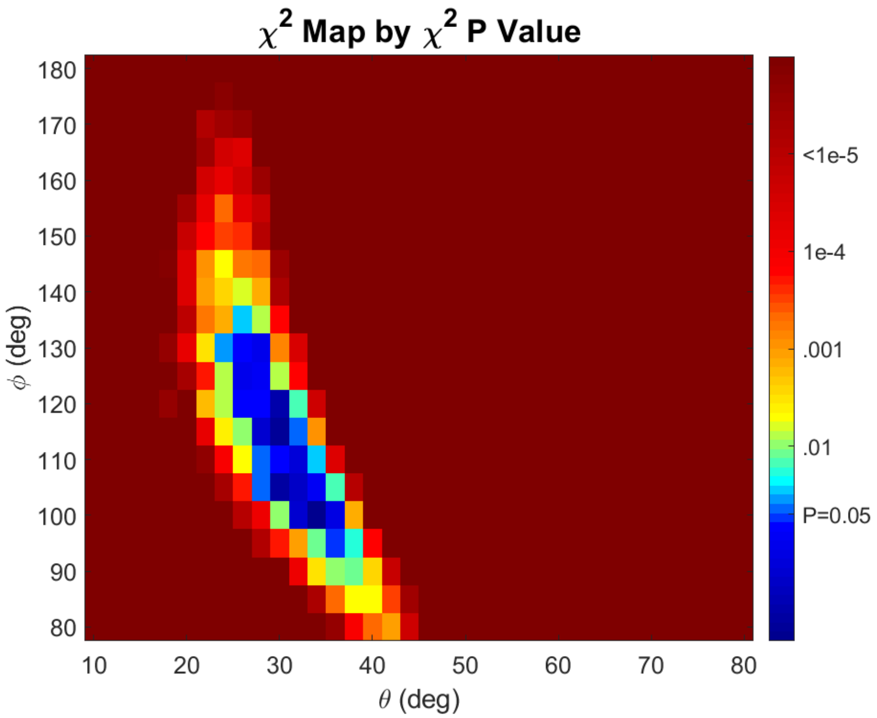


Figure 4.6. χ^2-map of TIM4 Membrane-Bound Orientation. This represents a single reflectivity experiment. Each pixel is the confidence level that its corresponding θ and ϕ values are the "true" values. This confidence is quantified as a *p*-value computed for the χ^2 distribution with two degrees of freedom, corresponding to the increase in χ^2 from the best-fit value induced by varying the two parameters, θ and ϕ. Every θ and ϕ value colored blue are candidates for the true values within a 95% confidence.

While this approach worked well for TIM4, our difficulties with TIM1 and TIM3 prompted us to reconsider this approach. We therefore sought a more rigorous, statistical analysis. We first realized that **Equation 4.39** only applied to linear models, i.e. fitting functions of the form (33),

$$f(x_m, p_v) = \sum_{v}^{\mathrm{N}} p_v B_v(x_m) \tag{4.40}$$

The parameters, p_v, appear linearly in this equation as coefficients for functions $B_v(x_m)$, evaluated at the measured points, x_m. **Equation 4.40** provides the definition for the degrees of freedom as a quantification of the degree of linear independence of the functions, $B_v(x_m)$. If these functions are indeed completely linearly independent, then the number of degrees of freedom is given by the formula, $D_f = M - \mathrm{N}$, where M is the number of measurement points and N is the number of parameters.

Clearly, the reflectivity for the membrane-protein system does not admit a linear form. Consequently, any analysis that requires a concept of "degrees of freedom" is unavailable to us, including the reduced χ^2. For TIM1, we overcame this obstacle by evaluating the uncertainty in terms of the variance of the best-fit values between individual data sets. This strategy was ineffective for TIM3, as each individual dataset on its own had poor resolution. The TIM3 data had to be handled with a joint fitting method. This entails fitting all data sets simultaneously to the same set of parameters. Since lipid monolayers vary significantly from preparation to preparation, we allowed each data set to have its own set of lipid parameters and protein coverage. The protein orientation parameters were constrained to be the same value for every data set. This resulted in far stronger resolution of the bound state, as illustrated in **Figure 4.7**, which provides a χ^2-map of TIM4's orientation evaluated for every data set.

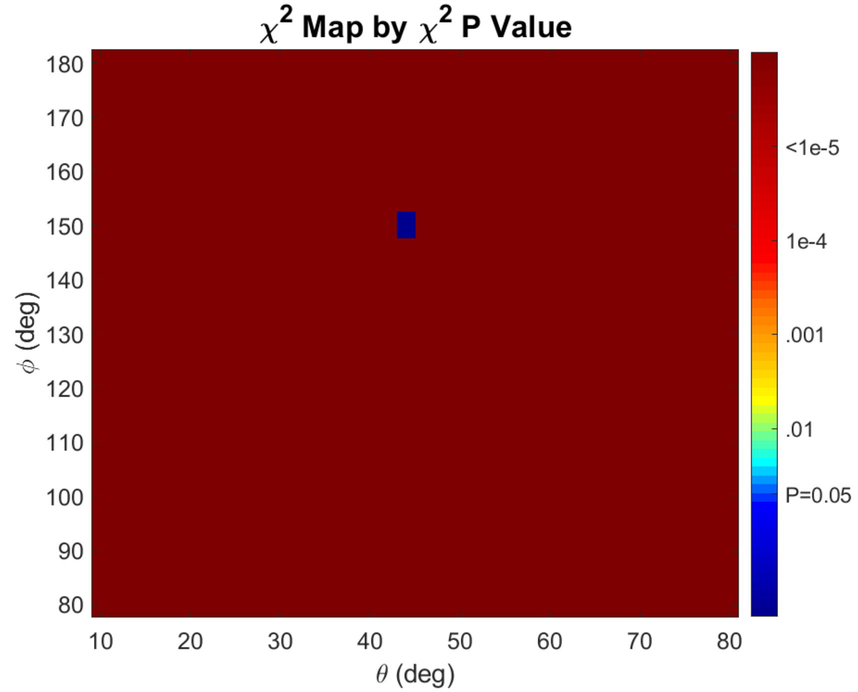


Figure 4.7. Joint χ^2-map of TIM4 Membrane-Bound Orientation. This represents every reflectivity experiment that was conducted for TIM4. The resolution of the bound state is starkly enhanced as compared to that resolved from a single data set, as depicted in **Figure 4.6**. The pixels are color-coded using **Equation 4.39**, despite its discussed inapplicability of non-linear models, for comparison with those in **Figure 4.6**.

While joint fitting results in greatly enhanced resolution of the bound state, we forfeited our ability to characterize the error through the analysis of individual data sets. We instead turned to bootstrapping methods (34), which empirically estimate the error by joint fitting subsets of the data sets, and determining the probability distribution of the best-fit values for each of those subsets. More explicitly, if there are a set of *K* measured reflectivity curves, $\{R_k(Q_{z,m})\}_{k\in\{1,\ldots,K\}}$, we calculate the joint, best-fit parameters by optimizing,

$$\chi^2 = \sum_{k=1}^{K} \sum_{m_k=1}^{M_k} \frac{\left(R_k(Q_{z,m_k}) - R_T(Q_{z,m_k}; \theta, \phi, d_{p,k}, C_k, \rho_{head,k}, \rho_{tail,k}, L_{head,k}, L_{tail,k})\right)^2}{{s_{k,m_k}}^2} \quad (4.41)$$

Here, θ and ϕ are constrained to be the same for all $R_k(Q_{z,m_k})$, while the remaining parameters have the index *k* appended to indicate that they are individually fit to each individual reflectivity data set. The index m_k represents the *k*th reflectivity experiment's m_kth measured point. This indexing reflects the fact that the various $R_k(Q_z)$ were not necessarily measured at the same set of Q_z values.

Bootstrapping involves the recalculation of **Equation 4.41** for subsets of the data sets. We introduce a family of indices, $n_{j,k}$, corresponding to a single bootstrap labeled by *j*,

$$\forall k, \forall k' \left(P(n_{j,k} = k') = \frac{1}{K} \right) \tag{4.42}$$

That is, each bootstrap, *j*, consists of a family of *K* indices, $n_{j,k}$, each of which is randomly assigned a particular data set, labeled by *k'*, according to a uniform probability distribution. Each bootstrap might have multiple copies of the same data sets assigned to its $n_{j,k}$ indices but there are always *K* total data sets within a given bootstrap.

The idea captured in **Equation 4.42** is that a single bootstrap, in subsampling the obtained data, simulates an actual experiment subsampling the set of all possible outcomes. Bootstrapping models the "set of possible outcomes" with the set of measured/observed outcomes. The sampling procedure is then simulated in the same way a real measurement samples the "set of possible outcomes." A given bootstrap, labeled by *j*, represents a joint fit over the subset of data sets tabulated by $\{n_{j,k}\}_j$, with a goodness of fit parameter given by,

$$\chi^2{}_j = \sum_{k=1}^{K} \sum_{m_{n_{j,k}}=1}^{M_{n_{j,k}}} \frac{\left(R_{n_{j,k}}\left(Q_{z,n_{j,k}}\right) - R_T\left(Q_{z,m_{n_{j,k}}}; \theta, \phi, d_{p,n_{j,k}}, \ldots\right)\right)^2}{s_{n_{j,k},m_{n_{j,k}}}{}^2} \tag{4.43}$$

The first sum is over *k*, since there are *K* total datasets within a bootstrap. The rest of the indices that were *k* in **Equation 4.42** are now replaced with $n_{j,k}$. For each bootstrap, *j*, a best-fit set of values θ_j and ϕ_j is obtained. By tabulating these values, we can determine their confidence intervals. Such a tabulation is presented in **Figure 4.8**, which resembles a χ^2-map except that the color corresponds to the probability of the values of θ and ϕ being the best-fit values for a bootstrap *j*.

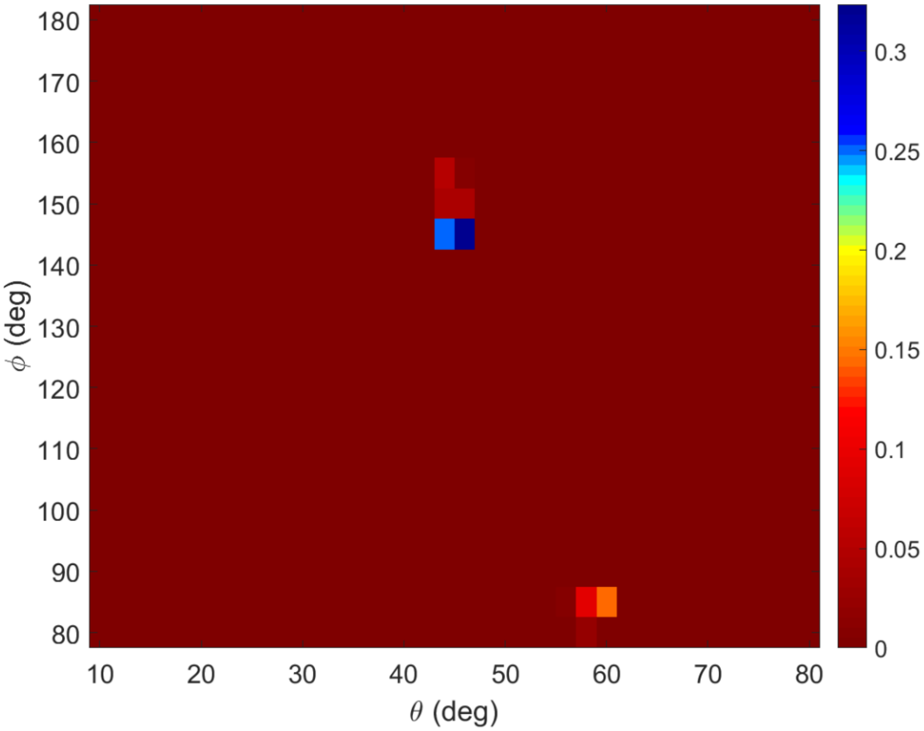


Figure 4.8. Probability Distribution of TIM4 Orientation Parameters in Bootstrapping. The color of each pixel represents the probability of the values of θ and ϕ being best-fit values within a sub-sampling of the complete data. Compared with **Figure 4.7**, the error is wider, suggesting that **Equation 4.39** underestimates the error.

This approach has several strengths over modeled error estimates, such as **Equation 4.39**, in that it is less sensitive to the estimation of experimental error. Reflectivity data only model the error of the x-ray detection process, which is Poisson-distributed for detecting photons. There are far more sources of error which contribute significantly and are nearly impossible to model, such as the preparation of the monolayer and its equilibration. Bootstrapping requires no such model for the error and will infer the influence of these error sources from the variation of the best-fit parameters among bootstraps. Additionally, if we assume the protein orientation is independent of monolayer conditions, such as surface pressure, and the amount of protein bound, several different experiments can be combined in the analysis to resolve the single bound state.

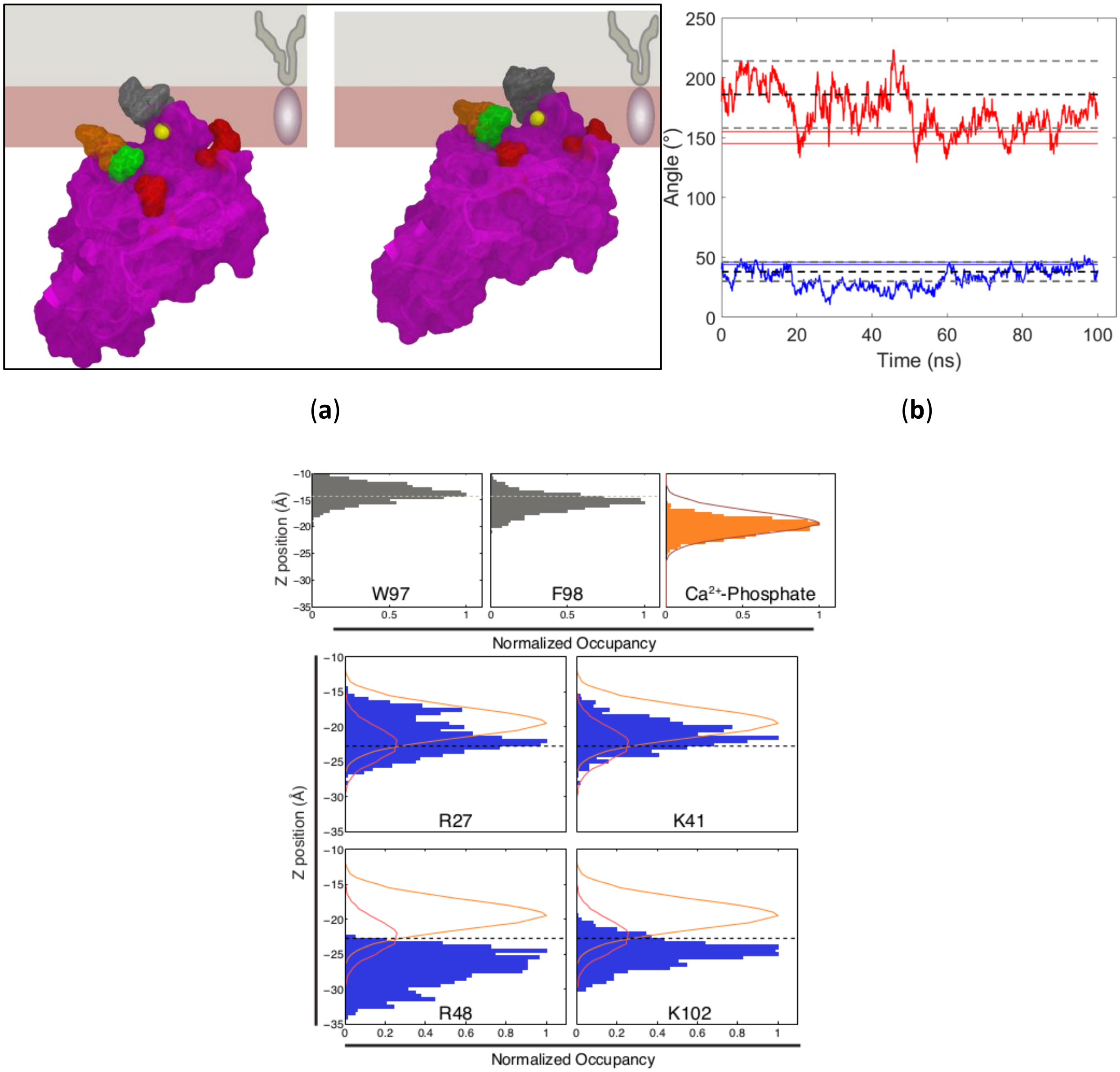
4.2.4 Comparison of Methods Using TIM4

Before applying these improved methods to TIM1 and TIM3 in Chapters 5 and 6, respectively, we applied them to the published model of TIM4. **Figure 4.9a** compares the published membrane-bound structure of TIM4 determined using the old and new methods described in this chapter: Rectangular area vs. encapsulating area, “addition of” vs. “replacement with” protein electron density to lipid electron density, and fitting of each data set with error measured by the variance of χ^2 vs. bootstrapping.

The membrane-bound structure resolved from the application of the new methods is qualitatively similar to the published model. However, the new methodology resolved an orientation with overall deeper insertion and nearly a 55% increase of surface burial. However, the interaction footprint of the protein, defined by the atoms in the protein that are within 3.5 Å of the monolayer, is increased by only 18%. This indicates that the same surface area is implicated by the two orientations, but that the new orientation directly buries more of this interaction surface into the membrane. Given the importance of these interfacial residues in the binding of TIM4 (21), the new analysis places them in a more plausible, deeply embedded configuration. Indeed, the insertion depths of the four basic residues highlighted in green (lysine) and red (arginine) in the reanalyzed orientation of **Figure 4.9a** are consistent with their insertion depths given by a docked MD simulation of a 7:3 POPC:POPS bilayer initiated with the protein associated in the published orientation. However, as the simulation progresses, the angles seemingly begin to drift away from the initial orientation and eventually towards the reanalyzed orientation (**Figure 4.9b**). Additionally, the simulation shows K41 and R48 embedded in the membrane (**Figure 4.9c**), consistent with the reanalyzed orientation, while the published orientation places them in the buffer. The center of mass of both the tryptophan and phenylalanine residues bury into the tail group region during the simulation but only the reanalyzed orientation places these residues in the tail group region. Finally, the bound Ca^{2+} ion is in-line with the phosphates throughout the simulation, which are located near the head-tail interface. The reanalyzed orientation places the Ca^{2+} ion much closer to this interface. Taken together, the MD more greatly supports the reanalyzed orientation.

The greater surface burial of the new orientation is likely due to the “replacement” method accommodating more protein electron density in the lipid layer. An equivalent orientation with the “addition” method would result in a much higher electron density in the monolayer. This is evidently not consistent with the data. The “addition” method instead withdraws and tilts the protein as to mostly maintain the electron density profile extending from the membrane as seen from the equal or nearly equal lengths of the buffer-exposed protein in **Figure 4.9** (published model extends 5 Å, ~12.5%, more from the membrane).

Furthermore, the improved statistical methods allowed us to utilize TIM4 data which were previously excluded from the publication on the basis of poor resolution. We were also able to simultaneously analyze data obtained for a variety of film and subphase conditions. The values corresponding to the published model and the fully refined model (right panel of **Figure 4.9**) are provided in **Table 4.1**.



(a)

(b)

(c)

Figure 4.9. Comparison of TIM4 Best-Fit Orientations Using Old and New Methods. (**a**) Left: Published TIM4 best-fit orientation using rectangular area (**Equation 4.**16), "addition" of protein electron density (**Equation 4.**28) and evaluated using the variation of χ^2 (**Equation 4.**39) (21). Right: TIM4 best-fit orientation analyzed with additional data using encapsulating area (**Equations 4.**18 and 4.19), "replacement" with protein electron density (**Equation 4.**27), and evaluated using bootstrapping (**Equation 4.**43). The structure is colored to highlight important residues, arginine in red, lysine in green, hydrophobic in silver, and polar in orange. (**b**) Values of θ (blue) and ϕ (red) over the course of a docked MD simulation of TIM4 with a 7:3 POPC:POPS bilayer. The original fit's best-fit angular values are overlaid as black dashed lines with confidence intervals as gray dashed lines. The 95% confidence intervals of the best-fit values of the orientation determined by the new methodology are plotted as blue and red solid lines.

Figure 4.9 cont. (**c**) Distribution of hydrophobic and basic residues during the simulation shown in panel **b**. Surface depth distribution of the center of mass of W97 (Left) and F98 (Center) relative to the start of the tail region (dashed gray line). (Right) Center-of-mass, surface depth distribution of the Ca^{2+}-bound, central PS-phosphate (horizontal orange bars) relative to the distribution of all other phosphates in the leaflet (burgundy line). All distributions are normalized to total number of sampled frames. (Below) Center-of-mass, surface depth distributions of side chain nitrogen atoms (cationic sites) for the lysine and arginine residues visualized in panel **a** (horizontal blue bars). These are shown relative to distributions of the two different species of anionic binding partners: phosphates (orange line; includes both PC and PS) and carboxyl moieties of PS head groups (burgundy line includes only PS). Both anion distributions are normalized to the total frames of the phosphate to show the relative abundance of phosphates (100% of lipids) versus carboxyl moieties (30% of lipids). The dashed black line denotes phosphate-dominated regions (above the line) versus carboxyl-dominated regions (below the line) of potential anionic interaction. Panels **b**, **c**, and their respective legends were adapted from ref. (21).

TIM4 Parameter	*Published Model*	*Reanalysis*
θ, °	38, [30, 46]	46, [44, 46]
ϕ, °	186, [158, 214]	145, [145, 155]
Protein Insertion d_p, Å	11.3, [10.4, 12.1]	12.9, [11.7, 14.1]
Surface Burial, Å^2	900	1400
Interaction Footprint, Å^2	700	830

Table 4.1. TIM4 X-ray Reflectivity Orientation Parameters. The left column contains the values previously published (21) and the right column contains the reanalyzed parameters obtained from the refinements to the fitting methodology. The lipid parameters and the coverage are not compared as the new approach combined several different film conditions in order to resolve the single, membrane-bound state. The parameters for the reanalyzed model are evaluated at the more likely peak identified in **Figure 4.8**. The surface burial is calculated from the solvent accessible area of the protein which buries into the monolayer, given by the protein insertion. The interaction footprint is defined as the two-dimensional area of the lipid monolayer under the protein which is within a 3.5 Å interaction distance. In brackets are the 95% confidence intervals of the fit parameters.

4.3 Beyond Crystal Structures: Fitting MD-Resolved Protein Structures to X-ray Reflectivity

4.3.1 Introduction

The methods developed thus far assume a given protein structure obtained from experiments. However, as discussed in section 4.1, the crystal structure can be insufficient for many reasons. In our work with TIM1 and TIM3, the crystal structures depict a “closed” state in which the loops that form the PS binding pocket are folded in on each other. Since x-ray reflectivity is highly sensitive to the interfacial electron density, we require a model that places these penetrating residues in the right place. We therefore employed MD simulations to yield membrane-bound structures of these proteins (see Chapter 3 for methodological details). Once we obtained trajectories, it was highly nontrivial to use them to analyze x-ray reflectivity data. In this section, the analysis of x-ray reflectivity data using MD simulations is developed.

MD simulations and x-ray reflectivity can potentially couple in many different ways. The membrane-bound structure obtained from reflectivity analyzed using a crystal structure can be used as the initial condition of a simulation. The equilibrated protein structure resulting from this simulation can subsequently be used to reanalyze the reflectivity data and improve the fit quality. Alternatively, MD approaches that simulate the binding process and do not rely on an initial docked orientation for the protein, can be utilized to determine the bound state for reflectivity analysis. In this thesis work, both methods were employed and are compared in Chapter 5.

Ideally, MD simulations would be conducted on unconstrained lipid monolayers at an air-water interface and the electron density would be explicitly calculated. Unfortunately, at the time of this writing, lipid monolayer simulations are highly contrived, requiring a force to keep water from escaping past the monolayer or requiring two monolayers with periodic boundary conditions that effectively form a bilayer (35). Advances in monolayer simulations are forthcoming and this method of analysis may be available to us in the future (36). Instead, we simulated bilayer systems and analyzed the protein in isolation. This structure was used to calculate an electron density, as described in section 4.2. The problem therefore reduces down to obtaining a single, representative structure from which we can model the electron density of the membrane-bound protein to combine with the two-slab model.

4.3.2 The Analysis and Alignment of MD Trajectories for X-ray Reflectivity Analysis

Our goal is to obtain a representative state of the membrane-bound protein. This is normally accomplished by characterizing states of the protein based on its internal configurations. MD simulations are routinely analyzed by aligning each frame of the protein to the backbone, using a simple and efficient method by Kabsch (37, 38). This inherently discards each frame's orientation and center of mass position information and only preserves the internal fluctuations of the protein. The states are effectively defined as perturbations to the internal structure only and not orientational changes.

However, reflectivity does not measure the protein in its center of mass frame but instead relative to a well-defined interface. The states of the protein must be analyzed from the perspective of the membrane. We therefore require an appropriate alignment algorithm that maintains this reference frame. Since the protein electron density appears in the reflectivity analysis as a supplement to the two-slab model, the center of the membrane effectively defines a reference frame for protein alignment. Fortunately, this reference frame does not require complete information of the protein's orientation and position to uniquely define a state. The membrane is isotropic in plane, meaning states of the protein that differ in their *x*-*y* center of mass or azimuthal orientation are equivalent. We take advantage of these unconstrained degrees of freedom and align each frame of the protein in the *x*-*y* plane. Specifically, this means that the fluctuations of the protein's center of mass motion in the *z* direction and the two Euler angles, θ and ϕ are retained through the alignment procedure, as illustrated in **Figure 4.10**.

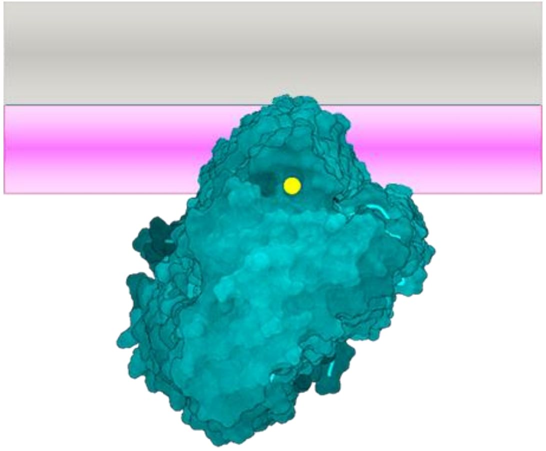


Figure 4.10. Characterization of Protein States from the Reference Frame of the Membrane. This cartoon illustrates a time average of membrane-bound TIM3 against a monolayer with head groups in pink and tail groups in gray. The states of the protein have been aligned by their *x*-*y* center of mass and azimuthal angle; parameters of which x-ray reflectivity is insensitive. The resulting aligned ensemble of protein states vary in tilt from the membrane normal and relative insertion, as seen by the black outline of protein states.

This alignment procedure is a special case of the alignment method of Kabsch (37, 38) in two dimensions and can be cast as the minimization of the function,

$$F\left[\left\{U_m\left(\{\vec{y}_n\}_{n\in\{1,\ldots,N\}}\right)\right\}_{m\epsilon\{1,\ldots,M\}}\right] = \frac{1}{2M}\sum_{m=1}^{M}\left(\sum_{n=1}^{N} w_n|U_m\cdot\vec{a}_{nm}-\vec{y}_n|^2 \right.$$
$$\left. +\mathrm{Tr}(L[U_m U_m^T - \mathbb{I}_2])\right) \tag{4.44}$$

Here, $\vec{a}_{nm}$ are the *x*-*y* coordinates of the protein for the frame *m* and the atom *n* aligned to the *x*-*y* center of mass for each frame m. The weights w_n determine the importance of each atom and the exact form of this function will be addressed later in this section. $\vec{y}_n$ are the *x*-*y* coordinates of a reference structure to which the frames are being aligned and U_m is an orthogonal matrix for the frame *m* which rotates the coordinates $\vec{a}_{nm}$. *L* is a matrix of Lagrange multipliers which constrain the matrices U_m to be orthogonal and $\mathbb{I}_2$ is the two-dimensional identity matrix. Since the rotation described by U_m is purely azimuthal (in-plane parallel to the membrane), each matrix has a corresponding azimuthal angle, ψ_m,

$$U_m\left(\{\vec{y}_n\}_{n\in\{1,\ldots,N\}}\right) = \begin{bmatrix}\cos\psi_m & -\sin\psi_m\\ \sin\psi_m & \cos\psi_m\end{bmatrix} \tag{4.45}$$

Minimizing **Equation 4.44** with respect to the orthogonal matrices, $\{U_m\}_{m\epsilon\{1,\ldots,M\}}$, determines the values of ψ_m which align $\vec{a}_{nm}$ as closely as possible to $\vec{y}_n$ without relatively moving any atom in the structure. The solution for these matrices is given by,

$$U_m\left(\{\vec{y}_n\}_{n\in\{1,\ldots,N\}}\right) = W_m V_m^T \tag{4.46}$$

where the orthogonal matrices W_m and V_m are, respectively, the matrices of right-singular vectors and left-singular vectors obtained from the singular value decomposition (SVD),

$$[\vec{y}_1 \quad \cdots \quad \vec{y}_N]\begin{bmatrix} w_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_N \end{bmatrix}\begin{bmatrix} \vec{a}_{1m} \\ \vdots \\ \vec{a}_{Nm} \end{bmatrix} = V_m S_m W_m^T \tag{4.47}$$

Recall that SVD represents the action of a matrix in terms of a rotation V_m, scaling S_m, and then another rotation W_m^T. The matrix of **Equation 4.47** describes the cross-covariance of a spread of atomic shifts in the *x*-*y* plane between a frame *m* and the reference structure. This covariance has two directions of maximal correlation with corresponding magnitudes reflecting the degree of correlation along those axes. SVD cleanly separates the direction from the magnitude into the rotation and scaling matrices, respectively. The magnitude of the scaling matrix contains information regarding the compression or expansion of the frame *m* relative to the reference structure. We do not want to scale the protein when aligning it. Therefore, if we retain only the rotation matrices, we align $\vec{a}_{nm}$ to the reference structure $\vec{y}_n$ along the axis of maximal correlation without stretching out the atoms. **Equation 4.46** is exactly this product of the SVD rotation matrices.

4.3.3 Obtaining a Representative Structure from Aligned Trajectories

With an alignment algorithm defined, we can now classify the membrane-bound states of the protein. We seek a representative structure that includes fluctuations in *z* and changes in tilt to the membrane normal in addition to internal rearrangements. Our alignment scheme automatically includes these components, but it requires modification to yield a representative state. Note that the optimization function of **Equation 4.44** can also be optimized with respect to the reference structure, $\{\vec{y}_n\}_{n\in\{1,\ldots,N\}}$. The reference structure is then an additional free-fitting parameter which we now recognize as the representative structure. **Equation 4.48** defines this optimization function,

$$F\left[\left\{U_m\left(\{\vec{y}_n\}_{n\in\{1,\dots,N\}}\right)\right\}_{m\epsilon\{1,\dots,M\}},\{\vec{y}_n\}_{n\in\{1,\dots,N\}}\right]=\frac{1}{2M}\sum_{m=1}^{M}\left(\sum_{n=1}^{N}w_n|U_m\cdot\vec{a}_{nm}-\vec{y}_n|^2\right.$$
$$\left.+\mathrm{Tr}(L[U_mU_m^T-\mathbb{I}_2])\right)\tag{4.48}$$

While it may seem that **Equation 4.48** is the complete answer, it contains a fatal flaw: The rotation matrices implicitly depend on the representative structure, $\vec{y}_n$. Indeed, the values for $\vec{y}_n$ that optimize **Equation 4.48** are given by solutions to the following partial differential equations (PDE),

$$w_n\vec{y}_n=\sum_{m=1}^{M}\nabla_{\vec{y}_n}\left(s_{m,1}\left(\{\vec{y}_n\}_{n\in\{1,\dots,N\}}\right)+s_{m,2}\left(\{\vec{y}_n\}_{n\in\{1,\dots,N\}}\right)\right)\tag{4.49}$$

The singular values s_m of the rotation matrices, U_m, are implicitly functions of all $\vec{y}_n$. Consequently, **Equation 4.49** describes a set of *N*-coupled PDEs. Furthermore, the singular values have no analytical form, meaning the PDEs are likely ill-defined as the derivatives may not exist. This problem must be approached numerically.

We define an iterative procedure in which we start with an assigned reference structure, and continually refine a representative structure with multiple optimization procedures,

$$F\left[\left\{U_m\left(\{\vec{x}_n\}_{n\in\{1,\dots,N\}}\right)\right\}_{m\epsilon\{1,\dots,M\}}\right] = \frac{1}{2M}\sum_{m=1}^{M}\left(\sum_{n=1}^{N} w_n|U_m\cdot\vec{a}_{nm}-\vec{x}_n|^2 \right.$$
$$\left. +\mathrm{Tr}(L[U_m U_m^T-\mathbb{I}_2])\right) \tag{4.50}$$

We choose $\{\vec{x}_n\}_{n\in\{1,\dots,N\}}$ to be one of the frames $\{\vec{a}_{nm}\}_{n\in\{1,\dots,N\}}$ at random and solve for the alignment matrices $\left\{U_m\left(\{\vec{x}_n\}_{n\in\{1,\dots,N\}}\right)\right\}_{m\epsilon\{1,\dots,M\}}$. At this point, we set the weight function equal to the number of electrons of atom *n*,

$$w_n = e_n \tag{4.51}$$

This choice is motivated by the fact that reflectivity is sensitive to the electron density and we want to prioritize the alignment of maximally scattering atoms. We proceed to align the coordinates with the starting reference frame $\{\vec{x}_n\}_{n\in\{1,\dots,N\}}$,

$$\vec{a}_{nm}^{(1)} = U_m\left(\{\vec{x}_n\}_{n\in\{1,\dots,N\}}\right)\cdot\vec{a}_{nm} \tag{4.52}$$

The aligned coordinates $\vec{a}_{nm}^{(1)}$ are then used to calculate a representative frame,

$$F\left[\left\{\vec{y}_n^{(1)}\right\}_{n\in\{1,\dots,N\}}\right] = \frac{1}{2M}\sum_{m=1}^{M}\sum_{n=1}^{N} w_n\left|\vec{a}_{nm}^{(1)}-\vec{y}_n^{(1)}\right|^2 \tag{4.53}$$

Optimization of **Equation 4.53** yields,

$$\vec{y}_n^{(1)} = \frac{1}{M}\sum_{m=1}^{M}\vec{a}_{nm}^{(1)} \tag{4.54}$$

We can then characterize the RMSD of the trajectory from this representative structure,

$$\mathrm{RMSD}^2 = \frac{1}{M}\sum_{m=1}^{M}\left|\vec{a}_{nm}^{(1)} - \vec{y}_n^{(1)}\right|^2 \tag{4.55}$$

Atoms that have larger RMSD, are inherently more spread out over the course of the trajectory, thus it is not as important to define a representative set of coordinates for these atoms. We choose to emphasize the positions of more rigid atoms, as they localize their corresponding electrons and define a sharper electron density profile. We then define the next weight function as,

$$w_n^{(1)} = \frac{e_n}{\mathrm{RMSD}^2} = \frac{e_n}{\frac{1}{M}\sum_{m=1}^{M}\left|\vec{a}_{nm}^{(1)} - \vec{y}_n^{(1)}\right|^2} \tag{4.56}$$

Taking the obtained representative structure from **Equation 4.54**, we re-align the trajectory to the new reference state,

$$F\left[\left\{U_m\left(\left\{\vec{y}_n^{(1)}\right\}_{n\in\{1,\dots,N\}}\right)\right\}_{m\in\{1,\dots,M\}}\right]$$

$$= \frac{1}{2M}\sum_{m=1}^{M}\left(\sum_{n=1}^{N} w_n^{(1)}\left|U_m\cdot\left(\vec{a}_{nm}^{(1)} - \frac{\sum_{n=1}^{N} w_n^{(1)}\vec{a}_{nm}^{(1)}}{\sum_{n=1}^{N} w_n^{(1)}}\right) - \vec{y}_n^{(1)}\right|^2 + \mathrm{Tr}(L[U_m U_m^T - \mathbb{I}_2])\right). \tag{4.57}$$

In **Equation 4.57**, the trajectory is centered according to the new weights, $w_n^{(1)}$. From this, we obtain a new representative structure,

$$\vec{y}_n^{(2)} = \frac{1}{M}\sum_{m=1}^{M} U_m\left(\left\{\vec{y}_n^{(1)}\right\}_{n\in\{1,\dots,N\}}\right)\left(\vec{a}_{nm}^{(1)} - \frac{\sum_{n=1}^{N} w_n^{(1)}\vec{a}_{nm}^{(1)}}{\sum_{n=1}^{N} w_n^{(1)}}\right). \tag{4.58}$$

This procedure is repeated until for some integer *p*,

$$\sum_{n=1}^{N} w_n^{(p-1)} \left| \vec{y}_n^{(p)} - \vec{y}_n^{(p-1)} \right| \leq \delta \tag{4.59}$$

$$\vec{y}_n^{(p)} = \frac{1}{M} \sum_{m=1}^{M} U_m \left(\left\{ \vec{y}_n^{(p-1)} \right\}_{n \in \{1,\ldots,N\}} \right) \left(\vec{a}_{nm}^{(p-1)} - \frac{\sum_{n=1}^{N} w_n^{(p-1)} \vec{a}_{nm}^{(p-1)}}{\sum_{n=1}^{N} w_n^{(p-1)}} \right). \tag{4.60}$$

δ is a threshold parameter set at a value for acceptable convergence. We use the value 10^{-6} Å by default as this gives sufficient confidence and is computationally feasible for ~100,000 frames.

The representative structure $\vec{y}_n^{(p)}$, is essentially the mean structure of the trajectory. As a result, the representative structure might be unphysical in the sense that no single state of the protein ever resembles it or is incompatible with the laws of physics. Since the mean is a first-order approximation to the membrane-bound state, we have multiple options to improve it. The most obvious improvement is to simply take the aligned trajectory $\vec{a}_{nm}^{(p)}$ and calculate the average electron density of the whole trajectory. This is equivalent to directly calculating the distribution to which the mean structure corresponds. However, this requires calculating the electron density described in section 4.2 *M* times. This is currently infeasible with our current algorithms and would take thousands of hours for ~30 data sets and ~10,000 frames.

Several alternative methods are computationally feasible: The first of which is to approximate the atom centers of **Equation 4.36** with those given by the mean structure of **Equation 4.60** to convolve the van der Waal's spheres of the atoms with the ellipsoidal covariance of the atoms obtained from the three-dimensional RMSD of the trajectory,

$$\frac{1}{M}\sum_{m=1}^{M}\left|\vec{a}_{3D,nm}^{(p)}-\vec{y}_{3D,n}^{(p)}\right|^{2} \tag{4.61}$$

This covariance can be used to redefine the indicator functions of **Equations 4.31** and **4.32** in terms of ellipsoids instead of spheres. The electron density is still constant within these ellipsoids. Alternatively, the van der Waal's spheres (**Equation 4.21**) can be convoluted with a multivariate Gaussian with a covariance given by **Equation 4.61**. Either approach acts as a second-order correction to the first-order approximation given by **Equation 4.60**.

However, recall that x-ray reflectivity is directly sensitive only to the gradient of the average electron density, and not the variance. Additionally, since the alignment procedure treats each atom independently, the mean and the variance are insensitive to the correlations between atoms and thus fails to capture them. These atomic correlations can be supplemented to the mean structure by using the trajectory itself. The constraints of MD ensure that each frame maintains the correlations between atoms, as these cross-dependencies arise from the laws of physics. Among the atomic coordinates $\vec{a}_{3D,nm}^{(p)}$, we determine which one best resembles the mean structure,

$$F\left[\left\{\vec{a}_{3D,nm}^{(p)}\right\}_{n\in\{1,\ldots,N\}}\right]=\frac{1}{2}\sum_{n=1}^{N}w_{3D,n}^{(p)}\left|\vec{a}_{3D,nm}^{(p)}-\vec{y}_{3D,n}^{(p)}\right|^{2} \tag{4.62}$$

Equation 4.62 determines a frame *m* as the representative structure for the whole trajectory. This frame is then used in the calculation of **Equation 4.36** when fitting reflectivity data.

We found this last approach to yield the best results. The second-order corrections to the mean significantly increased the computation time, offered no new information in the fitting of reflectivity data, and actually performed worse than the mean alone. **Table 4.2** compares the best-fit orientations and corresponding χ^2 values of these discussed approaches for TIM3, using the simulations described in Chapter 6. A comparison to the wrong model of TIM4 is included as a negative control for reference.

Values	*Best Frame*	*Mean*	*Covariance*	*Cov-Gaussian*	*TIM4 Crystal*
θ, °	30 [25, 35]	15 [15, 20]	20 [15, 25]	20 [15, 25]	30 [25, 35]
ϕ, °	140 [140, 150]	110 [100, 120]	130 [120, 140]	140 [120, 140]	130 [120, 130]
χ^2	5,472	5,537	5,978	5,933	6,214

Table 4.2. Comparison of TIM3 Fit Parameters for Different Representative Structures. The "Best Frame" method determines the representative frame from the trajectory that best fits the mean structure. "Mean" fits the mean structure directly to the data. "Covariance" assumes a uniform distribution of electron density over an ellipsoid given by the covariance of the trajectory. "Cov-Gaussian" assumes a Gaussian distribution of electron density given by the covariance of the trajectory. "TIM4 Crystal" uses the TIM4 crystal structure in the reflectivity analysis and is included here as a negative control. The analyzed data includes 1,287 independent measurements. If the model was linear, this would correspond to 1183 degrees of freedom. Brackets represent 95% confidence intervals.

4.4 Conclusions and Future Work

This chapter developed a work pipeline that processes MD trajectories for the purpose of fitting x-ray reflectivity data. This method overcomes a lot of obstacles, most of which are due to the inherent limitations of reflectivity as a technique, particularly its loss of phase information. This difficulty is mitigated by robust modeling that allows us to extract structures using statistical inference. There were plenty of alternative models available to us in both the fitting of x-ray reflectivity and the extraction of MD-derived structures. The choices regarding the former were explored by the analysis of TIM4 data, showing that a model which replaced lipid electron density with that of protein and more carefully treated the area of the protein yielded a more consistent and plausible structure. The choices regarding the latter were more plentiful and harder to distinguish based on physical principles. Analysis of TIM3 reflectivity data hinted at a clear preferred option. However, the development of better MD simulation methods and efficient algorithms for computing the electron density of trajectories may result in one of the other methods performing better.

It is worth considering the nature of the protein structure resolved by reflectivity, as the methodology presented here clearly implies a particular bias. The geometry of a reflectivity experiment and the two-slab model enforce a membrane-centric approach to modeling the protein structures. This induces a specific symmetry in the experiment, i.e. lateral isotropy of the membrane. The normal to the membrane has no such symmetry and while this allows us to infer the protein structure in the first place, it also defines a particular statistical ensemble biased by the membrane itself. Specifically, the membrane defines the coordinate system in which the protein structure is measured, meaning the ensemble of protein states is not defined in the center of mass frame of the protein as is typically the case. As seen in **Figure 4.10**, this ensemble has varying center of mass depending on the fluctuations of the protein along the membrane normal. The resulting ensemble average thus represents a protein structure that is “spread-out” along the direction of the membrane normal. Reflectivity-derived structures place protein residues based on their average positions relative to the membrane and not relative to the center of mass of the protein structure.

Such an analysis provides an unusual image, as most protein models are implicitly taken to represent the protein in terms of its own internal degrees of freedom. However, given the information we desire from a reflectivity experiment, this biased average is far more useful than the ensemble measured from the center of mass frame of the protein. This bias is more sensitive to the location of protein residues with respect to the membrane, providing better information for the specific physicochemical interactions that underlie membrane recognition.

The protein's average structure in its own reference frame provides different information, namely the fold of the protein itself while membrane-bound. On its own, reflectivity cannot probe the center of mass ensemble of the protein. Fortunately, in our investigations of the TIM proteins, we were not interested in the center of mass ensemble of the protein. However, when combined with MD, reflectivity can potentially probe this ensemble indirectly. It can be extracted from MD simulations that are shown to be consistent with reflectivity when analyzed from the reference frame of the membrane. The same trajectories can then be analyzed in terms of the center of mass. While such an analysis is not justified now, further development of techniques that use the reflectivity data to validate the entire trajectory will bring such methodologies to fruition. Eventually, as monolayer MD methods improve, the two-slab model might not be required at all. With such an analysis, the ensemble would be defined as fluctuations of the membrane and correlations between the protein states and the local membrane environment. Thus, the engagement of specific protein residues would be correlated with a response of the membrane. For now, determination of the membrane-bound structure of proteins more than suffices for our purposes.

4.5 References

1. Berman HM, et al. (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242.
2. Brunecky R, et al. (2005) Investigation of the binding geometry of a peripheral membrane protein. *Biochemistry* 44(49):16064–16071.
3. Wang G (2008) NMR of membrane-associated peptides and proteins. *Curr Protein Pept Sci* 9(1):50–69.
4. Schmidt-Krey I, Rubinstein JL (2011) Electron cryomicroscopy of membrane proteins: specimen preparation for two-dimensional crystals and single particles. *Micron* 42(2):107–116.
5. Caffrey M (2015) A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallogr Sect F Struct Biol Commun* 71(Pt 1):3–18.
6. Scott JL, Musselman CA, Adu-Gyamfi E, Kutateladze TG, Stahelin RV (2012) Emerging methodologies to investigate lipid-protein interactions. *Integr Biol (Camb)* 4(3):247–258.
7. Tamm LK, Tatulian SA (1997) Infrared spectroscopy of proteins and peptides in lipid bilayers. *Q Rev Biophys* 30(4):365–429.
8. Tatulian SA, Qin S, Pande AH, He X (2005) Positioning membrane proteins by novel protein engineering and biophysical approaches. *J Mol Biol* 351(5):939–947.
9. Málková S, et al. (2005) X-ray reflectivity studies of cPLA2{alpha}-C2 domains adsorbed onto Langmuir monolayers of SOPC. *Biophys J* 89(3):1861–1873.
10. Chen C-H, et al. (2009) Configuration of PKCalpha-C2 domain bound to mixed SOPC/SOPS lipid monolayers. *Biophys J* 97(10):2794–2802.
11. Akgun B, et al. (2013) Conformational transition of membrane-associated terminally acylated HIV-1 Nef. *Structure* 21(10):1822–1833.
12. Nanda H, et al. (2010) Electrostatic interactions and binding orientation of HIV-1 matrix studied by neutron reflectivity. *Biophys J* 99(8):2516–2524.
13. Tiemeyer S, Paulus M, Tolan M (2010) Effect of surface charge distribution on the adsorption orientation of proteins to lipid monolayers. *Langmuir* 26(17):14064–14067.
14. Wacklin HP, et al. (2016) Neutron reflection study of the interaction of the eukaryotic pore-forming actinoporin equinatoxin II with lipid membranes reveals intermediate states in pore formation. *Biochim Biophys Acta* 1858(4):640–652.

15. Heinrich F, Lösche M (2014) Zooming in on disordered systems: neutron reflection studies of proteins associated with fluid membranes. *Biochim Biophys Acta* 1838(9):2341–2349.

16. Kent MS, et al. (2010) Neutron reflectometry study of the conformation of HIV Nef bound to lipid membranes. *Biophys J* 99(6):1940–1948.

17. Weygand M, et al. (1999) Bacterial S-layer protein coupling to lipids: x-ray reflectivity and grazing incidence diffraction studies. *Biophys J* 76(1 Pt 1):458–468.

18. Czogalla A, Grzybek M, Jones W, Coskun U (2014) Validity and applicability of membrane model systems for studying interactions of peripheral membrane proteins with lipids. *Biochim Biophys Acta* 1841(8):1049–1059.

19. Brockman H (1999) Lipid monolayers: why use half a membrane to characterize protein-membrane interactions? *Curr Opin Struct Biol* 9(4):438–443.

20. Chen C-H, Málková Š, Cho W, Schlossman ML (2011) Configuration of membrane-bound proteins by x-ray reflectivity. *J Appl Phys* 110(10):102215.

21. Tietjen GT, et al. (2014) Molecular mechanism for differential recognition of membrane phosphatidylserine by the immune regulatory receptor Tim4. *Proc Natl Acad Sci U S A* 111(15):E1463–72.

22. Naughton FB, Kalli AC, Sansom MSP (2017) Modes of Interaction of Pleckstrin Homology Domains with Membranes: Toward a Computational Biochemistry of Membrane Recognition. *J Mol Biol*.

23. Rogaski B, Klauda JB (2012) Membrane-binding mechanism of a peripheral membrane protein through microsecond molecular dynamics simulations. *J Mol Biol* 423(5):847–861.

24. Nanda H, Heinrich F, Lösche M (2015) Membrane association of the PTEN tumor suppressor: neutron scattering and MD simulations reveal the structure of protein-membrane complexes. *Methods* 77-78:136–146.

25. Monje-Galvan V, Klauda JB (2016) Peripheral membrane proteins: Tying the knot between experiment and computation. *Biochim Biophys Acta* 1858(7 Pt B):1584–1593.

26. Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI (2007) The role of hydrophobic interactions in positioning of peripheral proteins in membranes. *BMC Struct Biol* 7:44.

27. Darré L, Iglesias-Fernandez J, Kohlmeyer A, Wacklin H, Domene C (2015) Molecular dynamics simulations and neutron reflectivity as an effective approach to characterize biological membranes and related macromolecular assemblies. *J Chem Theory Comput* 11(10):4875–4884.

28. Tikhonov AM, Pingali SV, Schlossman ML (2004) Molecular ordering and phase transitions in alkanol monolayers at the water-hexane interface. *J Chem Phys* 120(24):11822–11838.

29. Byrd RH, Schnabel RB, Shultz GA (1988) Approximate solution of the trust region problem by minimization over two-dimensional subspaces. *Math Program* 40-40(1-3):247–263.

30. Branch MA, Coleman TF, Li Y (1999) A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems. *SIAM J Sci Comput* 21(1):1–23.

31. Bevington PR, Robinson DK, Blair JM, Mallinckrodt AJ, McKay S (1993) Data reduction and error analysis for the physical sciences. *Comput Phys* 7(4):415.

32. Tanabashi M, et al. (2018) Review of particle physics. *Phys Rev D* 98(3):030001.

33. Andrae R, Schulze-Hartung T, Melchior P (2010) Dos and don’ts of reduced chi-squared. Available at: https://arxiv.org/abs/1012.3754 [Accessed June 23, 2019].

34. Davison AC (1997) *Bootstrap Methods And Their Application (cambridge Series In Statistical And Probabilistic Mathematics)* (Cambridge University Press). 1st Ed.

35. Liu B, Hoopes MI, Karttunen M (2014) Molecular dynamics simulations of DPPC/CTAB monolayers at the air/water interface. *J Phys Chem B* 118(40):11723–11737.

36. Olżyńska A, Zubek M, Roeselova M, Korchowiec J, Cwiklik L (2016) Mixed DPPC/POPC Monolayers: All-atom Molecular Dynamics Simulations and Langmuir Monolayer Experiments. *Biochim Biophys Acta* 1858(12):3120–3130.

37. Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Cryst A* 32(5):922–923.

38. Kabsch W (1978) A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst A* 34(5):827–828.

CHAPTER 5

THE MEMBRANE-BOUND STRUCTURE OF TIM1

5.1 Introduction

The membrane-bound structure of TIM1, like our studies with TIM4 (1), implicates the involvement of specific residues, and in doing so, provides a molecular model for the membrane-binding of TIM1. This structure was obtained by x-ray reflectivity and analyzed using the methods described in Chapter 4. Without these methods, MD simulations could only be qualitatively compared with reflectivity-derived structures on the basis of the orientation of the protein. If there were multiple MD-derived structures that predicted different residual contacts with lipids but otherwise shared the same orientation, these comparisons could not render any kind of judgment. However, the methods we have developed allowed us to compare multiple candidate membrane-bound structures and distinguish them based on their goodness of fit with the reflectivity data. This chapter focuses on using reflectivity to quantitatively distinguish four possible models of the membrane-bound state of TIM1; three of which are derived from MD and the remaining one is the crystal structure.

In addition to explaining the binding of TIM1, the membrane-bound structure of TIM1 also elucidates the binding of TIM4. To make these comparisons robust and sound, our analysis of TIM1 is preceded by a brief review of the resolved membrane-bound structure of TIM4 in section

5.2. Subsequently, the analysis of TIM1 based on its crystal structure and the inherent limitations of the approach is discussed in section 5.3. Section 5.4 describes the two MD modalities applied to resolve an improved, membrane-bound structure of TIM1. These two modalities are docked MD and HMMM, both described in Chapter 3. The molecular details of these simulations are extracted and analyzed as well. The fitting of these MD-derived structures to the reflectivity is covered in section 5.5. Finally, section 5.6 compares the obtained structure of TIM1 to that of TIM4. These molecular models are then extrapolated to explain their differential binding to PS-containing membranes.

5.2 The Membrane-Bound Structure of TIM4

TIM4 was co-crystalized with a short-chain PS (2), offering a potential model for membrane-bound TIM4, as shown in **Figure 5.1**. This model presents two main features that give high confidence for this model: (i) The entire PS tail is resolved, suggesting a uniform lipid-bound state that coordinates all moieties of the lipid. (ii) The PS is resolved squarely in between two loops (FG and CC’) that promote insertion. The FG loop contains a phenylalanine and tryptophan which, in this configuration, are free to insert into a lipid membrane. The CC’ loop contains polar and charged residues which might associate with polar lipid head groups. (iii) The side chains of the BC loop are oriented towards the PS binding site, possibly implicating interactions between its charged and polar residues with membrane PS.

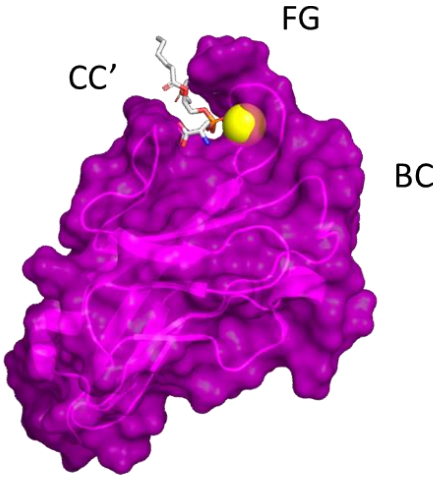
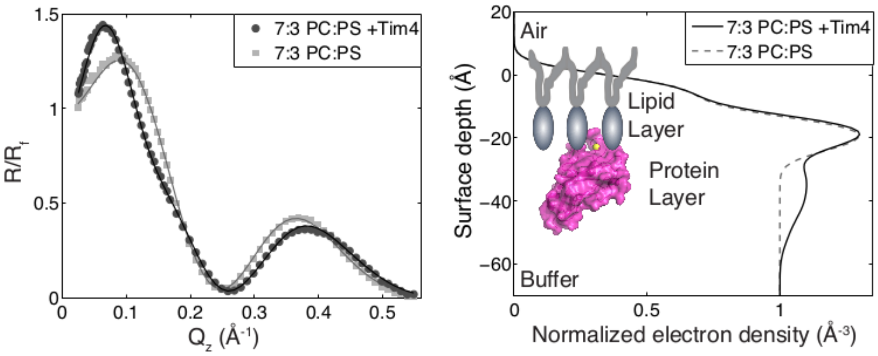
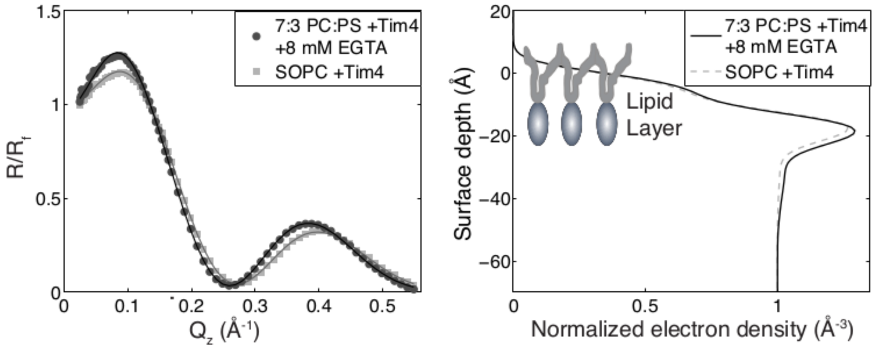


Figure 5.1. The crystal structure of PS-bound TIM4. The TIM proteins all share a similar architecture with loops labeled according to the diagram. The tip of TIM4’s FG loop contains tryptophan and phenylalanine, which partition into the lipid membrane more readily than in water. The CC’ loop contains polar and charged residues which could interact with lipid head groups. The BC loop also contains charged and polar residues that could conceivably interact with the membrane. Figure generated from ref. (2).

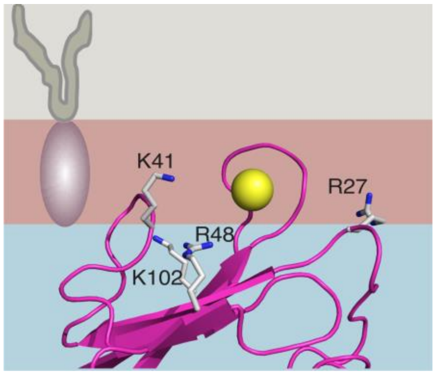
This crystal structure was therefore used to analyze x-ray reflectivity data obtained for a variety of membrane conditions. The left panel, **Figure 5.2a**, shows representative reflectivity curves of 7:3 SOPC:SOPS monolayers with and without TIM4, normalized to the reflectivity expected from a sharp air-water interface. The decay of the curves is due to thermal fluctuations of the interface (described in Chapter 3), while the other features indicate the presence of interfacial structure. TIM4 clearly shifts the peaks left, that the interfacial structure with TIM4 is longer than that without. The right panel of this figure shows the electron density profiles analyzed from the reflectivity curves and confirms that the protein forms an adsorbed layer. Since the TIM proteins all coordinate calcium in their binding of membranes, EGTA was administered to ensure that this signal was due to TIM4 and **Figure 5.2b** confirms that EGTA to a great extent (but not entirely) was able to reduce this adsorbed layer. Similarly, since the TIM proteins all require PS to bind membranes, incubation with SOPC-only monolayers showed no adsorbed layer of TIM4.



(**a**)



(**b**)



(**c**)

Figure 5.2. Reflectivity of TIM4. (**a**) Monolayers composed of 7:3 SOPC:SOPS with or without TIM4. The presence of TIM4 causes a left-shift of the peaks, indicative of a broadening of the interfacial structure. The electron density profile implicates the presence of an adsorbed TIM4 layer.

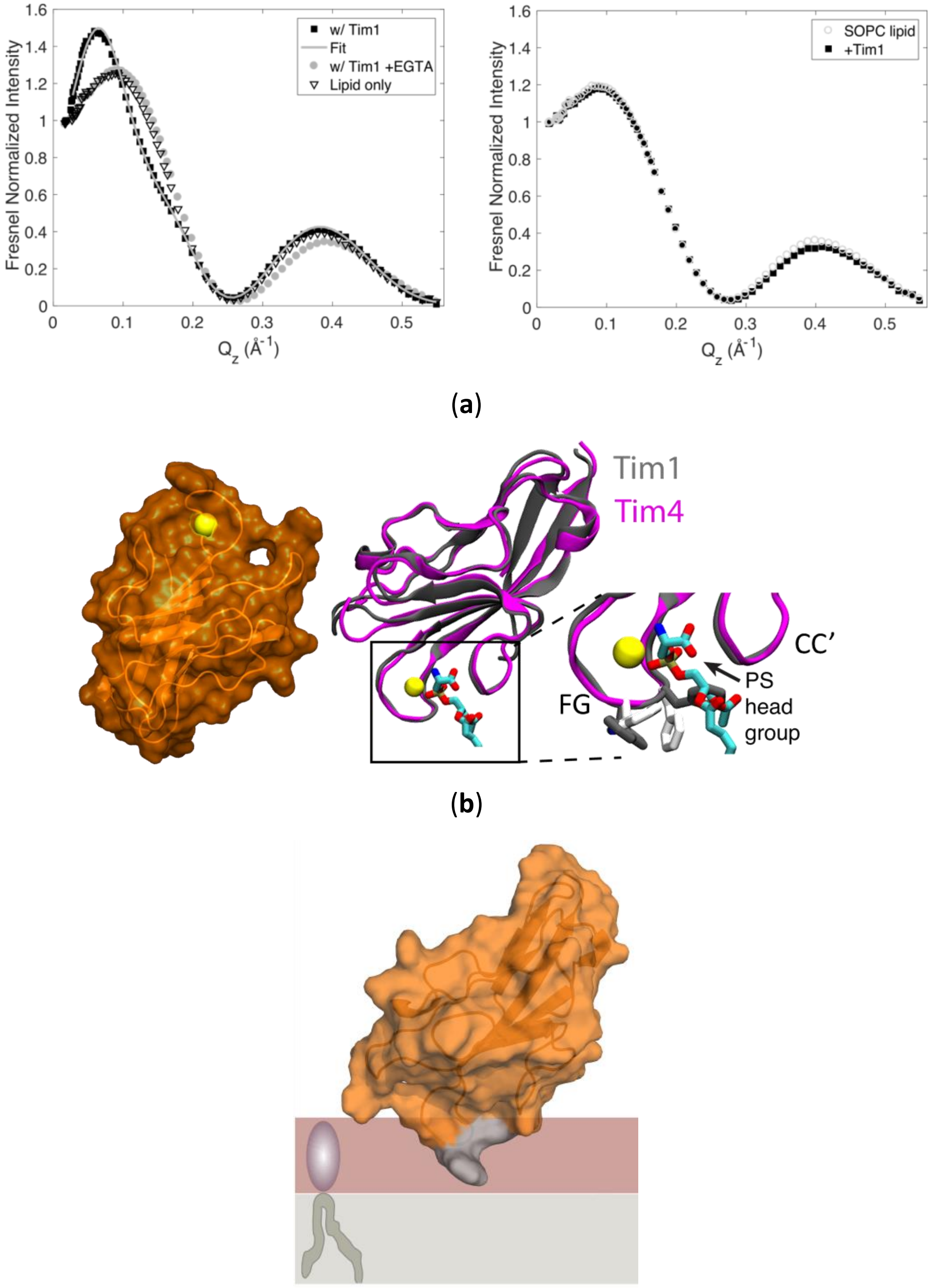
Figure 5.2 cont. (**b**) Administration of EGTA to 7:3 SOPC:SOPS monolayers incubated with TIM4 heavily reduced the adsorbed layer. Incubation of TIM4 with SOPC monolayers produced no measurable adsorbed layer. (**c**) The membrane-bound orientation of TIM4 which fit the reflectivity data best. Figure adapted from ref. (1).

Having confirmed that TIM4 was responsible for the adsorbed interfacial structure, the fitting described in Chapter 4 was conducted using the crystal structure. The membrane-bound structure is shown in **Figure 5.2c**, and implicates several basic residues at the interface. As shown in Chapter 4, MD simulations initiated with TIM4 docked with the membrane in this orientation implicated the same residues in making direct contact with peripheral PS. As will be shown in Chapter 7, these peripheral, basic residues confer TIM4 with a sensitivity for PS in the form of cooperative binding.

This structure also partly answers one of the questions raised in this thesis: How do the TIM proteins differ in their PS sensitivity if they share the same binding pocket? Crystallography implicated the same set of residues as coordinating calcium and PS in this central binding pocket, but x-ray reflectivity further implicates a whole other set of residues that interact with peripheral PS. If the TIM proteins differ in their peripheral contacts, then sharing a binding pocket would not be sufficient to grant them the same membrane sensitivity. Along these lines, we set out to resolve membrane-bound structures for TIM1 and TIM3 to investigate if their membrane contact surfaces perhaps presented different residues to recognize membrane conditions.

5.3 Analysis of TIM1 Reflectivity with its Crystal Structure

TIM1 reflectivity data were obtained from ChemMatCARS Sector 15 at the APS from two separate trips. We assayed a range of protein subphase concentrations and in total obtained 11 independent reflectivity scans corresponding to three independent 7:3 SOPC:SOPS monolayers. In some cases, the same monolayers were used for multiple protein subphase concentrations. In comparison, TIM4 was analyzed using 36 independent reflectivity scans across 11 independent monolayers. TIM1 did not require as much data as its signal was clear to resolve using the same conditions that optimized TIM4's signal. Like TIM4, TIM1 required SOPS in the monolayer for the peaks in the reflectivity to shift, as seen in **Figure 5.3a**. EGTA also eliminated most (but not all) binding, supporting that the adsorbed layer was due to the binding of TIM1.



(**a**)

(**b**)

(**c**)

Figure 5.3. Reflectivity of TIM1 analyzed using its crystal structure. (**a**) 7:3 SOPC:SOPS membranes with TIM1, with TIM1 +EGTA, and without TIM1 on the left, and SOPC-only with or without TIM1 on the right. The peak shift is dependent on the free calcium and SOPS, indicating that the association of TIM1 with the monolayer induced the observed peak shift. (**b**) Left: Crystal structure of TIM1 (3). Right: Unlike TIM4, the pocket of TIM1 is in a closed conformation, with the tryptophan and phenylalanine residues of the FG loop folded into the pocket interior.

Figure 5.3 cont. (**c**) Membrane-bound orientation of TIM1 crystal structure. The phenylalanine and tryptophan residues on the FG loop are in gray. The orientation parameters (defined in Chapter 4) are $\theta = 38 \pm 3°$, $\phi = 115 \pm 9°$, and the depth of insertion, $d_p = 9.4 \pm 0.7$ Å. Figure adapted from ref. (4).

The reflectivity data were analyzed using the crystal structure of TIM1, shown in **Figure 5.3b** (3). While TIM4 was co-crystalized with PS, TIM1 was not. As a result, the hydrophobic residues on the FG loop, homologous to those on TIM4 which held the PS tail group, are folded in to the pocket interior. Nonetheless, we continue with the analysis and the disposition of TIM1 inferred from the reflectivity data is shown in **Figure 5.3c**. This bound structure provides sensible angle parameters, orienting the pocket into the membrane. However, the aforementioned hydrophobic residues do not penetrate past the head group region. Given the strong contribution the residues have to binding (2), this seems highly unlikely.

The fit of the crystal structure of TIM1 highlighted how lucky we were to have a crystal structure of TIM4 in the right state. TIM1, however, represents the far more likely case for any peripheral membrane-binding protein, that if a crystal structure exists it is unlikely to depict the membrane-bound state. Since this analysis method requires a protein structure as an input and cannot generate any kind of correction or suggestion for an improved structure, we must turn elsewhere instead. The following sections will discuss how MD was used to obtain membrane-bound states of TIM1.

5.4 Docked and HMMM MD Simulations of TIM1

It is easy to say that we will use simulations to resolve an improved membrane-bound state of TIM1, however, upon even the slightest consideration the vagueness of this sentiment is immediately confronted by the overwhelming number of approaches using MD that are entirely inequivalent. It can be an art to determine which settings, forcefield, and parameters best represent a given system. Since our goal is to recover a structure from the simulations and not thermodynamics, we considered methods which had low resource consumption. For our first approach we employed a docked membrane simulation in which TIM1 was inserted into the membrane according to the parameters obtained from **Figure 5.3**. However, since the crystal structure did not even contain a calcium ion, we first ran a solution simulation of TIM1 with a calcium ion inserted to its binding pocket, in homology with TIM4 since both share the exact residues that coordinate the calcium.

Once we obtained the calcium-equilibrated TIM1 structure, we modeled a PS lipid in the pocket in homology with TIM4 as well and inserted this complex into a 7:3 POPC:POPS bilayer. This simulation was equilibrated for 300 ns, with the protein stably inserted for the entirety of the trajectory. A representative frame of this simulation is depicted in **Figure 5.4a**, while **Figure 5.4b** shows that the FG and CC’ loops open to accommodate a PS in the binding pocket as compared to the crystal structure. **Figure 5.4c** indicates that the tryptophan on FG loop does penetrate into the tail group region of the membrane, unlike the prediction of the crystal structure-analyzed reflectivity model. Notably, the docked model implicated deep burial of several positive peripheral residues into the headgroup and phosphate region of the membrane. Indeed, more of the peripheral basic residues of TIM1 were deeply buried than was the case for TIM4 (**Figure 4.9**). Given that TIM1 is less sensitive to PS surface density than TIM1 (1), this feature of the docked simulation is surprising.

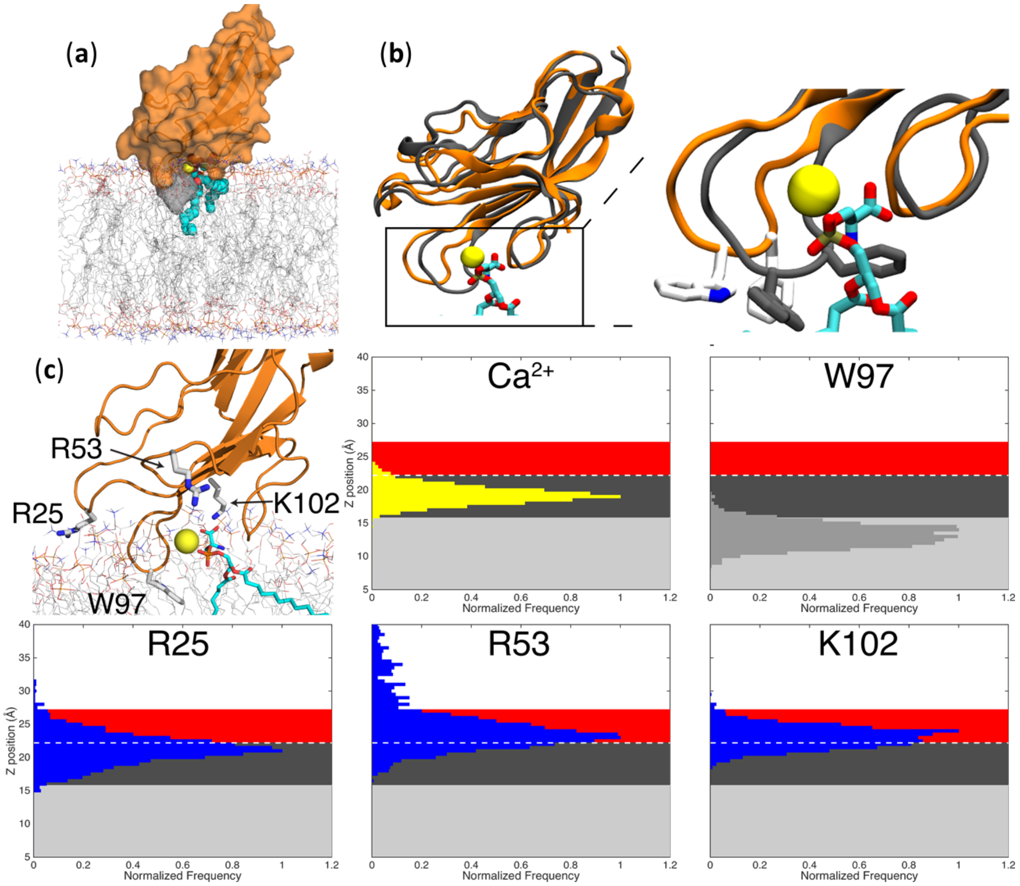


Figure 5.4. Docked Simulation of TIM1 in Complex with a 7:3 POPC:POPS Bilayer. (**a**) A representative frame of the membrane-bound TIM1. (**b**) In orange, the membrane-equilibrated TIM1 and in dark gray, the crystal structure of TIM1, showing that the hydrophobic residues have opened and cleared the pocket to make it available to bind PS. (**c**) In addition to the calcium ion, four residues were tracked during the simulation based on the importance of such residues for TIM4: The hydrophobic tryptophan, two peripheral arginine residues, and a single lysine. Histograms indicate the relative frequency that these residues were at several membrane depths, with the most likely location being normalized to 1. The regions of the membrane are colored as follows: *Red* is primarily the head group region, *dark gray* is dominated by phosphate groups, *light gray* marks the lipid tails and the *white dashed line* indicates the transition between the phosphate-dominated region and PS carboxyl dominated region. Figure adapted from ref. (4).

We suspected that since TIM1 started in solution state when we docked it, that perhaps this might have influenced the simulation to kinetically trap it or otherwise enable TIM1 to occupy a contrived state. To avoid the bias inherent in docking, we employed HMMM (described in Chapter 3) to simulate the membrane binding of TIM1 from solution (5, 6) to a membrane-mimetic composed of short-chained lipids floating on a layer of hydrophobic solvent, 1,1-dichloroethane (dClE). Using the same calcium-equilibrated state as before, we placed TIM1 adjacent to a membrane, roughly 10 Å away, in one of two orientations drawn from the distribution we obtained from the reflectivity fit using the crystal structure, illustrated in **Figure 5.5a**. There were seven replicates of each of the starting orientations and two of each resulted in the binding of TIM1 to a PS in the membrane. **Figure 5.5b** tracks the insertion of the bound calcium ion and the tryptophan residue on the FG loop. Based on these indicators, binding in these four simulations occurred within 50 ns. Furthermore, the initial orientation information was lost over the course of binding as indicated in **Figure 5.5c**. From here, these structures were equilibrated with membranes composed of full-length 7:3 POPC:POPS lipids for 300 ns. These four simulations equilibrated to two distinct states characterized by the mutually exclusive engagements of either a peripheral arginine or a peripheral lysine coordinated with the centrally bound PS (**Figure 5.5d**). Interestingly, these states can also be distinguished based purely on the coordination of the PS to the central calcium. The first state depicts a coordination of both the phosphate and serine moieties of the centrally bound PS while the second only features a coordination of the serine group. Furthermore, these two states do not correlate with the initial orientations.

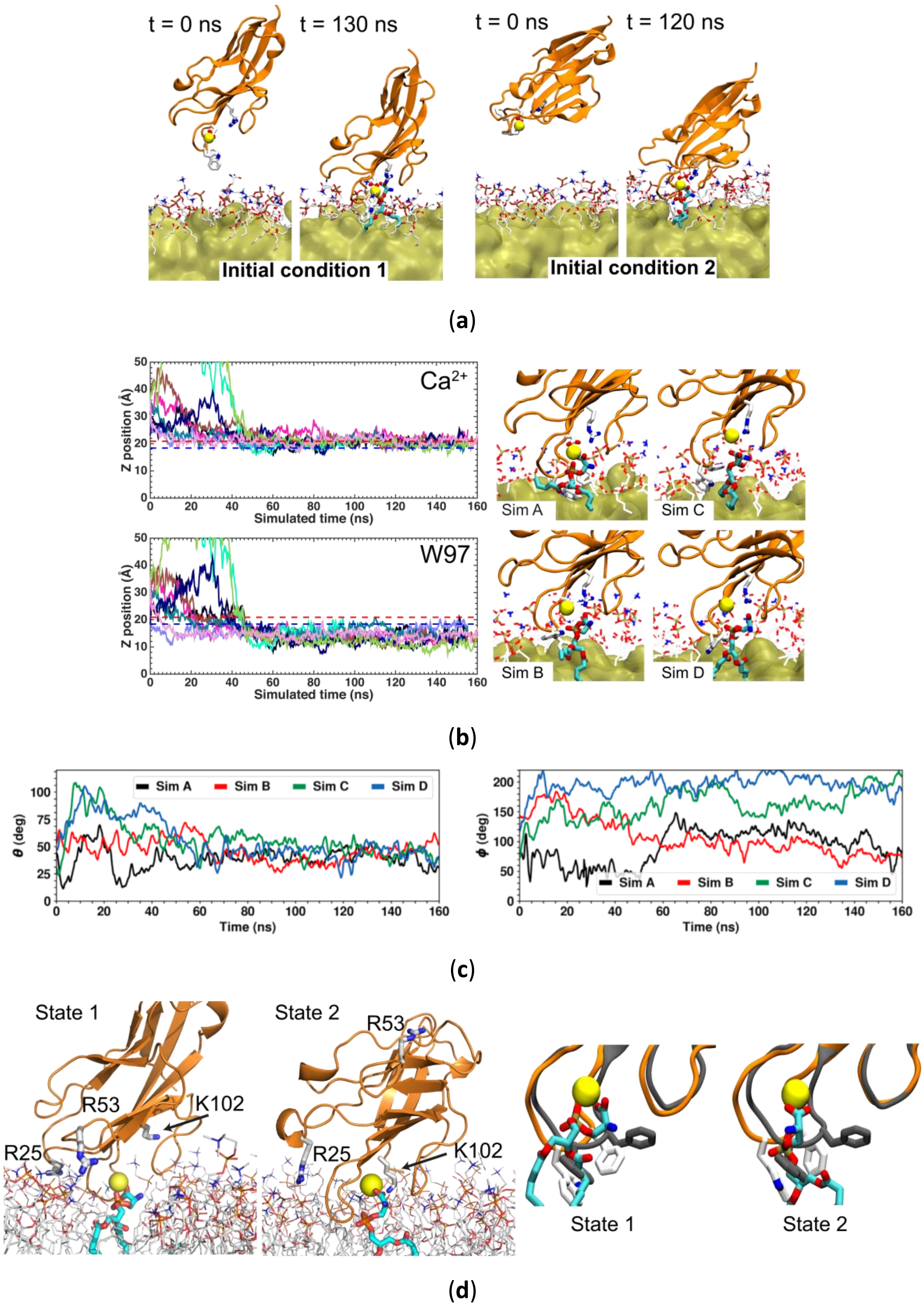


Figure 5.5. HMMM Simulations of TIM1 Binding to Membranes from Solution. (**a**) TIM1 was initiated in solution in two different orientations.

Figure 5.5 cont. (**b**) Of 14 simulations, four featured binding of TIM1 to PS in the membrane within 50 ns. This binding was tracked using the bound calcium ion and the tryptophan residue on the edge of the FG loop. Different colors represent different simulations. The red and blue dashed lines are the mean head group and phosphate positions, respectively. (**c**) Traces of the four simulations during the binding to the membrane. The initial orientation is eventually lost for all four simulations. (**d**) Upon equilibration with full-length lipid membranes, two distinct states were obtained, characterized by the engagement of the peripheral R53 or the coordination of K102 to the centrally bound PS. These states can also be distinguished based on the coordination with the calcium ion, as State 1 depicts a coordination of both the serine and phosphate groups of the PS while State 2 only depicts a coordination of the serine group. In dark gray on the right, the crystal structure disposition of the FG loop for comparison. Figure adapted from ref. (4).

The two obtained states occupy different orientations, as shown in **Figure 5.6a**. Interestingly, while the peripheral arginine and lysine residues clearly contact the membrane in **Figure 5.5d**, the contact map of TIM1 in these simulations indicate that these peripheral residues on the BC and CC’ loops preferentially associate with POPC over POPS (**Figure 5.6b**). In general, POPS is not significantly enriched within 3.5 Å of TIM1 (**Figure 5.6c**), as the ratio of peripheral POPC:POPS is near the value of 7:3 (discounting the PS in the central pocket).

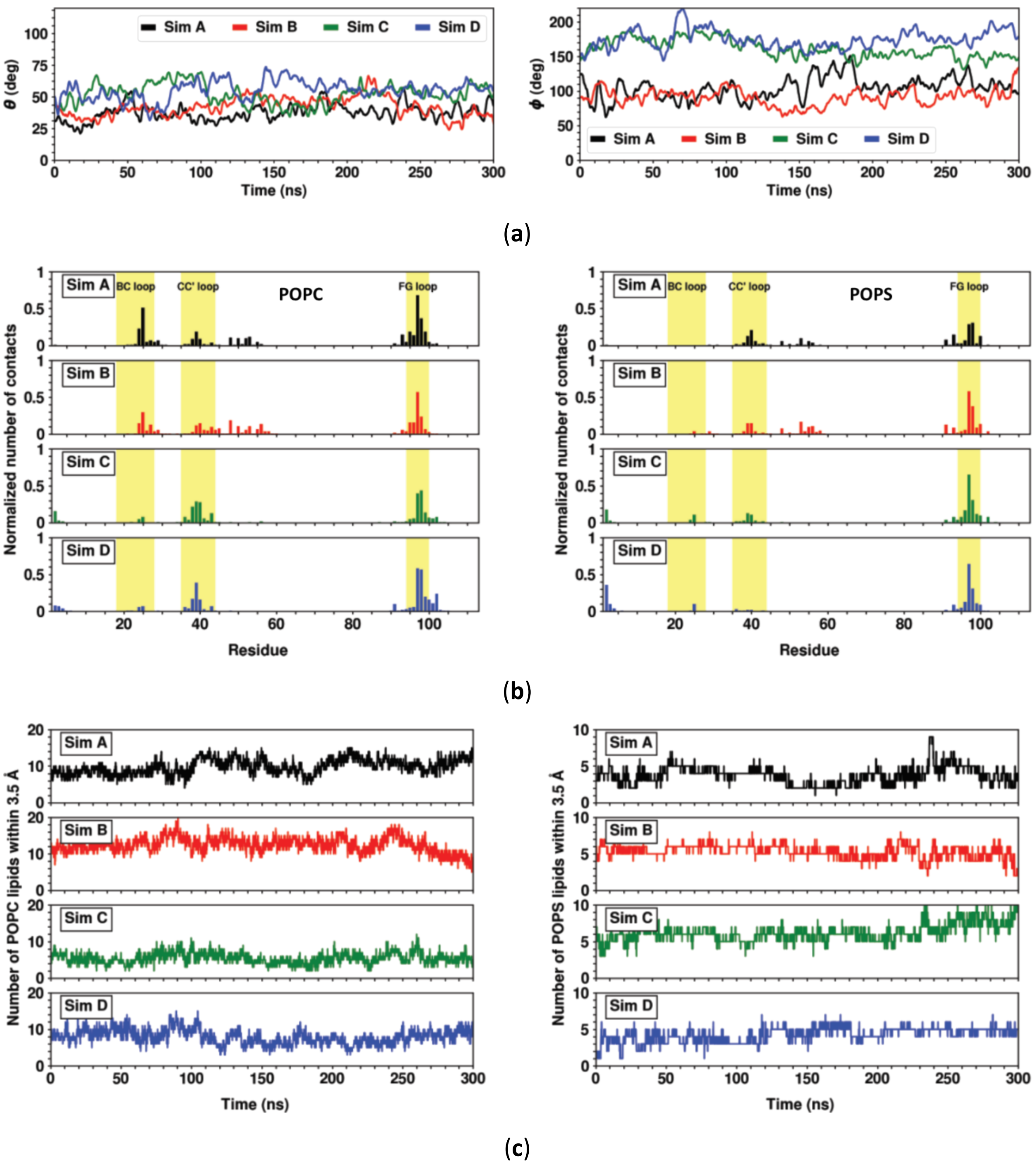


Figure 5.6. Molecular Details of the Two States Obtained from HMMM MD. (**a**) State 1 (Sims A and B) and State 2 (Sims C and D) are distinguished by their orientations. (**b**) Contact map of TIM1 across all four simulations with POPC contacts on the left and POPS contacts on the right. (**c**) Total number of POPC and POPS contacts with TIM1 within a 3.5 Å interaction distance. Figure adapted from ref. (4).

Figure 5.7 quantifies the differences between the states based on the residue positions. State 1 (simulations A and B) and State 2 (simulations C and D) position the tryptophan, W97, on the FG loop at nearly the same insertion, showing that the two states have the same hydrophobic anchor. However, the bound calcium ion sits at the level of the phosphates for State 1, while it occupies the head group region for State 2. Since State 1 coordinates both the serine and phosphate groups of the bound PS, State 1 features a slightly pulled out PS while State 2 represents a more passive, adsorptive state.

The peripheral basic residues, R25, R53, and K102 all associate with the membrane in State 1, while only R25 and K102 do in State 2. This led us to hypothesize that perhaps State 2 was a low PS binding mode that is populated when TIM1 is associated with regions that are low in PS. However, the likelihood of this hypothesis is low given that the number of PS under the protein's footprint is nearly the same for the two states (**Figure 5.6c**) and that evidently, these residues interact more with PC than PS (**Figure 5.6b**). Regardless, details of the simulation may be responsible for these discrepancies and it is possible that State 2 has a low, but finite probability of being occupied with high PS content but a significantly higher probability when PS is low.

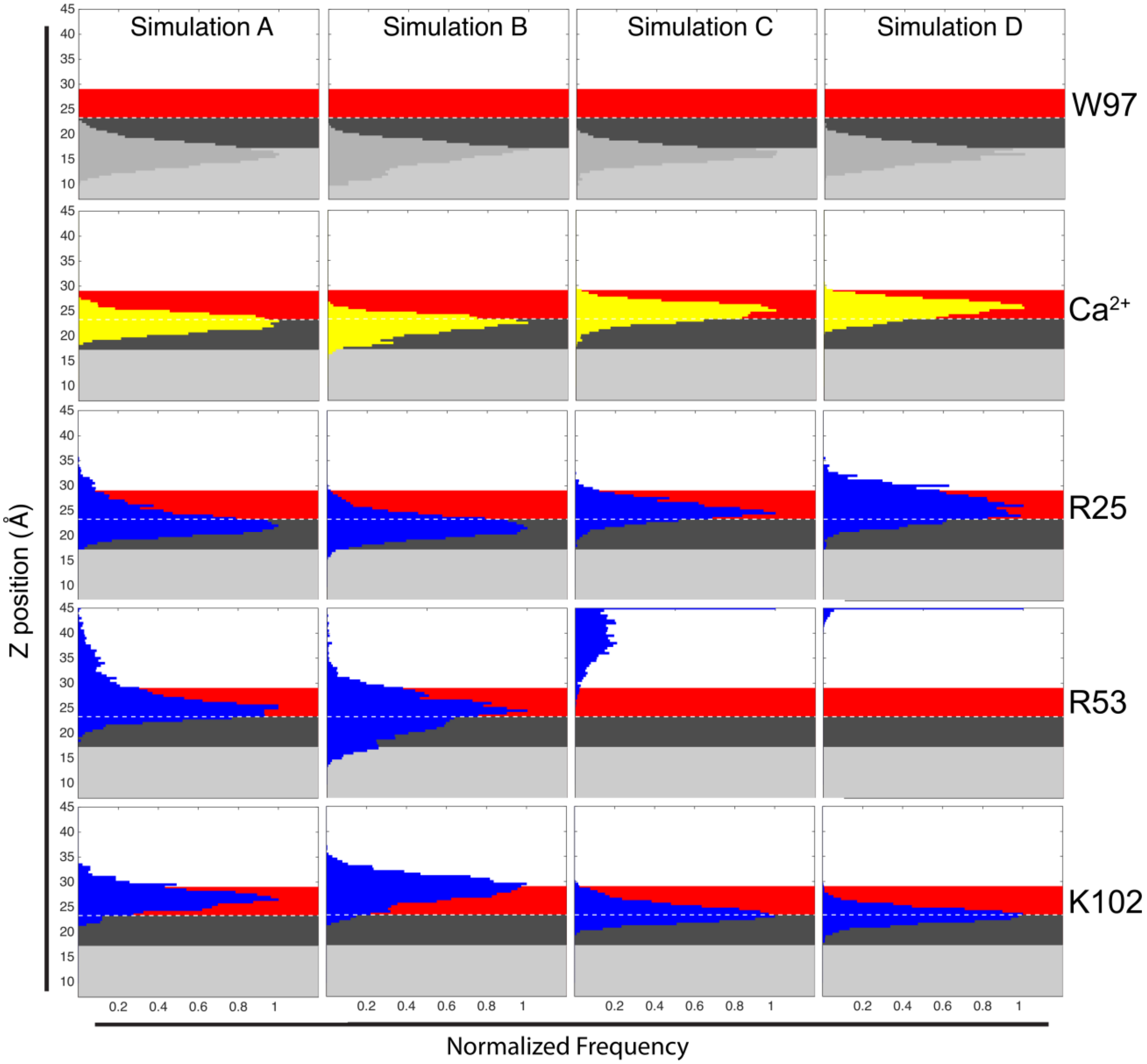


Figure 5.7. Distribution of Residues of TIM1 Relative to Membrane Features. Simulations A and B comprise State 1 while simulations C and D comprise State 2. The tryptophan at the edge of the FG loop, W97, and bound calcium ion are tracked to determine insertion. Three basic residues, R25, R53, and K102, are tracked to determine peripheral membrane association. The regions of the membrane are colored as follows: *Red* is primarily the head group region, *dark gray* is dominated by phosphate groups, *light gray* marks the lipid tails and the *white dashed line* indicates the transition between the phosphate-dominated region and PS carboxyl dominated region.

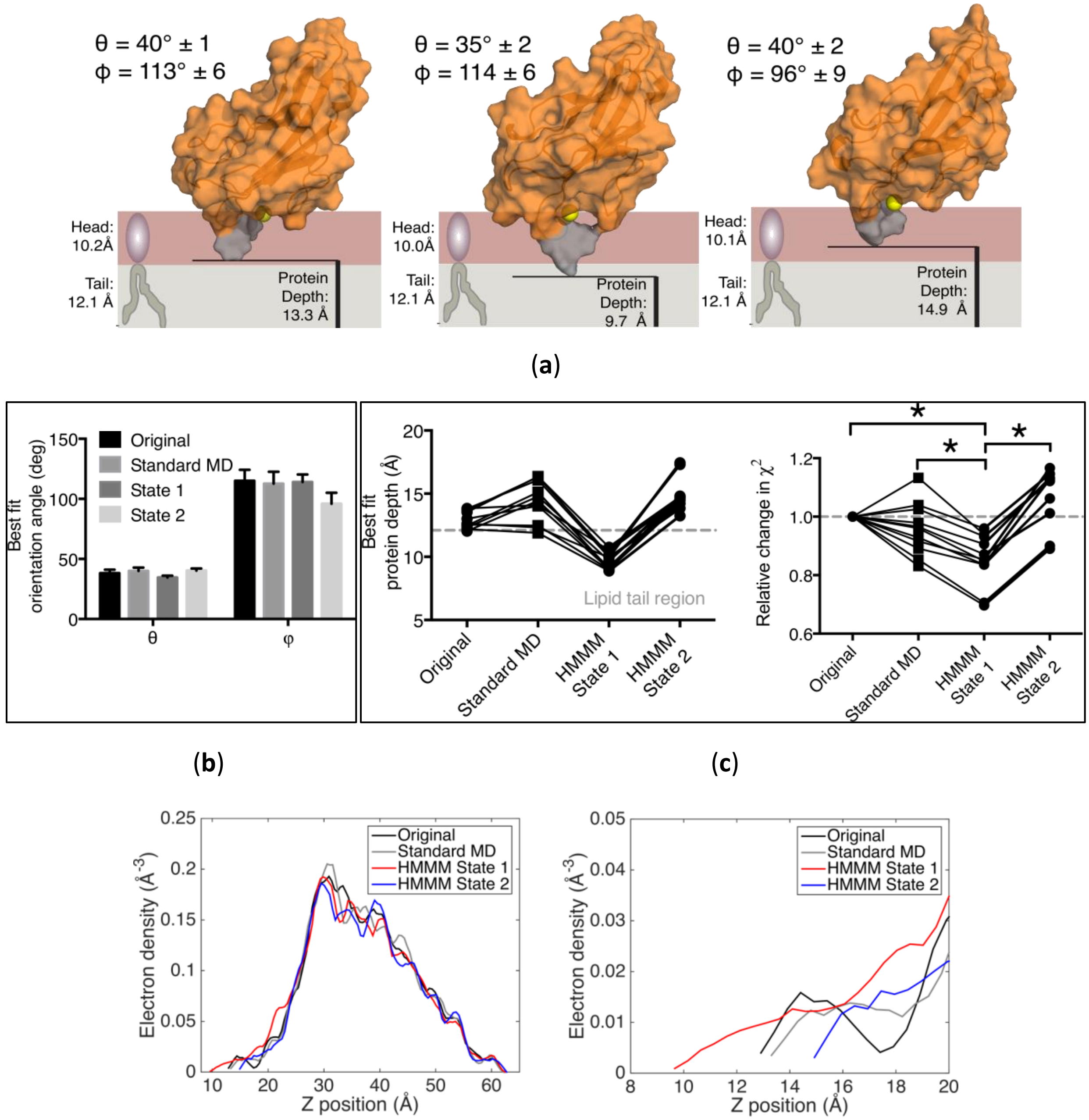
5.5 X-ray Reflectivity Fits by MD-Derived Structures

5.5.1 Introduction

With three candidate membrane-bound states obtained from MD, we sought to use the newly leveraged integration of MD with reflectivity, described in Chapter 4, to distinguish which of these models presented better fits. However, at the time that this work was published, not all of the refinements described in Chapter 4 were introduced. Namely, the “encapsulating area” and bootstrapping methods were not implemented. Following the review of the published data, the fit of the best MD-derived structure with these refinements will be discussed in section 5.6.

5.5.2 Reflectivity Can Distinguish Different Membrane-Bound Models

We fit the same data of **Figure 5.3** to the structure obtained from the docked simulations of **Figure 5.4** and the structures of the two states obtained from HMMM simulations in **Figure 5.5**. The dispositions of these structures are shown in **Figure 5.8a**. The best-fit structures for all four models (including the crystal structure) select for the same orientation, as demonstrated in **Figure 5.8b**. The largest difference is due to the insertion depth of the relative models (**Figure 5.8c**), with HMMM State 1 achieving the greatest insertion and is the only model which places the hydrophobic residues in the tail group region. Evaluating the models based on their goodness of fit, χ^2, HMMM State 1 is also the most consistent with the data. The preference for a given model seems complex to parse, but in this case **Figure 5.8d** demonstrates that HMMM State 1 has the most extended electron density profile at the monolayer interface. This is due to the full opening of its hydrophobic residues on the FG loop.



(**a**)

(**b**) (**c**)

(**d**)

Figure 5.8. Reflectivity Fits of MD-Derived TIM1 Structures. (**a**) The orientation and insertion of the structures, from left to right: Docked MD, HMMM State 1, and HMMM State 2. (**b**) Comparison of the fit angle values for all four fit structures. (**c**) Left panel: Spread of penetration depths for the four models over all of the data sets. Right panel: A comparison of goodness of fit of the four models across all data sets. Each dataset is normalized to the χ^2 of the crystal structure (original). This plot shows the improvement of fit due to a change in model for each data set. (**d**) Best-fit electron density profiles for the four protein models (in the absence of membrane or buffer). The right panel zooms in on the edge of the protein electron density that would penetrate into the monolayer. Figure adapted from ref. (4).

5.5.3 Summary

Using the reflectivity data as a criterion, we were able to select between multiple models yielded from MD with unambiguous determination of HMMM State 1. This state conforms to our physical intuition regarding the insertion of hydrophobic residues and statistically provides a better fit to the data. This result proves that x-ray reflectivity is a powerful tool for comparing many proposed models. TIM1 underwent a minor rearrangement (RMSD ~4 Å) from its crystal structure to its bound state, yet reflectivity could clearly distinguish between them. Previously, x-ray reflectivity was used to distinguish proposed models of membrane-bound proteins on the basis of orientation (7, 8), but our work here expands the scope of this technique to evaluate models of proteins on the basis of conformational changes.

That HMMM MD resolved two different orientations was a surprising finding. It is tempting to believe that both structures may represent different binding modes. However, HMMM State 2 categorically did not fit the data well, performing worse than the crystal structure (**Figure 5.8c**). Moreover, the orientation of HMMM State 2 determined from reflectivity did not match its orientation in the simulations (**Figures 5.6a** and **5.8b**). We went to great lengths to determine the reality of this state (work not shown), including a reformulation of the reflectivity analysis methods of Chapter 4 to include two different protein states. These efforts were fruitless as the fits always returned a better fit with HMMM State 1 alone than combined in any way with HMMM State 2.

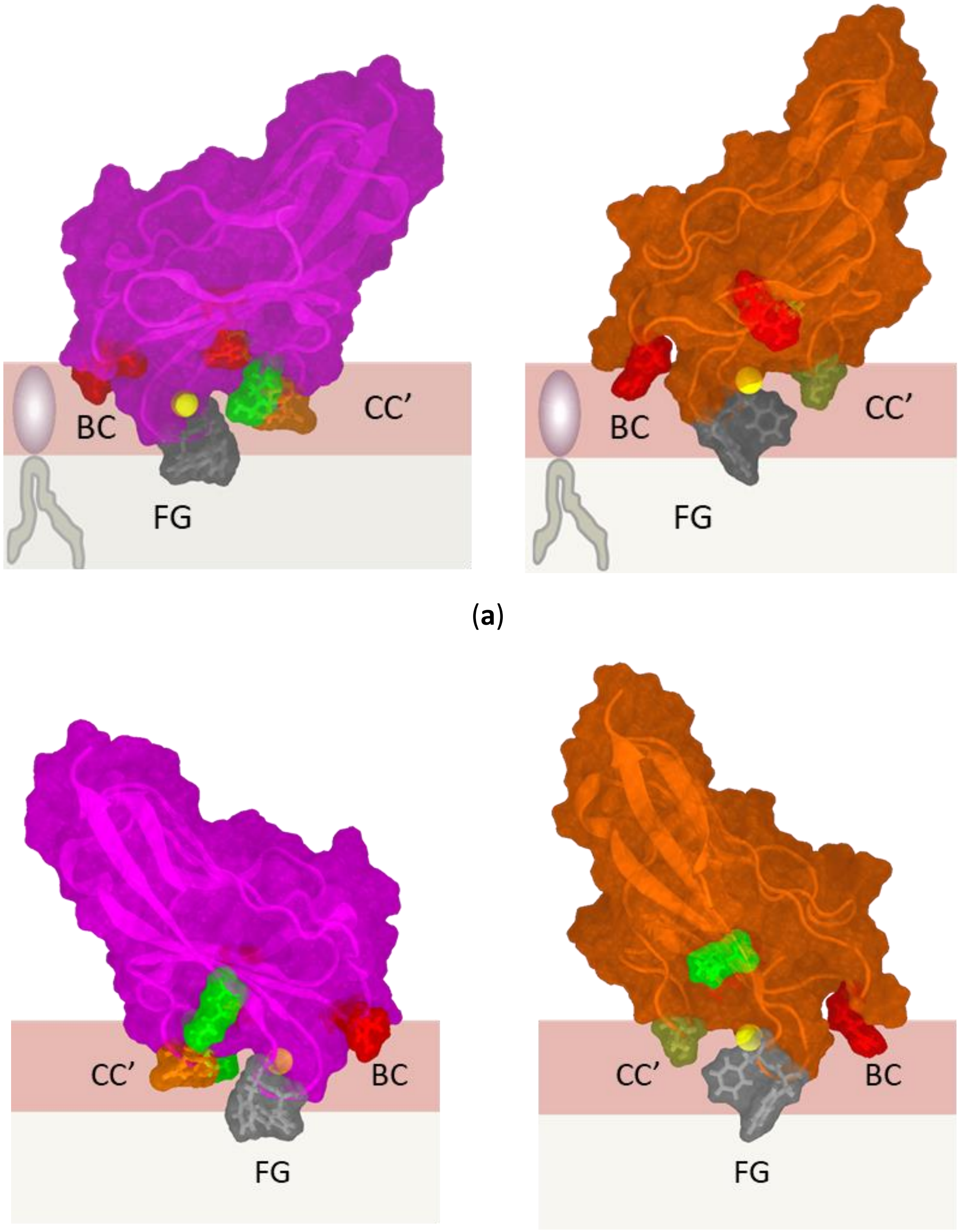
Instead, we believe that HMMM State 2 represents an infrequently populated bound state at equilibrium. The membrane-bound state is an ensemble of binding modes in which protein residues make a variety of different kinds of contacts with membrane lipids. Since reflectivity is sensitive to only the average electron density, we can only view an ensemble average of the membrane-bound state. Therefore, HMMM State 1 is interpreted as a better model for the ensemble average, while HMMM State 2 might be a microstate in this ensemble that is unrepresentative. Indeed, the MD evidence presented here suggests that HMMM State 2 is a transition state that is not fully equilibrated. The hydrophobic residues are closed and withdrawn compared to even the crystal structure (**Figure 5.8d**). Its calcium ion does not penetrate past the head groups to the phosphate layer, in stark contrast to HMMM State 1 and the docked MD which bury the calcium ion in the middle of these two layers (**Figures 5.4c** and **5.7**). It seems that HMMM State 2 is anchored only by its tryptophan (**Figure 5.7**) and peripheral lysine. It is likely that HMMM State 2 is in the midst of molecular rearrangements on a time scale longer than the simulation time, appearing as an equilibrium state. These two states indicate that MD is used most effectively with reflectivity if it tries to model the ensemble average, and not one particle state that is populated at equilibrium.

Another surprising finding in the MD was that the peripheral loops of TIM1 preferentially contact POPC lipids as opposed to POPS lipids. In particular the BC loop, which contains a peripheral arginine, almost exclusively contacts POPC and not POPS (**Figure 5.6b**). TIM1 was shown to be relatively insensitive to membrane PS (4) and this result offers a plausible explanation. Perhaps future NMR experiments could resolve if this is real and not an artifact of simulation or sampling bias.

Lastly, the findings here validate the use of HMMM as a model both to fit reflectivity and obtain *in silico* membrane-bound states. That HMMM produced a model which significantly outperformed alternative models speaks to the potential of this approach. Furthermore, the use of HMMM MD over docked MD provides another element of corroboration, as the orientations obtained in MD are independent of the reflectivity parameters and conversely the reflectivity is analyzed with this information discarded from the MD structures. Convergence of both techniques therefore further strengthen the confidence of the obtained structures.

5.6 Conclusions: Comparison of TIM1 with TIM4

As stated previously, the TIM1 analysis provided here was conducted with analysis techniques that were not yet fully developed. Specifically, the “area encapsulating” and bootstrapping components of the analysis were not utilized. In order to compare the TIM proteins on equal footing, we reanalyzed the structure of TIM1 using the aforementioned refinements. **Figure 5.9** compares the structures of TIM4 and TIM1 both obtained from this complete analysis.



(**a**)

(**b**)

Figure 5.9 Reanalyzed Membrane-Bound Structures of TIM4 and TIM1. The full set of refinements described in Chapter 4 were applied to both TIM4 (left panels) and TIM1 (right panels) to yield comparable structures. Two 180°-rotated views of the proteins are presented in (**a**) and (**b**) to emphasize the engagement of various residues. Peripheral lysine residues are shown in green, arginine in red, hydrophobic residues in silver, and polar residues in gold.

While TIM1 and TIM4 present similar contact surfaces, the penetration of the two proteins is significantly different. The bound calcium ion is deeply embedded in the head group region for TIM4 but is barely immersed for TIM1. In general, TIM1 buries significantly less into the membrane. This might be due to the presence of two lysine residues, one on the CC’ loop and one flanking the FG loop, on TIM4 that encourage the engagement of its CC’ loop (which contains polar residues as well). TIM1 contains only the lysine residue which flanks the FG loop, and thus does not strongly promote insertion of the CC’ loop. The BC loops of TIM4 and TIM1 both bury to similar insertion depths, however the MD simulations of TIM1 implicated the interactions of this loop to be primarily with PC rather than PS.

Taken together, these structures imply that the peripheral, basic residues of TIM1 associate less weakly with the membrane than those of TIM4. Furthermore, TIM1 presents the same hydrophobic surface, achieving similar insertion of these residues but with reduced total surface burial. The increased surface burial of TIM4 may be a consequence of the engagement of its peripheral residues. The energy gained from those Coulombic interactions should ostensibly offset the energy barrier of displacing lipids in the membrane to accommodate these associations. In conclusion, these models provide a molecular rationale for their differential binding; the number and degree of penetration of peripheral, basic residues seem to correlate with the cooperativity of these proteins for PS.

5.7 References

1. Tietjen GT, et al. (2014) Molecular mechanism for differential recognition of membrane phosphatidylserine by the immune regulatory receptor Tim4. *Proc Natl Acad Sci U S A* 111(15):E1463–72.
2. Santiago C, et al. (2007) Structures of T cell immunoglobulin mucin protein 4 show a metal-Ion-dependent ligand binding site where phosphatidylserine binds. *Immunity* 27(6):941–951.
3. Santiago C, et al. (2007) Structures of T Cell immunoglobulin mucin receptors 1 and 2 reveal mechanisms for regulation of immune responses by the TIM receptor family. *Immunity* 26(3):299–310.
4. Tietjen GT, et al. (2017) Coupling X-Ray Reflectivity and In Silico Binding to Yield Dynamics of Membrane Recognition by Tim1. *Biophys J* 113(7):1505–1519.
5. Ohkubo YZ, Pogorelov TV, Arcario MJ, Christensen GA, Tajkhorshid E (2012) Accelerating membrane insertion of peripheral proteins with a novel membrane mimetic model. *Biophys J* 102(9):2130–2139.
6. Qi Y, et al. (2015) CHARMM-GUI HMMM Builder for Membrane Simulations with the Highly Mobile Membrane-Mimetic Model. *Biophys J* 109(10):2012–2022.
7. Chen C-H, et al. Supporting Material' ' Configuration of PKC -C2 Domain Bound to Mixed SOPC/SOPS Lipid Monolayer.
8. Málková S, et al. (2005) X-ray reflectivity studies of cPLA2{alpha}-C2 domains adsorbed onto Langmuir monolayers of SOPC. *Biophys J* 89(3):1861–1873.

CHAPTER 6

THE MEMBRANE-BOUND STRUCTURE OF TIM3

6.1 Introduction

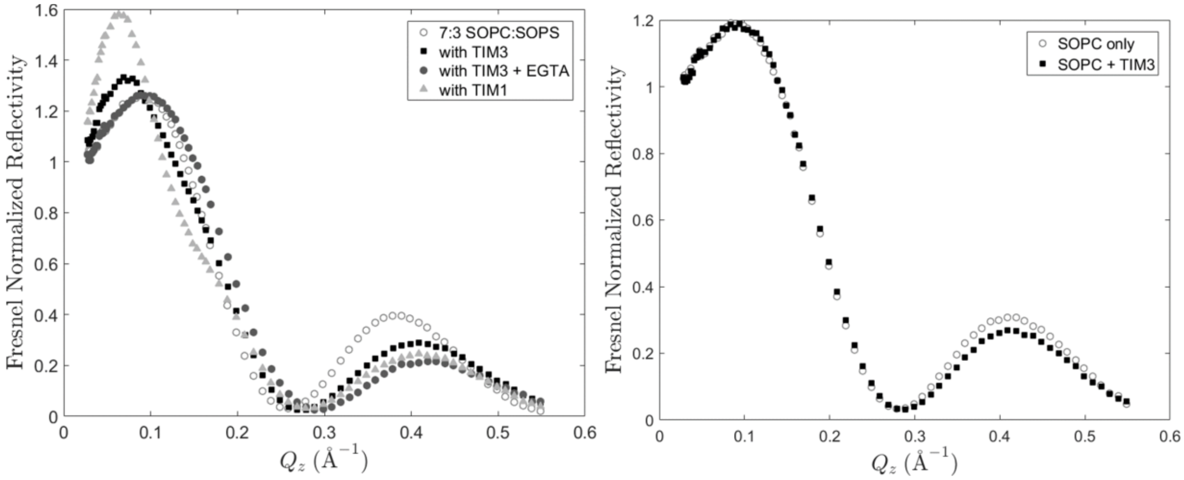
While the analysis of TIM1 extended our understanding of the kind of structure x-ray reflectivity probes, TIM3 presented an entirely different problem. TIM3 has a much lower affinity for PS than TIM1 or TIM4 [1, 2], which resulted in saturation of protein with poor coverage. Consequently, the reflectivity signal obtained for TIM3-bound films indicated a weak presence of the protein. It has previously been demonstrated that the resolution of a membrane-bound structure is highly dependent on the protein coverage [3]. Low signal conditions place the onus on the accuracy of the model to resolve the underlying structure. While TIM3 was co-crystallized with a short-chain PS, the structure evidently was insufficient to fit the data perhaps due to a solution state not adequately representing a membrane-bound state. We therefore had to leverage the MD apparatus we developed for TIM1 but with no clear, guiding intuition for what a "better" membrane-bound structure would look like.

This chapter describes a multi-pronged approach to resolve the structure of TIM3. First, section 6.2 covers the fitting of the TIM3 crystal structure to reflectivity data, which motivated our utilization of joint fitting and bootstrapping methods described in Chapter 4. Section 6.3 presents the results of HMMM MD in determining a membrane-bound structure and reviewing its fit to the reflectivity data. In addition to a direct computational approach, we also opted for an experimental solution that increased resolution. Section 6.4 describes reflectivity experiments conducted under higher binding but less physiologically relevant conditions. Corroboration with MD shows that the low ionic membrane-bound state is nearly identical under normal ionic strength conditions. The final structure is reviewed in section 6.5, leveraging the insights gained from the computational and experimental approaches. This obtained structure makes a number of predictions, one of which seemingly contradicts previous studies with an allelic variant [2]. Section 6.6 analyzes the membrane-bound structure of this allelic variant using MD and offers a resolution to the apparent inconsistency. The myriad results are consolidated in section 6.7.

6.2 The Orientation of the TIM3 Crystal Structure is not Resolved

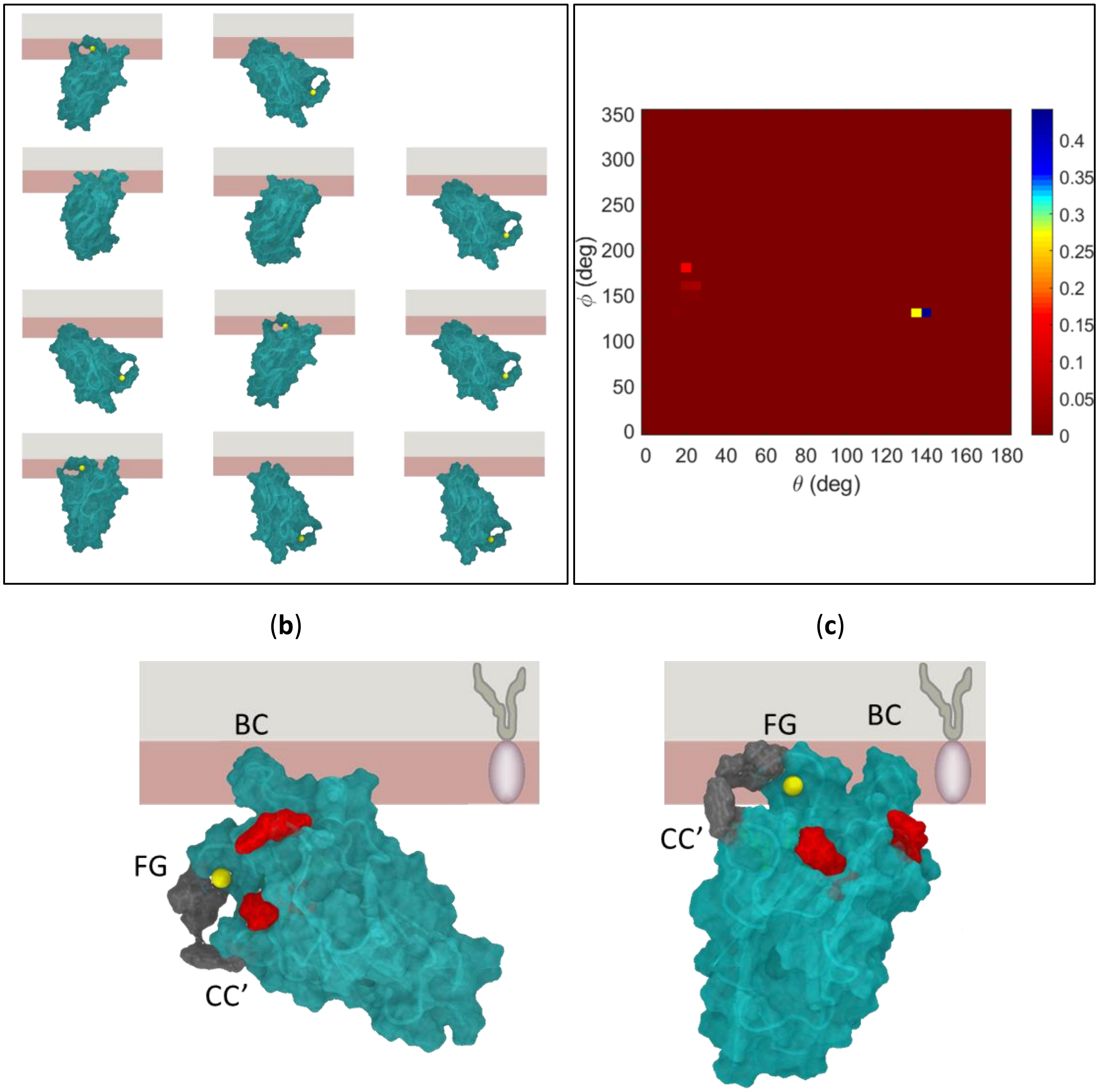
The low affinity of TIM3 resulted in a low bound signal in reflectivity conducted on 7:3 SOPC:SOPS films with 1 mM $CaCl_2$ and 150 mM NaCl in the buffer. **Figure 6.1a** compares a representative TIM3 reflectivity signal to that of TIM1, which has higher affinity for PS. The peak shift due to TIM3 is far less pronounced than it is for TIM1, barely distinguishable from lipid alone. Experiments with EGTA and SOPC confirm that this small adsorbed layer is indeed TIM3 as both reduced the peak shift. Increasing the protein concentration of the subphase corresponded to no increase in signal beyond 1.5 μM and thus our experiment was conducted at a saturating protein concentration. Application of the same analysis that was applied to TIM4 and TIM1 resulted in a poor resolution of the bound orientation of the TIM3 crystal structure. Each dataset returned a different orientation as shown for a subset of the data in **Figure 6.1b**. The data represent six independently prepared monolayers at a surface pressure of 25 mN/m and a total of seventeen scans. While three of these data sets resolved sensible orientations, there was no experimental basis to exclude the remaining data sets from the analysis.

We employed a joint fit method to resolve a single orientation across all reflectivity data sets. The error of this joint fitting method was characterized using bootstrapping, which is shown in **Figure 6.1c**. From this, two orientations were identified at 75% and 25% confidence level (**Figure 6.1d**). The former corresponds to an orientation in which the pocket is not engaged in the membrane while the latter places the pocket in the head group region along with the FG, CC', and BC loops which inserted in the corresponding TIM4 and TIM1 structures. This indicates that we can potentially recover a signal buried in the noise of both the model and the data to obtain a sensible orientation.



(**a**)

Figure 6.1. The TIM3 Crystal Structure Fails to Resolve a Membrane-Bound Orientation. (**a**) Left: Representative reflectivity curves of 7:3 SOPC:SOPS monolayers without TIM3, with TIM3, with TIM3 and EGTA, or with TIM1. Right: Reflectivity of SOPC monolayers with or without TIM3.



(**b**) (**c**)

(**d**)

Figure 6.1 cont. (**b**) Analysis of individual datasets resulted in a collage of best-fit orientations. The pink represents the head group region of the monolayer and gray represents the tail group region. (**c**) Joint fitting of all data sets converged on an orientation in which the pocket is not buried in the membrane. Each pixel represents the confidence level of a given orientation of best fitting the data. The corresponding orientation parameters are $\theta = 130 \pm 5°$, $\phi = 140 \pm 5°$, and $d_p = 7.6 \pm 1$ Å with 75% confidence. Error characterization using bootstrapping revealed a secondary model at $\theta = 20 \pm 10°$, $\phi = 180 \pm 20°$, and $d_p = 8.4 \pm 0.5$ Å with 25% confidence. (**d**) The orientations identified in panel **c**. The FG, CC’, and BC loops are all shown to insert for TIM4 and TIM1 and are indicated in the figure. The hydrophobic residues are in gray and peripheral arginine residues are in red. Left: The 75% confidence orientation. Right: The 25% confidence orientation.

We therefore pursued two independent strategies to reveal the bound structure. First, we used HMMM MD to resolve a better membrane-bound structure as the lipid-bound crystal structure seems to be insufficient for this low signal quality. It is worth noting that several residues of TIM3 were not resolved in the crystal structure, so we added these residues into the structure prior to our simulations. Second, we boosted the signal by increasing protein binding. We accomplished this by reducing the buffer ionic strength. These approaches are described in sections 6.3 and 6.4, respectively.

6.3 A HMMM MD-Derived Structure Resolves a Consistent Orientation

6.3.1 MD Characterization of Membrane-Bound State of TIM3

As described in Chapter 3, we employed HMMM MD to obtain a membrane-bound structure especially since we could not resolve one from reflectivity alone. We initiated the protein in solution 10-15 Å away from the HMMM membrane in two different orientations. We ran ten replicate simulations, among which four bound PS in the membrane. In Chapter 5, we described how the HMMM-equilibrated structures were used to initiate a full-length lipid membrane simulation. Specifically, we used the last frame of the HMMM simulations as the initial condition. However, based on our results with TIM1 in which we identified two possible states, we realized that the HMMM portion of the simulation could accelerate transitions between these states. We therefore ran each of the four trials for 220 ns each. We were able to classify five distinct states, some of which occurred across multiple independent simulations, based on the coordination of

PS with the central pocket. This resulted in eight full-length lipid simulations, which were each run for 220 ns. **Figure 6.2** shows representative structures of these states. The first state depicts a coordination of the phosphate group of the PS with the calcium ion and the serine group with a peripheral arginine, R54. The second state depicts a coordination of the serine group of the PS with the bound calcium and coordination of the phosphate by a polar residue, Q95. The third state depicts the converse coordination, with the phosphate and serine groups switching places. We also observed two states in which two lipids were coordinated in the pocket. The first of these depicts a coordination of two PS and the second shows a coordination of one PS and one PC.

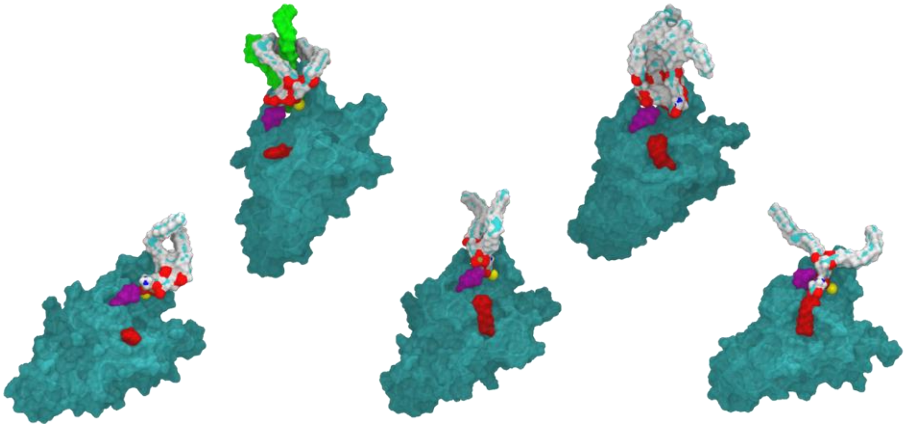
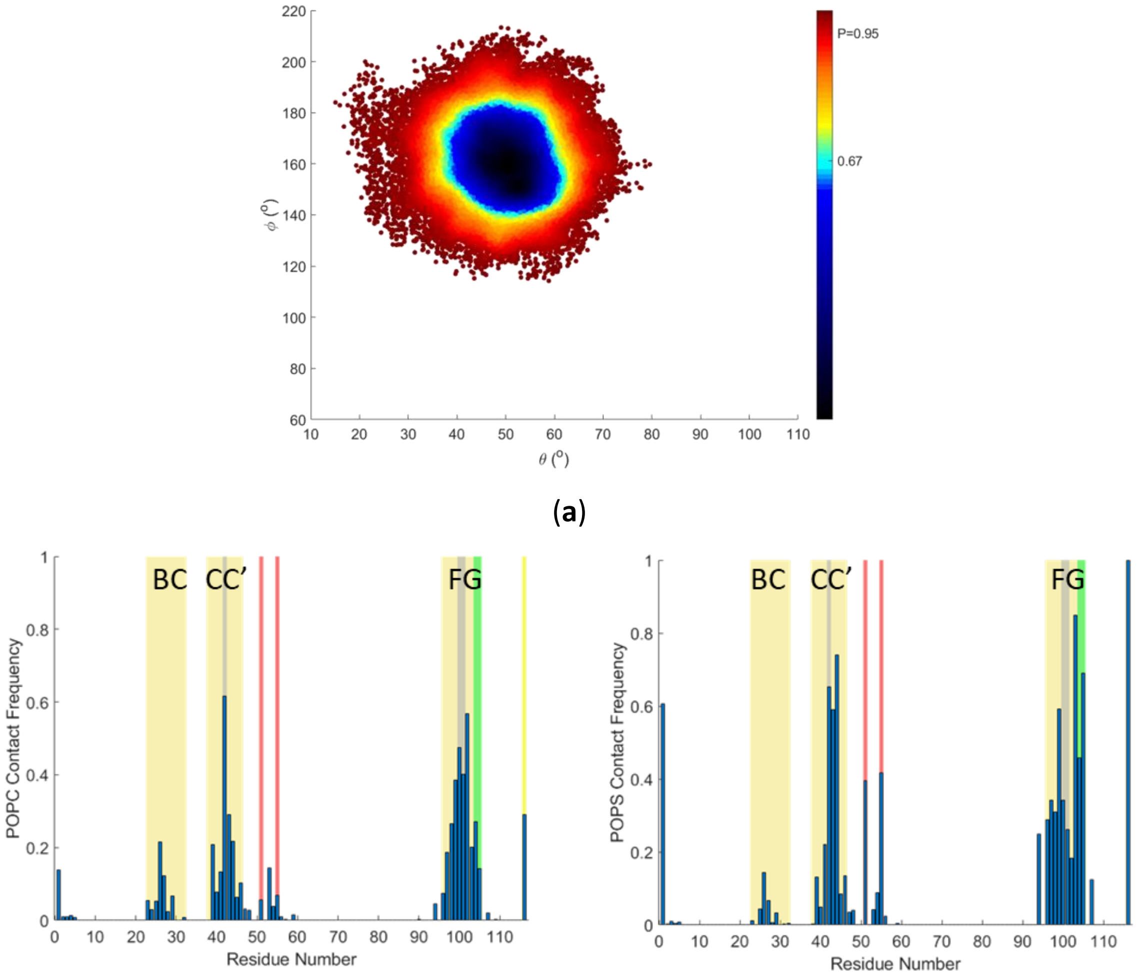


Figure 6.2. HMMM MD Resolved Multiple Membrane-Bound States of TIM3. Several distinct states were identified in the HMMM portion of the simulations on the basis of their coordination of bound lipids. The bottom row depicts states bound to a single lipid and the top row depicts states bound to multiple lipids. POPS is colored in white with red and blue portions denoting negative and positive charges, respectively, while POPC is colored strictly in green. Two residues that distinguished these binding modes are highlighted on TIM3, R54 in red and Q95 in purple.

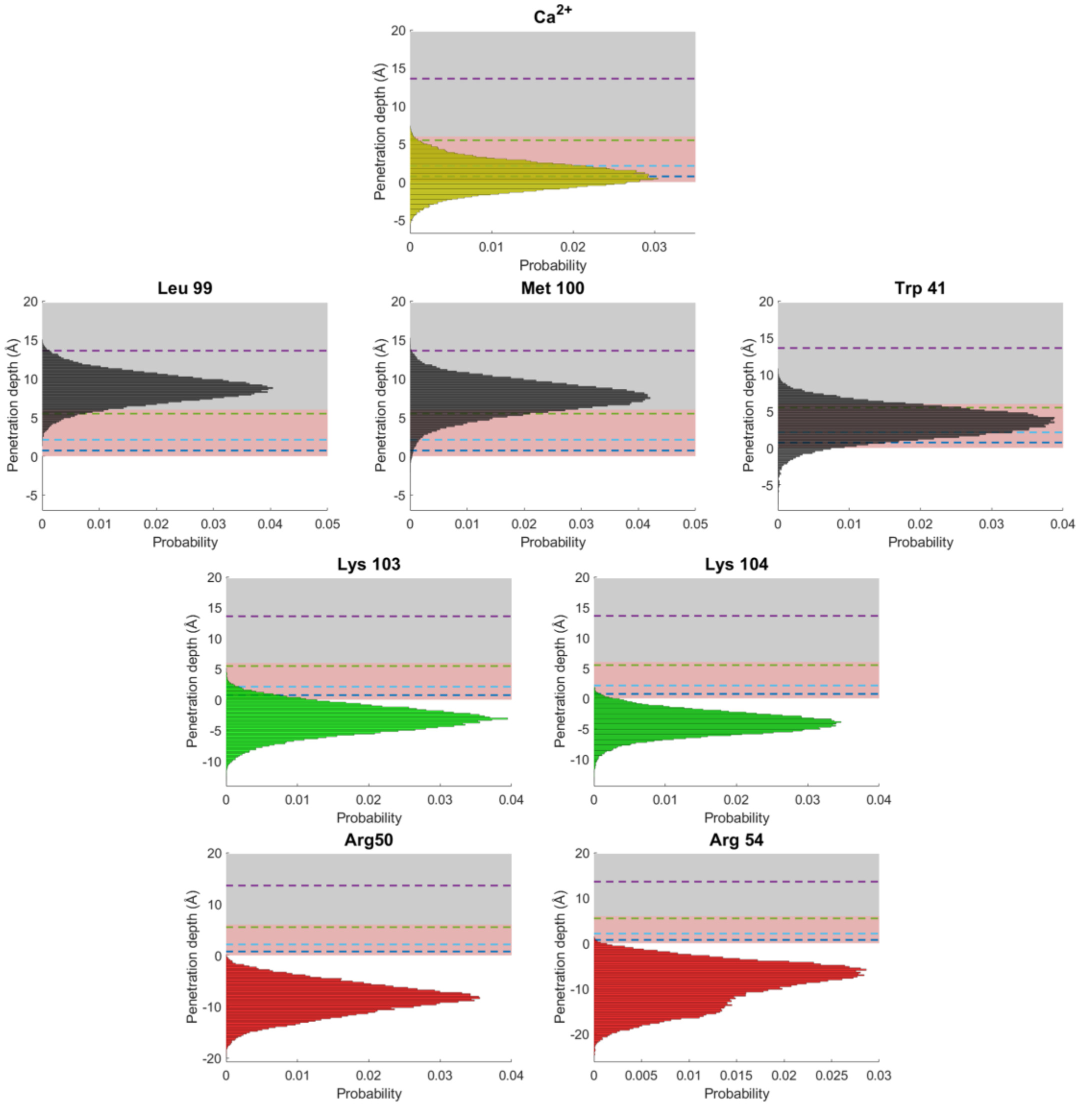
From our experience with TIM1, we knew that not all of these states would best represent an ensemble average of the bound state. We therefore decided to analyze all of the simulations simultaneously, treating them each as a subsample of the broader ensemble of bound states. Indeed, the orientation of TIM3 over all of the simulations approximately forms a Gaussian distribution, as seen in **Figure 6.3a**, indicating the collection of states form a single ensemble. Across these simulations, the calcium ion is in constant contact with POPS, as seen in **Figure 6.3b**. Unlike TIM1, the FG and CC' loops display a clear preference for POPS over POPC. Four peripheral residues homologous to those identified to be instrumental in the binding of TIM4 also showed preferential interactions with POPS (highlighted red and green in **Figure 6.3b**). In contrast, the hydrophobic residues seem to have no preference for POPS, consistent with the fact that they mainly associate with the lipid tails (highlighted gray in **Figure 6.3b**).



(**a**)

(**b**)

Figure 6.3. The Membrane-Bound State of TIM3 is a Single Ensemble. The eight full-length lipid membrane simulations were combined for analysis.



(c)

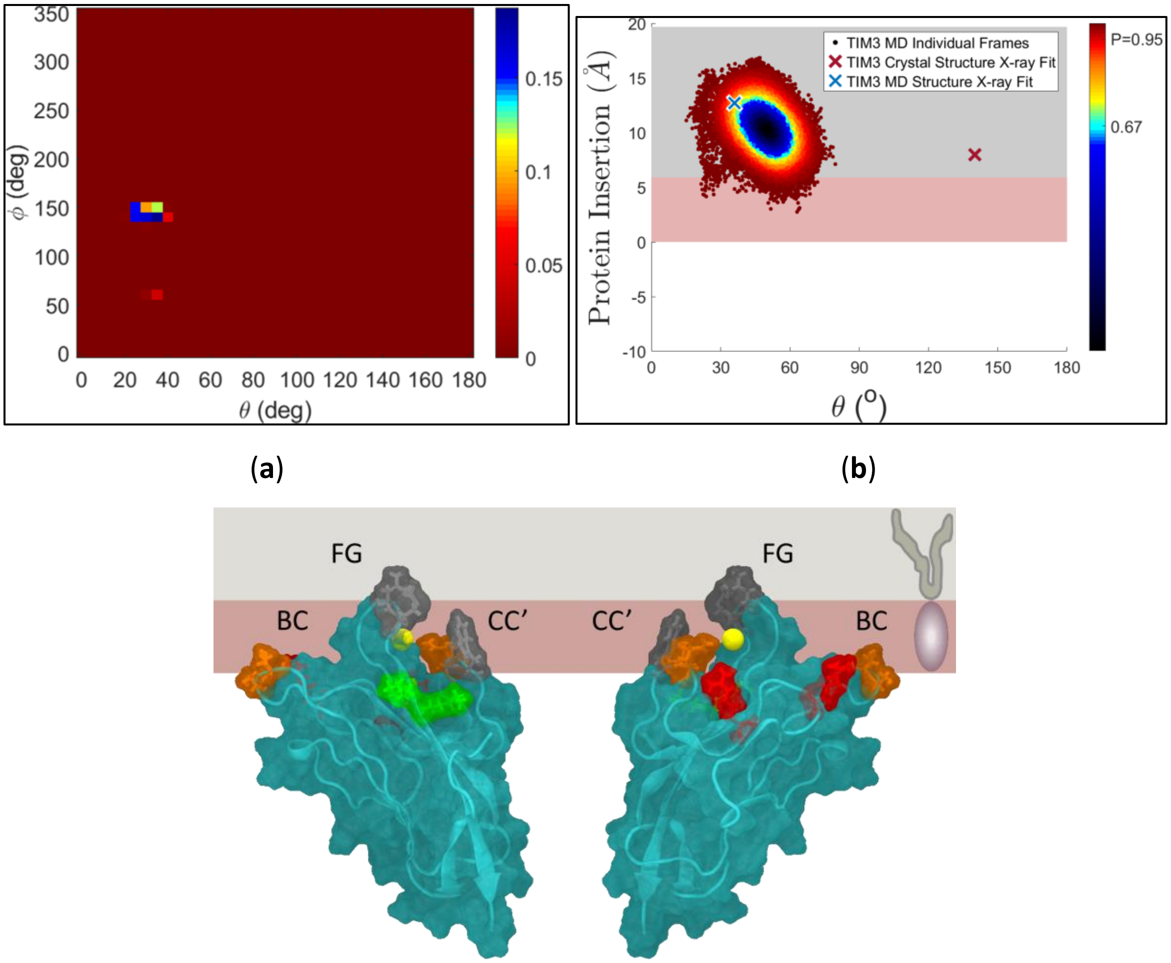
Figure 6.3. cont. (**a**) Scatter plot of observed orientations of TIM3 over the eight simulations. Each point is colored according to its cumulative probability density, measured as a series of contours representing equal density, as calculated by a kernel smoothing function. For example, points contained within a cyan contour represent 67% of the total points in the plot. (**b**) Contact maps of the residues of TIM3 with either POPC (left panel) or POPS (right panel). The y-axis represents the frequency for which a given residue contacts a lipid over the course of the simulations. A value of 1 indicates 100% engagement of that residue. The residues belong to BC, CC’, and FG loops are highlighted in yellow, the hydrophobic residues that facilitate insertion are highlighted in gray, two peripheral arginine residues homologous to those shown to facilitate binding for TIM4 are highlighted in red, two peripheral lysine residues which are also homologous to those in TIM4 are highlighted in green, and the rightmost column represents the contacts of the calcium ion, highlighted in gold for the POPC panel. This calcium makes contact with POPS for the entire set of simulations.

Figure 6.3. cont. (**c**) Distribution of insertion of the highlighted peripheral and hydrophobic residues, along with the bound calcium, in panel **b**. The electron density of the entire system was computed as to fit a two-slab model (directly to the electron density) to visualize this insertion. The gray box corresponds to the tail group region and the pink the head group region. The dashed lines indicate several membrane features: The average head group location in blue, the average phosphate location in cyan, the average carboxyl linker in green, and the average center of the tails in purple.

We also analyzed the insertion depths of the calcium ion and several key residues in **Figure 6.3c**. Interestingly, the mean positions of the calcium ion and the tryptophan of the CC' loop (W41) are rather shallow, with the latter mostly residing in the head group region. On the contrary, the leucine and methionine residues of the FG loop (L99 and M100) bury into the tail group region. Meanwhile, the peripheral arginine and lysine residues homologous to those of TIM4 (R50, R51, K103, and K104) are surprisingly withdrawn from the membrane despite their many lipid contacts (**Figure 6.3b**). Since these figures track the center of mass of these residues, the interaction is mostly adsorptive and not inserting.

6.3.2 Reflectivity Analysis Using the HMMM MD-Derived Membrane-Bound State

After obtaining a representative frame of this ensemble, we fit the same reflectivity data in **Figure 6.1** with the MD-derived structure. One orientation was unequivocally determined to be the best-fit (**Figure 6.4a**). The upside-down orientation did not appear at all in the bootstrapping process, indicating that the MD-derived structure is a much better model than the crystal structure. Furthermore, the χ^2 is reduced by 1000 (with 1287 independent measurement points) for the MD-derived model as compared with the crystal structure. **Figure 6.4b** shows that the best fit orientation of the MD-derived structure closely agrees with the orientation observed in the simulations. The resolved orientation places the pocket squarely in the membrane (**Figure 6.4c**) with the positions of the hydrophobic residues aligning with those of the MD simulations (**Figure 6.3c**). However, this orientation places the peripheral basic residues nearer to the membrane than they were during the simulations. The BC loop is also significantly buried in the head group region while the simulations indicated minimal membrane contacts (**Figure 6.3b**). This reflected in a slight discrepancy of the tilt and insertion from the MD, as shown in **Figure 6.4b**. This disagreement could arise from an artifact of either analysis. Its resolution is contingent on the results of the low ionic strength measurements presented in the next section.



(**a**) (**b**)

(**c**)

Figure 6.4. Fit of MD-Derived TIM3 Structure to Reflectivity. (**a**) Similar to **Figure 6.1c** this plots the level of confidence that a given orientation fits the data, represented by the color of a pixel. > 95% of the probability is clustered around one area, with a minor population slightly below it. (**b**) A comparison of the MD fit to the MD shows correspondence within 95% of the orientation-insertion distribution of the MD. Similar to **Figure 6.3a**, the color corresponds to the probability density of that orientation and insertion during the simulation. The blue and red "X" markers represent the best-fit values of the MD and crystal structures, respectively. The membrane was fit to a two-slab model as in **Figure 6.3c** as to directly compare the insertion (the d_p parameter) of the MD to that of the reflectivity fits. The head group region is in pink and the tail group region is in gray. (**c**) The most likely orientation from panel **a**. Important residues are highlight as follows: Hydrophobic in gray, polar in orange, peripheral arginine in red, and peripheral lysine in green. The corresponding parameters are $\theta = 32°$, $[24, 40]$, $\phi = 145°$, $[135, 150]$, and $d_p = 14.0$ Å, $[13.6, 14.4]$ where brackets contain 95% confidence intervals.

6.4 Low Ionic Strength Buffer Boosts Reflectivity Signal

6.4.1 Low Ionic Strength Buffer Increases Membrane Affinity of TIM3

The pocket residues of the TIM proteins are polar and charged, as well as the various peripheral residues, which suggests that the binding of the TIM proteins would be inhibited by ionic screening. We tested this hypothesis with a series of calcium dependence binding experiments using the tryptophan fluorescence method described in Chapter 3 (and explained in more depth in Chapter 7), and the results are shown in **Figure 6.5a**. We found that the binding of TIM3 was significantly enhanced when the effects of ionic screening were reduced by lowering the monovalent salt concentration of the buffer. We used a buffer with no added NaCl containing only 10 mM Hepes and a small amount of NaOH to bring it to pH 7.4. We estimate that the effective ionic strength of this buffer to be equivalent to 5 μM of a monovalent ion. The reflectivity signal on a 7:3 SOPC:SOPS film on this buffer showed a pronounced peak shift (**Figure 6.5b**) as compared with those obtained for monolayers on subphases with 150 mM NaCl. We obtained twelve scans on six independently prepared monolayers compressed to a surface pressure in the range of 20 mN/m to 30 mN/m.

We analyzed the reflectivity data using both the crystal structure and MD-derived structures. As shown in **Figure 6.5c**, both structures resolved the same orientation. **Figure 6.5d** shows the obtained orientation for the MD-derived structure, which had an improved χ^2 by nearly 500 (for 1224 independent measurement points).

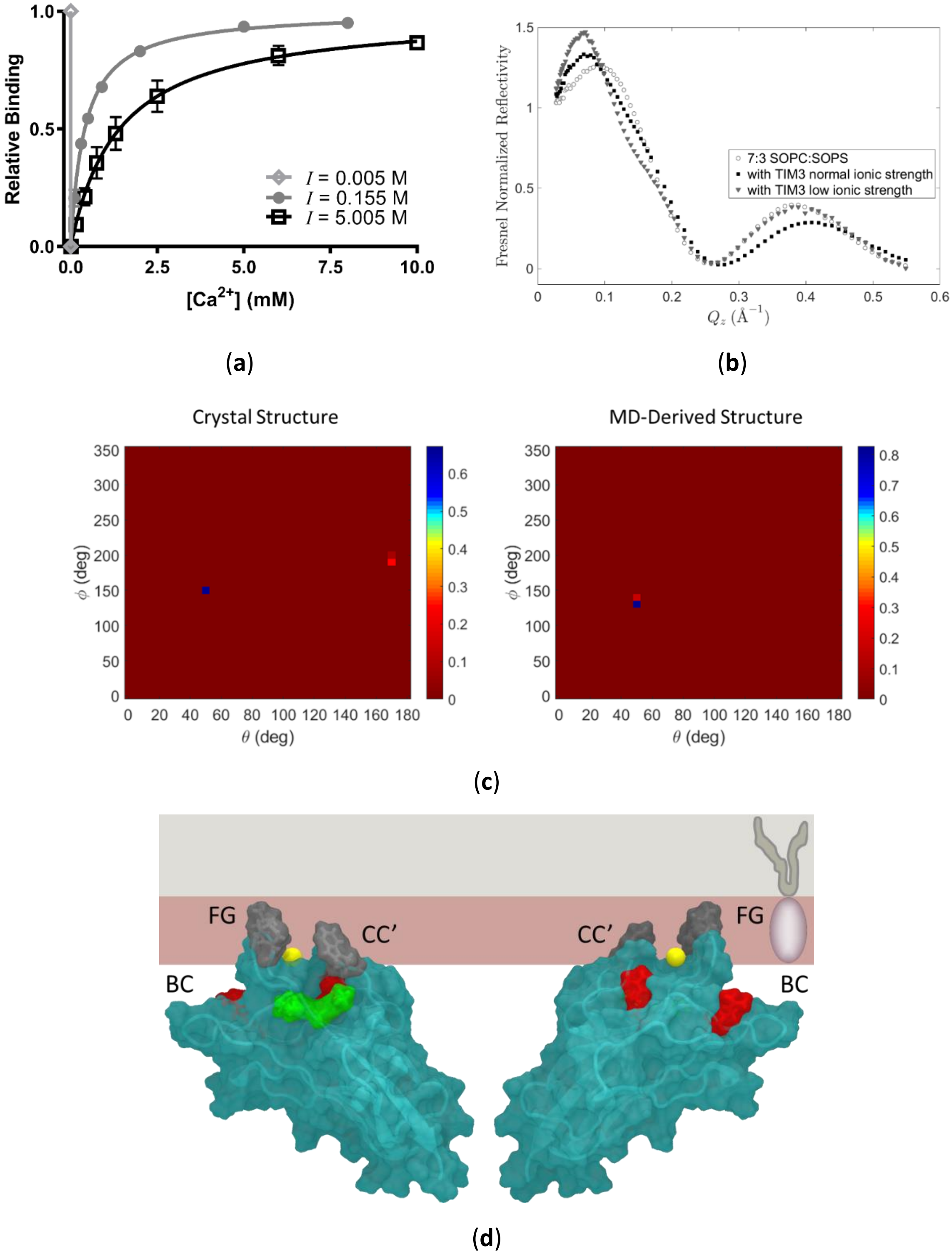
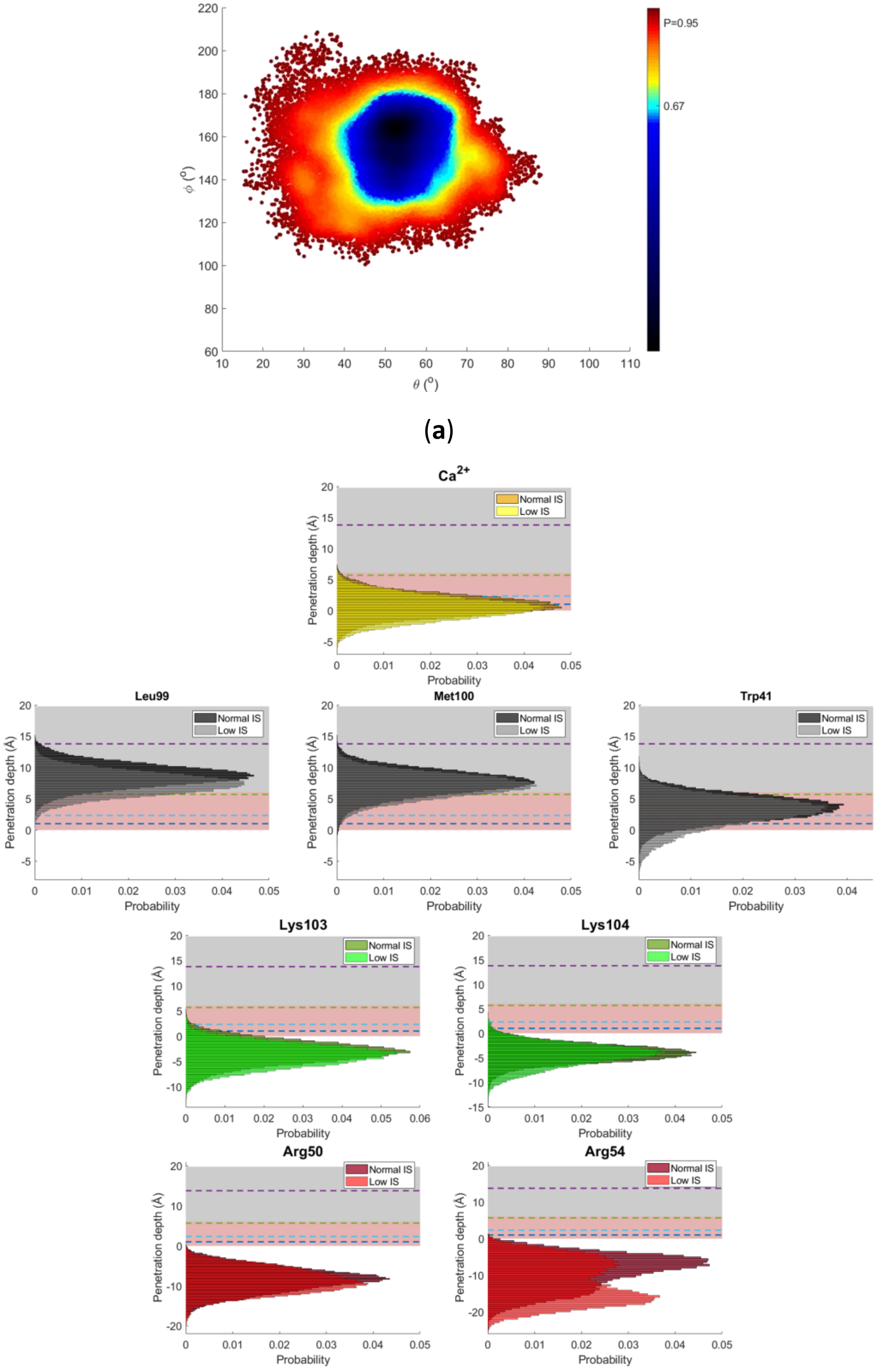


Figure 6.5. Low Ionic Strength Buffer Increases the Association of TIM3 with PS. (**a**) A calcium dependence experiment conducted on 7:3 POPC:POPS vesicles in the presence of 0 mM, 150 mM, and 5 M NaCl. The affinity for the membrane, as measured by the membrane-binding induced by calcium, is significantly affected by the buffer ionic strength.

Figure 6.5 cont. The y-axis represents the bound fraction of protein, scaled to the bound fraction at saturating calcium for that membrane and buffer condition. The ionic strength of the 10 mM Hepes buffer was calculated to be 5 µM, hence the offsets of the ionic strength from the NaCl concentrations. (**b**) The reflectivity signal for TIM3 under the 5 mM low ionic strength buffer was significantly enhanced compared to its signal at 155 mM ionic strength. (**c**) The orientation confidence maps of the crystal structure (left panel) and the MD-derived structure (right panel). The best-fit parameters for the crystal structure are $\theta = 50°,\ [50, 52]$, $\phi = 155°,\ [145, 155]$, and $d_p = 9.2$ Å, $[9.0, 9.4]$ and those for the MD-derived structure are $\theta = 50°,\ [50, 52]$, $\phi = 135°,\ [130, 140]$, and $d_p = 10.1$ Å, $[9.9, 10.3]$. (**d**) The best-fit orientation of the MD-derived structure with hydrophobic residues in gray, peripheral arginine residues in red, and peripheral lysine residues in green.

6.4.2 Low Ionic Strength Simulations Implicate the Same Membrane-Bound Ensemble

As compared with the orientation at normal ionic strength conditions, the low ionic strength orientation is less deeply inserted and more flattened. The hydrophobic residues do not even penetrate the tail group region. Thus, this structure is inconsistent with the MD simulations of **Figure 6.3** for different reasons than the normal ionic strength structure. It is possible that the membrane state is an entirely different ensemble when the buffer conditions are different. To explore this possibility, we ran HMMM simulations with water that contained only neutralizing ions and no additional salt. Similarly, we organized the HMMM portion of the simulations by state and obtained six full-length simulations in which TIM3 bound PS in its central pocket. **Figure 6.6a** shows the orientation distribution of these simulations, agreeing with that of the normal ionic strength simulations of **Figure 6.4a**. Furthermore, the distributions of the bound calcium, hydrophobic residues, and peripheral basic residues almost completely overlap (**Figure 6.4b**). Therefore, the membrane-bound ensembles of the two buffer conditions do not seem to substantially differ.



(a)

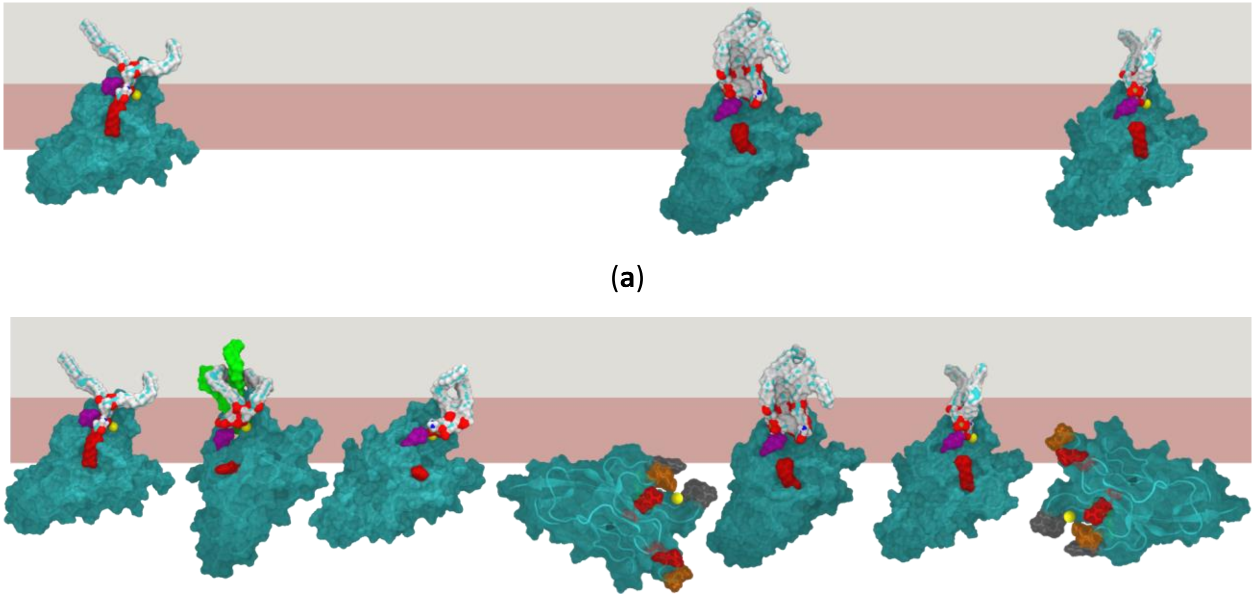
(b)

Figure 6.6. Low Ionic Strength Simulations Identify the Same Membrane-Bound Ensemble. (a) The distribution of orientation angles is scaled to that of **Figure 6.3a** to show that the ensembles are the same.

Figure 6.6 cont. (**b**) Distribution of insertion depths of the calcium ion, hydrophobic residues, and peripheral basic residues. The distributions from the normal ionic strength simulations are in a darker color and those from the low ionic strength simulations are overlaid in a lighter color.

6.5 A Single Consistent Structure for Normal and Low Ionic Strength Conditions

The two reflectivity datasets thus provide their own models that are each inconsistent with the simulations for different reasons. The low ionic strength discrepancies are possibly explained by the non-specific association of TIM3 in addition to the primary binding mode. Over the course of the HMMM portion of the simulations, the proteins were stably associated to the membrane in upside down orientations for an appreciable time before the central binding pocket was engaged. It is therefore possible that this non-specifically membrane-associated state is substantially populated under conditions of low ionic strength. The binding experiments using tryptophan fluorescence would be blind to these states, as the tryptophan residue does not insert into the membrane. However, reflectivity would be sensitive to them as they provide an interfacial structure. **Figure 6.7** illustrates this in a cartoon comparing the interfaces of normal ionic strength and low ionic strength monolayers.

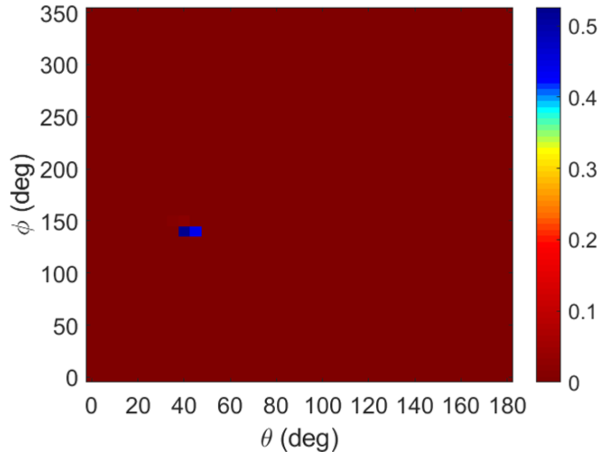


(**a**)

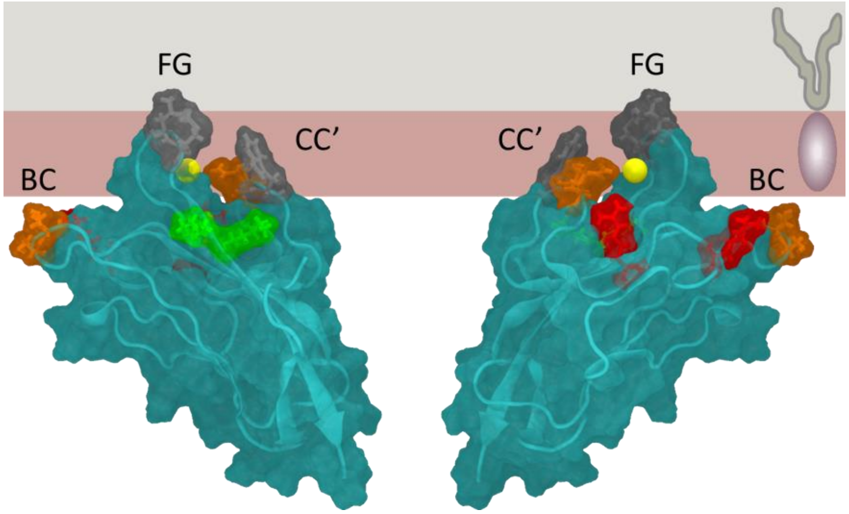
(**b**)

Figure 6.7. Membrane-Associated TIM3 Ensembles. (**a**) The proposed distribution of membrane associated states for normal ionic strength is predominantly populated by TIM3 that has specifically bound PS. (**b**) The low ionic strength ensemble of membrane-associated TIM3 has a higher population of non-specifically associated states despite the increased population of specifically PS-bound states as compared to normal ionic strength conditions.

Guided by the simulations which implicate the same ensemble, we elected to analyze the full set of reflectivity conditions simultaneously, mandating that a single orientation should fit all conditions. The obtained fit is shown in **Figure 6.8a** with the corresponding structure in **Figure 6.8b**. The orientation is highly resolved and is similar to the normal ionic strength orientation, except that the BC loop is more resolved. Additionally, the calcium ion and hydrophobic and peripheral basic residues reside in the membrane at the same position given by the distributions of **Figures 6.3c** and **6.6b**. The calcium ion is shallowly inserted into the head group region while the tryptophan, W41, sits near the head-tail interface but on average resides in the head group region. L99 and M100 however insert into the tail group region. The peripheral basic residues are withdrawn from the head group region as well. The arginine and lysine residues are close enough to form Coulombic interactions with adjacent PS, but are not buried in the membrane. Similarly, the BC loop is not buried in the interface but is close enough to mediate the contacts predicted by the MD simulations.



(**a**)



(**b**)

Figure 6.8. Joint Fitting of Normal and Low Ionic Strength Reflectivity. (**a**) The orientation map shows a significantly high level of confidence in a single orientation. (**b**) The best fit orientation has the parameter values $\theta = 42°$, $[38, 46]$, $\phi = 145°$, $[135, 150]$, and $d_p = 12.5\,\text{Å}$, $[12.2, 12.8]$. Hydrophobic residues are shaded gray, polar residues in orange, peripheral arginine in red, and peripheral lysine in green.

This orientation highlights the importance of considering the ensemble when analyzing reflectivity data. The reflectivity-resolved structure of **Figure 6.9b** is an average structure only, it does not contain the variance. Indeed, the orientation probability map of **Figure 6.8a** does not resemble the orientation occupation maps for the simulations in **Figures 6.3a** and **6.6a**. The variance of the former map is related to the standard error of the mean structure while the latter is a variance corresponding to the population standard deviation. Since x-ray reflectivity is only sensitive to the average electron density, the variance is minimally probable. However, the low ionic strength reflectivity data indicate this variance can affect the estimate of the mean structure. This is because the average electron density is in general a function of all of the moments of the molecular positions, not just the mean and the variance i.e.

$$\langle \rho(z) \rangle = \sum_i A_i f_i \left(\langle \vec{a}_i \rangle \right) + \sum_i \sum_j B_{ij} f_{ij} \left(\langle (\vec{a}_i - \langle \vec{a}_i \rangle) \otimes \left(\vec{a}_j - \langle \vec{a}_j \rangle \right) \rangle \right) + \cdots$$

Here, $\vec{a}_i$ is the position of atom *i*, f_i, f_{ij} are arbitrary functions, and A_i and B_i are coefficients for these functions. With improved models, it might be possible to probe the variance of the protein through the mean electron density, but the coefficients B_{ij} are likely small, especially for small globular proteins such as the TIM proteins. However, this also places higher demands on the simulations, as characterizing the correct variance requires much more simulation time and more resources to attain as much accuracy as the mean. In Chapter 4, our attempts to analyze the data presented here using the variance obtained from MD did not result in improved fits.

In context with the other TIM proteins, TIM3 clearly has many more adjacent PS contacts than TIM1 does, as shown in the positions of residues inferred from both the reflectivity data and the contact map of **Figures 6.3b**. However, these peripheral residues do not bury as deeply as the homologous residues of TIM4 do. This is consistent with TIM3 having an intermediate sensitivity for PS between that of TIM1 and TIM4. However, the TIM3 structure also suggests that the contributions of these peripheral residues could be enhanced on membranes that either accommodate great surface burial of TIM3 or increase the surface anionic charge. Both of these conditions are explored in Chapter 7.

6.6 The Role of the BC Loop is Revealed by the Allelic Variant TIM3 HBA

While our final membrane-bound structure of TIM3 is the most consistent with simulations, previous work with TIM3 has implicated the BC loop as a significant contribution to binding [2]. Specifically, an allelic variant, TIM3 HBA (the variant studied so far is labeled TIM3 Balb/c, both named after the mouse models in which they were identified [4]), which contains seven amino acid substitutions highly reduced the binding and phagocytosis effect by TIM3. Three of these residues located in the BC loop are sufficient to reduce the binding. None of these substitutions change the character of the amino acids; the BC loop is populated with polar residues and no charged or hydrophobic residues. In our work, the reflectivity-derived structure shows this loop withdrawn from the membrane and MD simulations only implicate a few lipid contacts, with no preference for POPS over POPC (**Figure 6.3b**). However, the orientation resolved from the normal ionic strength conditions places this loop well within the head group region, potentially consistent with the study of TIM3 HBA.

We therefore sought to model the binding of TIM3 HBA to determine the role of the BC loop. **Figure 6.9a** shows the orientation distribution for a series of five full-length lipid membrane simulations obtained from initial HMMM simulations conducted similarly to the previous ones described here. TIM3 HBA has a much broader ensemble of occupied orientations. It appears there are possibly three distinct states but it is more likely that five simulations were not enough to fully sample this distribution. Surprisingly, the mean of the two distributions is nearly the same, indicating that the ensembles potentially have highly similar representative bound structures. The breadth of bound orientations of TIM3 HBA was correlated with a higher average and variance of contact surface area, defined as the surface area of the protein within 3.5 Å of the membrane at any given time. Conversely, the correlation was the opposite for the volume enclosed by the central binding pocket, as seen in **Figure 6.9b**.

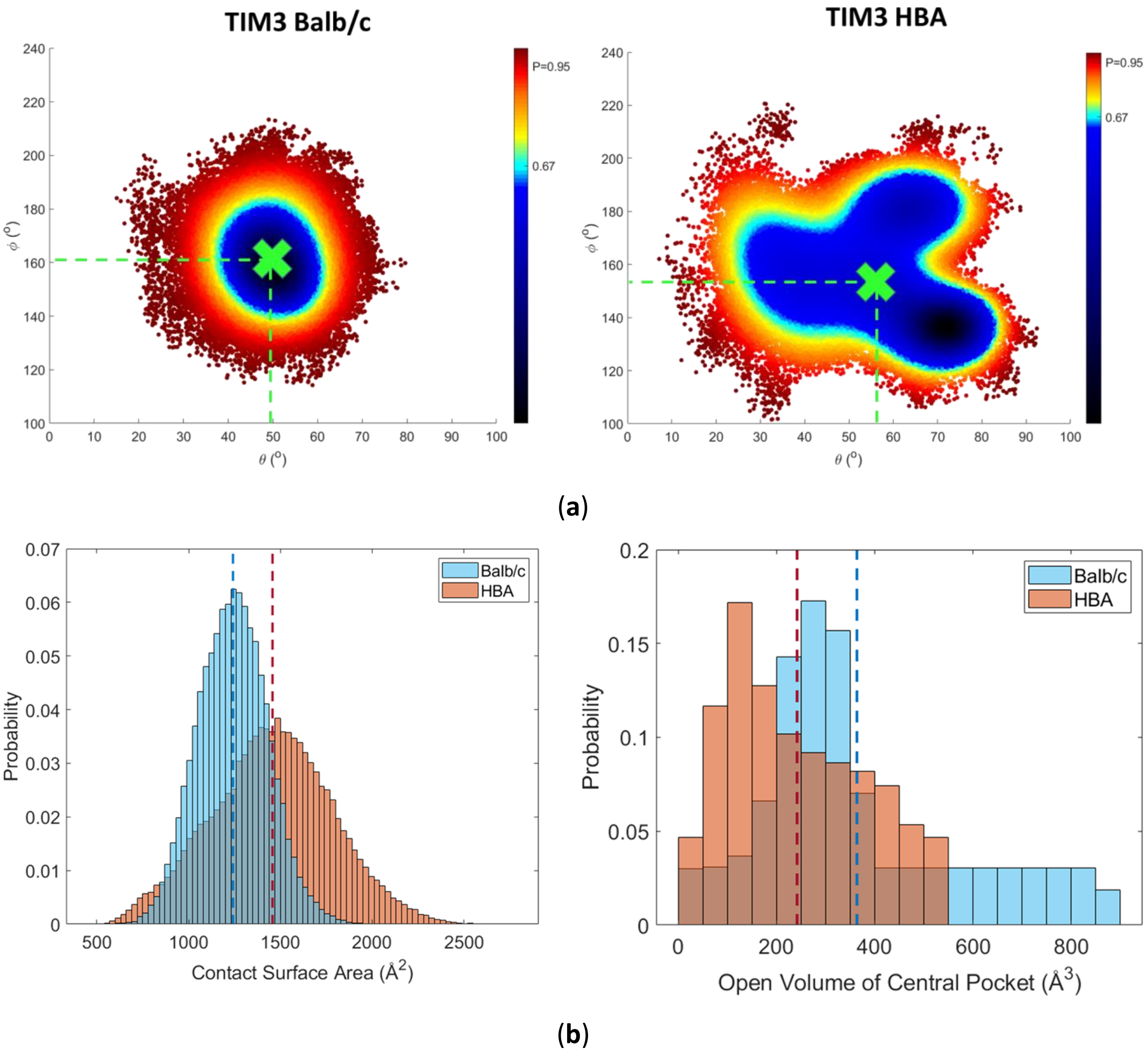
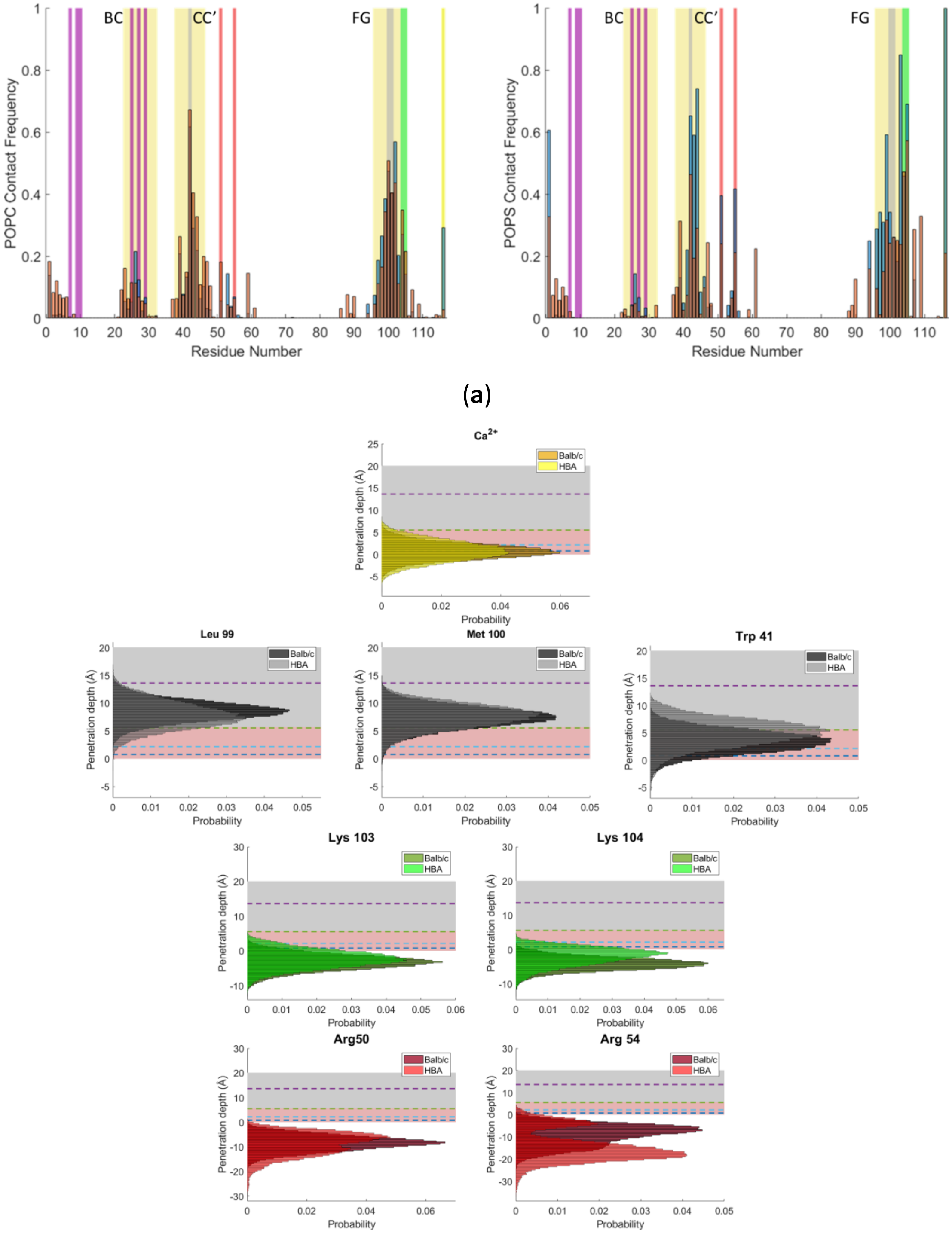


Figure 6.9. The Membrane-Bound Ensemble is Significantly Wider for TIM3 HBA. (**a**) The orientation distributions for TIM3 Balb/c and HBA are compared where the green “X” markers and dashed lines indicate the mean values. (**b**) Probability distributions of the contact surface area (left), defined as the surface area of protein within 3.5 Å of the membrane, and the open volume of the central pocket (right). In addition to having a lower mean value, the distribution of the contact surface area for TIM3 Balb/c is narrower than for TIM3 HBA, correlating with the tighter angular distribution in panel **a**. The central pocket volume distributions follow the opposite trend, with TIM3 HBA having a lower mean and a narrower distribution while TIM3 Balb/c having a long tail.

These trends indicate that the pocket of TIM3 HBA has considerably less configurational entropy, as evidenced by the narrower, and on average lower, distribution of central pocket volumes. The FG and CC’ loops seem to be more constrained for TIM3 HBA. The lack of this configurational entropy is seemingly compensated by an increase of entropy for the center of mass orientation of the protein and several surface-exposed residues. This is supported by examining the contact maps for the residues of TIM3 HBA compared to TIM3 Balb/c in **Figure 6.10a**. TIM3 HBA makes fewer contacts with key pocket residues located on the FG loop, namely G98 and D102. Consistently, the calcium ion is bound to a PS for only a fraction of the time for TIM3 HBA, while it was in contact with at least one PS for every frame of the TIM3 Balb/c simulations. This indicates that the pocket is significantly hampered in its ability to coordinate PS.

Besides the central pocket, the hydrophobic and peripheral basic residues of TIM3 HBA display a similar frequency of lipid contacts to those of TIM3 Balb/c. In general, more residues of TIM3 HBA make contact with lipids than those of TIM3 Balb/c, but with less frequency. Notably, the sites of the substitutions in the BC loop indicate little change in the contact frequency. More residues of the BC loop on TIM3 HBA contact lipids than the corresponding residues on TIM3 Balb/c. The HBA BC residues also have a slight preference for PC lipids over PS lipids. Regardless, these differences are minimal and do not support the notion that the BC loop confers a significant amount of energy to the bound state. Moreover, the insertion depths of the calcium ion, hydrophobic, and peripheral basic residues do not significantly differ between TIM3 Balb/c and TIM3 HBA. Minor differences seem to roughly cancel out (deeper insertion of W41, but more shallow insertion of L99, closer association of K104 but more distant association of R54), as seen in **Figure 6.10b**.



(**a**)

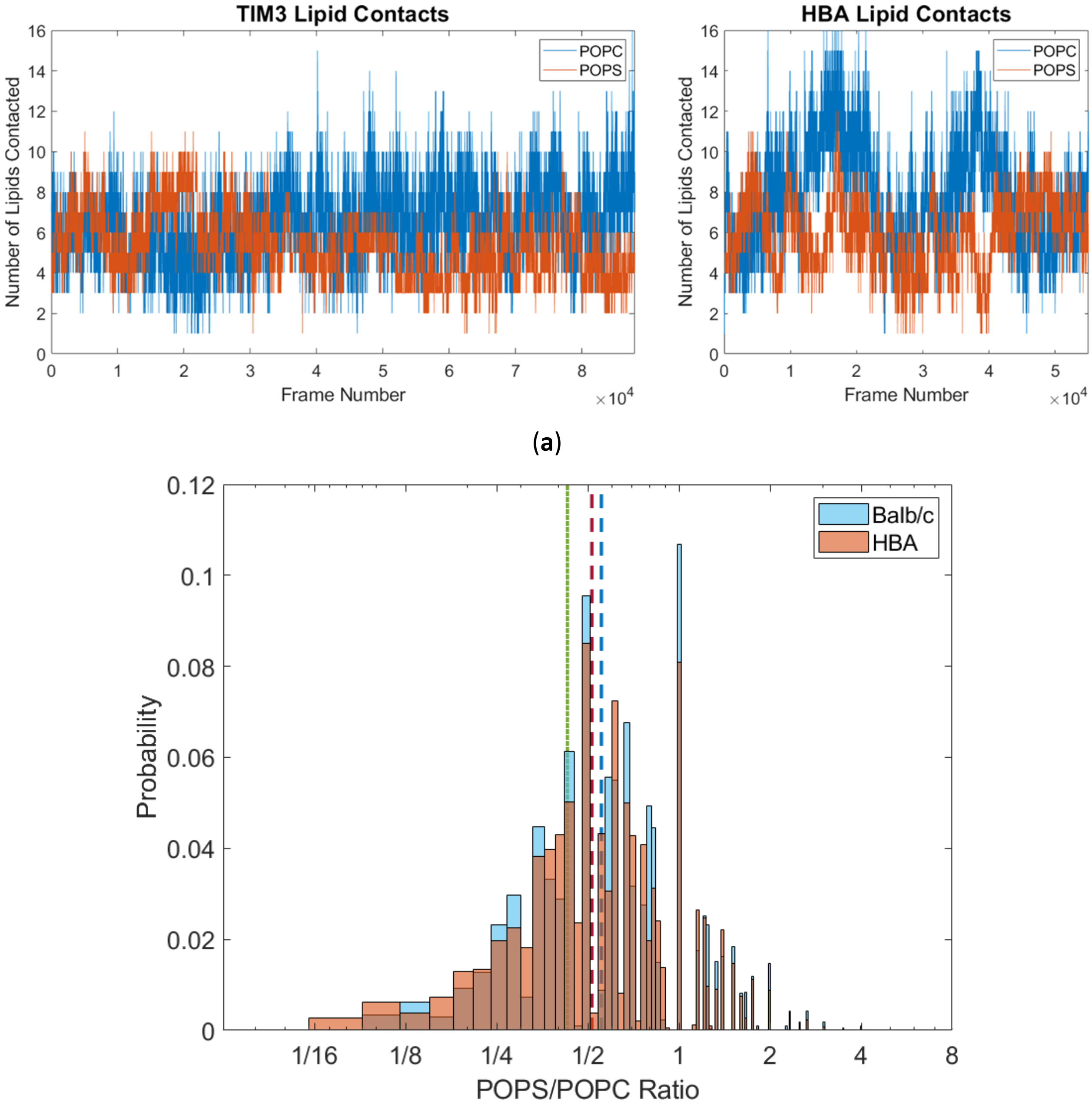
(**b**)

Figure 6.10. The Interaction of TIM3 HBA Residues with Membrane Lipids. (**a**) POPC (left) and POPS (right) residue contact maps similar to **Figure 6.3b**. TIM3 HBA is overlaid in orange over TIM3 Balb/c contact frequencies. Yellow highlights mark the BC, CC’, and FG loops. Surface-exposed hydrophobic residues are highlighted in gray while peripheral arginine and lysine residues are highlighted in red and green respectively.

Figure 6.10 cont. The calcium ion is the rightmost column, highlight in yellow on the POPC contact map. The residues that differ between TIM3 Balb/c and TIM3 HBA are highlighted in magenta. (**b**) Distribution of membrane insertion of key residues and the calcium ion. TIM3 Balb/c is plotted in darker colors and TIM3 HBA is overlaid in lighter colors.

The last remaining mechanism for influencing binding through the BC loop is if TIM3 HBA somehow associates with less PS than TIM3 Balb/c. **Figure 6.11a** shows the number of lipids contacted by either variant of TIM3 within 3.5 Å as a function of time. More residues of TIM3 HBA contact the membrane than those of TIM3 Balb/c, consistent with its larger average surface area. Despite the larger surface contact of TIM3 HBA, the two TIM3 variants display the same slight preference for POPS over POPC (**Figure 6.11b**).

The simulations therefore show that TIM3 HBA does not make fewer peripheral contacts either directly or as a result of its BC loop. Instead, it seems that TIM3 HBA has a much weaker binding pocket, with the calcium and central PS association significantly compromised compared to TIM3 Balb/c. This predicts that TIM3 HBA will have weaker affinity but similar cooperativity to that of TIM3 Balb/c. Binding experiments of TIM3 Balb/c and HBA are compared in Chapter 7 and bear out this prediction.



(**a**)

(**b**)

Figure 6.11. Number of Lipids Associated with TIM3 Balb/c and TIM3 HBA. (**a**) Traces of the number of POPC and POPS lipids within 3.5 Å of TIM3 Balb/c (left) or TIM3 HBA (right). (**b**) Probability distributions for the ratio of POPS and POPC lipids within 3.5 Å of either TIM3 Balb/c (cyan) or TIM3 HBA (orange) with corresponding means as dashed lines of the same color. The green dashed line marks the global 7:3 ratio of the membrane. The PS in the central pockets were subtracted from this analysis since this one PS is required for the binding of TIM3 and thus represents a biased sample of the membrane. The y-axis represents the probability that a protein is associated with that lipid ratio at a given time.

6.7 Conclusions

The story presented in this chapter is convoluted and labor intensive just to parse. A combination of experimental, computational, and mutational studies was used to develop and validate a molecular model. In the end, the membrane-bound model for TIM3 is highly consistent with MD simulations and our previous characterization of its binding on POPC:POPS membranes [1]. Similar to TIM1, the challenges in resolving this structure further pushed our understanding of the membrane-bound structure of the TIM proteins. While TIM1 forced us to consider the kind of average structure reflectivity probes, TIM3 forced us to consider the nature of the entire ensemble and the very definition of a bound state. With an expanded sampling of states in the MD simulations of TIM3, more than two bound states were identified. In our work with TIM1, we considered each “state” as a candidate for the ensemble average. However, MD simulations are inherently limited by sampling time and thus it is highly unlikely that any one simulation will resolve the ensemble average. In the work presented here, we combined all such states, recognizing them as biased subsamples of a larger ensemble. This bias could be corrected for with repeated sampling and in doing so we obtained well-resolved ensembles (**Figures 6.3** and **6.6**). Naturally, these well-characterized ensembles borne out well-resolved ensemble averages that unambiguously fit the reflectivity data in low signal conditions.

The insights gained from this study are not confined only to reflectivity. The binding of peripheral membrane-binding proteins to lipids is almost always modeled as a series of ligand interactions assuming distinct binding sites. Our work here reveals that the concept of a distinct binding site does not apply straightforwardly to this class of proteins. **Figure 6.2** shows many different binding modes of TIM3 in which a supposed single binding site is occupied by multiple ligands, seemingly contradicting the tenet that there are distinct, non-overlapping binding sites. The equations used to describe binding are employed in solvent conditions, in which the likelihood of encountering two ligands crowded together is low. However, the membrane is a highly crowded environment. Because of the size of the TIM proteins, they are guaranteed to have a footprint of ~15 lipids at a time. Thus, when the protein is bound, the pocket is wedged in between several lipids. While there may be a preferred configuration of a single lipid in the pocket in solution (such as those resolved in the crystal structures of the TIM proteins), there are potentially many nearly equivalent configurations when additional lipids are involved in the interaction. Therefore, thinking of the TIM proteins as containing a "specific" binding site is a misleading oversimplification. Depending on the lipid composition of the surrounding membrane, the binding site might have a completely different configuration of binding PS.

The analysis here suggests that the best concept for binding is to regard PS as a kind of anchor in which the TIM proteins probe the crowded membrane environment. The presence of more PS in the local environment further encourages binding by offering more coordination modes of the proteins with the lipids. For example, both TIM3 Balb/c and TIM3 HBA coordinate PS through D1, even though both moieties carry an overall negative charge. However, D1 specifically coordinates with the positively charged amine group of the PS. In a membrane with a low amount of PS, this interaction mode might not be highly populated in the bound state. Similarly, the binding pocket offers a set of charged and polar residues that have a mixed set of charged moieties. There are multiple lipid configurations that satisfy these charges. The more lipid configurations that accommodate PS there are, the larger the response to membrane PS will be. Therefore, it seems most likely that the phenomenon of cooperative binding arises not from conformational changes but from a combination of a highly crowded lipid environment and the configurational entropy of the protein with PS.

Therefore, cooperativity would represent a balance of several components that act to increase or decrease binding affinity on their own. For example, the insertion of a residue into the membrane provides an anti-cooperative component as the energy for displacing the lipid acts against binding. However, if that residue prefers to partition into the hydrophobic core or if it interacts with an opposite charge, then the interaction could overall enhance binding provided that its energy is greater than that associated with displacing lipid. In this case, the interaction would provide a positively cooperative component as the hydrophobic burial and charge coordination promote binding. With regards to the TIM proteins, their peripheral basic residues interact preferentially with PS and other anionic head groups and therefore contribute cooperatively to the binding. Their association with anionic moieties is energetically favorable and increases the configurational entropy of the bound state of the protein.

Chapter 7 uses these insights to formulate a comprehensive binding equation that captures the details of this confined, two-dimensional interaction and takes into account the complicated interplay of association with calcium and PS.

6.8 References

1. Tietjen, G. T., Gong, Z., Chen, C.-H., Vargas, E., Crooks, J. E., Cao, K. D., Heffern, C. T. R., Henderson, J. M., Meron, M., Lin, B., Roux, B., Schlossman, M. L., Steck, T. L., Lee, K. Y. C., and Adams, E. J. "**Molecular Mechanism for Differential Recognition of Membrane Phosphatidylserine by the Immune Regulatory Receptor Tim4.**" *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 15 (2014): E1463–72. doi:10.1073/pnas.1320174111,

2. DeKruyff, R. H., Bu, X., Ballesteros, A., Santiago, C., Chim, Y.-L. E., Lee, H.-H., Karisola, P., Pichavant, M., Kaplan, G. G., Umetsu, D. T., Freeman, G. J., and Casasnovas, J. M. "**T Cell/transmembrane, Ig, and Mucin-3 Allelic Variants Differentially Recognize Phosphatidylserine and Mediate Phagocytosis of Apoptotic Cells.**" *Journal of Immunology* 184, no. 4 (2010): 1918–1930. doi:10.4049/jimmunol.0903059,

3. Tietjen, G. T. "**Immune Recognition of Membrane Phosphatidylserine by the Tim Proteins: A Multidisciplinary Study**" (2013):

4. McIntire, J. J., Umetsu, S. E., Akbari, O., Potter, M., Kuchroo, V. K., Barsh, G. S., Freeman, G. J., Umetsu, D. T., and DeKruyff, R. H. "**Identification of Tapr (an Airway Hyperreactivity Regulatory Locus) and the Linked Tim Gene Family.**" *Nature Immunology* 2, no. 12 (2001): 1109–1116. doi:10.1038/ni739,

CHAPTER 7

THE BINDING OF THE TIM PROTEINS TO PHOSPHOLIPID MEMBRANES

7.1 The TIM Proteins are Sensitive to the Membrane Context

7.1.1 Introduction

The TIM proteins have been shown to recognize PS-exposing cell membranes (see Chapter 2), but their sensitivity to different forms of PS exposure has not been sufficiently established. Apoptotic PS exposure is accompanied by the scrambling of other inner leaflet lipids, looser membrane packing while non-apoptotic exposure occurs in an otherwise intact outer leaflet. The TIM proteins are involved with but not necessarily restricted to the recognition of apoptotic cells, as TIM4 has been shown to bind PS on activated T-cells. Since PS exposure is ubiquitous and not a unique signal for apoptosis, the TIM proteins (and other PS-receptors in general) must bind a subset of PS-exposing membranes.

Proteins have multiple dimensions to distinguish apoptotic from non-apoptotic PS-exposure. This suggests that specific context-sensitive mechanisms allowing the TIM proteins to recognize apoptotic PS exposure would also underlie their sensitivity to particular non-apoptotic PS-containing membranes. For example, if one of the TIM proteins was sensitive to the scrambling of additional anionic lipids from the inner leaflet, non-apoptotic PS exposure that also scrambled anionic lipids would be similarly recognized. If, instead, one of the TIM proteins was entirely insensitive to these other apoptotic membrane features, that protein would recognize all PS-containing membranes equally.

This chapter addresses the following question: Given that the TIM proteins are seemingly specialized for distinct physiological roles, do they differentially recognize different PS-containing membranes? In answering this question, we developed a biochemical model of TIM binding based on studies of the *in vitro* binding of the TIM proteins to model membranes. Tryptophan fluorescence, as described in Chapter 3, was used in the work described throughout this chapter. The justification of this technique for the TIM proteins was provided in Chapters 5 and 6, which showed tryptophan insertion of each of the TIM proteins in their membrane-bound states.

The next section, 7.1.2, reviews previous work of the Lee Lab which investigated the sensitivity of the TIM proteins to PS surface density using the techniques described in Chapter 3 (1). In section 7.1.3, the types of lipids used in the model membranes as well as the physicochemical features of cell membranes they model are reviewed. Section 7.2 develops a mathematical framework to describe lipid-binding proteins, with a particular focus on the TIM proteins. Subsequently, section 7.3 uses the derived equations to analyze the binding of TIM proteins to several model membranes. The last section, 7.4, then interprets the analysis of TIM binding in terms of the structure and energetics involved. The physiological ramifications of these results are summarized in Chapter 8.

7.1.2 Peripheral Basic Residues Confer TIM4 with Sensitivity to PS Surface Density

Characterization of the membrane context requires preparations of membranes with varying compositions and properties. Liposomes prepared with varying amounts of PS can be used to probe the influence of PS surface density on PS-receptor binding. **Figure 7.1a** shows the result of experiments with TIM4 performed with liposomes containing 30 mol% PS and 10 mol% PS, at various vesicle concentrations. At the same concentration of total PS, the 30 mol% PS liposomes bound more TIM4 than did the 10 mol% PS liposomes as illustrated in **Figure 7.1b**. If the TIM proteins followed a single-site binding model for PS, the binding curve would be hyperbolic. Instead, we observed pronounced sigmoidicity for TIM4 as seen in **Figure 7.1c**, which depicts experiments conducted at the same total lipid concentration (600 µM) but with varying PS surface density. TIM1 and TIM3 displayed slight sigmoidal character in their binding curves compared to TIM4, with the sigmoidicity of each binding curve quantified using a Hill model for cooperative binding. The values of the Hill coefficient for each TIM protein are shown in **Figure 7.1d**. Since the Hill coefficients are above 1, the binding of the three TIM proteins are cooperative and thus sensitive to PS surface density (1).

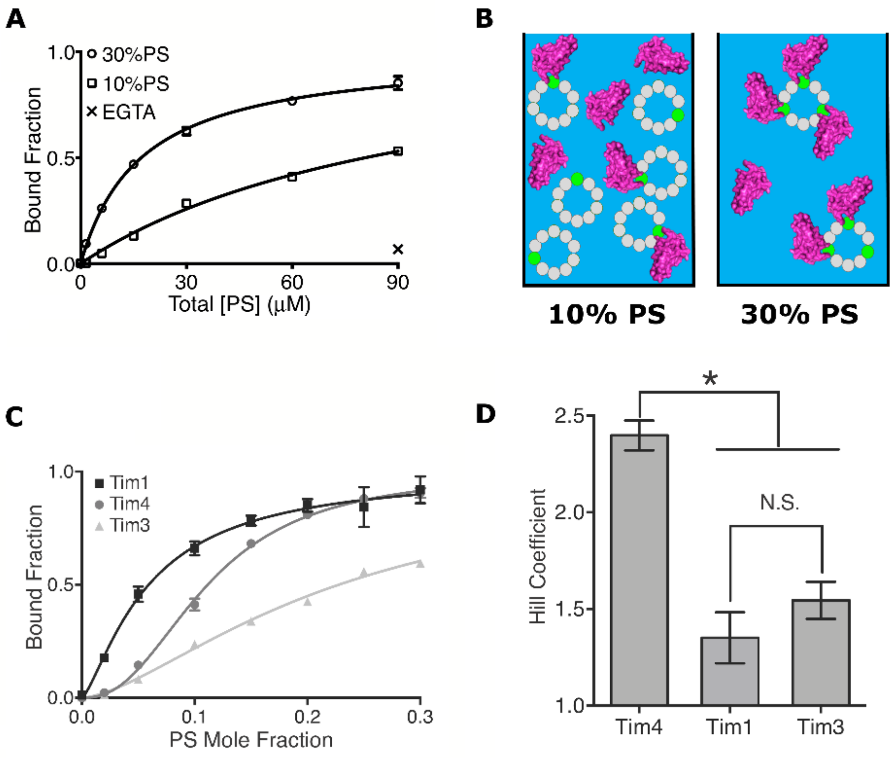


Figure 7.1. TIM4 is Sensitive to PS Surface Density. (**a**) At the same overall concentration of PS, TIM4 has a higher affinity for 30 mol% PS-containing vesicles than 10 mol% PS-containing vesicles. (**b**) Cartoon depicting result in A. Green circles represent PS while gray circles represent phosphatidylcholine (PC), a neutral zwitterionic lipid that the TIM proteins do not bind. The two cells have the same number of PS lipids but the 30 mol% sample on the right has them distributed on fewer vesicles. The two cells also have the same number of protein molecules but the 30 mol% sample has more of them bound to vesicles than the 10 mol% sample on the left. (**c**) Titration series of the mol % of PS-containing vesicles where the overall lipid concentration of PC + PS is held constant. The TIM proteins all have sigmoidal dependence on the PS surface density of vesicles. (**d**) TIM4 has a higher Hill coefficient than TIM1 or TIM3. TIM4 and PS have a 1:3 stoichiometry while TIM1 and TIM3 each have a 1:2 stoichiometry with PS. Figure adapted from *Tietjen et al.* (2014) (1) and *Kerr et al.* (2018) (2).

Unlike the cooperativity measured for Annexin V, the PKC family, or Protein Z, the cooperativity of the TIM proteins cannot be explained by multiple PS or Ca^{2+} binding sites (3–5). Alternative mechanisms can underlie cooperative binding such as the protein's non-specific electrostatic interactions with neighboring PS. These interactions would contribute more to the free energy of binding with higher PS surface density, presenting as cooperative binding. However, these contributions do not necessarily correspond to structural “binding sites” on the protein that are specific for PS. In this scenario, the TIM proteins are only peripherally sensitive to neighboring PS with the PS surface density serving as a membrane-contextual cue, especially in the case of TIM4.

The membrane-bound structures obtained for TIM4 and TIM1 (described in Chapter 5) suggest that the sensitivity for PS surface density is directly conferred by non-specific electrostatic interactions of peripheral basic residues with adjacent PS. The bound structure of TIM4 shows two arginine and two lysine residues embedded in the head group region of the film (**Figure 7.2a**). TIM1 has only a single arginine residue buried in the head group region and one arginine and one lysine close to the membrane, which are relatively more withdrawn than the equivalent residues in TIM4 (**Figure 7.2b**). These positively charged, peripheral residues likely interact with adjacent PS to stabilize the bound state and account for the differential sensitivity to PS surface density between TIM1 and TIM4. Since TIM4 has greater surface burial than TIM1, it incurs a greater energetic penalty for displacing lipids to accommodate its insertion. At high PS surface densities, TIM4’s peripheral residues are likely to be interacting with adjacent PS which would contribute more energy to compensate for TIM4’s surface burial. At low PS surface densities, TIM4 would not be able to employ this mechanism and would be burdened by the full energetic penalty of its surface burial.

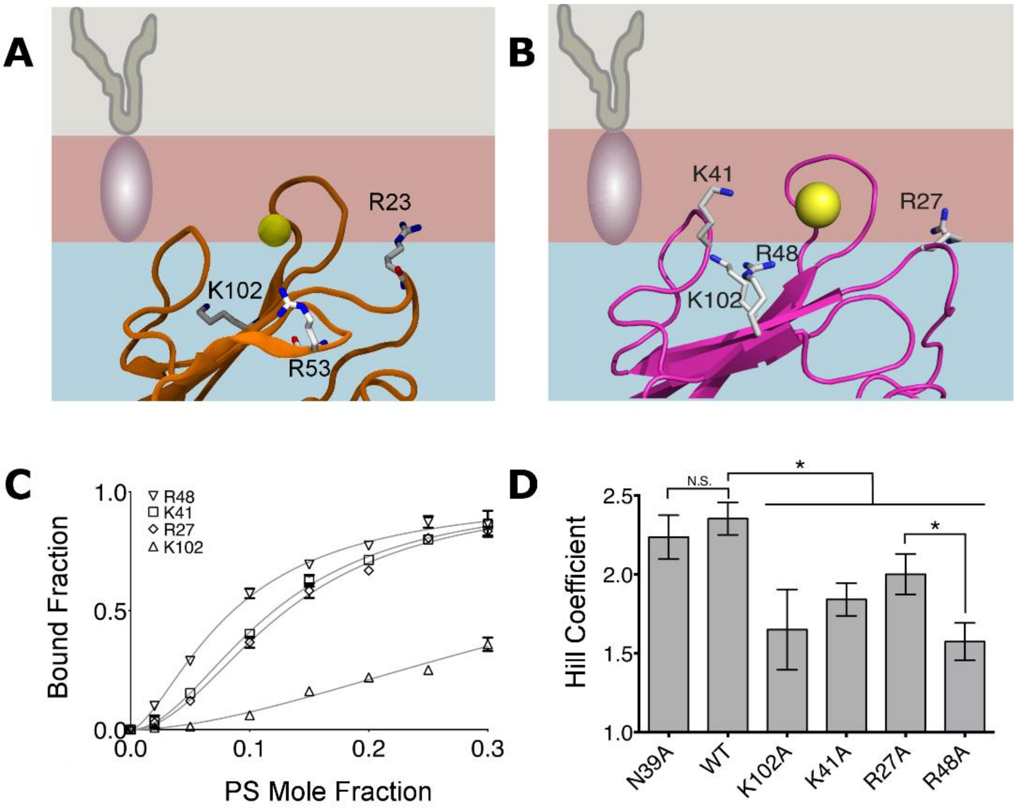


Figure 7.2. The Membrane-Bound Structures Implicate Peripheral Residues That Might Confer Sensitivity to PS Surface Concentration. (**a**) TIM1's membrane-bound structure shows insertion of the PHE and TRP residues into the tail group region and embedding of a single, peripheral ARG into the head group region of the membrane. (**b**) The membrane-bound TIM-4 structure shows similar burial of these hydrophobic residues but also contains insertion of two ARG and two LYS residues into the headgroup region of the membrane. Red: ARG, Green: LYS, Gray: Hydrophobic, Yellow: Polar. (**c**) Alanine mutants of the peripheral basic residues depicted in **b** result in diminished sigmoidicity of TIM4's binding curve. (**d**) The alanine mutants reduce the Hill coefficient of TIM4 substantially indicating these residues are important for sensing PS surface density. Figure adapted from (1, 2) and (6).

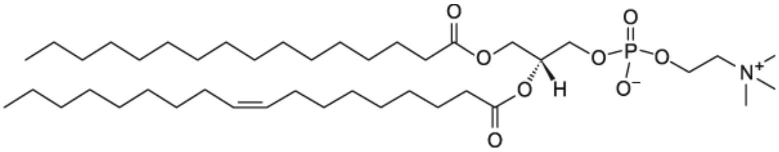
The peripheral basic residues provide a structural basis for the increased sensitivity of TIM4 to PS surface density. **Figures 7.2c** and **7.2d** show the binding curves of four TIM4 single alanine substitution mutations at the two lysine and two arginine residues described above. These mutants have reduced Hill coefficients compared to the wild type, confirming that these residues confer a sensitivity to PS surface density. The elimination of a single peripheral basic residue in TIM4 reduced the Hill coefficient from 2.4 to as low as 1.6. Presumably double mutants would have reduced Hill coefficients comparable to TIM1's value of 1.35 while a quadruple mutant would be expected to exhibit single site binding (1). Peripheral residues surrounding the binding sites of other PS-receptors may also confer a sensitivity to PS exposure distinct from multi-site cooperativity.

7.1.3 Model Membranes Capture Physicochemical Properties of PS-Exposing Membranes

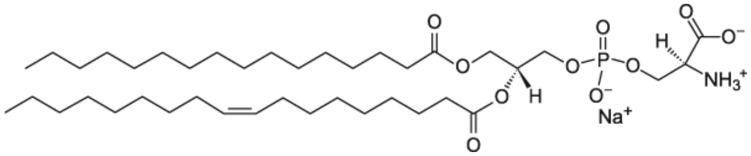
Having established that PS surface density, one of many features that distinguish apoptotic from non-apoptotic PS exposure, is differentially recognized by the TIM proteins, we decided to investigate whether other membrane features could be similarly detected. These features were modeled by using a broad range of lipid compositions; **Figure 7.3** shows the structures of several lipids used in the studies discussed in this chapter. Traditionally, lipids with mixed tails of saturated and mono-unsaturated acyl chains are used to model plasma membrane lipids, based on classic studies of acyl chain distribution of rat plasma membranes (7, 8). Due to its large abundance in plasma membranes and its wide commercial availability, 1-palmitoyl-2-oleoyl (PO) tails are commonly used as plasma membrane models and is the primary tail group used in the studies discussed in this chapter unless otherwise indicated. Indeed, this was also the lipid tail group used in the studies discussed in the previous section.

However, PO membranes capture the physicochemical properties of plasma membranes of healthy cells and not apoptotic membranes. Apoptotic membranes have lower lipid packing, which can be modeled using lipids with larger areas per lipid. Compared with PO, 1,2-dioleoyl (DO) tail groups introduce a significantly larger area per lipid (68 $Å^2$ vs. 72 $Å^2$, respectively, for PC headgroups at 30°C (9)). We also used saturated 1,2-dimyristoyl (DM) tail groups to test the other extreme of smaller area per lipid (60 $Å^2$ at 30°C).

Since the sensitivity to PS surface density of TIM4 was found to originate from its peripheral residues, it is likely that these interactions would be enhanced by adjacent anionic lipids besides PS. Indeed, the inner leaflet contains several anionic lipid species that scramble indiscriminately during apoptosis, as discussed in Chapter 2. We therefore studied membranes containing phosphatidylglycerol (PG) and phosphatidic acid (PA) head groups.

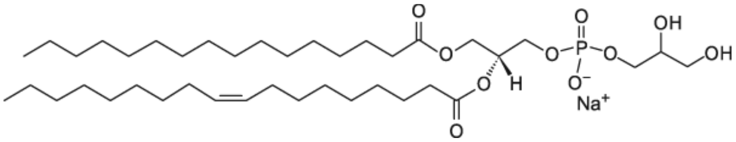


1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC)

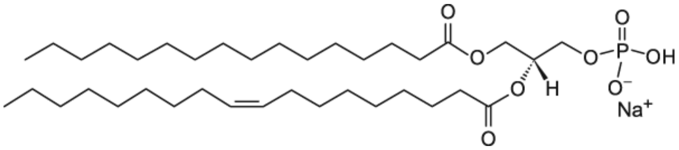


1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-L-serine (POPS)

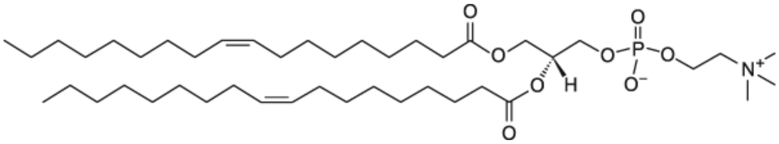
Figure 7.3: Structures of the Lipids Used in Experiments. Lipids with PO, DO, or DM tails and PC, PS, PG or PA headgroups are shown.



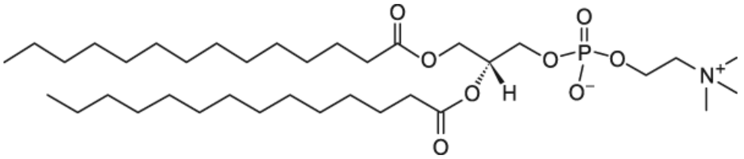
1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-L-glycerol (POPG)



1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphate (POPA)



1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC)



1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC)

Figure 7.3 cont. Lipids with PO, DO, or DM tails and PC, PS, PG or PA headgroups are shown.

7.2 Master Equation for the Binding of Anionic Phospholipids and Divalent Cations

7.2.1 Introduction

In extending our studies to a broader set of membrane conditions, it was important to establish the theory underlying TIM binding. As seen in **Figure 7.1**, the TIM proteins show hyperbolic dependence on total lipid concentration, [L], but Hill dependence for PS mole fraction, $[PS]/[L]$. However, since there are two ligands for the TIM proteins, PS and Ca^{2+}, there are three ligand dependence experiments that can be conducted: (i) Dependence on total vesicle concentration with fixed surface PS mole fraction and $[Ca^{2+}]$, (ii) Dependence on surface PS mole fraction with fixed $[Ca^{2+}]$ and total lipid concentration, [L], and (iii) Dependence on $[Ca^{2+}]$ at fixed total lipid concentration and surface PS mole fraction. Since there is a single Ca^{2+} binding site, the third experiment is also described by hyperbolic binding. These experiments are therefore described by three inherently incompatible equations with no underlying model. In this section, we construct a self-consistent, mathematical model to unify the analysis of these three experiments under one framework.

Inspired by the tryptophan shift of the TIM proteins upon engagement of their pockets, we define a set of "central" binding sites which elicit measurable binding. The simulations of the TIM proteins in Chapters 5 and 6 indicated that the central binding pocket was capable of coordinating multiple PS with the assistance of several peripheral residues. However, these residues were not "binding sites" in the strict sense of the term, rather they contributed to the number of PS-bound conformations the proteins were able to adopt. The engagement of these residues is not directly detectable with tryptophan fluorescence; their contributions can only be inferred from their influence on the binding of the central pocket, namely the insertion of the tryptophan. We therefore assume the protein has two types of coordination with PS, those which are observable through a direct signal and those which are unobservable. We label the latter "peripheral associations" to distinguish them from the "central" observable ones. The former concept generalizes the peripheral basic residues of the TIM proteins and the latter generalizes the central pocket of the TIM proteins, which was defined in Chapter 2. While the engagement of the "central" site of the TIM proteins is inferred from the state of their tryptophan residues, the specific "binding pockets" of other peripheral proteins may not generally be correlated with the state of the observable central sites. We further assume that the observable sites are indistinguishable; that is, these states do not have separate outputs in the experiment such as different fluorescence spectra. However, the intensity of that signal does not have to be the same for each central site. For example, two central sites might have the same fluorescence spectrum but one site fluoresces more intensely.

7.2.2 Definitions of Reactions and Bound Complexes

We assume that the protein can interact with *M* divalent cations, *R* phospholipids of a given species within its central binding sites, and associate with *N* additional, peripheral phospholipids. With regards to the TIM proteins, the divalent cation is Ca^{2+} and the phospholipid is PS. Without any loss of generality, we will use Ca^{2+} and PS throughout this section as our labels for the divalent cation and phospholipid, respectively. In this treatment, the protein can interact with an arbitrary amount of each ligand. After binding the first lipid, the reactions take place in two dimensions, confined to the plane of the membrane (10). In our development, we follow the methods of Segel (11).

We consider a set of reactions involving the protein, P, up to *R* centrally-coordinated PS and *N* peripherally-associated PS, and up to *M* divalent cations Ca^{2+} where complexes with the same number of ligands have equal affinity (this assumption greatly simplifies the derivation and does not change the form of the final, approximated equation that we derive). We use lower case subscripts to denote partial occupancy, i.e., $0 \leq m \leq M$, $0 \leq n \leq N$, and $0 \leq r \leq R$. Brackets represent three-dimensional molar concentrations with three-dimensional dissociation constants K and curly brackets represent two-dimensional molar concentrations with two-dimensional dissociation constants κ. These dissociation constants have four subscripts denoting the number and type of reactants involved. The order of the subscripts is: Ca^{2+}, central PS, peripheral PS, and the number of vesicles. For example, $K_{mC,S1,0S,V1}$ is the dissociation constant for the reaction between P-$(Ca^{2+})^m$ and PS in a central binding site. This reaction adds a single PS and a single vesicle to a complex with *m* Ca^{2+} and 0 peripheral PS. The addition is represented by the number following the symbol "S" in the second subscript and following the symbol "V" in the fourth subscript while the unchanging reactants are represented by the number preceding the symbols "C" and "S" in the first and third subscripts respectively.

Interaction of Ca^{2+} with PS

$$[Ca^{2+}] + [PS] \overset{K_{CS}}{\rightleftharpoons} [Ca^{2+}PS]$$

Interactions of P-Ca^{2+} with PS in central sites

$$\left[\mathrm{P}Ca^{2+^{m}}\right] + [PS] \overset{K_{mC,S1,0S,V1}}{\rightleftharpoons} \left[\left(\mathrm{P}Ca^{2+^{m}}PS\right)\right]_1$$

$$\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r-1}\right)PS^{n}\right]_{v-1} + [PS] \overset{K_{mC,Sr,nS,Vv}}{\rightleftharpoons} \left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]_v$$

$$\left\{\left(\mathrm{P}Ca^{2+^{m}}PS^{r-1}\right)PS^{n}\right\}_v + \{PS\} \overset{\kappa_{mC,Sr,nS,vV}}{\rightleftharpoons} \left\{\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right\}_v$$

Interactions of P-Ca^{2+} with PS in peripheral sites

$$\left[\mathrm{P}Ca^{2+^{m}}\right] + [PS] \overset{K_{mC,0S,S1,V1}}{\rightleftharpoons} \left[\left(\mathrm{P}Ca^{2+^{m}}\right)PS\right]_1$$

$$\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n-1}\right]_{v-1} + [PS] \overset{K_{mC,rS,Sn,Vv}}{\rightleftharpoons} \left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]_v$$

$$\left\{\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n-1}\right\}_v + \{PS\} \ldots \overset{\kappa_{mC,rS,Sn,vV}}{\rightleftharpoons} \left\{\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right\}_v$$

Interactions of all bound protein states with Ca^{2+}

$$\left[\left(\mathrm{P}Ca^{2+^{m-1}}PS^{r}\right)PS^{n}\right]_v + [Ca^{2+}] \overset{K_{Cm,rS,nS,vV}}{\rightleftharpoons} \left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]_v$$

Interactions of protein with Ca^{2+}-PS

$$\left[\mathrm{P}Ca^{2+^{m-1}}\right]+[Ca^{2+}PS]\overset{K_{Cm,S1,0S,V1}}{\rightleftharpoons}\left[\left(\mathrm{P}Ca^{2+^{m}}PS\right)\right]_1$$

$$\left[\mathrm{P}Ca^{2+^{m-1}}\right]+[Ca^{2+}PS]\overset{K_{Cm,S1,0S,V1}}{\rightleftharpoons}\left[\left(\mathrm{P}Ca^{2+^{m}}PS\right)\right]_1$$

$$\left[\left(\mathrm{P}Ca^{2+^{m-1}}PS^{r-1}\right)PS^{n}\right]_{v-1}+[Ca^{2+}PS]\overset{K_{Cm,Sr,nS,Vv}}{\rightleftharpoons}\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]_v$$

$$\left\{\left(\mathrm{P}Ca^{2+^{m-1}}PS^{r-1}\right)PS^{n}\right\}_{v}+\{Ca^{2+}PS\}\overset{\kappa_{Cm,Sr,nS,vV}}{\rightleftharpoons}\left\{\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right\}_v$$

$$\left[\left(\mathrm{P}Ca^{2+^{m-1}}PS^{r}\right)PS^{n-1}\right]_{v-1}+[Ca^{2+}PS]\overset{K_{Cm,rS,Sn,Vv}}{\rightleftharpoons}\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]_v$$

$$\left\{\left(\mathrm{P}Ca^{2+^{m-1}}PS^{r}\right)PS^{n-1}\right\}_{v}+\{Ca^{2+}PS\}\overset{\kappa_{Cm,rS,Sn,vV}}{\rightleftharpoons}\left\{\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right\}_v$$

The two-dimensional concentrations can be interpreted as the number of moles of PS per area on the vesicle surface and have the units of $\mathrm{moles}/\mathrm{m}^2$. Some of the reactions occur both in three-dimensions and two-dimensions. The former corresponds to binding of new vesicles and the latter corresponds to the association of protein with PS on a vesicle to which it has already bound. This is reflected in the subscripts of the brackets labeled with *v*, i.e., $[X]_v$ is the concentration of the X complex that is bound to *v* vesicles.

This set of reactions is accompanied by the following mass conservation equations:

$$[\mathrm{P}]_{tot}=\sum_{v=0}^{V}\sum_{r=0}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N}\left[\left(\mathrm{P}{Ca^{2+}}^{m}PS^{r}\right)PS^{n}\right]_{v} \tag{7.1}$$

$$[PS]_{tot}\cong[PS]+[Ca^{2+}PS] \tag{7.2}$$

$$[Ca^{2+}]_{tot}\cong[Ca^{2+}]+[Ca^{2+}PS] \tag{7.3}$$

Equations 7.2 and **7.3** implicitly assume that both PS and Ca^{2+} are in great excess compared to P. This allows us to derive tractable, analytic formulae. For this development, we assume that K_{CS} is very large and thus the association of Ca^{2+} and PS without protein is negligible. If this latter assumption is invalid, then the derived equation at the end of this section can be solved together with **Equations 7.2** and **7.3** numerically, with $[Ca^{2+}PS]$ given by:

$$[Ca^{2+}PS]=\frac{1}{2}\Big(K_{CS}+[Ca^{2+}]_{tot}+[PS]_{tot} \\ -\sqrt{(K_{CS}+[Ca^{2+}]_{tot}+[PS]_{tot})^{2}-4[Ca^{2+}]_{tot}[PS]_{tot}}\Big) \tag{7.4}$$

7.2.3 Derivation of Master Equation

To yield a binding equation with practical value, we must carefully consider the bound signal detected in an experiment. For simplicity, we first consider an interaction of the protein with only one vesicle, returning to the case of multiple vesicles later in the text. The coefficient $C_{r,n}$, describes the contribution to the signal from bound protein with *r* centrally bound phospholipids in complex with *n* peripheral lipids. We assume that $C_{r,n}$ is a monotonically increasing function of *r* and *n*, as more lipid association corresponds to more bound signal. The inclusion of the *n* index is motivated by the possibility that engagement of peripheral PS might affect the central binding signal by locally influencing the environment or, for example, by facilitating deeper insertion of a tryptophan. $C_{r,n}$ is assumed to be independent of both Ca^{2+} and the number of vesicles bound, depending specifically on the membrane environment. Since we can only detect protein that have their central sites engaged, we are interested in the following quantity:

$$\frac{[\mathrm{P}]_{bound}}{[\mathrm{P}]_{tot}} = \frac{\sum_{r=1}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N} C_{r,n}\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]}{\sum_{r=0}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N}\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]} \tag{7.5}$$

The numerator contains protein complexes that have bound at least one PS in a central site as required for detection. If Ca^{2+} is required for membrane binding or for the experimental readout of the bound state, then the numerator's sum should start from $m = 1$ instead of $m = 0$. If the r central binding sites are distinguishable, then the r terms of the sum should be considered independently,

$$\frac{[\mathrm{P}]_{r}}{[\mathrm{P}]_{tot}} = \frac{\sum_{m=0}^{M}\sum_{n=0}^{N} C_{r,n}\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]}{C_{R,N}\sum_{r=0}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N}\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]} \tag{7.6}$$

Since **Equation 7.6** is a term in **Equation 7.5**, we proceed to analyze the latter. Our goal is to find a tractable expression for **Equation 7.5** in terms of the ligand concentrations. We start by expressing $\left[\left(PCa^{2+^{m}}PS^{r}\right)PS^{n}\right]$ in terms of the ligand concentrations and the equilibrium constants governing their reactions. Assuming detailed balance, every binding pathway is equivalent at equilibrium, we thus choose to decompose the reaction first as a binding of $r + n$ PS and then a binding of m Ca^{2+}, the latter is expressed in **Equation 7.7**. This choice allows us to cleanly separate the PS and Ca^{2+} binding events since, while the latter is straightforward to decompose, the former requires us to consider the two-dimensional reactions.

$$\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right] = \frac{B_{m,r,n}[(\mathrm{P}PS^{r})PS^{n}][Ca^{2+}]^{m}}{\prod_{i=1}^{m} K_{Ci,rS,nS,1V}} \tag{7.7}$$

The coefficient $B_{m,r,n}$ reflects the various conformations of *m*-bound Ca^{2+} ions, *r*-centrally bound and *n*-peripherally bound PS with (assumed) equivalent affinity. In the case that the protein has distinct binding sites, $B_{m,r,n}$ is the product of binomial coefficients $\binom{M}{m}\binom{R}{r}\binom{N}{n}$.

We now consider the first reaction of the protein with PS alone:

$$\{(\mathrm{P}PS^{r})PS^{n}\} = \frac{\{(\mathrm{P}PS)\}\{PS\}^{r+n-1}}{\prod_{j=1}^{r-1} \kappa_{mC,Sj,0S,1V} \prod_{q=1}^{n} \kappa_{mC,rS,Sq,1V}} \tag{7.8}$$

To convert these two-dimensional concentrations to three-dimensional concentrations, we will take advantage of the fact that in a given experiment, the ratio of PS to total lipid, L, is equal to a constant both in bulk solution and within each vesicle.

$$\frac{[PS]}{[L]} = \frac{\{PS\}}{\{L\}} \tag{7.9}$$

$\{L\}$ is the inverse of the area per lipid, which minimally varies as a function of $\{PS\}$. Since PS is in large excess to protein, the protein minimally covers the surface of a given vesicle, thus the area per lipid is approximately constant over the course of protein binding. With this relationship, we can rewrite **Equation 7.8** in terms of three-dimensional concentrations.

$$[(\mathrm{P}PS^r)PS^n] = \frac{B_{m,r,n}[(\mathrm{P}PS)][PS]^{r+n-1}}{\prod_{j=1}^{r-1}\kappa_{P,mC,Sj,0S,1V}\prod_{q=1}^{n}\kappa_{P,mC,rS,Sq,1V}}\left(\frac{\{L\}}{[L]}\right)^{r+n-1} \tag{7.10}$$

We define a unitless equilibrium constant, $\mathrm{K}_{P,mC,Sj,0S} = {}^{\kappa_{P,mC,Sj,0S}}\!/_{\{L\}}$, which allows us to rewrite **Equation 7.10** as,

$$[(\mathrm{P}PS^r)PS^n] = \frac{B_{m,r,n}[(PPS)]\left(\frac{[PS]}{[L]}\right)^{r+n-1}}{\prod_{j=1}^{r-1}\mathrm{K}_{P,mC,Sj,0S,1V}\prod_{q=1}^{n}\mathrm{K}_{P,mC,rS,Sq,1V}} \tag{7.11}$$

$$[(\mathrm{P}PS^r)PS^n] = \frac{B_{m,r,n}[P][L]\left(\frac{[PS]}{[L]}\right)^{r+n}}{K_{P,0C,S1,0S,V1}\prod_{j=1}^{r-1}\mathrm{K}_{P,mC,Sj,0S,1V}\prod_{q=1}^{n}\mathrm{K}_{P,mC,rS,Sq,1V}} \tag{7.12}$$

In the final line, we have substituted the reaction for $[(\mathrm{P}PS)]$, yielding an expression solely in terms of the components of the interaction. This equation can be interpreted as the binding of P to a single vesicle (with a concentration given by $[L]$) with a lipid composition of ${}^{[PS]}\!/_{[L]}$.

We can extend this model to allow binding of multiple vesicles. Specifically, we assume that the protein associates with *V* vesicles at most. *V* must be less than or equal to the number of total PS that can associate with the protein, i.e. $V \leq R + N$. In this case, $[(\mathrm{P}PS^r)PS^n]$ represents a combination of proteins bound to *n+r* PS on a single vesicle, *n + r - 1* PS on one vesicle and a single PS on another, one PS on *v = n+ r* vesicles, and everything in between. If this is the case, then **Equation 7.12** becomes:

$$[(\mathrm{P}PS^{r})PS^{n}] = \sum_{v=1}^{\min(V,n+r)} \binom{\min(V,n+r)}{v} B_{m,r,n,v}[(\mathrm{P}PS^{r})PS^{n}]_{v} \tag{7.13}$$

$$[(\mathrm{P}PS^{r})PS^{n}]_{v} = \frac{B_{m,r,n,v}[\mathrm{P}][L]^{v}\left(\frac{[PS]}{[L]}\right)^{n+r}}{K_{v,n,r}} \tag{7.14}$$

$$K_{v,n,r}{}^{-1} = \sum_{t=1}^{v}\sum_{s=0}^{t-1}\left(K_{P,0C,S(t-s),Ss,Vt}\prod_{j=1}^{r-t-1}\mathrm{K}_{P,mC,Sj,0S,tV}\prod_{q=1}^{n-t}\mathrm{K}_{P,mC,rS,Sq,tV}\right)^{-1} \tag{7.15}$$

$$K_{P,0C,S(t-s),Ss,Vt} = \prod_{w=1}^{t-s}\left(K_{P,0C,Sw,0S,Vw}\prod_{u=1}^{s}K_{P,0C,tS,Su,V(t-s+u)}\right) \tag{7.16}$$

Equation 7.13 represents the group of complexes with *r*-centrally bound and *n*-peripherally bound PS as a sum of complexes each associated with *v* vesicles. The maximum number of vesicles associated with this group of complexes is given by the smaller of *V* or *n* + *r*, as the number of bound vesicles is limited by the number of PS in the complex. Each of these complexes is, in turn, a group of complexes each associated with the same number of vesicles but through a different combination of central and peripheral interactions. We consider this as just one complex with the affinity constant given by **Equation 7.15**, which reflects the multiple ways *v* vesicles can be bound by *v* total central and peripheral PS associations.

We can now return to **Equation 7.7**, where substitution with **Equations 7.13** and **7.14** yields a complete expression of $\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right]$ solely in terms of three-dimensional concentrations of the interaction components.

$$\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right] = \sum_{v=1}^{\min(V,n+r)} B_{m,r,n,v}\binom{\min(V,n+r)}{v}\frac{[\mathrm{P}][Ca^{2+}]^{m}[L]^{v}\left(\frac{[PS]}{[L]}\right)^{n+r}}{K_{v,n,r}\prod_{i=1}^{m}K_{Ci,rS,nS,1V}} \qquad (7.17)$$

With this we recover the desired result, an expression relating the population of detectable membrane-bound protein to the concentration of reactants,

$$\frac{[\mathrm{P}]_{bound}}{[\mathrm{P}]_{tot}} = \frac{\sum_{r=1}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N}C_{r,n}\sum_{v=1}^{\min(V,n+r)}B_{m,r,n,v}\binom{\min(V,n+r)}{v}\frac{[\mathrm{P}][Ca^{2+}]^{m}[L]^{v}\left(\frac{[PS]}{[L]}\right)^{n+r}}{K_{v,n,r}\prod_{i=1}^{m}K_{Ci,rS,nS,vV}}}{C_{R,N}\sum_{r=0}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n+r)}B_{m,r,n,v}\binom{\min(V,n+r)}{v}\frac{[\mathrm{P}][Ca^{2+}]^{m}[L]^{v}\left(\frac{[PS]}{[L]}\right)^{n+r}}{K_{v,n,r}\prod_{i=1}^{m}K_{Ci,rS,nS,vV}}} \qquad (7.18)$$

Since **Equation 7.18** has upwards of $\frac{M\times R\times N\times(N+R)\times(N+R+1)}{2}$ unique terms, each with its own product of dissociation constants, there are too many parameters to reasonably fit to any binding experiment. We wish to make an approximation, similar to the Hill model, that reduces **Equation 7.17** to a function with as few terms as possible. However, considering there are three different exponents in this expression, we must be careful in implementing this approximation. First, we review the standard Hill model before proceeding.

7.2.4 The Hill Model for Multiple Binding of a Single Ligand Species in Bulk

The Hill equation arises as an approximation to the binding of a protein with multiple, equivalent ligand (denoted by $[L]$) binding sites. For example, consider the Adair equation for a protein with four equivalent ligand binding sites (12):

$$\frac{[\mathrm{P}]_{bound}}{[\mathrm{P}]_{tot}} = \frac{\frac{4[L]}{K_1} + \frac{6[L]^2}{K_2K_1} + \frac{4[L]^3}{K_3K_2K_1} + \frac{[L]^4}{K_4K_3K_2K_1}}{1 + \frac{4[L]}{K_1} + \frac{6[L]^2}{K_2K_1} + \frac{4[L]^3}{K_3K_2K_1} + \frac{[L]^4}{K_4K_3K_2K_1}} \tag{7.19}$$

In the standard approach, each term in the numerator of **Equation 7.19** (11, 12) would have an additional factor equal to the exponent of the ligand (i.e. $6 \cdot 2$ for the second term, $4 \cdot 3$ for the third, and 4 for the last). These extra factors are a consequence of measuring the binding in terms of the saturation of each binding site. However, our experiment does not probe the binding of multiple sites, but rather only the protein as a whole. Therefore, these coefficients are absent in **Equation 7.19**. For the case of no cooperativity where $K_1 = K_2 = K_3 = K_4$, assuming the bound complexes have the exact same signal, this binding equation does not reduce to a single-site Michaelis-Menten equation. The equation is not sigmoidal either, instead it resembles a steeper hyperbola than the Michaelis-Menten equation.

$$\left.\frac{[\mathrm{P}]_{bound}}{[\mathrm{P}]_{tot}}\right|_{non-coop} = \frac{4\frac{[L]}{K} + 6\left(\frac{[L]}{K}\right)^2 + 4\left(\frac{[L]}{K}\right)^3 + \left(\frac{[L]}{K}\right)^4}{1 + 4\frac{[L]}{K} + 6\left(\frac{[L]}{K}\right)^2 + 4\left(\frac{[L]}{K}\right)^3 + \left(\frac{[L]}{K}\right)^4} = \frac{\left(1 + \frac{[L]}{K}\right)^4 - 1}{\left(1 + \frac{[L]}{K}\right)^4} \tag{7.20}$$

Following the procedure of the Hill approximation, we assume that the reaction is highly cooperative. This entails that the binding of the second ligand is enhanced by the first ligand already being bound, i.e. $K_4K_3K_2K_1 \ll K_3K_2K_1 \ll K_2K_1 \ll K_1$, where K_n is the equilibrium constant of the *n*th binding site given that *n-1* binding sites have bound ligand already. This assumption allows us to neglect terms with lower powers of ligand concentration.

$$\frac{[\mathrm{P}]_{bound}}{[\mathrm{P}]_{tot}} \cong \frac{\frac{[L]^4}{K_4K_3K_2K_1}}{1 + \frac{[L]^4}{K_4K_3K_2K_1}} = \frac{[L]^4}{K_4K_3K_2K_1 + [L]^4} \tag{7.21}$$

The approximation holds for proteins with an arbitrary number of ligand binding sites. However, even if the highly cooperative assumption does not hold, the exponent can be replaced with the Hill coefficient, *h*, and the product of equilibrium constants can be replaced with a new equilibrium constant, K_D:

$$\frac{[\mathrm{P}]_{bound}}{[\mathrm{P}]_{tot}} \approx \frac{[L]^h}{K_D^h + [L]^h} \tag{7.22}$$

There are values of these parameters which produce a sigmoidal curve that aligns very closely with the full expression, even in the case of a non-sigmoidal "true" binding curve (resulting from a non-cooperative interaction). The Hill coefficient then takes on fractional values, with the interpretation that its round-up value provides a lower bound for the number of ligand binding sites of a protein, e.g. if *h*=1.3, then there are at least two ligand binding sites. Additionally, if it is the case that for our four-ligand example, one of the binding sites is not identical to the other three and has much higher affinity for the ligand compared to the other three (even in the absence of cooperativity), the Hill coefficient will likely be, $1 < h < 2$. Consequently, the Hill coefficient can drastically underestimate the number of ligand binding sites.

7.2.5 Hill Approximation of Master Equation

Applying the Hill approximation to **Equation 7.18** requires us to carefully consider how cooperative the ligands are for each other. In general, PS might be cooperative for more PS but not cooperative for more Ca^{2+}, and vice versa. We therefore consider the *m* = 0, and *r* = *n* = 0 terms separately from the rest of the sums, which, respectively, refer to the interaction of protein with PS alone and Ca^{2+} alone. We now turn our attention to the r = *n* = 0 terms to determine the binding in the absence of lipid:

$$\sum_{m=1}^{M}\left[\mathrm{P}Ca^{2+^{m}}\right] = \sum_{m=1}^{M} B_{m,0,0,0}\frac{[\mathrm{P}][Ca^{2+}]^{m}}{\prod_{i=1}^{m} K_{Ci,rS,nS,vV}} \cong \frac{[\mathrm{P}][Ca^{2+}]^{M}}{\prod_{i=1}^{M} K_{Ci,rS,nS,1V}} \approx \frac{[\mathrm{P}][Ca^{2+}]^{h'_{Ca}}}{K_{PC}^{h'_{Ca}}} \qquad (7.23)$$

A Hill approximation replaces the whole sum with the *m* = *M* term resulting from the assumption of high cooperativity. We replace *M* with h'_{Ca} and the equilibrium constant with the new equilibrium constant (with units of three-dimensional molar concentration), K_{PC}, to emphasize that the high cooperativity assumption might not hold. This term represents the multi-Ca^{2+} bound state of the protein in the absence of PS.

We now consider the *m* = 0 terms, treating *r > 1* and *r* = 0 separately since the former are directly detectable while the latter complexes are not:

$$\sum_{r=1}^{R}\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n+r)}[(\mathrm{P}PS^{r})PS^{n}] = \sum_{r=1}^{R}\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n+r)} B_{0,r,n,v}\binom{\min(V,n+r)}{v}\frac{[\mathrm{P}][L]^{v}\left(\frac{[PS]}{[L]}\right)^{n+r}}{K_{v,n,r}}$$

$$\cong \sum_{v=1}^{\mathrm{V}} B_{0,r,n,v}\binom{\mathrm{V}}{v}\frac{[\mathrm{P}][L]^{v}\left(\frac{[PS]}{[L]}\right)^{n+r}}{K_{v,N,R}} \cong \frac{[\mathrm{P}][L]^{V}\left(\frac{[PS]}{[L]}\right)^{N+R}}{K_{V,N,R}} \qquad (7.24)$$

Since the number of vesicle binding sites, *V*, is not necessarily equal to the number of total associated PS, we must assign them different exponents:

$$\sum_{r=1}^{R}\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n+r)}[(\mathrm{P}PS^{r})PS^{n}] \approx \frac{[\mathrm{P}][L]^{h'_{V}}\left(\frac{[PS]}{[L]}\right)^{h'_{PS}}}{K_{PS}^{h'_{V}}} \qquad (7.25)$$

Similar to **Equation 7.23**, **Equation 7.25** describes the multi-vesicle/multi-PS bound state of the protein. We have introduced new exponents h'_{PS} and h'_V, and equilibrium constant (with units of three-dimensional molar concentration), K_{PS}, to again denote that the high cooperativity assumption may not hold. These two exponents capture different aspects of the PS cooperativity; h'_V describes the cooperativity for multiple vesicles and h'_{PS} describes the cooperativity for PS within a vesicle.

Now we analyze the *m* = 0 and *r* = 0 terms, corresponding to undetectable peripheral associations with PS:

$$\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n+r)}[(\mathrm{P})PS^n]=\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n)}B_{0,0,n,v}\binom{\min(V,n)}{v}\frac{[\mathrm{P}][L]^v\left(\frac{[PS]}{[L]}\right)^n}{K_{v,n,0}}$$

$$\cong\sum_{v=1}^{\mathrm{V}}B_{0,0,n,v}\binom{\mathrm{V}}{v}\frac{[\mathrm{P}][L]^v\left(\frac{[PS]}{[L]}\right)^{n+r}}{K_{v,N,0}}\cong\frac{[\mathrm{P}][L]^V\left(\frac{[PS]}{[L]}\right)^N}{K_{V,N,0}}\tag{7.26}$$

In case the high cooperativity assumption does not hold, we write **Equation 7.26** as:

$$\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n+r)}[(\mathrm{P})PS^n]\approx\frac{[P][L]^{h'_{V,F}}\left(\frac{[PS]}{[L]}\right)^{h'_{PS,F}}}{K_{PS,F}^{h'_{V,F}}}\tag{7.27}$$

To distinguish the parameters from those corresponding to complexes centrally bound to PS, the subscript "*F*" has been appended.

We now turn our attention to the nonzero *m* and *n* terms with *r* > 0, which describe protein complexes with Ca^{2+} and centrally bound PS. The high cooperativity assumption applies to both ligands, meaning they are interdependent. Only the *M*, *R*, and *N* terms survive:

$$\sum_{r=1}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N}\left[\left(\mathrm{P}{Ca^{2+}}^{m}PS^{r}\right)PS^{n}\right]$$

$$=\sum_{r=1}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n+r)} B_{m,r,n,v}\binom{\min(V,n+r)}{v}\frac{[\mathrm{P}][Ca^{2+}]^{m}[L]^{v}\left(\frac{[PS]}{[L]}\right)^{n+r}}{K_{v,n,r}\prod_{i=1}^{m}K_{Ci,rS,nS,vV}}$$

$$\cong\sum_{v=1}^{\mathrm{V}}\binom{\mathrm{V}}{v}\frac{[\mathrm{P}][Ca^{2+}]^{M}[L]^{v}\left(\frac{[PS]}{[L]}\right)^{R+N}}{K_{v,N,R}\prod_{i=1}^{M}K_{Ci,RS,NS,vV}}\cong\frac{[\mathrm{P}][Ca^{2+}]^{M}[L]^{V}\left(\frac{[PS]}{[L]}\right)^{R+N}}{K_{V,N,R}\prod_{i=1}^{M}K_{Ci,RS,NS,VV}} \qquad (7.28)$$

Similar to the case of binding with PS alone, the exponents for vesicle cooperativity and PS cooperativity must be different.

$$\sum_{r=1}^{R}\sum_{m=0}^{M}\sum_{n=0}^{N}\left[\left(\mathrm{P}{Ca^{2+}}^{m}PS^{r}\right)PS^{n}\right]\approx\frac{[\mathrm{P}][Ca^{2+}]^{h_{Ca}}[L]^{h_{V}}\left(\frac{[PS]}{[L]}\right)^{h_{PS}}}{K^{h_{Ca}+h_{V}}} \qquad (7.29)$$

We introduce new exponents as well as an equilibrium constant for the fully bound complex (with units of three-dimensional molar concentration), *K*, to emphasize that the high cooperativity assumption may not hold. The exponents are unprimed here as not to confuse them with the Ca^{2+} or PS only complexes. The difference in these exponents stems from the possibility that the degree of cooperativity for PS may be modulated by Ca^{2+} or vice versa.

Lastly, we consider the terms with nonzero *m* and *n* with *r* = 0:

$$\sum_{m=0}^{M}\sum_{n=0}^{N}\left[\left(\mathrm{P}Ca^{2+^{m}}\right)PS^{n}\right] = \sum_{m=0}^{M}\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n)} B_{m,0,n,v}\binom{\min(V,n)}{v}\frac{[\mathrm{P}][Ca^{2+}]^{m}[L]^{v}\left(\frac{[PS]}{[L]}\right)^{n}}{K_{v,n,0}\prod_{i=1}^{m}K_{Ci,0S,nS,vV}}$$

$$\cong \sum_{v=1}^{\mathrm{V}}\binom{\mathrm{V}}{v}\frac{[\mathrm{P}][Ca^{2+}]^{M}[L]^{v}\left(\frac{[PS]}{[L]}\right)^{N}}{K_{v,N,0}\prod_{i=1}^{M}K_{Ci,RS,NS,vV}} \cong \frac{[\mathrm{P}][Ca^{2+}]^{M}[L]^{V}\left(\frac{[PS]}{[L]}\right)^{N}}{K_{V,N,0}\prod_{i=1}^{M}K_{Ci,0S,NS,VV}} \tag{7.30}$$

Again, we replace these exponents to highlight that the assumption of high cooperativity may not hold,

$$\sum_{m=0}^{M}\sum_{n=0}^{N}\left[\left(\mathrm{P}Ca^{2+^{m}}\right)PS^{n}\right] \approx \frac{[\mathrm{P}][Ca^{2+}]^{h_{Ca,F}}[L]^{h_{V,F}}\left(\frac{[PS]}{[L]}\right)^{h_{PS,F}}}{K^{h_{Ca,F}+h_{V,F}}} \tag{7.31}$$

The added subscript, "F," again denotes complexes with solely peripheral PS associations.

Note that $C_{r,n}$ coefficients were omitted from the above analysis. In this approximation, they are effectively absorbed into the binding parameters. However, these coefficients only appear in the numerator and thus their values skew the interpretation of the resulting fit parameters. If $C_{r,n} = r$, as is the case when measuring the fractional saturation of binding sites, then the exponents are exactly the statistical variance of the number of centrally bound PS (13, 14). If $C_{r,n} = r + \epsilon(r,n)$, as is likely the case for tryptophan fluorescence, then the Hill coefficients lack such a clear interpretation. The values of $C_{r,n}$ are not easily determined, so one must be aware that the values of the binding parameters are effectively dependent on the experiment used to infer them.

We now turn back to **Equation 7.18**, rearranged in a form to highlight the division of separately cooperative terms outlined above,

$$\begin{gathered}\frac{[\mathrm{P}]_{bound}}{[\mathrm{P}]_{tot}} = \frac{\sum_{r=1}^{R}\sum_{n=0}^{N} C_{r,n}\sum_{v=1}^{\min(V,n+r)}([(\mathrm{P}PS^{r})PS^{n}] \ldots}{C_{R,N}\left(\sum_{m=0}^{M}\left[\mathrm{P}Ca^{2+^{m}}\right] + \sum_{r=0}^{R}\sum_{n=0}^{N}\sum_{v=1}^{\min(V,n+r)}([(\mathrm{P}PS^{r})PS^{n}] \ldots\right.} \\ \frac{\ldots + \sum_{m=1}^{M}\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right])}{\left.\ldots + \sum_{m=1}^{M}\left[\left(\mathrm{P}Ca^{2+^{m}}PS^{r}\right)PS^{n}\right])\right)}\end{gathered} \tag{7.32}$$

The observable complexes, in the numerator, are clearly separated into Ca^{2+}-containing and Ca^{2+}-lacking complexes. In the case that Ca^{2+} is required for binding, the first term in the numerator should be omitted. Moreover, the numerator only contains terms which have at least one centrally-bound PS, while the denominator contains remaining terms (with *r* = 0). If the peripheral associations are highly cooperative, then the *r* = 0 states should be mostly unpopulated. However, if the central sites do not bind with high cooperativity, these terms will contribute significantly and delay the saturation in low Ca^{2+} conditions. Finally, we arrive at our Hill-like approximation by substitution of **Equations 7.23**, **7.25**, **7.27**, **7.29** and **7.31** into **Equation 7.32**:

$$\begin{gathered}\frac{[\mathrm{P}]_{bound}}{[\mathrm{P}]_{tot}} \approx \frac{\dfrac{[L]^{h'_V}\left(\dfrac{[PS]}{[L]}\right)^{h'_{PS}}}{K_{PS}^{h'_V}} \ldots}{1 + \dfrac{[Ca^{2+}]^{h'_{Ca}}}{K_{PC}^{h'_{Ca}}} + \left(\dfrac{[L]^{h'_V}\left(\dfrac{[PS]}{[L]}\right)^{h'_{PS}}}{K_{PS}^{h'_V}} + \dfrac{[L]^{h'_{V,F}}\left(\dfrac{[PS]}{[L]}\right)^{h'_{PS,F}}}{K_{PS,F}^{h'_{V,F}}}\right) \ldots} \\ \frac{\ldots + \dfrac{[Ca^{2+}]^{h_{Ca}}[L]^{h_V}\left(\dfrac{[PS]}{[L]}\right)^{h_{PS}}}{K^{h_{Ca}+h_V}}}{\ldots\left(\dfrac{[Ca^{2+}]^{h_{Ca}}[L]^{h_V}\left(\dfrac{[PS]}{[L]}\right)^{h_{PS}}}{K^{h_{Ca}+h_V}} + \dfrac{[P][Ca^{2+}]^{h_{Ca,F}}[L]^{h_{V,F}}\left(\dfrac{[PS]}{[L]}\right)^{h_{PS,F}}}{K^{h_{Ca,F}+h_{V,F}}}\right)}\end{gathered} \tag{7.33}$$

This equation applies to a broad range of peripheral membrane-binding proteins including those that do not require divalent cations, multi-domain proteins that simultaneously bind multiple vesicles, proteins that bind multiple phospholipids on a single vesicle, and any combination. These equations can seem obtuse at first sight, but their utility will be demonstrated in their application to the TIM proteins.

7.3 Application of Binding Equation to TIM Proteins

7.3.1 Reduction of Binding Equation to TIM Protein System

We consider that a TIM protein can bind to PS-containing bilayers through both ternary calcium coordination and one or more binary (electrostatic) links. The TIM proteins each contain only a single PS and Ca^{2+} binding pocket, we therefore set the exponents describing vesicular and Ca^{2+} association to one (also confirmed by fitting these parameters without constraint). In fitting **Equation 7.32**, we determined that there was no detectable, peripheral association with PS. We also were unable to saturate the Ca^{2+}-independent binding, thus, for simplicity, we assume equality of the exponents for Ca^{2+}-independent and Ca^{2+}-dependent binding. This assumption is valid for low saturation of the Ca^{2+}-independent binding mode since the binding is approximately linear in this regime and thus insensitive to the true value of its Hill coefficient. We therefore rewrite **Equation 7.32** specifically for the TIM proteins.

$$\frac{b}{b_{tot}} = \frac{\left(\frac{[L]}{K_{TS}} + \frac{[Ca^{2+}][L]}{K^2}\right)\left(\frac{[PS]}{[L]}\right)^h}{1 + \frac{[Ca^{2+}]}{K_{TC}} + \left(\frac{[L]}{K_{TS}} + \frac{[Ca^{2+}][L]}{K^2}\right)\left(\frac{[PS]}{[L]}\right)^h} \tag{7.34}$$

Here, *b* is the bound protein and b_{tot} is its binding at saturation with Ca^{2+} and PS. All dissociation constants and reactant concentrations are molar. [PS]/[L] is the mole fraction of PS in the vesicles. K is the dissociation constant for the complex; it potentially involves both the ternary calcium link and binary interactions between basic residues and anionic polar head groups. K_{TC} is the dissociation constant for a membrane-free binary TIM-Ca^{2+} complex. K_{TS} is the dissociation constant for binary interaction(s) of the protein with PS (taken as unimolecular regardless of their multiplicity) absent Ca^{2+}. The exponent h resembles a Hill coefficient; it describes the high-order dependence of the binding of a protein making multiple links to membrane PS. [While *h* can vary with reactant concentrations, it is treated here as constant (11–13, 15). [Ca^{2+}], [PS] and [L] are in great excess over protein, and the data suggest that binary PS-Ca^{2+} complexes are negligible in magnitude; thus, total reactant concentrations are good approximations of the free species (16). We also assume that the ligands do not alter membrane properties so as to the change binding constants.

7.3.2 Analysis of TIM4 Binding to Membranes

We sought to apply **Equation 7.34** to obtain numerical values for the association of TIM4 with membranes. However, our data were insufficient for this analysis. This problem was overcome by using the exponent obtained previously for the [PS] dependence of TIM4 binding: $h = 2.35$

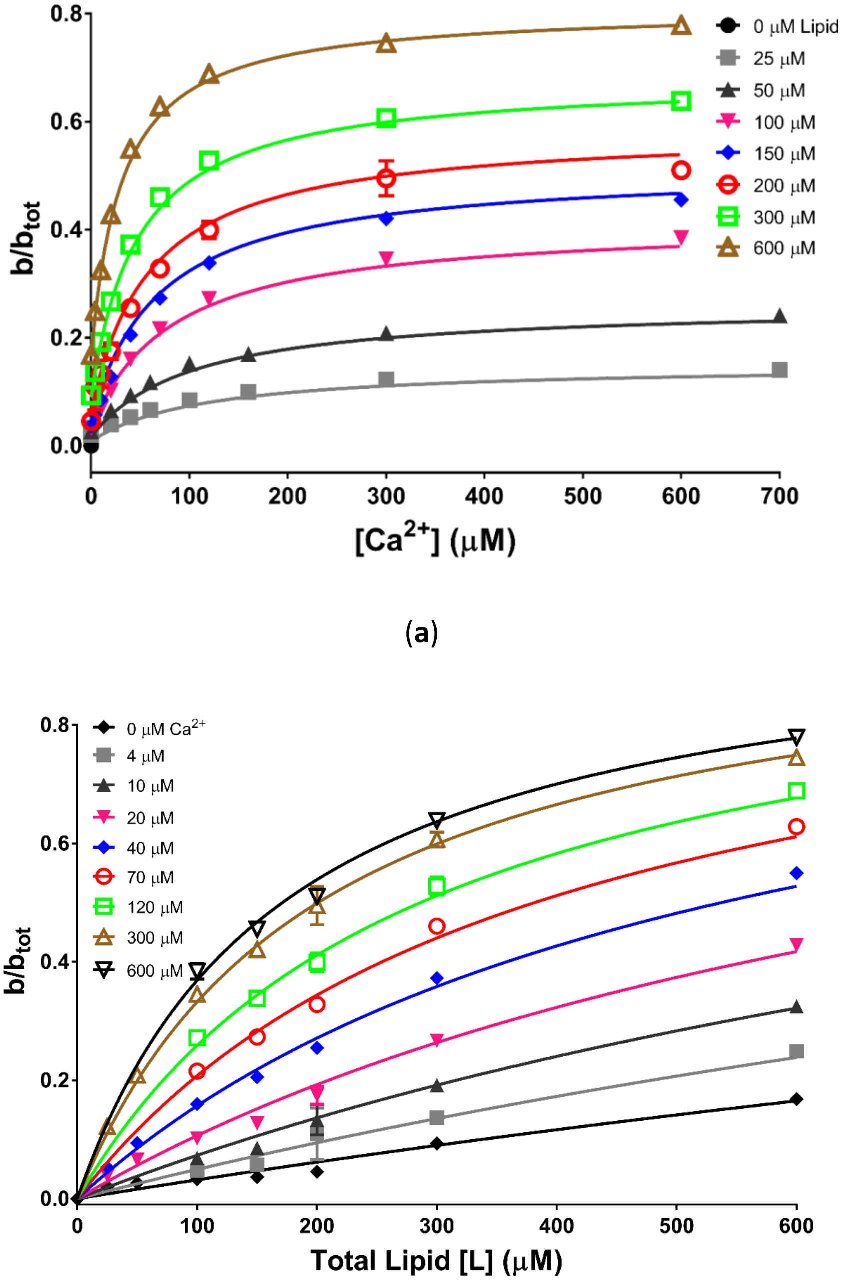
(1). We had to show that an empirical value for h can be rigorously applied in our analysis. For this, we derived an expression for $[PS]_{½}$ by setting *b/b*$_{tot}$ = ½ in **Equation 7.34**:

$$[PS]_{½}{}^{h} = \frac{[L]^{h-1}\left(1 + \frac{[Ca^{2+}]}{K_{TC}}\right)}{\frac{1}{K_{TS}} + \frac{[Ca^{2+}]}{K^{2}}} \tag{7.35}$$

[PS]$_{½}$ can be seen to be set by the four binding parameters at a given [Ca^{2+}] and [L]. The relationship in **Equation 7.35** shows that **Equation 7.34** has the form of a Hill expression; namely,

$$\frac{b}{b_{tot}} = \frac{[PS]^{h}}{[PS]_{½}{}^{h} + [PS]^{h}} \tag{7.36}$$

Because $[PS]_{½}$ serves as a dissociation constant in **Equation 7.36**, it justifies the use of an empirical value for the exponent, h, in **Equation 7.34** for the analysis of data for TIM4. The values so obtained for the binding parameters are given in **Table 7.1**. These values were then used to draw the curves in **Figures 7.4a** in terms of [Ca^{2+}] dependence and in **Figures 7.4b** in terms of [L] dependence. The goodness of fit of the theoretical curves to the data serves to validate both the model and the primary data.



(**a**)

(**b**)

Figure 7.4. Dependence of TIM4 Binding on [Ca^{2+}]: Influence of Total Lipid Concentration. (**a**) TIM4 binding to 70:30 POPC:POPS vesicles of the indicated total lipid concentration was determined as a function of [Ca^{2+}]. The data contained 64 triplicate measurements. *b* is bound protein and b_{tot} is binding at infinite [Ca^{2+}] and [L] (hence, also infinite [PS]). The curves were generated from a non-linear least-squares best fit to **Equation 7.34**, yielding the values in **Table 7.1**. (**b**) The data in panel **a** were replotted to show their dependence on [L] at different [Ca^{2+}] values. For both panels **a** and **b**, zero [Ca^{2+}] indicates zero added Ca^{2+}.

7.3.3 Analysis of TIM3 Binding to Membranes

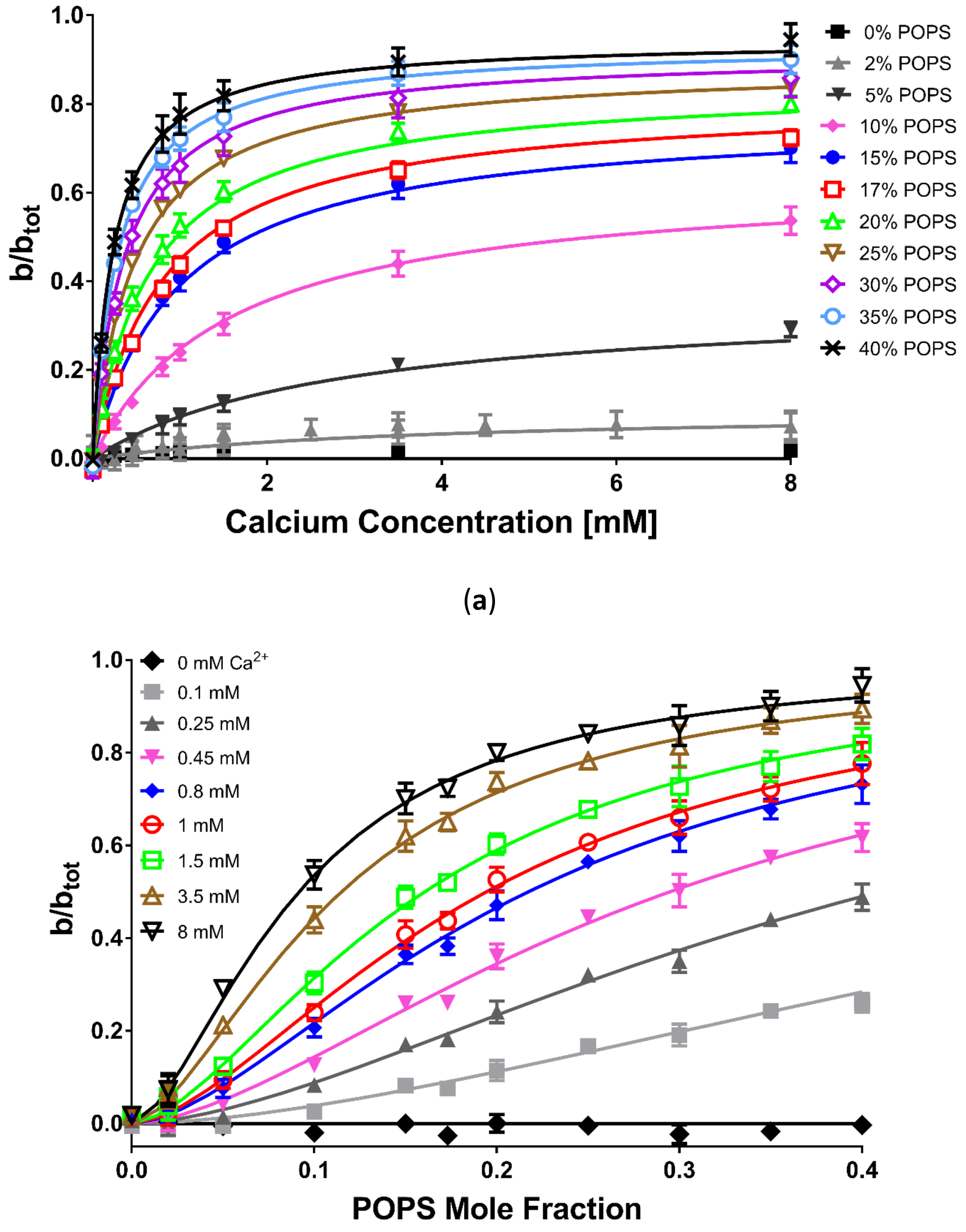
TIM3 association with PS-containing membranes is a special case of that described above for TIM4 in that its direct binding to PS is not detected. Omitting this interaction from **Equation 7.34** yields

$$\frac{b}{b_{tot}} = \frac{[Ca^{2+}][L]\left(\frac{[PS]}{[L]}\right)^h}{K^2\left(1+\frac{[Ca^{2+}]}{K_{TC}}\right)+[Ca^{2+}][L]\left(\frac{[PS]}{[L]}\right)^h} \tag{7.37}$$

At constant [Ca^{2+}] and [L], **Equation 7.37** takes the form of a Hill equation for [PS] as given by **Equation 7.35**. In this case, $[PS]_{½}$ is given by

$$[PS]_{½}{}^h = \frac{K^2[L]^{h-1}}{[Ca^{2+}]}\left(1+\frac{[Ca^{2+}]}{K_{TC}}\right) \tag{7.38}$$

As with TIM4, this justifies using h as a Hill coefficient for TIM3. We therefore fit the data in **Figure 7.5** to **Equation 7.37** to obtain values for the binding parameters given in **Table 7.1**. The curves generated by **Equation 7.37** match those binding data well thus supporting the validity of the model and quality of the results. Therefore, for the remainder of this text, we consider the binding of TIM3, as described by **Equation 7.37**, as a special case of **Equation 7.34** where $1/K_{TS} = 0$.



(**a**)

(**b**)

Figure 7.5. Dependence of TIM3 Binding on [Ca^{2+}] and [PS]. (**a**) TIM3 binding to 300 μM vesicles of the indicated POPS content was determined as a function of [Ca^{2+}]. The data contained 122 triplicate (or more) measurements. The curves were generated by non-linear least-squares best fits to **Equation 7.37**, yielding the values in **Table 7.1**. b is bound protein and b_{tot} is binding at infinite [Ca^{2+}] and [PS]. (**b**) This is a replot of the data in panel **a**. The curves in panels **a** and **b** were generated from the values given in **Table 7.1** using **Equation 7.34**.

7.3.4 Comparison of TIM3 and TIM4 Binding Parameters

All of the parameters significantly differed between TIM3 and TIM4, as shown in **Table 7.1**. In particular, TIM4 binding showed both a higher cooperativity (*h*) and stronger binding to the membrane (*K*). Furthermore, TIM4 bound Ca^{2+} independent of PS far more strongly than did TIM3 (K_{TC}). It was also the case that TIM4 but not TIM3 had a distinctive membrane affinity in the absence of Ca^{2+} (K_{TS}). Technical limitations prevented a direct determination of K_{TS} with PS in the absence of Ca^{2+}. These results indicate that TIM4's higher affinity for PS-containing membranes and Ca^{2+} are deeply intertwined. The structural and energetic implications in these differences are explored in section 7.4.

Binding Parameter	***TIM3***	***TIM4***
Hill coefficient, h	1.65 [1.59, 1.72]	2.35 [2.14, 2.57]
K for TIM-Ca^{2+}-PS dissociation	127 µM [120, 136]	34 µM [32, 35]
K_{TC} for Ca^{2+}-TIM binary complexes	4300 µM [3600, 5200]	136 µM [123, 150]
K_{TS} for PS-TIM binary complexes	Immeasurably weak	218 µM [202, 236]

Table 7.1. Values for TIM3 and TIM4 Binding Parameters to PS-Containing Vesicles. Estimates for TIM3 were obtained by fitting the data from **Figure 7.5** to **Equation 7.37**. Estimates for TIM4 were obtained by fitting the data in **Figure 7.4** to **Equation 7.34** using the *h* value from **Figure 2** in ref. (1). Brackets give 95% confidence intervals. All of the values for the parameters for TIM3 are significantly different from TIM4 at >95% confidence as calculated using a z test.

7.3.5 [Ca^{2+}]$_{½}$ is a Useful Measure of Protein Affinity

We sought a theoretical justification for the use of [Ca^{2+}]$_{½}$ as measure of relative binding strength in experimental results. However, [Ca^{2+}]$_{½}$ was not the [Ca^{2+}] value that yielded b/b_{tot} = ½, because the Ca^{2+}-dependence curves reached a plateau at saturating [Ca^{2+}] less than b_{tot}. Moreover, this plateau value decreased with diminishing [PS] as seen in **Figures 7.4a** and **7.5a**. We therefore rearranged **Equation 7.34** to capture its Ca^{2+} dependence:

$$\frac{b}{b_{tot}} = \left(\frac{[L]\left(\frac{[PS]}{[L]}\right)^h}{\frac{K^2}{K_{TC}} + [L]\left(\frac{[PS]}{[L]}\right)^h}\right)\left\{\frac{[Ca^{2+}]}{[Ca^{2+}]_{½} + [Ca^{2+}]}\right\}\cdots$$

$$+\left(\frac{[L]\left(\frac{[PS]}{[L]}\right)^h}{K_{TS} + [L]\left(\frac{[PS]}{[L]}\right)^h}\right)\left\{\frac{[Ca^{2+}]_{½}}{[Ca^{2+}]_{½} + [Ca^{2+}]}\right\} \tag{7.39}$$

Here [Ca^{2+}]$_{½}$ is given by

$$[Ca^{2+}]_{½} = \frac{1 + \frac{[L]}{K_{TS}}\left(\frac{[PS]}{[L]}\right)^h}{\frac{1}{K_{TC}} + \frac{[L]}{K^2}\left(\frac{[PS]}{[L]}\right)^h} \tag{7.40}$$

The first parenthetical term in **Equation 7.39** (in round brackets) describes a rectangular hyperbola expressing the dependence of TIM binding on [L] at fixed [PS] and a sigmoidal dependence on [PS] at fixed [L]. This term importantly defines the short-fall of the binding plateau from the total (i.e., b_{tot}) that is achieved when [Ca^{2+}] is saturated. We call this plateau binding b_{max}, the maximal protein bound at infinite [Ca^{2+}] but limited by [PS] and/or [L]. This relationship is defined by **Equation 7.41**,

$$\frac{b_{max}}{b_{tot}} = \frac{[L]\left(\frac{[PS]}{[L]}\right)^h}{\frac{K^2}{K_{TC}} + [L]\left(\frac{[PS]}{[L]}\right)^h} \tag{7.41}$$

Equation 7.41 resembles a Hill equation in which the association of TIM-Ca^{2+} with PS has a de facto dissociation constant of ${K^2}/{K_{TC}}$. This expression also shows how the binding plateaus observed in most of our experiments fall as [PS] or [L] is reduced relative to ${K^2}/{K_{TC}}$.

Equation 7.39 also describes protein binding at zero [Ca^{2+}], at which point the first curly bracketed term is zero and the second is unity. We call this value *b$_0$*, defined by this special case of **Equation 7.39** and written here as **Equation 7.42**,

$$\frac{b_o}{b_{tot}} = \frac{[L]\left(\frac{[PS]}{[L]}\right)^h}{K_{TS} + [L]\left(\frac{[PS]}{[L]}\right)^h} \tag{7.42}$$

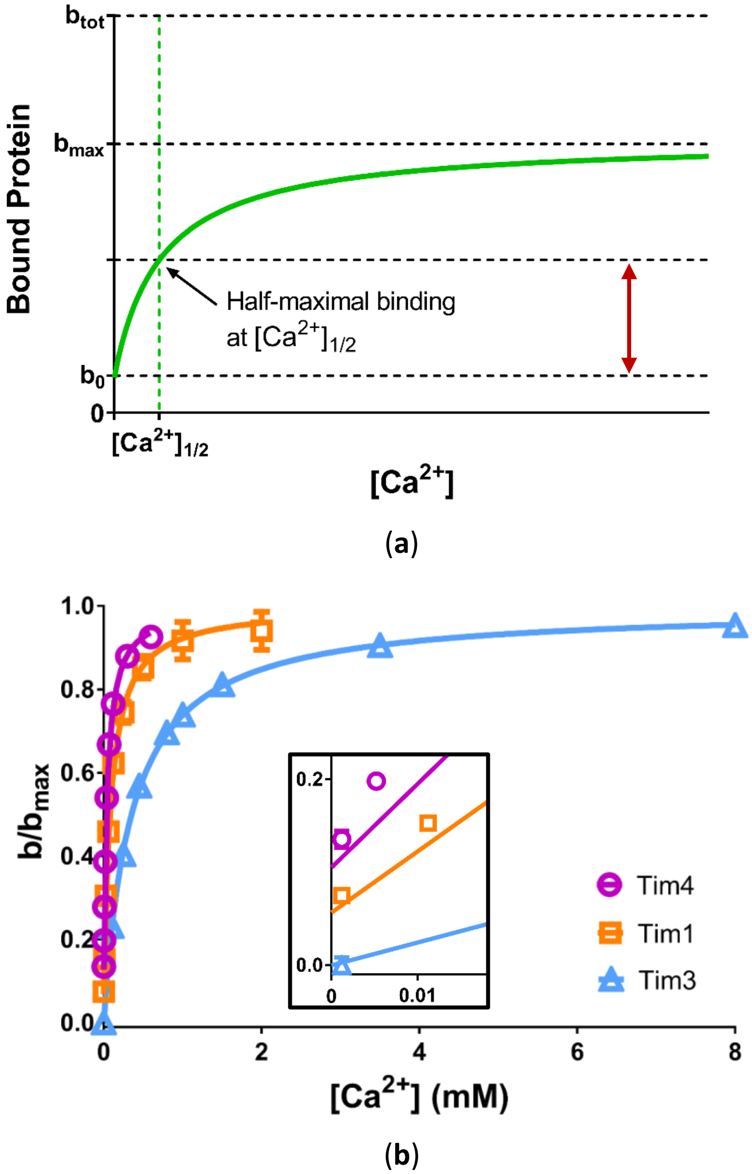
Protein binding in the absence calcium is not simple. Rather, **Equation 7.42** has the form of a Hill equation in which the Ca^{2+}-free, direct association of a TIM is hyperbolic in [L] and sigmoidal in [PS] with a dissociation constant of K_{TS}.

Substituting **Equations 7.41** and **7.42** into **Equation 7.39**, we obtain the Ca^{2+}-dependence binding equation we sought:

$$\frac{b}{b_{max}} = \frac{[Ca^{2+}]}{[Ca^{2+}]_{½} + [Ca^{2+}]} + \frac{b_O}{b_{max}}\frac{[Ca^{2+}]_{½}}{[Ca^{2+}]_{½} + [Ca^{2+}]} \tag{7.43}$$

This equation holds a rectangular hyperbola plus a simple hyperbolic decay term. The first hyperbolic term describes how Ca^{2+} drives the formation of the TIM-Ca^{2+}-PS complex. Such hyperbolic dependence on [Ca^{2+}] is visible in **Figures 7.4a** and **7.5a**. The second term in **Equation 7.43** describes how the Ca^{2+}-independent association of TIM with PS boosts its membrane binding at low [Ca^{2+}]. **Figure 7.6a** provides an illustrative guide for the parameters of **Equation 7.43**.

Equation 7.43 thus describes the dependence of protein binding on both Ca^{2+}-dependent and Ca^{2+}-independent interactions. However, the two hyperbolic terms in this expression do not describe two independent (parallel) reactions; rather, they are interlocked because each is influenced by all three dissociation constants as well as the PS and L concentrations for a given reaction. (This is clear from **Equation 7.39**.) Thus, there is a unique $[Ca^{2+}]_{½}$ value for the binding of a TIM protein to membranes of a given composition and concentration. Practically speaking, comparisons of different TIM proteins such as in **Figure 7.6b** are valid only when performed on identical vesicle preparations.

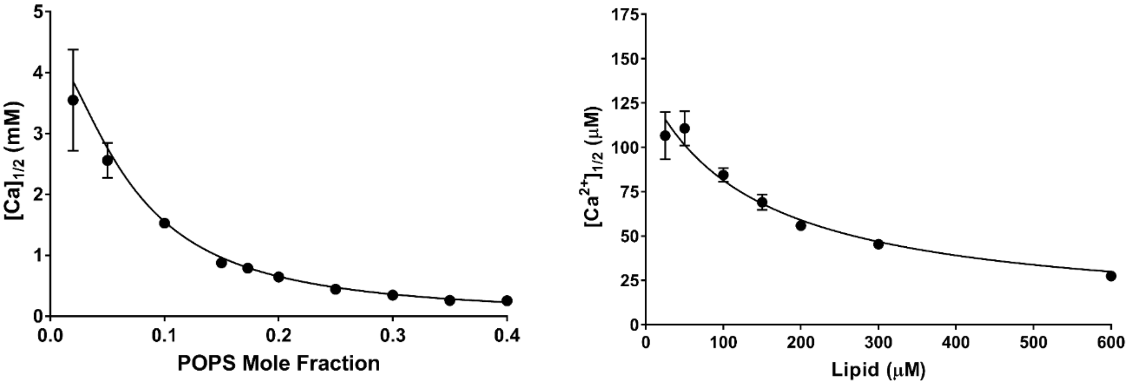


(a)

(b)

Figure 7.6. Ca^{2+} Dependence of TIM1, TIM3 and TIM4 Binding to Membranes. (**a**) A diagram illustrating the parameters for TIM protein binding. b is bound protein, b_{max} is maximal binding at saturating $[Ca^{2+}]$, b_{tot} is the value at saturating $[Ca^{2+}]$ and [PS], and b_0 is Ca^{2+}-independent binding. The red arrow denotes half-maximal binding in the Ca^{2+}-dependent isotherm at $[Ca^{2+}]_{½}$. (**b**) The vesicles (300 µM) were composed of 70% POPC and 30% POPS. The curves were analyzed by non-linear least-squares best fits to **Equation 7.43** and scaled to b_{max}. The $[Ca^{2+}]_{½}$ values obtained for TIM1, TIM3 and TIM4 in units of µM with 95% confidence intervals were 89, [74, 104], 351, [330, 372] and 45, [40, 51], respectively. The ratios of b_0 (fluorescence at $[Ca^{2+}] = 0$) to b_{max} (fluorescence at $[Ca^{2+}] = \infty$), were 0.06 [0.04, 0.07], 0 [0, 0.01], and 0.11 [0.09, 0.12] for TIM1, TIM3, and TIM4, respectively. These values are the intercepts of the fitted curves at the ordinate, while the first data points are at $[Ca^{2+}] = 1.2$ µM arising from the calcium in the buffer. Inset: Blowup of the y-axis intercept.

Our confidence in this analysis is supported by **Figures 7.7a** and **7.7b** in which the dependence of TIM3 and TIM4 binding on $[Ca^{2+}]_{1/2}$ at varied [PS] and [L] concentrations were well fit by theoretical curve generated by **Equation 7.40** using the values in **Table 7.1**.



(**a**) (**b**)

Figure 7.7. Dependence of TIM3 and TIM4 $[Ca^{2+}]_{1/2}$ Values on Membrane Composition and Lipid Concentration. (**a**) Dependence of TIM3 $[Ca^{2+}]_{1/2}$ on [POPS]. $[Ca^{2+}]_{1/2}$ values were obtained for each curve in **Figure 7.5** from non-linear least-squares best fits to **Equation 7.43**. The $[Ca^{2+}]_{1/2}$ values were plotted and the curve was generated from **Equation 7.40** using the values in **Table 7.1**. (**b**) TIM4 $[Ca^{2+}]_{1/2}$ values for each lipid concentration were obtained from the data in **Figure 7.4** by non-linear least-squares best fits to text **Equation 7.43**. The curve was generated from the values in **Table 7.1** using **Equation 7.40**.

7.3.6 Interaction of TIM Proteins with PS and Additional Anionic Phospholipids

While neither PG nor PA supported the binding of TIM3 and TIM4 to bilayers in the absence of PS, substituting either of these phospholipids for a portion of PS strengthened the binding of both TIM3 and TIM4 (**Figure 7.8**). We postulated that these proteins could form complexes through the ternary Ca^{2+} link described above plus Ca^{2+}-independent associations with PS plus PG or PA (represented below as X). **Equation 7.44** was derived in a similar method to that which yielded **Equations 7.33** and **7.34** just by expanding the list of reactions to include association with another anionic lipid. This derivation is omitted here as it follows just the same rationale as the derivation presented in section 7.2. The obtained equation simplified for the TIM proteins is,

$$\frac{b}{b_{tot}} = \frac{\left(\frac{[L]}{K_{TX}} + \frac{[Ca^{2+}][L]}{K_{TXC}^2}\right)\left(\frac{[PS]}{[L]}\right)^{g_{PS}}\left(\frac{[X]}{[L]}\right)^{g_X} \dots}{1 + \frac{[Ca^{2+}]}{K_{TC}} + \left(\frac{[L]}{K_{TX}} + \frac{[Ca^{2+}][L]}{K_{TXC}^2}\right)\left(\frac{[PS]}{[L]}\right)^{g_{PS}}\left(\frac{[X]}{[L]}\right)^{g_X} \dots}$$

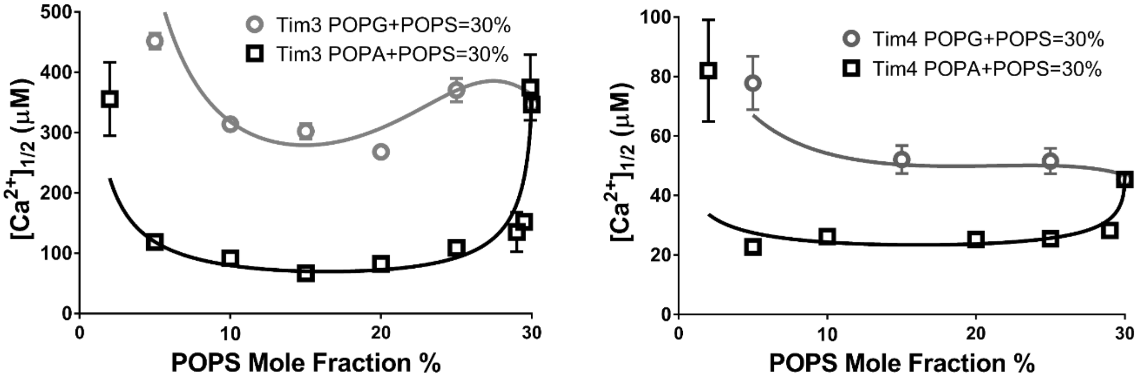
$$\frac{\dots + \left(\frac{[L]}{K_{TS}} + \frac{[Ca^{2+}][L]}{K^2}\right)\left(\frac{[PS]}{[L]}\right)^{h}}{\dots + \left(\frac{[L]}{K_{TS}} + \frac{[Ca^{2+}][L]}{K^2}\right)\left(\frac{[PS]}{[L]}\right)^{h}} \qquad (7.44)$$

The dissociation constants K_{TXC} corresponds to a X-TIM-Ca^{2+}-PS complex and K_{TX} corresponds to a X-TIM-PS complex. **Equation 7.44** omits interactions between the protein and PA or PG in the absence of PS. The exponent, $g_{PS,}$ embodies the multiple cooperative interactions that depend on [PS]. The exponent g_x is limited to cooperative interactions between the protein and [PG] or [PA]. These g exponents do not have the same meaning as the Hill coefficient, h, and cannot be compared to them, because they capture both the cooperative reinforcement and competition (that is negative cooperativity) between PS and PG or PA for the basic residues in the protein.

Equation 7.44 can be rearranged into a form that highlights the hyperbolic dependence of protein binding on [Ca^{2+}], similar to **Equation 7.39**. The resulting expression for [Ca^{2+}]$_{½}$ is given by **Equation 7.45**,

$$[Ca^{2+}]_{½} = \frac{1 + \frac{[L]}{K_{TX}}\left(\frac{[PS]}{[L]}\right)^{g_{PS}}\left(\frac{[X]}{[L]}\right)^{g_X} + \frac{[L]}{K_{TS}}\left(\frac{[PS]}{[L]}\right)^{h}}{\frac{1}{K_{TC}} + \frac{[L]}{K_{TXC}^2}\left(\frac{[PS]}{[L]}\right)^{g_{PS}}\left(\frac{[X]}{[L]}\right)^{g_X} + \frac{[L]}{K^2}\left(\frac{[PS]}{[L]}\right)^{h}} \tag{7.45}$$

In any given experiment, $[Ca^{2+}]_{½}$ behaves as a constant that is specified by all five of the dissociation constants, all three exponents and all of the ligands for the protein.



(**a**) (**b**)

Figure 7.8. Effect of PG and PA on the Membrane Binding of TIM3 and TIM4. The calcium dependence of protein binding was determined using 300 μM vesicles composed of 70 mol% POPC and 30 mol% anionic phospholipids distributed between PS and PA or PS and PG. $[Ca^{2+}]_{½}$ values were estimated from non-linear least-squares fits to Ca^{2+}dependence curves as in **Figures 7.4**, **7.5**, and **7.6**. (**a**) $[Ca^{2+}]_{½}$ for TIM3 binding at different lipid compositions. Data points contained at least 128 triplicate measurements for PG and 299 triplicate measurements for PA. (**b**) $[Ca^{2+}]_{½}$ for TIM1 and TIM4 binding at different lipid compositions. Data points contained 26 triplicate measurements for TIM4 with PG, 48 triplicate measurements for TIM4 with PA, and 24 measurements for TIM1 with PA. The curves for TIM3 and TIM4 were generated by **Equation 7.45** using the values listed in **Table 7.2**.

Equations 7.44 and **7.45** were used to extract the values for the parameters involving PA or PG from the data in **Figure 7.8**. The figure shows how well the model predicts the complex dependence of protein binding on [PS]. **Table 7.2** suggests that TIM3 and TIM4 interact with PA with a lower cooperativity than with PG, as manifested by the exponent g_x. Additionally, the values for the exponent g_{PS} indicate that TIM3 and TIM4 interact with PS with a lower cooperativity in the presence of PA as compared with PG. While both TIM3 and TIM4 have a similar change in cooperativity upon the introduction of PG or PA, only the affinity of TIM3 affinity is largely increased with these lipids (K_x) as compared to its affinity for complexes involving solely PS (K in **Table 7.1**).

Binding Parameters	***TIM3 PS+PG=0.3***	***TIM3 PS+PA=0.3***	***TIM4 PS+PG=0.3***	***TIM4 PS+PA=0.3***
PS Exponent g_{PS}	$1.5^{\dagger}$ [0.8, 2.2]	$0.81^{\dagger}$ [0.65, 0.97]	$0.9^{\dagger}$ [0.5, 1.9]	$0.36^{\dagger}$ [0.29, 0.43]
PG or PA Exponent g_X	$2.0^{\dagger}$ [1.2, 3.0]	$0.76^{\dagger}$ [0.62, 0.93]	$1.4^{\dagger}$ [0.7, 3.5]	$0.49^{\dagger}$ [0.42, 0.56]
K_{TXC} ***for PG- or PA-TIM-*** Ca^{2+}***-PS*** *Dissociation*	13 μM [3, 56]	34 μM [25, 46]	20 μM [0, 5]	41 μM [34, 47]
K_{TX} ***for PG- or PA-TIM-PS*** *Dissociation*	Immeasurably weak	Immeasurably weak	140 μM [1, 790]	790 μM [550, 1030]

Table 7.2. Values for the Binding Parameters of TIM3 and TIM4 to PA- and PG-Containing Vesicles. Values for TIM3 and TIM4 were obtained by fitting the data from **Figure 7.8a** and **Figure 7.8b** to **Equation 7.**44 utilizing the values given in **Table 7.1**. Brackets give 95% confidence intervals. † indicates >94% confidence that the values are different between PA and PG as calculated using a z test.

7.4 Structural and Energetic Analysis of TIM Binding

7.4.1 Introduction

The binding experiments depicted in section 7.3 and quantified by the equations developed in section 7.2 to provide a quantitative basis for the comparison of the TIM proteins. In this section, these and additional experiments are analyzed through the lens of structural and energetic mechanisms of binding. In order to extract this information from experiments with the various lipid compositions described in section 7.1.3, we require a set of parameters that are globally defined for each protein and membrane condition. The parameters of **Equations 7.33** or **7.34** offer exhaustive descriptions but are difficult to compare for a broad set of different lipids for which there would be additional dissociation constants for each additional lipid species. Instead, **Equations 7.40** and **7.45** suggest that $[Ca^{2+}]_{½}$ can always be formulated in terms of the relevant dissociation constants for a given membrane, regardless of the specific components of that membrane. We therefore use this parameter to compare the binding of the TIM proteins with various membranes and analyze the energetic contributions that give rise to their different affinities.

7.4.2 Hydrophobic Interactions Are Requisite Drivers of TIM Binding

In order to compare the contribution of hydrophobic burial and lipid packing, we conducted experiments varying the acyl chains of membrane lipids. The $[Ca^{2+}]_{½}$ values obtained for the experiments with 7:3 POPC:POPS in **Figure 7.6** indicate that TIM3 bound about four times more weakly than did TIM1 and about eight times more weakly than did TIM4. This characterization is consistent with others using the same model membranes (1, 17), further supporting the utility of $[Ca^{2+}]_{½}$ as a general binding parameter. These experiments also gave values for the calcium-independent binding of the proteins. b_0/b_{max} was 0.06 for TIM1, 0 for TIM3 and 0.11 for TIM4.

We next determined the binding of TIM3 to bilayers composed of phospholipids with 7:3 PC:PS headgroups and varied fatty acyl chains. In particular, we compared phospholipid vesicles with monounsaturated PO tail groups against their more expanded di-unsaturated DO and more condensed saturated DM counterparts. As seen in **Figure 7.9a**, substituting DM for PO reduced protein binding below the limit of detection. Substituting DO lipids for PO lipids caused a ten-fold increase in the TIM3 binding affinity; that is, $[Ca^{2+}]_{½}$ decreased from 405 µM to 41 µM. The effect of the di-unsaturated form of phospholipid on the $[Ca^{2+}]_{½}$ varied linearly with its abundance in the membrane (**Figure 7.9b)**. This linear dependence implies that the acyl chains act as a two-dimensional buffer, changing the values of all dissociation constants contained implicitly in **Equation 7.32**. These lipids do not interact with TIM3 in a specific liganded association; otherwise the dependence would solely be on the amount of DOPS in the membrane and not the amount of DO tails in general.

We also determined the sensitivity of TIM4 to lipid packing by comparing the effect of DO with PO tails. As seen in **Figure 7.9c**, 7:3 DOPC:DOPS lipids bound TIM4 much more avidly than did their 7:3 POPC:POPS counterpart. In particular, the $[Ca^{2+}]_{½}$ for the data in **Figure 7.9c** decreased from 48 µM, 95% [44, 52] to 15 µM, 95% CI [12, 20]. Of note, the calcium-independent fraction of the binding of Tim4, b_0/b_{max}, increased from 11% [9, 12] to 38% [35, 41] for the less saturated phospholipid species.

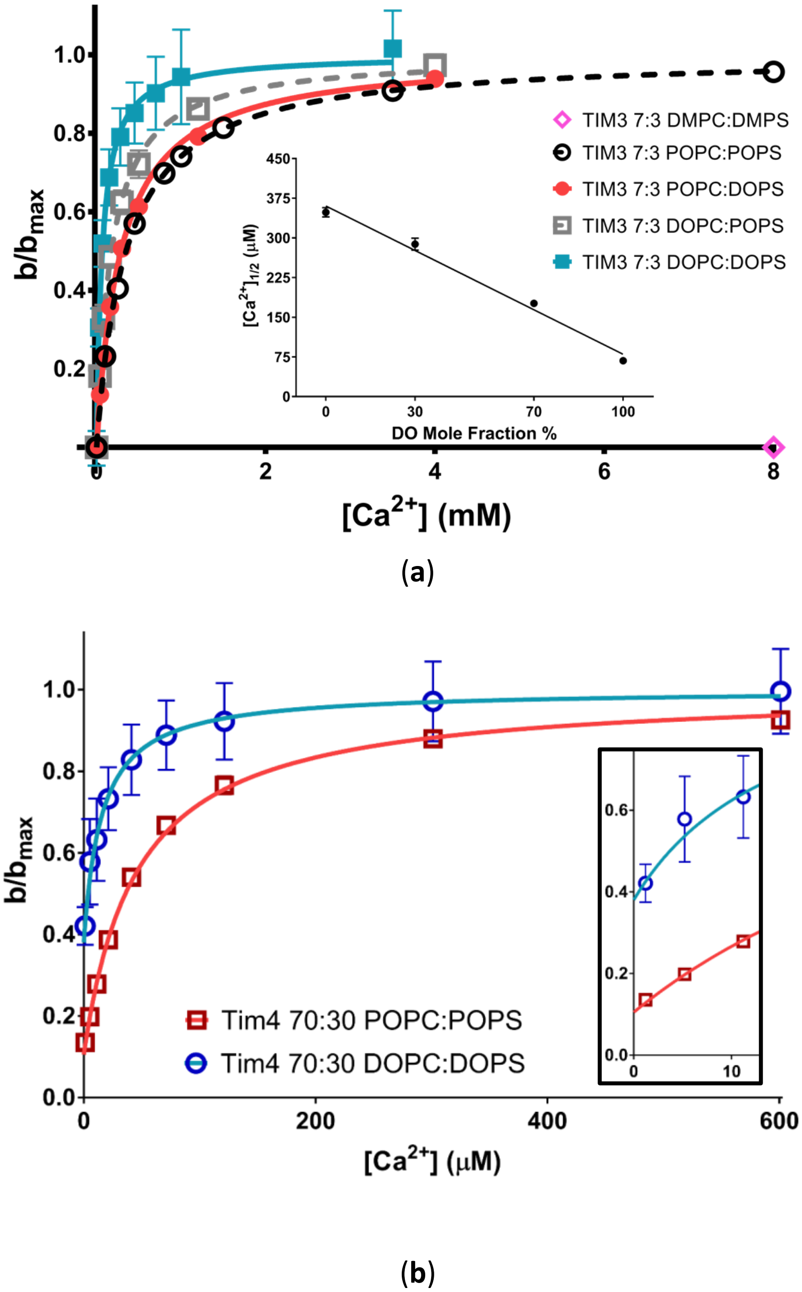


Figure 7.9. Effect of Phospholipid Unsaturation. (**a**) TIM3 binding to 300 μM vesicles composed of 7:3 PC:PS containing either saturated DM, mono-unsaturated PO or di-unsaturated DO phospholipids. The data were analyzed by non-linear least-squares best fits to **Equation 7.43** and the bound protein signal, *b*, scaled to its value at infinite [Ca^{2+}], b_{max}. Inset: TIM3 [Ca^{2+}]½ values obtained from the data were plotted as a function of the DO content of the bilayers.

Figure 7.9 cont. (**b**) TIM4 binding to 300 µM vesicles composed of 7:3 PC:PS containing either PO or DO phospholipids. The $[\mathrm{Ca}^{2+}]_{½}$ decreased from 48 µM, 95% CI [44, 52] with the 70:30 POPC:POPS to 15 µM, 95% CI [12, 20] with the 70:30 DOPC:DOPS. The first data points were at $[\mathrm{Ca}^{2+}]$ = 1.2 µM arising from calcium in the buffer. Estimated as in **Figure 7.6**, the b_0/b_{max} values (fraction bound at $[\mathrm{Ca}^{2+}]$ = 0) were 0.11 [0.09, 0.12] for 7:3 POPC:POPS and 0.38 [0.35, 0.41] for 7:3 DOPC:DOPS. Inset: Blowup of the y-axis intercept.

These results are consistent with the greater hydrophobic interaction of TIM1 and TIM4 than TIM3 revealed by the spectra in **Figure 3.2** of Chapter 3 as well as their membrane-bound states in Chapters 5 and 6. While the spectra make it clear that all three proteins make contact with the bilayer core, the increased binding of TIM1 and TIM4 in the absence of Ca^{2+}may be promoted by their stronger hydrophobic associations. This analysis is consistent with mutational data showing that alanine substation of these hydrophobic residues heavily reduce binding to PS (17–19). Taken together, these results imply that the hydrophobic interactions constitute a necessary component of membrane-binding and primarily underlie the differences in membrane affinity of the TIM proteins.

7.4.3 Electrostatic Interactions of Peripheral Basic Residues Underlie Cooperativity

We sought to characterize the cooperativity of the TIM proteins for anionic lipids through analysis of $[\mathrm{Ca}^{2+}]$ dependence experiments. As a baseline, we investigated the dependence of $[\mathrm{Ca}^{2+}]_{½}$ on PS alone before considering additional lipids. **Figure 7.5a** shows that $[\mathrm{Ca}^{2+}]$ dependence was hyperbolic at multiple [PS] values, confirming the assumptions underlying **Equation 7.34**. The $[\mathrm{Ca}^{2+}]_{½}$ values corresponding to this experiment, plotted in **Figure 7.7a**, illustrates that the strength of protein binding to the membrane increased with [PS], as shown by the decreasing

$[Ca^{2+}]_{½}$ with increasing [PS]. The rate of decay is exactly inversely proportional to $\left({[PS]}/{[L]}\right)^{h}$, as predicted by **Equation 7.43**, showing that $[Ca^{2+}]_{½}$ explicitly contains a complete description of the cooperativity of the binding.

A complementary experiment examined the dependence of the binding of TIM4 on the total lipid concentration of the vesicles, as shown in **Figure 7.4a**. As was the case in **Figure 7.6**, **Figure 7.4b** reveals a small but reproducibly significant binding of TIM4 in the absence of calcium. Furthermore, a plot of the dependence of $[Ca^{2+}]_{½}$ on total lipid concentration (**Figure 7.7b**) revealed that, as was the case for TIM3, $[Ca^{2+}]_{½}$ values for TIM4 varied with reactant concentration.

Addressing the influence of additional anionic lipids, we examined the binding of TIM3 and TIM4 to vesicles of POPC and POPS augmented with either POPG or POPA, keeping the total anionic phospholipid content at 30 mol%. Membranes in which POPS was entirely replaced by either POPG or POPA were not bound by the TIM proteins (not shown). However, vesicles containing small amounts of either of these other two lipids bound TIM3 more strongly than those simply containing POPC and POPS (**Figures 7.8a**).

Notably, POPA dramatically decreased the $[Ca^{2+}]_{½}$ for TIM3. One mol% POPA reduced the $[Ca^{2+}]_{½}$ from 347 µM to 138 µM, and 10 mol% PA decreased the $[Ca^{2+}]_{½}$ to 84 µM. (One mol% corresponds to ~9 POPA in the bilayer for each Tim3 added to the reaction.) The gap in membrane affinity between TIM3 and TIM4 or TIM1 is significantly decreased with the addition of PA to the membrane (**Figure 7.10**). The saturation of PA stimulation at ~5 mol% suggests that it interacts specifically with TIM3 rather than affecting the bulk properties of the bilayer. POPA had no appreciable effect on TIM1 (**Figure 7.10**); the binding of TIM4 was modestly augmented by POPA but not by POPG (**Figures 7.8b**).

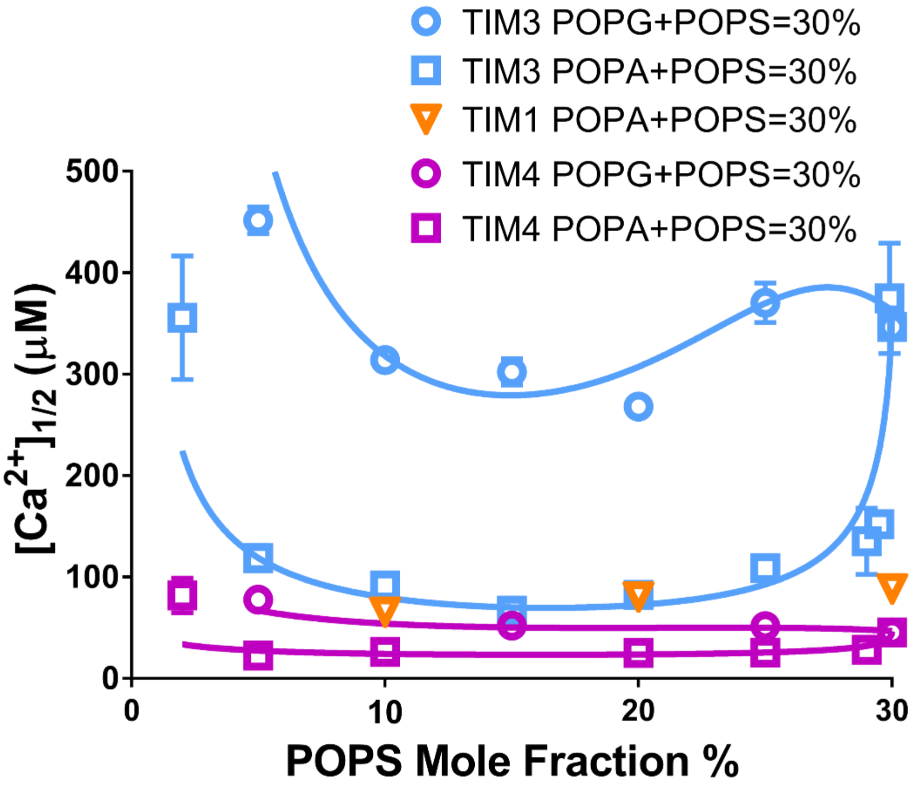


Figure 7.10. Effect of PG and PA on the Membrane Binding of TIM1, TIM3, and TIM4. This is a replot of **Figure 7.8** with additional TIM1 experiments. The calcium dependence of protein binding was determined using 300 μM vesicles composed of 70 mol% POPC and 30 mol% anionic phospholipids distributed between PS and PA or PS and PG. TIM1 data points contained 24 measurements. The curves for TIM3 and TIM4 were generated by **Equation 7.45** using the values listed in **Table 7.2**.

Note that the binding of TIM3 and TIM4 diminished at [PS] < 10 mol%. We speculate that 10 mol% POPS was sufficient to fulfill the calcium coordination requirement. Increasing the total anionic phospholipid beyond 10 mol% with POPS, POPG or POPA strengthened binding through electrostatic interactions with peripheral basic residues. Presumably, the extra charge on PA molecules (approaching -2 under our experimental conditions) increases its affinity for the basic residues in the protein but, in doing so, decreases the magnitude of the cooperativity. Thus, the binding of TIM3 or TIM4 to PA likely reduces the molecularity of the parallel reactions with multiple singly-charged PG or PS molecules.

7.4.4 The Ca^{2+} Link of the Central Pocket Contributes to Affinity but not Necessarily Cooperativity

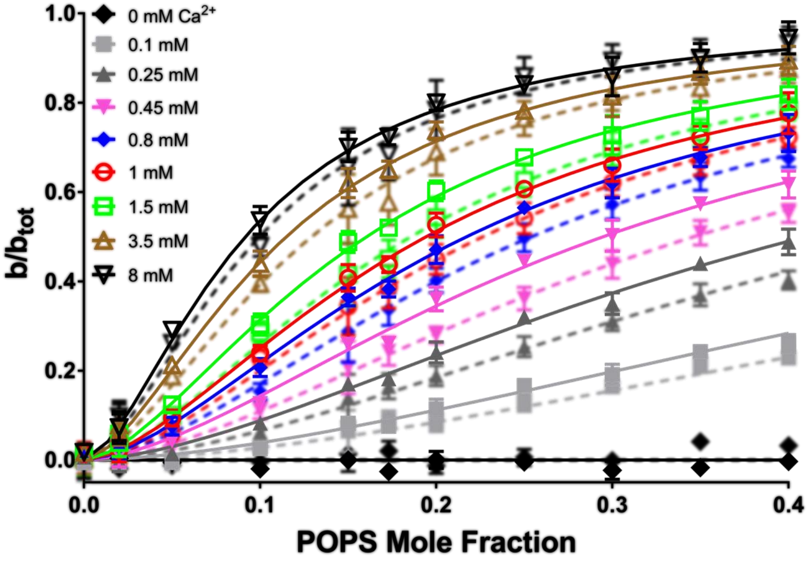
While $[Ca^{2+}]_{½}$ allows us to probe the affinity and cooperativity of the interaction, it does not directly reveal the molecular interactions governing these thermodynamic quantities. While we were able to use various membranes to isolate the contributions of hydrophobic and peripheral basic residues, the contributions of the pocket are not so cleanly separated. The pocket involves the coordination of polar and charged residues with PS and Ca^{2+}, there is no distinct set of interaction forces which specifically report on these energetic contributions.

In Chapter 6, we compared the structures of two allelic variants of TIM3, TIM3 Balb/c (the variant used in the experiments throughout this chapter) and TIM3 HBA. The latter variant was previously shown to exhibit weaker affinity to PS-containing membranes as well as a reduced ability to effect phagocytosis when expressed on the surface of macrophages (17). The authors attributed this weaker binding to a peripheral loop that contains only polar residues (and is the site of the point mutations between the two variants). Our MD simulations revealed that this loop infrequently makes direct contacts with the membrane and instead the effect of these mutations is to weaken the Ca^{2+} and PS link of the central pocket. Furthermore, the hydrophobic and peripheral residues of TIM3 HBA had the same membrane penetration and association as those of TIM3 Balb/c. Thus, the simulations indicate the two differ only in the central pocket.

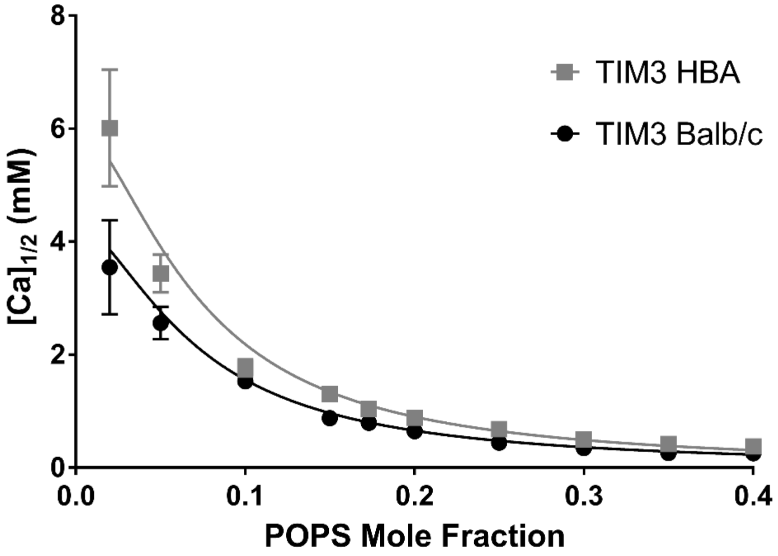
We performed a series of Ca^{2+} and PS dependence experiments for TIM3 HBA on the exact same vesicles samples as we did for TIM3 Balb/c in **Figure 7.5**. These two experiments are compared in **Figure 7.11** and the obtained parameter values in **Table 7.3**. The PS dependence curves are nearly identical for the two variants with TIM3 HBA displaying consistently lower binding at the PS mole fraction. However, the sigmoidicity of the two families of curves looks nearly the same. This is reflected in the dependence of their respective $[Ca^{2+}]_{½}$ values displaying similar curvature.

The values of the parameters reveal that both variants share the same value of the Hill coefficient. This is consistent with the data presented in Chapter 6, as the equivalent cooperativity of the two variants correspond to their equivalent membrane association of peripheral basic residues. However, the dissociation constants for the TIM3-Ca^{2+} and TIM3-Ca^{2+}-PS complexes significantly differ. The weaker central pocket of TIM3 HBA, implicated in Chapter 6, manifests as a lower affinity for both Ca^{2+} and the membrane.

These results indicate that the ability of the central pocket to form a tight coordination of Ca^{2+} and PS is a significant component of the binding as well and correlates with the overall affinity of the interaction. In principle, the pocket could contribute to the cooperativity by affecting the number of multi-PS bound states. However, it just happens to be the case that the TIM3 HBA variant seemingly retains its ensemble of multi-PS bound states despite a weaker pocket. Extrapolating these findings to TIM4 and TIM1 leads us to conclude that their higher affinity is, in part, due to stronger central pockets. The strength of the central pocket is likely entangled with the insertion of the hydrophobic residues as the degree of insertion determines the ability of these proteins to sequester PS into their pockets.



(**a**)



(**b**)

Figure 7.11. Dependence of TIM3 Balb/c and TIM3 HBA Binding on [Ca^{2+}] and [PS]. (**a**) TIM3 Balb/c (solid lines) and TIM3 HBA (dashed in the background) binding to 300 µM vesicles of the indicated [Ca^{2+}] concentrations was determined as a function of [PS]. The data contained 122 triplicate (or more) measurements. The curves were generated by non-linear least-squares best fits to **Equation 7.37**, yielding the values in **Table 7.3**. *b* is bound protein and b_{tot} is binding at infinite [Ca^{2+}] and [PS]. The curves were generated from the values given in **Table 7.3** using **Equation 7.34**. (**b**) Dependence of TIM3 Balb/c (black circles) and TIM3 HBA (gray squares) [Ca^{2+}]$_{½}$ on [POPS]. [Ca^{2+}]$_{½}$ values were obtained for each curve in **Figure 7.5** and panel **a** for TIM3 Balb/c and TIM3 HBA respectively from non-linear least-squares best fits to **Equation 7.43**. The [Ca^{2+}]$_{½}$ values were plotted and the curve was generated from **Equation 7.40** using the values in **Table 7.3**.

Binding Parameter	***TIM3 Balb/c***	***TIM3 HBA***
Hill coefficient, h	1.65 [1.59, 1.72]	1.69 [1.63, 1.75]
K for TIM-Ca^{2+}-PS dissociation	127 μM [120, 136]	145 μM [135, 155]
K_{TC} for Ca^{2+}-TIM binary complexes	4300 μM [3600, 5200]	6100 μM [5100, 7300]

Table 7.3. Values for TIM3 Balb/c and TIM3 HBA Binding Parameters to PS-Containing Vesicles. Estimates for TIM3 Balb/c and TIM3 HBA were obtained by fitting the data from **Figures 7.5** and **7.11** respectively to **Equation 7.37**. Brackets give 95% confidence intervals. The values of the dissociation constants for TIM3 Balb/c and TIM3 HBA differ at >95% confidence as calculated using a z test.

7.4.5 A Heuristic Model for the Energetic Contributions to TIM Binding

The binding of the TIM proteins to the membranes investigated in this chapter indicate multiple energetic components. Here we summarize these components into three broad categories. These are (a) the calcium coordination between the central pocket of the protein and PS; (b) the electrostatic interaction of basic residues neighboring the central pocket with anionic phospholipids; (c) the membrane insertion of their hydrophobic residues and the accommodation of protein burial by the packing of bilayer phospholipids. These features are discussed below.

(a) Ca^{2+} coordination with PS is clearly a prominent driver of TIM binding**.** That is, association with bilayers was not observed for the TIM proteins in the absence of PS while Ca^{2+} is required for the binding of TIM3 and significantly enhances the affinity of TIM4 for PS. Ca^{2+} is also hyperbolic for all of the TIM proteins, supporting the predictions of the membrane-bound structures that only the central pocket is involved in the coordination of Ca^{2+}. We used the affinity for [Ca^{2+}] to measure the binding of the remaining energetic components, indicating the interdependence of Ca^{2+} on electrostatic and hydrophobic components of binding. This was further confirmed in our investigations of TIM3 HBA, which displayed a weaker membrane affinity due to a weaker central pocket.

(b) The electrostatic interaction between the basic residues on TIM3 and TIM4, and anionic phospholipids was manifested by the sigmoidal dependence of their binding on [PS]. This phospholipid comprises ~10 mol% of the bilayer of many mammalian plasma membranes. Presumably, half of the PS moves to the outer leaflet of the scrambled plasma membrane bilayer of apoptotic cells. This concentration of PS is in the low range of the sigmoidal rise in both TIM3 and TIM4 binding at physiological [Ca^{2+}]. Other cytoplasmic-surface anionic phospholipids, namely PA, PG, phosphatidylinositol and its derivatives, should also scramble to the outer surface of triggered plasma membrane bilayers. While neither PA nor PG supported the TIM-Ca^{2+} link to the bilayer, both outperformed PS in strengthening TIM3 binding once the Ca^{2+} linkage was saturated with PS. POPG was instructive: while only a minor constituent of mammalian plasma membranes, it has the same unitary net negative charge and nearly the same cross-sectional area as POPS yet differs in its interaction with TIM3 and TIM4. It appears that the polar head groups of these phospholipids differ in (i) their ability to form the Ca^{2+} coordination complex; (ii) their electrostatic association with basic residues; (iii) their molecular geometry; and (iv) their intimate stereochemical interactions with the membrane contact surface of the protein.

(c) A third feature of TIM binding is through its hydrophobic interaction with the bilayer core. This feature was made clear by the spectral shift of tryptophan fluorescence that reports its transfer to a less polar environment (**Figure 3.2** in Chapter 3). This was also evidenced by the increased binding of TIM3 and TIM4 to membranes containing DO phospholipids. It has been shown elsewhere that increasing the unsaturation of the phospholipid acyl chains promotes the membrane association of peripheral proteins through the creation of packing defects (20, 21).

Our study offers some insights into how these four factors might contribute to differences in the membrane affinity of the various TIM proteins (**Table 7.4**). First, both the overall bilayer affinity of TIM3 and its affinity for calcium are weaker than those of TIM1 and TIM4 (**Figure 7.6** and **Table 7.1**). The weak calcium affinity of TIM3 (K_{TC} = 4.3 mM) raises the possibility that, in vivo, its membrane binding varies linearly with available calcium (~1 mM). In contrast, the strong affinity of TIM4 (K_{TC} = 0.134 mM) suggests that this protein is always saturated with calcium *in vivo*. Secondly, the contrasting disposition of the peripheral basic residues on TIM3 and TIM4 suggest that their energetic contribution to binding might be smaller for TIM3 than TIM4. Our data quantified this difference: the K_{TS} value was smaller (stronger) for TIM4 and its *h* value was greater (also stronger), implicating a larger energy in their association with PS as compared with of TIM3. In fact, it may be that TIM4 and TIM1 respond predominantly to membrane PS and is minimally sensitive to the other factors investigated in this work. In contrast, TIM3 is highly sensitive to conditions that accompany apoptotic PS exposure, namely the scrambling of additional anionic phospholipids.

	Inferred contribution to physiological binding		
Tim protein	**Ca^{2+} saturation**	**Basic residues**	**Hydrophobic residues**
1	Strong	Weak	Strong
3	Weak	Moderate	Weak
4	Strong	Strong	Strong

Table 7.4. Comparison of Energetic Contributions to the Binding of the TIM proteins. "Ca^{2+} saturation" refers to the ternary Ca^{2+} coordination in the central binding pocket. "Basic residues" refers to electrostatic interactions with anionic phospholipids. "Hydrophobic residues" refers to partitioning into the core of the bilayer, related to the relative resistance to bilayer insertion due to the saturation (packing) of the phospholipid chains.

In addition to its contributions to the electrostatic component of binding, PA was a strong ligand for TIM3 compared to PS and PG. Certain molecular features can account for its potency. At our experimental pH of 7.5, the diacidic headgroup of PA has a negative charge of ~1.5 while PS and PG have a single negative charge and PC is net neutral (22–25). The association of PA with the basic residues on peripheral proteins can induce further deprotonation, bringing its charge close to -2 (22). Variation in ambient pH could therefore affect Tim3 binding to cells *in vivo*, perhaps pathophysiologically. Note that the PA molecule is distinguished by a conical contour, and its head group is small and lies close to the nonpolar stratum. It could also be that the positive charge on the headgroup of PS reduces its electrostatic interaction with the basic residues on the protein, while this would not be the case for PA or PG. All of these features would facilitate the access of the hydrophobic residues on TIM3 to the bilayer core (26, 27).

7.5 Conclusions

Many studies of plasma membrane binding proteins such as the Tim family have inferred Hill coefficients and apparent affinity constants. However, we suggest that the underlying significance of these and other features is best appreciated through a formal analysis such as the mathematical model presented in the Supplement. This model provides values for the relevant binding parameters and allows quantitative comparisons of the sister proteins. Furthermore, it lends insight into and makes use of such features as the calcium-independent binding of TIM1 and TIM4 in **Figure 7.6**. The model also explains the characteristic variations in the plateau values of the [Ca^{2+}], [PS] and [L]-dependent binding curves (**Figures 7.4a**, **7.4b**, and **7.5a**). [The reason is that the ligand held constant limits binding no matter how great the independent variable (see **Equation 7.**41)]. Most importantly, the model lays bare the complex composition of the apparent dissociation constants as well as the useful parameter, $[Ca^{2+}]_{½}$. This approach not only sheds light on earlier studies (e.g., (28–30)) but also provides hypotheses for future investigations. The mathematical model could serve as a starting point for analyzing the membrane association of other proteins utilizing multiple molecular links.

7.6 References

1. Tietjen GT, et al. (2014) Molecular mechanism for differential recognition of membrane phosphatidylserine by the immune regulatory receptor Tim4. *Proc Natl Acad Sci U S A* 111(15):E1463–72.
2. Kerr D, et al. (2018) Sensitivity of peripheral membrane proteins to the membrane context: A case study of phosphatidylserine and the TIM proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1860(10):2126–2133.
3. Tait JF, Gibson DF, Smith C (2004) Measurement of the affinity and cooperativity of annexin V-membrane binding under conditions of low membrane occupancy. *Anal Biochem* 329(1):112–119.
4. Appelt U, et al. (2005) Viable, apoptotic and necrotic monocytes expose phosphatidylserine: cooperative binding of the ligand Annexin V to dying but not viable cells and implications for PS-dependent clearance. *Cell Death Differ* 12(2):194–196.
5. Sengupta T, Manoj N (2016) Phosphatidylserine and Phosphatidylethanolamine Bind to Protein Z Cooperatively and with Equal Affinity. *PLoS ONE* 11(9):e0161896.
6. Tietjen GT, et al. (2017) Coupling X-Ray Reflectivity and In Silico Binding to Yield Dynamics of Membrane Recognition by Tim1. *Biophys J* 113(7):1505–1519.
7. Yabuuchi H, O’Brien JS (1968) Positional distribution of fatty acids in glycerophosphatides of bovine gray matter. *J Lipid Res* 9(1):65–67.
8. Wood R, Harlow RD (1969) Structural analyses of rat liver phosphoglycerides. *Arch Biochem Biophys* 135:272–281.
9. Kucerka N, Tristram-Nagle S, Nagle JF (2005) Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J Membr Biol* 208(3):193–202.
10. Mosior M, Newton AC (1998) Mechanism of the apparent cooperativity in the interaction of protein kinase C with phosphatidylserine. *Biochemistry* 37(49):17271–17279.
11. Segel IH (1975) Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems.
12. Adair GS, Others (1925) The hemoglobin system VI. The oxygen dissociation curve of hemoglobin. *Journal of Biological Chemistry* 63(2):529–545.
13. Wyman J (1964) LINKED FUNCTIONS AND RECIPROCAL EFFECTS IN HEMOGLOBIN: A SECOND LOOK. *Adv Protein Chem* 19:223–286.
14. Abeliovich H (2005) An empirical extremum principle for the hill coefficient in ligand-protein interactions showing negative cooperativity. *Biophys J* 89(1):76–79.

15. Stefan MI, Le Novère N (2013) Cooperative binding. *PLoS Comput Biol* 9(6):e1003106.

16. London WP, Steck TL (1969) Kinetics of enzyme reactions with interaction between a substrate and a (metal) modifier. *Biochemistry* 8(4):1767–1779.

17. DeKruyff RH, et al. (2010) T cell/transmembrane, Ig, and mucin-3 allelic variants differentially recognize phosphatidylserine and mediate phagocytosis of apoptotic cells. *J Immunol* 184(4):1918–1930.

18. Santiago C, et al. (2007) Structures of T cell immunoglobulin mucin protein 4 show a metal-Ion-dependent ligand binding site where phosphatidylserine binds. *Immunity* 27(6):941–951.

19. Ballesteros Morcillo Á (2012) Caracterización de las protenas TIM como receptores de fosfatidilserina. Doctoral dissertation.

20. Pande AH, Qin S, Tatulian SA (2005) Membrane fluidity is a key modulator of membrane binding, insertion, and activity of 5-lipoxygenase. *Biophys J* 88(6):4084–4094.

21. Vanni S, et al. (2013) Amphipathic lipid packing sensor motifs: probing bilayer defects with hydrophobic residues. *Biophys J* 104(3):575–584.

22. Kooijman EE, et al. (2007) An electrostatic/hydrogen bond switch as the basis for the specific interaction of phosphatidic acid with proteins. *J Biol Chem* 282(15):11356–11364.

23. Shin JJH, Loewen CJR (2011) Putting the pH into phosphatidic acid signaling. *BMC Biol* 9:85.

24. Loew S, Kooijman EE, May S (2013) Increased pH-sensitivity of protein binding to lipid membranes through the electrostatic-hydrogen bond switch. *Chem Phys Lipids* 169:9–18.

25. Putta P, et al. (2016) Phosphatidic acid binding proteins display differential binding as a function of membrane curvature stress and chemical properties. *Biochim Biophys Acta* 1858(11):2709–2716.

26. Kwolek U, et al. (2015) Effect of phosphatidic acid on biomembrane: experimental and molecular dynamics simulations study. *J Phys Chem B* 119(31):10042–10051.

27. Dickey A, Faller R (2008) Examining the contributions of lipid shape and headgroup charge on bilayer behavior. *Biophys J* 95(6):2636–2646.

28. Bazzi MD, Nelsestuen GL (1991) Highly sequential binding of protein kinase C and related proteins to membranes. *Biochemistry* 30(32):7970–7977.

29. Jeppesen B, Smith C, Gibson DF, Tait JF (2008) Entropic and enthalpic contributions to annexin V-membrane binding: a comprehensive quantitative model. *J Biol Chem* 283(10):6126–6135.

30. Tran HT, Anderson LH, Knight JD (2019) Membrane-Binding Cooperativity and Coinsertion by C2AB Tandem Domains of Synaptotagmins 1 and 7. *Biophys J* 116(6):1025–1036.

CHAPTER 8

CONCLUSIONS AND FUTURE DIRECTIONS

In this work, we have resolved the membrane-bound structures of TIM proteins to a high level of detail by combining and corroborating x-ray reflectivity with MD simulations. These structures were then explained the binding of the TIM proteins on a variety of phospholipid membranes, as measured by tryptophan fluorescence. Furthermore, the combination of these techniques resolved interactions between specific residues and lipids, providing a blueprint for how membranes are specifically selected by proteins. This work also provides a much-needed basis for studies that design membrane-binding proteins from first principles or otherwise target proteins to specific membranes.

As claimed in the Introduction, the structures of the TIM proteins correlate with their differential sensitivity for phospholipid membranes. **Figure 8.1** compares the membrane-bound structures of TIM1, TIM3, and TIM4 obtained in Chapters 5, 6, and 4, respectively. Side-by-side, it seems evident that the geometry of the pocket is a major determinant of the orientation and insertion of the bound state. The three proteins conserve the same calcium- and PS-coordinating residues in the central pocket. These residues are the most important for binding and together represent the strongest PS coordination the proteins can make. The pocket is most likely oriented to maximize the strength of this interaction but minimally disturb the alignment of the bound PS with the membrane. The second largest determinant for the bound state is the FG loop. This loop contains hydrophobic residues that, in solution, are solvent-exposed. The burial of these residues is highly energetically favorable, but their insertion necessitates the burial of a greater protein surface area. Despite the CC’ loop of TIM3 containing a tryptophan residue, the reflectivity-derived structure indicates that the loop resides predominantly in the head group region. The MD simulations reveal that this tryptophan infrequently partitions into the tail group region. The surface burial required to place the tryptophan in the tail group region, is evidently too energetically unfavorable. The importance of the hydrophobic residues and surface burial is demonstrated in the experiments with DO lipids in which the energetic penalty of surface burial is reduced.

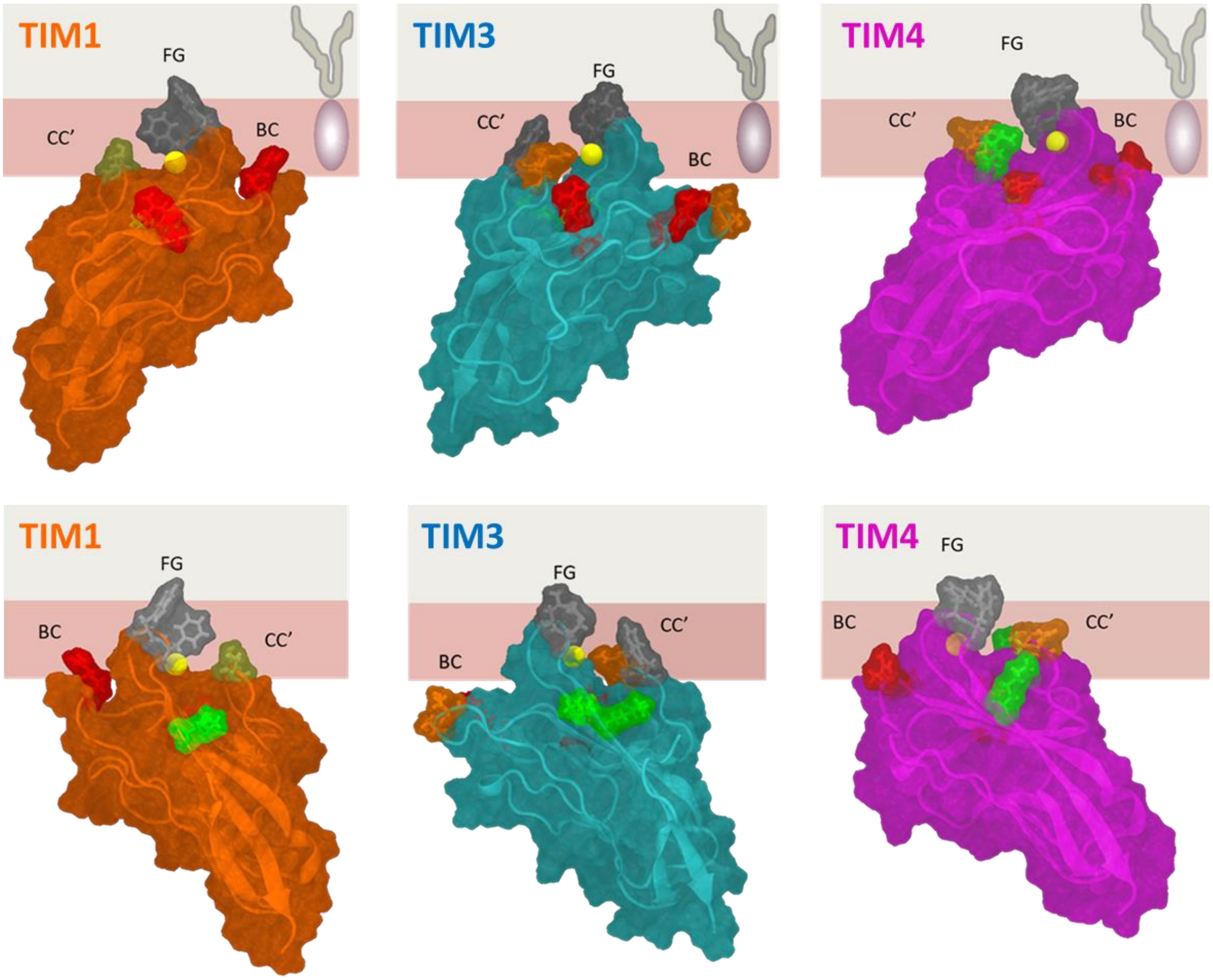


Figure 8.1. Membrane-Bound Structures of the TIM Proteins. The structures for TIM1, TIM3, and TIM4 were obtained from Chapters 5, 6, and 4, respectively. Hydrophobic residues are in gray, arginine in red, lysine in green, and polar residues in gold.

In addition to the pocket and hydrophobic residues, the peripheral, basic residues have some influence on the overall bound state. TIM1 contains the lowest number and least engagement of these residues at the membrane interface. The bound-state seems to be optimally oriented to maximize insertion of the hydrophobic residues on the FG loop and minimize overall surface burial. TIM4 contains the same pocket and hydrophobic tail but has an additional lysine residue near its CC' loop. The bound state of TIM4 places the hydrophobic residues at the same depth as those of TIM1, but the rest of the protein is significantly more buried as a result of the shallower tilt of the protein. The interactions of its peripheral residues with anionic lipids is apparently sufficient to overcome the cost of increased surface burial of the protein. On the contrary, the peripheral residues of TIM1 are either fully satisfied at a lower insertion or the energy gained from further insertion is not sufficient to outweigh the penalty of increased burial. This tracks with the significantly lower Hill coefficient of TIM1, as TIM1 displays the most hyperbolic dependence to membrane PS of the TIM proteins.

TIM3 represents somewhat of an aberration compared to TIM1 and TIM4. Firstly, its central binding pocket is weaker as evidenced by its lower affinity for calcium in solution and its inability to bind PS in the absence of calcium. Its weaker pocket is most likely a consequence of its weaker hydrophobic contact. This is demonstrated in its less pronounced tryptophan spectral shift upon binding, shown in Chapter 3. The studies with the allelic variant, TIM3 HBA, suggest that other distant residues might also contribute to this weaker pocket. The peripheral basic residues of TIM3 are also more withdrawn than those of TIM4 and TIM1. Its two arginine residues are located on more distant β-strands as opposed to the surface-adjacent BC loop. As a result, the peripheral residues are less able to effect surface burial on par with TIM4. The MD simulations reveal that these residues are in near-constant engagement with the lipid headgroups and, unlike TIM1, have a high preference for PS over PC. This suggests that TIM3 requires more anionic charge to overcome its weaker hydrophobic and calcium-coordinating components. Confirming this speculation, the binding of TIM3 was greatly enhanced on PG- and, especially, PA-containing membranes while the binding of TIM4 and TIM1 was minimally enhanced.

In addition to providing a molecular basis for membrane sensitivity, the obtained structures also informed a change in our paradigm for peripheral membrane-binding proteins. The notion of cooperativity is generally mathematically defined as an increased or decreased affinity for a ligand after some ligand-binding has already occurred. The structural basis for this change in affinity does not have an equally general conception and is instead based on the study of just one protein, hemoglobin. The concept of cooperativity is often applied to proteins based on the same model of protein conformational changes upon subsequent ligand binding. On the contrary, the binding of the TIM proteins introduces minimal conformational changes to the protein and instead induces the largest change on their interface with the membrane. The most obvious demonstration of this is the three- to two-dimensional transition of the search space that occurs after the first association with the membrane. Trivially, the affinity of the central pocket is much higher for PS once the protein is already at the membrane, as the dimensionality of its search for PS is reduced. However, this is far from the only contribution. The crowded environment of the membrane induces multiple lipid-bound configurations of the protein, as we saw for the simulations of TIM3 in Chapter 6. If the engagement of PS by any subset of these residues is greater when multiple PS are engaged, then this will manifest as positive cooperativity. However, the cooperativity is positive only up to a point, as the increased interface with the membrane upon binding PS necessitates an increased surface burial. As discussed in the previously, the surface burial is an energetic penalty that offsets the energy gained from these lipid associations.

This analysis suggests that cooperativity is best conceptualized as a balance of the proclivity for the protein to bind more lipids but at the cost of increased surface burial. Previously, we compared the cooperativity of the TIM proteins purely in terms of their Hill coefficients for PS, but this is a reductive approach that can miss the cooperative contributions of membrane features such as membrane charge or lipid packing. If we apply this new notion of cooperativity to the TIM proteins, we see that the Hill coefficients for PS are explainable in terms of structural features. With the highest Hill coefficient, TIM4 also the maximum benefit-cost ratio as it has the highest degree of engagement of peripheral basic residues and the greatest surface burial. Having the lowest Hill coefficient, TIM1 represents the opposite extreme with the least engagement of peripheral basic residues and the lowest surface burial. TIM3 represents a nuanced case of cooperativity in that increased amounts of PS further promote its binding not necessarily enough to overcome its surface burial. The anionic lipids PG and PA seemingly increased the benefit (or minimized the cost) incurred from surface burial of TIM3. Viewed in the context of our newly refined notion of cooperativity, the effect of PG and PA are not surprising and can be explained in terms of their respective structural features. TIM3 is no longer classified as moderately cooperative as the addition of a small amount of PG or PA significantly increases the affinity for PS. In this analysis, TIM3 has the highest cooperativity of the TIM proteins despite its intermediate Hill coefficient for PS alone.

With this structure-informed paradigm, we were able to derive a set of binding equations that dramatically improved our description of TIM binding. Several inconsistencies which plagued our analysis were transformed into informative features that allowed us to delineate the mutual influence of calcium and PS. These equations defined parameters that provide a quantification of these myriad components of binding and allowed us to compare the binding of the TIM proteins in detail.

This thorough characterization of the structures and binding of the TIM proteins have several physiological implications. To start with, our data support the premise that the TIM proteins do not normally interact with quiescent cells. Rather they are tailored to recognize the exposure of phospholipids normally sequestered in their inner leaflets during apoptosis. The linkage of these three Tim proteins through calcium coordination would be sensitive to both the level of PS and that of the free calcium in the aqueous environment. The basic residues on these TIM proteins would interact electrostatically with this PS and other newly-exposed anionic phospholipids, including PA which is present in most membranes at ~1-2%. The burial of the hydrophobic residues of the TIM proteins would add to the free energy of binding.

The disordering and surface expansion of target plasma membrane bilayers could serve as a signal for the binding of TIM proteins as well as that of certain other peripheral proteins. In fact, apoptosis renders the lipids in the outer leaflet of the plasma membrane bilayer less packed . With similar consequences, the scrambling of plasma membrane bilayers brings phospholipids with greater fatty acyl chain unsaturation to the cell surface, and the triggering of lymphocytes increases the unsaturation of their plasma membrane phospholipids. Such alterations could help phagocytes distinguish apoptotic from unperturbed cells.

The work here suggests that a major physiological role of the TIM proteins, especially TIM3, is to recognize apoptotic cells. They are most sensitive to membrane properties that maximally distinguish apoptotic membranes from those of healthy cells. The sensitivity to lipid packing corresponds to the looser packing of apoptotic membranes induced by several mechanisms such as the decoupling of the membrane from the cytoskeleton as well as increased populations of unsaturated and oxidized lipids. The broad sensitivity to anionic charge, and not solely PS, correlates with the overall scrambling of inner leaflet lipids in apoptotic membranes, including phosphoinositides, cardiolipin, and PA. In particular, TIM3 is highly sensitive to low percentages of PA, which would allow TIM3 to highly distinguish between healthy PS-signaling and apoptotic PS.

Conversely, if proteins were insensitive to apoptotic features, this would suggest a possible role in a non-apoptotic environment. For example, TIM4 cannot distinguish between PS and anionic lipids as well as TIM3 does. Indeed, as stated in the Introduction, TIM4 has been implicated in the non-apoptotic recognition of PS-exposing T cells. TIM1 is minimally sensitive to both PS and anionic lipids in general and has been implicated as an anchoring protein in non-immunological contexts. While TIM1 is still heavily involved in the recognition of apoptotic cells, it is seemingly the most versatile of protein family since it has the most wide-ranging physiological contexts for PS recognition.

There are several features of physiological membranes that we did not investigate. Future studies should elucidate the role of other inner leaflet membranes which are not necessarily charged, such as phosphatidylethanolamine, as these are also scrambled to the outer leaflet during apoptosis and thus further distinguish apoptotic from non-apoptotic membranes. In this work, we investigated the role of several lipids on a PC membrane "background." In general, it would be much more valuable to determine these dependences on a physiological membrane background including such components as PE, cardiolipin, sphingomyelin and cholesterol.

It is also worth investigating if the human TIM proteins share a similar set of sensitivities to those of the murine versions studied here. These experiments present a worthy challenge as human TIM3 lacks an interfacial tryptophan to track binding. However, the crystal structures of these proteins have recently been solved and, with the tools developed and refined in this thesis, are ripe for full characterization.

Along these lines, the TIM proteins contain other domains that tether them to a cell membrane. Studies that determine the role of these domains in PS recognition would be invaluable. Additionally, the TIM proteins contain many other ligands besides PS which possibly modify or out-compete their interactions with PS. The simplest of these other ligands to study is CEACAM-1, which has been shown to interact with the soluble domain of TIM3.

In terms of the techniques used in this work, there are several paths forward. The binding equations developed in Chapter 7 demonstrated a dependence of the analysis on the chosen experimental technique. Therefore, it would be fruitful to model these differences for a set of techniques and determine how the values of the parameters depend on the chosen technique. For example, ITC may return a different Hill coefficient than fluorescence anisotropy and it would be important to account for these differences from modeling alone. As it stands now, most fluorescence techniques are likely to be self-consistent but other techniques may result in different values of the parameters. As peripheral membrane-binding proteins are characterized with a vast array of techniques, this would allow a greater degree of comparison between them.

X-ray reflectivity required the use of lipid monolayers as a model for the outer leaflet of the cell membrane. However, such a model restricts the set of proteins that can be characterized as proteins that insert past the hydrophobic core cannot be studied on monolayers. Developing x-ray reflectivity for solid-liquid interfaces with supported bilayers would allow characterization on a better model membrane.

While HMMM MD was instrumental in independently resolving the membrane-bound state of the TIM proteins, it has a few limitations as well. Since the model fluidizes the membrane, it cannot adequately capture the differences in binding that might occur as a result of changes of the acyl lipid tails. It is entirely possible that DO membranes induce a different membrane-bound ensemble of states. Perhaps HMMM MD can be manipulated in a way to provide such a model, but there are also promising alternatives such as coarse-grained simulations. The latter require all-atom simulations to parameterize, so perhaps the best way forward is to use both of them.

Lastly, the set of techniques and analyses developed here are not limited to the TIM proteins. This full description of binding represents the culmination of a structure-guided approach. More important than determining the structures of specific proteins, these results provide a paradigm which promises to streamline the investigation of the broader class of peripheral membrane-binding proteins.