Molecular and Structural Basis of Anthrax Lethal Toxin Pore Formation and Translocation

By

Alexandra Machen
M.P.H., Emory University, 2013
B.S., University of Kansas, 2011

Submitted to the graduate degree program in Biochemistry and Molecular Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Committee Chair: Bret D. Freudenthal, PhD

_____
Alexey S. Ladokhin, PhD

_____
Chad Slawson, PhD

_____
Aron W. Fenton, PhD

_____
Joe Lutkenhaus, PhD

Date Defended: 17 June 2020

The dissertation committee for Alexandra Machen certifies that this is the approved version of the following dissertation:

Molecular and Structural Basis of

Anthrax Lethal Toxin Pore Formation and Translocation

_____

Chair: Bret D. Freudenthal, PhD

_____

Graduate Director: Aron W. Fenton, PhD

Date Approved: June 29 2020

Abstract

The morbidity and mortality of anthrax disease are associated with the anthrax toxin, which is generated by the gram-positive bacterium *Bacillus anthracis*. The anthrax toxin is an AB toxin composed of two components: an active (A) moiety named lethal factor (LF) and a binding (B) moiety termed protective antigen (PA). In order for the anthrax toxin to elicit its cytotoxic effect, the LF component must first enter the cell. To accomplish this, the PA component binds to a target host cell receptor and forms a pore (PA_{pore}) that translocates LF into the cytosol. In this work, we explore PA binding the host cell receptor, the interactions between PA and LF during pore formation, and the translocation of LF through the PA_{pore}. We hypothesize PA dissociates from its cellular receptor to facilitate pore formation. To test this, in Chapter 2, we investigated the anthrax toxin receptor CMG2 binding capabilities to the PA_{pore} under endosomal conditions. Our results provide evidence for receptor release prior to pore formation. In Chapter 3, we hypothesized the LF N-terminal tail travels down the pore lumen and interacts with the narrowest part of the pore. To test this, we characterized the structure of three LF_N bound to PA_{pore} at neutral pH. Our results indicate the N-terminal tail of LF remains flexible in the translocation incompetent neutral pH environment and underscores the necessity of using physiologically relevant conditions. In Chapter 4, we hypothesize LF begins to refold inside the PA_{pore} during translocation. To test this, we captured intermediates of LF translocation through PA_{pore} using cryoEM. Our results support the hypothesis that initial refolding of LF structural elements occurs in the PA_{pore} beta barrel during translocation. Cumulatively, we have made significant contributions to our understanding of the anthrax intoxication mechanism at multiple biologically relevant steps.

Dedication

To my parents, my cat, and my husband

for

their unwavering support, comfort, and love

For Mark…

…Scientist

…Mentor

…Friend

Acknowledgements

This work was financially supported by National Institutes of Health R01AI090085, R35-GM128562, and R03-AI142361, as well as KUMC Bridging Funds and the Madison and Lila Self Graduate Fellowship.

There are no conflicts of interest to report.

Table of Contents

List of Abbreviations

2D: two dimensional
3D: three dimensional
ADP: adenosine diphosphate
ATP: Adenosine Triphosphate
BLI: biolayer interferometry
CCD: charge couple device
CMG2: capillary morphogenic gene-2
cryoEM: cryogenic electron microscopy
cryoSPARC: cryoEM single particle *ab initio* reconstruction and classification
DED: direct election detector
DTT: Dithiothreitol
EDC: 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide)
EF: edema factor
EM: electron microscopy
FSC: Fourier shell correlation
GuHCL: guanidine hydrochloride
HPLC: high pressure liquid chromotography
HT: high throughput
LF: lethal factor
LF_N: N-terminal domain of lethal factor
MALDI-TOF: matrix-assisted laser desorption/ionization time of flight
MAPKK: Mitogen-activated protein kinase kinase
MS: mass spectrometry
MSP: membrane scaffold protein
Nd: nanodisc
NHS: N-hydroxysuccinimide
Ni-NTA: nickel nitrilotriacetic acid
NS: negative stain
PA: protective antigen
PA_{pore}: heptamperic protective antigen prore
$PA_{prepore}$: heptameric protective antigen prepore
PCR: polymerase chain reaction
PDB: protein data base
PDEA: 2-(2-pyridinyldithio)ethanamine
POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPG: 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol)
SGD: stochastic gradient descent
SPARX: single particle analysis for resolution extension
SPR: surface plasmon resonance
SUMO: small ubiquitin like modifier
TB: terabyte
TEM: transmission electron microscopy
TEM8: tumor endothelium marker-8

Chapter 1: Introduction

Anthrax Virulence and Outbreaks

Anthrax disease in humans is rare, but can cause serious illness as a result of zoonotic (Fasanella, Galante et al. 2010), accidental (Meselson, Guillemin et al. 1994), or deliberate (Bush, Abrams et al. 2001) exposure to *Bacillus anthracis*. *B. anthracis*, the causative agent of anthrax disease, is a gram-positive rod-shaped bacterium that produces two key virulence factors: an anti-phagocytic polyglutamic acid capsule and a proteinaceous toxin termed the anthrax toxin (Turnbull, Kramer et al. 1991). Anthrax poses danger to the public through outbreaks, which are most often the result of zoonotic exposure from the environment (Martin and Friedlander 2010). Adding to its danger, virulent spores of *B. anthracis* are found widespread within the environment and can even remain in the soil for decades (Halvorson 1997). Moreover, as climate change thaws the arctic regions of Greenland, Alaska, Russia, China, and Eastern Europe there has been increasing concern regarding the potential release of anthrax spores from the permafrost (Timofeev, Bahtejeva et al. 2019). The spores that make anthrax stable for long periods of time in nature also make it well suited for aerosolized release, as was the case in the 2001 Amerithrax bioterrorism attack. In this attack, anthrax spores were shipped using the United States Postal Service to American political figures and resulted in five deaths (Khan 2011). The lethality of anthrax disease is due, in part, to the anthrax toxin, a protein toxin secreted by the bacterium. In order for the anthrax toxin to elicit its lethality, it must first enter the cytosol of host cells and then initiate cell death (Wang and Roehrl 2005, M Beierlein and C Anderson 2011). In this work, we explore the molecular and structural details of the anthrax toxin cell entry process and aim to provide a better understanding of the anthrax intoxication mechanism. Importantly, a more complete understanding of this mechanism will serve to aid in our ability to prevent and/or manipulate this process and may ultimately benefit human health.

Anthrax Toxin Intoxication Overview

The anthrax toxin is an AB toxin which consists of a two-component system with an active (A) moiety and a binding (B) moiety. It has two alternative A components: lethal factor (LF), a mitogen-activated protein kinase kinase (MAPKK) protease, and edema factor (EF), an adenylate cyclase. In order to perform their physiological functions, LF and EF must be translocated into the cytosol of the cell (Pannifer, Wong et al. 2001). Cellular entry is accomplished by the B component of the toxin termed protective antigen (PA). To accomplish this, PA binds a target host cell receptor and forms a pore to translocate LF or EF into the cytosol (Santelli, Bankston et al. 2004). Of note, toxin complexes consisting of PA and LF are referred to as anthrax lethal toxin, while toxin complexes composed of PA and EF are referred to as anthrax edema toxin. In this work, we will focus on the anthrax lethal toxin that is composed of PA and LF.

An overview of the anthrax intoxication mechanism is shown as a schematic in **Figure 1.1** and described here (Young and Collier 2007). After secretion from *B. anthracis*, the 83 kDa PA monomer (PA_{83}) first binds to its target host cell receptor. Then, the pro-domain of PA is cleaved leaving the 63 kDa PA (PA_{63}) to oligomerize into a heptameric prepore ($PA_{prepore}$) (Santelli, Bankston et al. 2004, Kintzer, Thoren et al. 2009). This $PA_{prepore}$ heptamer can bind up to three LF and/or EF components (Mogridge, Cunningham et al. 2002, Kintzer, Thoren et al. 2009, Antoni, Quentin et al. 2020). This AB toxin complex, consisting of $PA_{prepore}$ bound to LF or EF, is subsequently endocytosed through clatherin mediated endocytosis (Abrami, Liu et al. 2003) and trafficked to the late endosome. As the endosome acidifies, the $PA_{prepore}$ undergoes a conformational change to form a pore (Miller, Elliott et al. 1999). This heptameric PA pore (PA_{pore}) inserts into the endosomal membrane to form a channel. LF or EF unfold and translocate

through the PA_{pore} channel and into the cytosol (Krantz, Finkelstein et al. 2006). The pH gradient between the endosome and the cytosol facilitates rapid translocation *in vivo*. Upon entering the cytosol, the natively refolded LF and/or EF are then able to alter cell signaling pathways ultimately leading to apoptotic cell death (Duesbery, Webb et al. 1998). Each of these steps are discussed in more detail in the subsequent sections.

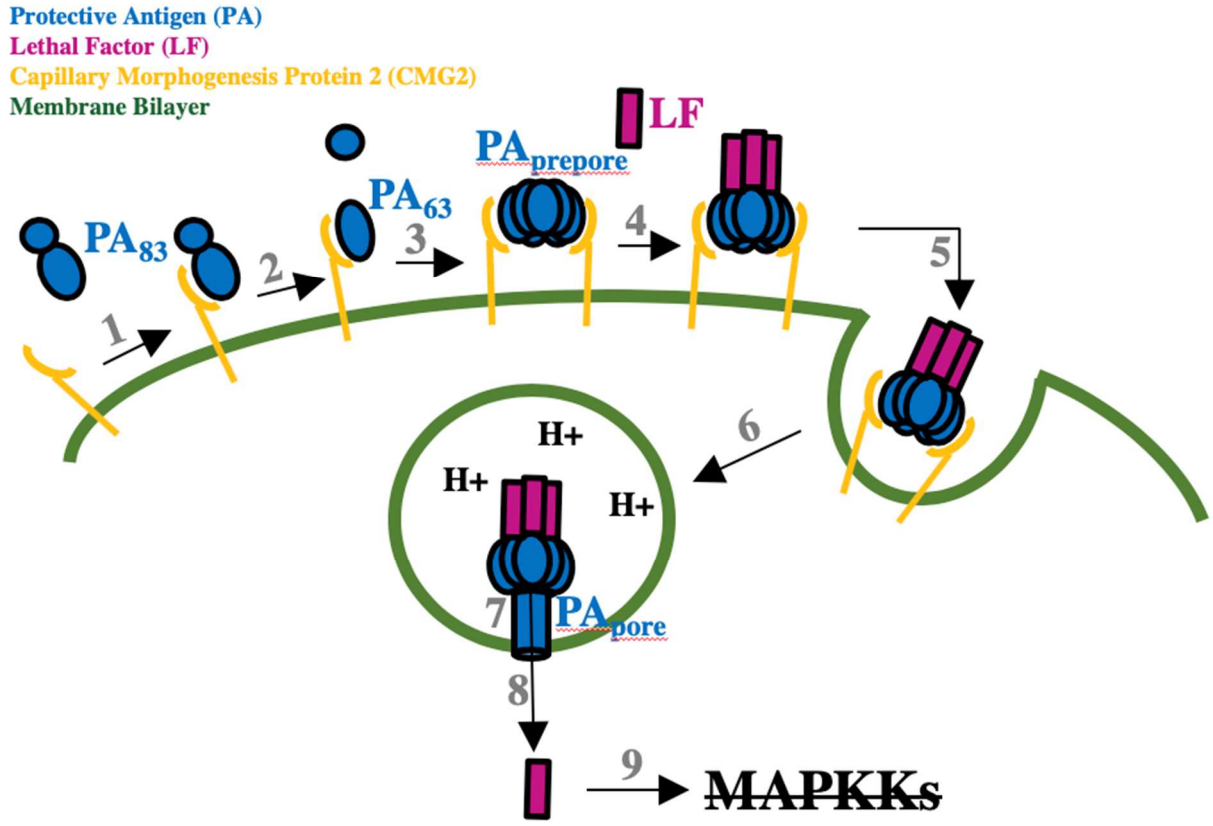


Figure 1.1. Overview of anthrax intoxication mechanism. (1) PA_{83} (blue) binds to a host cell receptor (yellow). (2) PA_{83} is cleaved leaving PA_{63} bound to receptor. (3) PA_{63} self-associations forming heptameric $PA_{prepore.}$ (4) LF (magenta) binds to $PA_{prepore}$ (5) The co-toxin complex is endocytosed via ubiquitin, actin, and clatherin dependent mechanisms and (6) trafficked to the late endosome. (7) The low pH of late endosome induces a prepore to pore conformational change. (8) The endosomal pH and pH gradient between the endosome and cytosol facilitate LF unfolding and translocation across membrane. (9) Refolded LF proteolytically inactivates MAPKKs. Adapted from (Sellman, Mourez et al. 2001).

Formation of the Protective Antigen Heptameric Prepore ($PA_{prepore}$) Complex

PA is secreted from *B. anthracis* as an 83 kDa monomer (PA_{83}) that binds to host cell receptors (**step 1 , Figure 1.1**) (Petosa, Collier et al. 1997). The PA_{83} monomer is activated by cleavage of the pro-domain, leaving a 63 kDa PA monomer (PA_{63}) that will self-assemble into a heptameric $PA_{prepore}$ complex capable of binding LF and EF (**steps 2-4, Figure 1.1**). The PA monomer consists of four domains (**Figure 1.2**). These domains are each discussed in more detail below. Briefly, domain 1 plays a role in both oligomerization and LF/EF binding; domain 2 facilitates PA pore formation; domain 3 is involved in oligomerization of PA from a monomer to a heptamer; and domain 4 is the receptor binding domain.

The crystal structure of PA_{83} is shown in **Figure 1.2.A**. In order for PA to oligomerize into the $PA_{prepore}$ complex, residues 1-167 of domain 1 are cleaved by either extracellular proteases or a membrane-bound furin-like protease leaving PA_{63} bound to the cell receptor (**step 2, Figure 1.1)**. The resulting cleaved domain 1 of PA_{63} is denoted as domain 1'. After cleavage, domain 1' of PA_{63} consists of a β-sandwich with a jelly-roll topology, as well as four α helices stabilized by two calcium ions (**Figure 1.2.B**) (Ezzell and Abshire 1992, Petosa, Collier et al. 1997, Young and Collier 2007). Next, PA_{63} self assembles into a heptameric prepore consisting of seven PA protomers (**step 3, Figure 1.1**). The structure of the $PA_{prepore}$ is shown in **Figure 1.2.C**. To oligomerize, PA_{63} protomers come together, like pie wedges, facilitated by the charged residues of domains 1′ and 2, to form a hollow ring with an outer diameter of 160 Å and an inner diameter of 35 Å. After its formation, either LF or EF bind domain 1' of $PA_{prepore}$ (**step 4, Figure 1.1)**. Importantly, LF and EF are unable to bind the uncleaved domain 1 of PA due to steric clashing. Domain 2 of the $PA_{prepore}$ consists of residues 259 to 487 and is folded up against Domain 4, burying the hydrophobic 2β2–2β3 loop of Domain 2. This loop will ultimately form the pore

component of PA_{pore} in the endosome (**step 6, Figure 1.1 and Figure 1.2D**) (Lacy, Wigelsworth et al. 2004, Young and Collier 2007, Jiang, Pentelute et al. 2015). Domain 4 consists of residues 596 to 735 and forms a β-sandwich and an immunoglobulin-like fold (Young and Collier 2007) that binds the von Willebrand factor A (vWF-A) domain of the anthrax toxin receptor (Wigelsworth, Krantz et al. 2004).

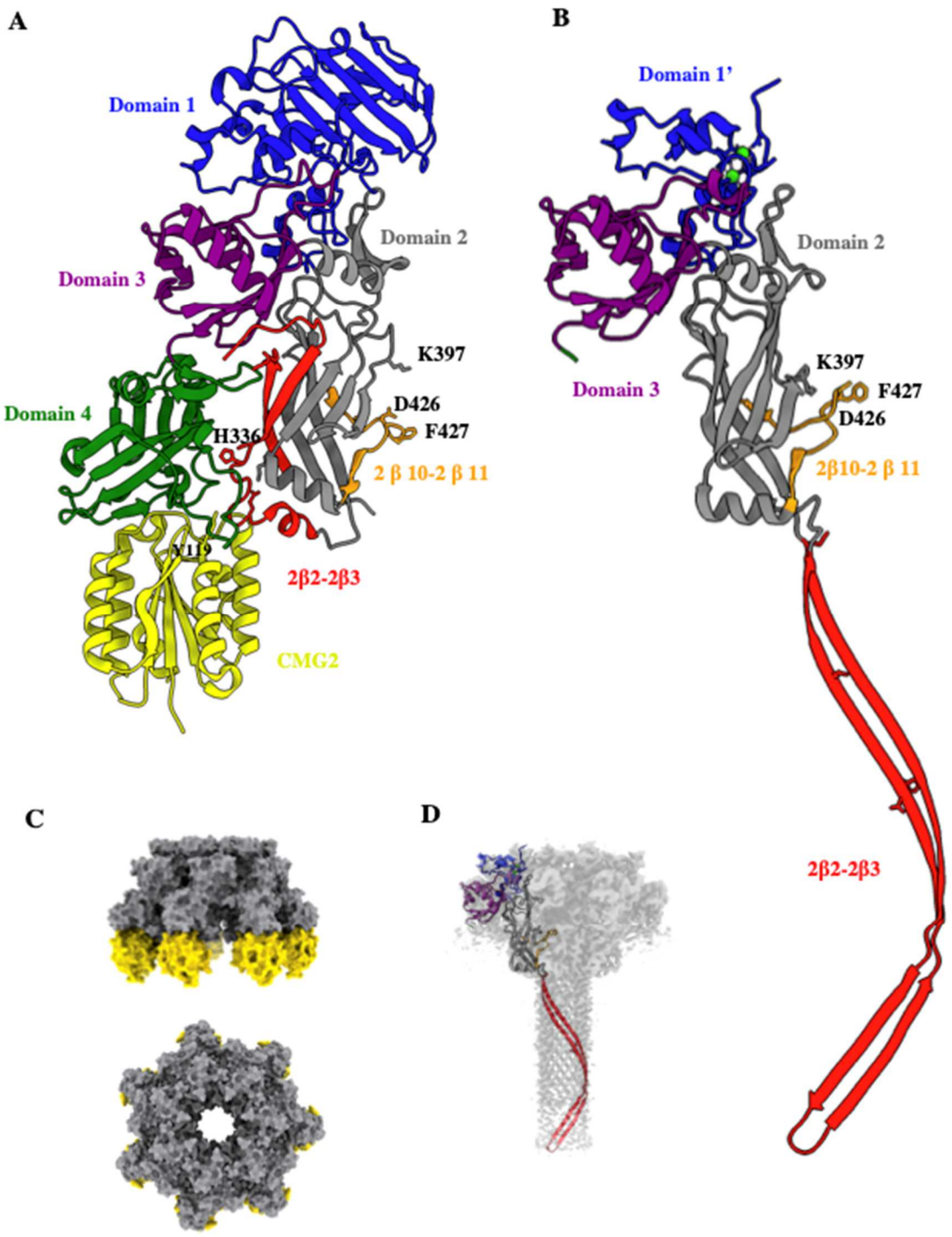


Figure 1.2 Anthrax protective antigen prepore and pore protomer structure. (A) Crystal structure of PA_{83} bound to receptor (capillary morphogenesis protein 2 (CMG2)) (PDB 1T6B) compared to (B) Cryogenic electron microscopy (cryoEM) structure of PA_{63} pore protomer (domain 4 was not resolved) (PDB 3J9C). (C) Side and top view of $PA_{prepore}$ (grey) bound to CMG2 (yellow) (PDB 1TZN). (D) Side view of PA_{pore} cryoEM density map (EMDB 6224, grey) with PA_{pore} protomer highlighted (PDB 3J9C).

Anthrax Toxin Receptors: CMG2 and TEM8

The formation of the $PA_{prepore}$ heptamer described in the previous section occurs when PA is bound to an anthrax toxin receptor. Two known anthrax toxin PA host cell receptors are the capillary morphogenesis protein 2 (CMG2) and tumor endothelial marker-8 (TEM8) (Bradley, Mogridge et al. 2001, Scobie, Rainey et al. 2003). CMG2 and TEM8 are ubiquitously expressed and hypothesized to regulate angiogenic processes (Croix, Rago et al. 2000, Bonuccelli, Sotgia et al. 2005, Rmali, Puntis et al. 2005). There are several protein isoforms of TEM8 and CMG2 due to alternative mRNA splicing (Bradley, Mogridge et al. 2001, Liu and Leppla 2003, Scobie, Rainey et al. 2003). PA is able to bind to the long (489 amino acids) isoform of CMG2 and the long (564 amino acids) and medium (368 amino acids) isoforms of TEM8. Each of these three isoforms have the extracellular vWF-A fold that is considered a hallmark of anthrax toxin receptors.

The crystal structure of PA_{83} bound to CMG2 is shown in **Figure 1.2.A** with CMG2 in yellow. PA monomer and $PA_{prepore}$ protomer bind CMG2 or TEM8 in a 1 to 1 ratio (**Figure 1.2.A,C)**. Interestingly, the PA binding affinities for CMG2 and TEM8 differ by orders of magnitude: PA binds to CMG2 with picomolar affinity whereas TEM8 binds in the micromolar range. PA-receptor equilibrium dissociation rate constants are shown in **Table 1.1**.

Table 1.1. Anthrax toxin receptor binding affinity for protective antigen.

	K_d, Equilibrium Dissociation Constant		
Receptor	Mg2+	Ca2+	
CMG2	170 pM	780 pM	(Wigelsworth, Krantz et al. 2004)
TEM8	1.1 uM	130 nM	(Scobie, Thomas et al. 2005)

The difference in the binding affinities for CMG2 and TEM8 to PA can be explained by the way in which they bind to PA. Specifically, the PA monomer-CMG2 complex has approximately 700

Å^2 additional buried protein surface compared to the PA monomer-TEM8 complex (Santelli, Bankston et al. 2004) (Scobie, Marlett et al. 2007).

What is the Roll of TEM8 and CMG2 in PA Pore Formation?

The anthrax toxin receptors have been hypothesized to act as a molecular clamp, preventing premature pore formation (Wimalasena, Janowiak et al. 2010) by binding to PA domain 4 as well as PA domain 2 (**Figure 1.2.A**) (Lacy, Wigelsworth et al. 2004, Santelli, Bankston et al. 2004). This pH-dependent receptor clamp mechanism was indicative of a histidine titration. Indeed, PA residue H336, located on the pore forming loop and contacting domain 4, was identified as a potential pH sensor (Kintzer, Tang et al. 2012). In addition, the conserved receptor residue H121, located at the PA binding interface, has been implicated (Lacy, Wigelsworth et al. 2004, Santelli, Bankston et al. 2004). Interestingly, pH of pore formation varies depending on which receptor PA is bound to. When PA is bound to TEM8, pore formation of PA occurs at pH 6.2. However, when PA is bound to CMG2, pore formation does not occur until pH 5.2 (Rainey, Wigelsworth et al. 2005, Wolfe, Krantz et al. 2005). This difference lead to the hypothesis that pore formation occurs within distinct endosomal compartments, depending on which receptor is present (Maxfield and McGraw 2004, Young and Collier 2007). Specifically, when bound to TEM8, PA pore formation is predicted to occur in the sorting endosome (pH range 5.9–6.0). When bound to CMG2, PA pore formation is predicted to occur in the late endosome (pH 5.0–6.0). The biological ramifications of this difference are not well understood. However, the molecular basis for this difference is hypothesized to be due to residue Y119 of CMG2 (Scobie, Marlett et al. 2007). This is based on mutagenesis studies, which revealed that when Y119 is mutated to either an alanine or phenylalanine the pH at which pore formation occurs shifts by 0.6-0.8 towards a neutral pH.

While it is clear that CMG2 and TEM8 are important in regulating PA pore formation in the context of the endosome, it is not known how these receptors interact with PA_{pore}. It was previously proposed that CMG2 or TEM8 would remain bound to the pore form of PA as a structural support (Santelli, Bankston et al. 2004). However, structural insight into pore formation supports a model involving release of PA domain 2 and 4 from the receptor. We hypothesize the receptor dissociates from PA during PA pore formation to facilitate domain 2 rearrangement and pore formation (Rainey, Wigelsworth et al. 2005). In **Chapter 2**, we investigate whether CMG2 remains bound to PA *in vitro* when PA transitions from the prepore to pore conformation.

Structural Characterization of PA Pore Formation

One key step that must occur during the anthrax intoxication mechanism is pore formation of PA in the endosome (**step 7, Figure 1.1**). Early biophysical characterization of PA established this pore forming capability (Blaustein, Koehler et al. 1989). However, high resolution structural information regarding the pore remained elusive for several decades due in large part to the challenging nature of membrane protein structural biology (Jiang, Pentelute et al. 2015). Moreover, well defracting PA_{pore} crystals could not be obtained for X-ray crystallography and the size of the PA_{pore} was prohibitive for structural determination through nuclear magnetic resonance. Therefore, structural biologists turned instead to electron microscopy (EM), which requires relatively low concentrations of protein and is amenable to large complexes. Unfortunately, aggregation of the PA_{pore} remained a significant challenge and creative solutions were needed to circumvent aggregation of this membrane protein. One such solution involved the use of the GroEL chaperonin as a molecular scaffold (Katayama, Janowiak et al. 2008). The original concept for using GroEL was that the large hydrophobic cavity in this molecular

chaperone would bind to the hydrophobic tip of the PA_{pore} β barrel, thus preventing aggregation. However, a serendipitous seven-fold symmetric electrostatic interaction with PA domain 1' resulted in GroEL associating with the LF/EF binding interface instead of the β barrel (**Figure 1.3.A**). This interaction slowed aggregation long enough to visualize GroEL-PA_{pore} complexes with negative stain electron microscopy. The resulting low-resolution PA_{pore} structure revealed a mushroom-shaped pore with a 125 Å diameter cap and a 100 Å stem (Katayama, Janowiak et al. 2008). An important caveat to this study is that the GroEL-PA_{pore} structure is not a biological relevant complex, because PA_{pore} is physiologically found in a lipid membrane bilayer at low pH. Two years later, a PA_{pore} structure was published that incorporated PA into a lipid bilayer using lipid nanodiscs (**Appendix A**) and vesicles, in combination with cryogenic electron microscopy (cryoEM) to reveal the narrow pore lumen (**Figure 1.3.B**) (Katayama, Wang et al. 2010). This work did not, however, consider the low pH of the endosomal environment. The most recent PA_{pore} structure to date was obtained under endosomal conditions (**Figure 1.3.C**) (Jiang, Pentelute et al. 2015). This structure is the result of depositing $PA_{prepores}$ on cryoEM carbon support grids and washing the prepores in a low pH buffer. The low pH buffer allowed the prepores to transition into pores, with the β barrel extended up from the carbon support. The use of grid adsorption avoided the aggregation at low pH problem, but it failed to account for the membrane bilayer. Future structural work of the PA_{pore} would do well to incorporate both physiologically relevant elements of low pH and a membrane bilayer. The work described in **Chapter 4** reports, to our knowledge, the only PA_{pore} structure inserted in to a membrane bilayer obtained under endosomal pH conditions.

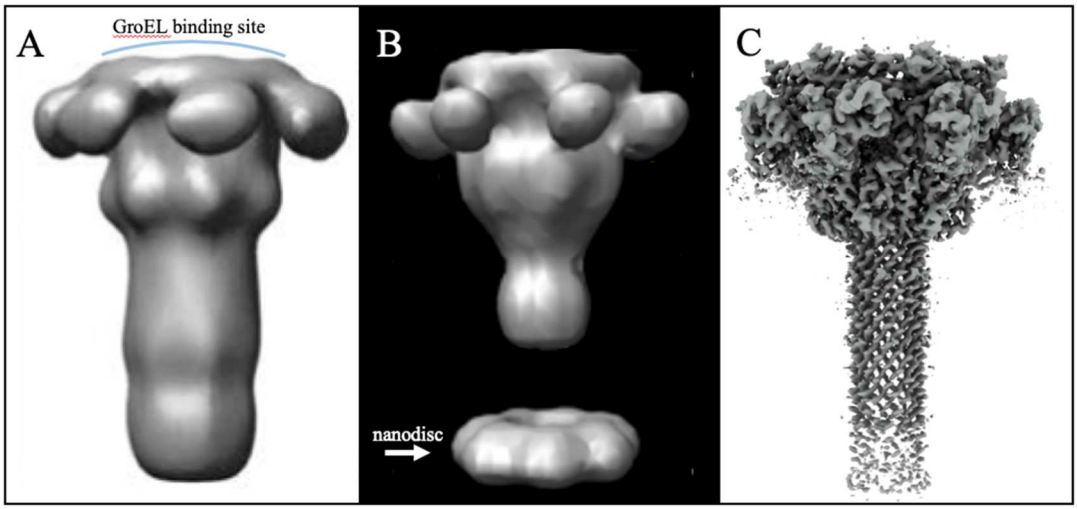


Figure 1.3. Evolving cryoEM densities of PA_{pore}. (A) 22 Å PA_{pore} density resulting from GroEL as a molecular scaffold adapted from (Katayama, Janowiak et al. 2008). (B) 22 Å PA_{pore} inserted into a lipid nanodiscs at pH 7.5 (Katayama, Wang et al. 2010). (C) 2.9 Å PA_{pore} transition from $PA_{prepore}$ on cryoEM grid carbon support (PDB 3J9C) (Jiang, Pentelute et al. 2015).

The α Clamp, the Φ Clamp, and the Charge Clamp of the Protective Antigen Pore

The overall structure of the PA_{pore} can be divided into two regions: the funnel and the channel (**Figure 1.4.A**). The funnel facilitates binding and unfolding of LF. LF binds to the top of the PA_{pore} funnel and travels down the narrowing structure as it is unfolded. The second region of PA_{pore}, the channel, is a β barrel that extends from the funnel and spans the endosomal membrane. Three nonspecific PA_{pore} clamps aid in the translocation of LF: The α clamp, the Φ clamp, and the charge clamp (**Figure 1.4.A**). The α clamp, located at the PA_{pore} funnel rim, is formed by adjacent PA protomers and binds helical portions of LF to position them towards the pore lumen. The Φ clamp is a ring of seven phenylalanine residues that maintain the pH gradient between the endosome and the cytosol (Krantz, Melnyk et al. 2005). The Φ clamp is essential to anthrax intoxication, with mutations of the phenylalanines abrogating translocation (Sellman, Nassi et al. 2001, Mourez, Yan et al. 2003). The charge clamp is located within the β barrel of PA_{pore}. The charge clamp deprotonates acidic side chains of LF and ensures unidirectional

movement of the polypeptide (Pentelute, Sharma et al. 2011, Brown, Thoren et al. 2015). The mechanism of translocation is discussed in more detail in the Brownian Ratchet Mechanism section below.

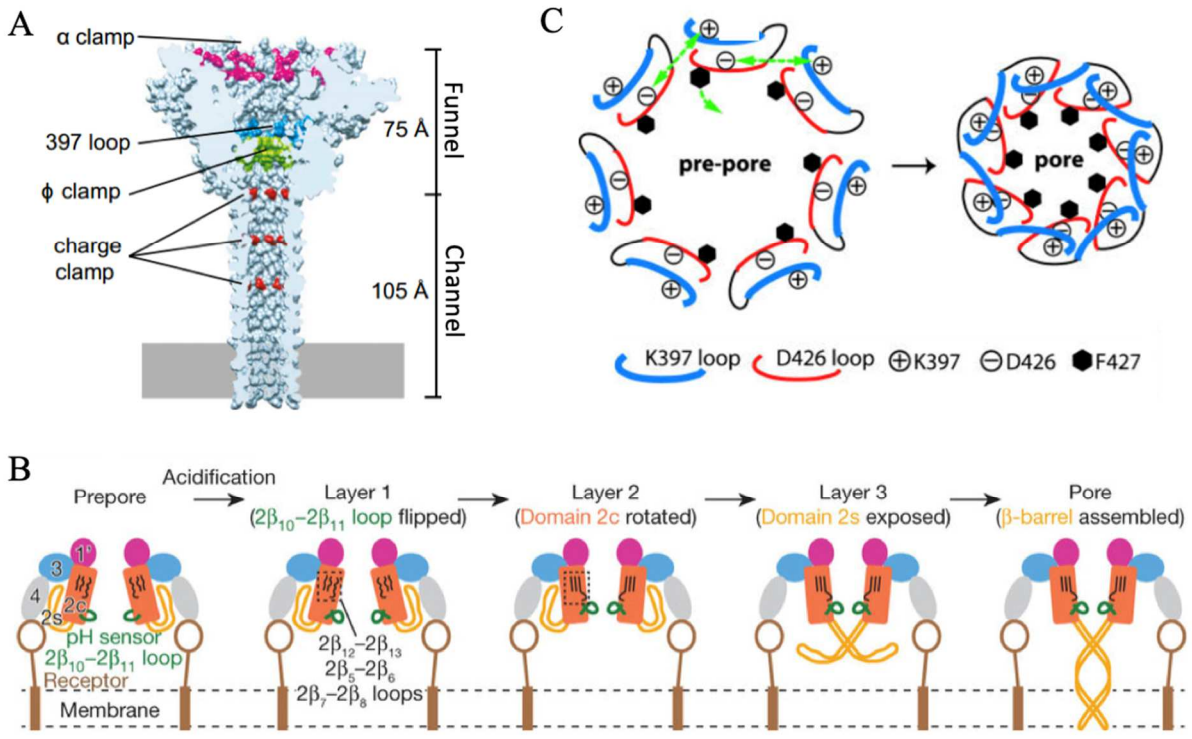


Figure 1.4. Protective antigen pore formation. (A) Molecular model of PA_{pore} (PDB 3J9C) w clamp sites highlighted (Das and Krantz 2016). (B) Proposed structural changes during PA_{pore} formation (Jiang, Pentelute et al. 2015). (C) Salt bridge formation between key PA loops upon pore formation (Melnyk and Collier 2006) Copyright 2006 National Academy of Sciences.

Insights into the Conformational Changes going from PA Prepore to Pore

The current model of pore formation was developed from comparisons of prepore crystallographic structures and pore cryoEM structures (**Figure 1.2**) (Lacy, Wigelsworth et al. 2004, Jiang, Pentelute et al. 2015). A schematic of the hypothesized conformational changes is shown in **Figure 1.4.B**. The first PA conformational change is the flipping of the 2β10-2β11 loop from one side to another, followed by the straightening of three antiparallel loops within

domain 2. This results in a 15-degree rotation of domain 2 and the formation of a salt-bridge between K397 and D426 of the adjacent protomer (**Figure 1.4.C**). This inter-protomer loop interaction causes the pore to constrict, and positions F427 of the 2β10-2β11 loop to face the pore lumen forming the Φ clamp (Melnyk and Collier 2006). The final step in PA pore formation is the release and refolding of the 2β2-2β3 loop to form the β-barrel (Jiang, Pentelute et al. 2015). The role of the cellular receptor, which is hypothesized to act as a molecular clamp on domain 4 and domain 2, in pore formation from prepore is not known. One possibility is that the receptor, CMG2 or TEM8, would remain bound to PA domain 4. Another possibility is that the receptor completely dissociates. Prior to this work, evidence for either possibility was lacking. In **Chapter 2**, we present evidence that the receptor completely dissociates during pore formation.

Stoichiometry of LF and EF Binding to the $PA_{prepore}$

Both LF and EF bind to $PA_{prepore}$ with nanomolar affinity (Elliott, Mogridge et al. 2000), thus competing for the same binding site in a mostly electrostatic interaction stabilized by two calcium ions (Lacy, Mourez et al. 2002, Gupta, Chandra et al. 2003). The LF/EF binding site involves the interface of two PA protomers (**Figure 1.5.A**) (Feld, Thoren et al. 2010). Due to steric clashing and the odd number of PA protomers in heptameric $PA_{prepore}$, only three LF or EF are able to simultaneously bind to the $PA_{prepore}$. This results in one PA protomer unbound to LF or EF (**Figure 1.5., grey protomer**) (Lacy, Lin et al. 2005). Upon binding, the first α helix of LF or EF rotates away from the enzyme and binds to an amphipathic cleft formed between two PA protomers (**Figure 1.5.A**) (Feld, Thoren et al. 2010). This cleft, termed the α clamp for its ability to nonspecifically bind α helical regions, positions LF or EF towards the pore lumen for translocation.

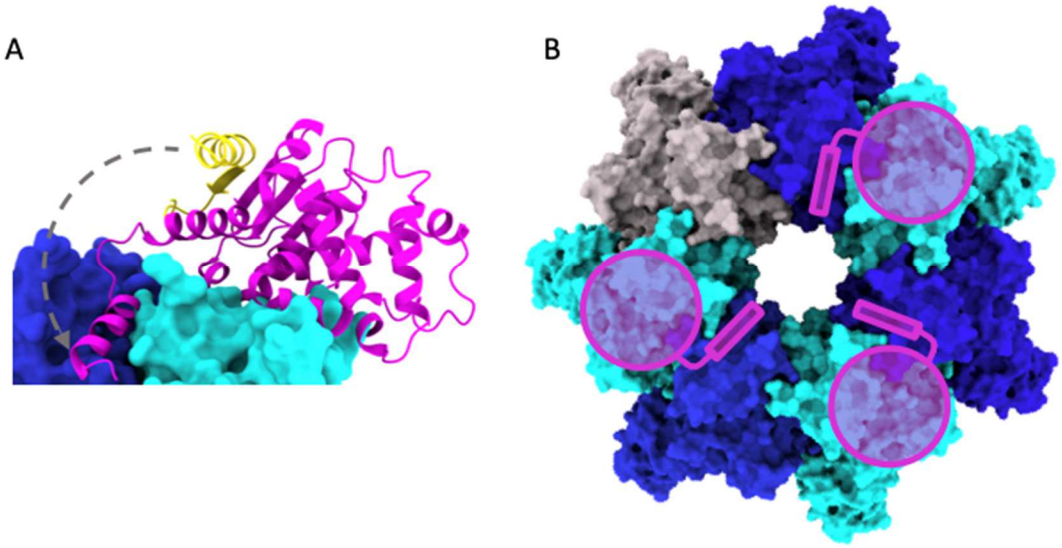


Figure 1.5. Symmetry mismatch of LF-PA$_{prepore}$ binding sites. (A) Ribbon structure of bound LF$_N$ (PDB 3KWV) (magenta) with unbound LF$_N$ α helix 1 (yellow) overlaid. Grey arrow indicating structural rearrangement of α helix 1 upon binding. (B) Surface model top view of PA$_{prepore}$ heptamer (PDB 1TZO) with protomers in cyan, blue, and grey showing predicted binding sites of LF$_N$ (magenta).

Prior to the onset of this work, little structural information existed regarding how LF or EF interact with the pore form of PA. This resulted in several unanswered questions. Does LF facilitate structural changes to the α or Φ clamp within the PA$_{pore}$ lumen? What path does LF take down the pore lumen? Can and how does the PA$_{pore}$ accommodate multiple LF N-terminal tails in the pore lumen? Previous cryoEM analysis of LF$_N$ bound to the pore was unable to address these questions due to low particle number, thick ice, and data collection on charged-coupled devices (CCD). These limitations resulted in no discernable density for the enzymatic component (LF$_N$) (Akkaladevi, Hinton-Chollet et al. 2013, Gogol, Akkaladevi et al. 2013). These studies did, however, lay the groundwork for assembly of LF$_N$-PA$_{pore}$-nanodisc complexes (Akkaladevi, Mukherjee et al. 2015). In **Chapter 3**, we examine the pore form of the toxin co-complex and, using new data analysis tools, are able to distinguish density of three bound LF$_N$ to PA$_{pore}$.

Brownian Ratchet Translocation Mechanism

In order for LF or EF to disrupt cell signaling, they must enter the cytosol of the cell by translocating through the β barrel of the PA_{pore} (Abrami, Liu et al. 2003, Young and Collier 2007). The pH gradient between the endosome and the cytosol facilitates this PA_{pore} proton-protein symporter (Krantz, Finkelstein et al. 2006). Prior to translocation, the N-terminal domain of LF (or EF) is bound to the top of the PA_{pore} with the first LF α helix bound to the PA_{pore} α clamp (**Figure 1.5**) (Feld, Thoren et al. 2010). Here, LF and EF are destabilized by the acidic environment of the endosome, facilitating enzyme unfolding and translocation (Krantz, Trivedi et al. 2004). Transport is unidirectional, from the N to C terminus with α helical regions unfolding more easily than β sheet regions (Pentelute, Sharma et al. 2011, Das and Krantz 2016, Das and Krantz 2017). The α clamp funnels unfolded polypeptides towards the narrowest part of the pore lumen: the Φ clamp. A ring of seven F427 residues in the Φ clamp forms a hydrophobic seal between the endosome and the cytosol (Krantz, Melnyk et al. 2005). This 6 Å diameter clamp is too narrow for secondary structural elements, such as α helices to pass (**Figure 1.6.A**). Therefore, it is hypothesized that peptide substrates must completely unfold and refold in order to translocate through the PA_{pore} and enter the cytosol of the cell (Jiang, Pentelute et al. 2015), (Thoren and Krantz 2011). Importantly, translocation of the unfolded polypeptide through the PA_{pore} is hypothesized to be mediated by a proton gradient driven Brownian ratchet mechanism (**Figure 1.6.B**) (Krantz, Finkelstein et al. 2006, Wynia-Smith, Brown et al. 2012). In this mechanism, the unfolded LF peptide passes the Φ clamp mediated by Brownian motion. At the charge clamp site of the PA_{pore}, the acidic LF side chains that were protonated in the low pH environment of the endosome, are now deprotonated by residues D276, E343, and E335 (Wynia-Smith, Brown et al. 2012). Deprotonated LF residues prevent retrograde transfer back through

the Φ clamp resulting in unidirectional transport of LF into the cytosol. Interestingly, the diameter of the PA_{pore} β barrel is large enough to accomidate an α helix (**Figure 1.6.C**), which would allow for initial refolding to occur inside the PA_{pore} prior to LF entering the cytosol. However, translocation intermediates have not yet been structural characterized. In **Chapter 4**, we generate snapshot structures of what we predict is PA_{pore} translocating the N-terminal domain of LF leading to new insight into translocational unfolding and refolding.

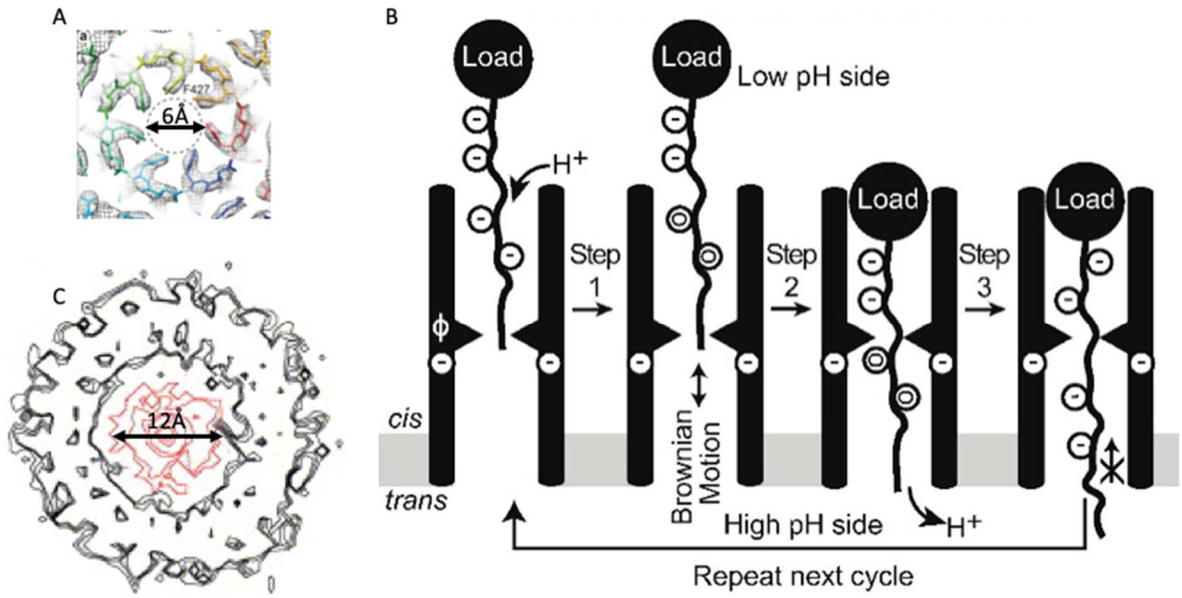


Figure 1.6. Hypothesized Brownian ratchet translocation mechanism. (A) Narrow Φ clamp of PA_{pore} (Jiang, Pentelute et al. 2015) (B) Proposed translocation model with LF (load) acidic residues protonated in the endosome, passing the narrow Φ clamp mediated by Brownian motion, followed by deprotonation in the β barrel charge clamp (Das and Krantz 2017). (C) PA_{pore} β barrel (black) diameter is wide enough to accommodate an α helix (red) (Krantz, Trivedi et al. 2004).

Structure and Enzymatic Function of EF and LF in the cytosol

EF is an 89 kDa calmodulin-dependent adenylate cyclase that alters cellular cAMP production. EF domain 1 is responsible for PA binding and is similar in structure to LF_N (Lacy, Mourez et al. 2002). EF domain 2 and 3 are adenylate cyclase domains (Drum, Yan et al. 2002). EF Domain 4 is a helical domain. Functionally, EF disrupts cellular water homeostatic by significantly

increasing cAMP and upregulating the protein kinase A signal pathway. Interestingly, EF has been associated with an upregulation of TEM8 and CMG2, the anthrax toxin receptors (Maldonado-Arocho, Fulcher et al. 2006).

LF is a 90 kDa zinc-dependent metalloprotease with four domains (Pannifer, Wong et al. 2001). LF domain 1 is responsible for PA binding. LF domain 2 and 3 are important for substrate recognition and proteolytic activity, and LF domain 4 contains the enzyme active site (Pannifer, Wong et al. 2001). LF proteolytically cleaves members of the mitogen activated protein (MAP) kinase kinase (MEK) family. It specifically cleaves the N-terminal amino acids of the MAPKKs abrogating their ability to interact with MAPKs, disrupting the cell signal pathway, and eventually leading to apoptosis (Park, Greten et al. 2002, Chopra, Boone et al. 2003). This activity of LF is hypothesized to help *B. anthracis* evade the immune system by targeting macrophages (Hanna, Acosta et al. 1993). Throughout this work, we use a truncated form of LF consisting of domain 1 that we refer to as LF_N or the N-terminal domain. We chose to focus on LF interaction with PA in our work for several reasons: as its name suggests, LF is responsible for cellular death (Duesbery, Webb et al. 1998); LF_N is the component used for drug delivery systems; and the N-terminal binding domain of LF and EF are structurally similar (Pannifer, Wong et al. 2001, Drum, Yan et al. 2002).

Conclusion

In this work, we address some of the remaining structural and mechanistic questions regarding the anthrax intoxication mechanism (**Figure 1.7**). These questions include: Does PA remain bound to its cellular receptor when transitioning from prepore to pore? How does the LF N-terminal tail interact with the PA_{pore} lumen? Do LF structural elements begin to refold inside the PA_{pore} β barrel during translocation? To address receptor binding during pore formation, we

investigated the anthrax toxin receptor CMG2 binding capabilities to the PA_{pore} under endosomal conditions using biolayer interferometry (BLI), mass spectrometry (MS), and electron microscopy (EM). Our results provide evidence for receptor release prior to pore formation consistent with the hypothesized receptor clamp mechanism whereby the receptor prevents premature pore formation by binding to both the PA receptor binding domain and the pore forming domain (**Chapter 2**). We then characterize the structure of three LF_N bound to PA_{pore} at neutral pH using stochastic gradient decent algorithms to sort heterogenous cryoEM data. Our results indicated the N-terminal tail of LF remains flexible in the translocation incompetent neutral pH environment. This underscores the necessity of using physiologically relevant conditions (**Chapter 3**). To address the structural details of LF translocation, we captured snapshots of LF translocation intermediates using cryoEM, thus allowing us to characterize LF-PA_{pore} interactions at key clamp sites culminating in a proposed LF unfolding and refolding mechanism (**Chapter 4**).

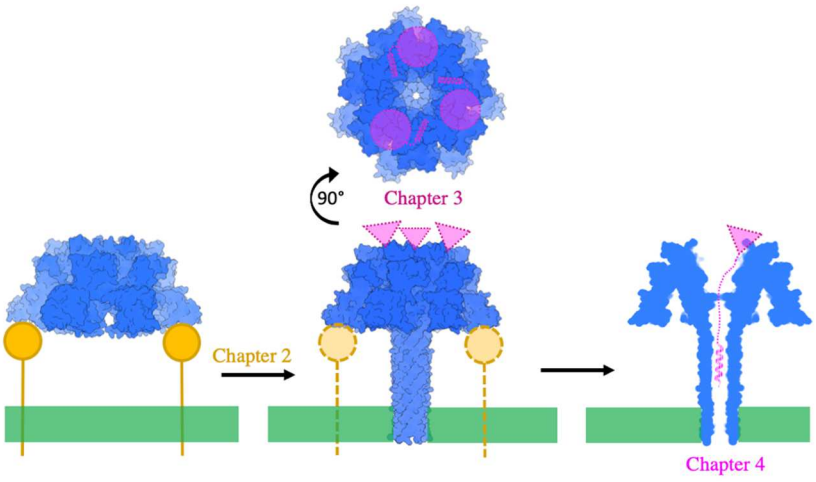


Figure 1.7. Schematic of gaps in anthrax intoxication pore formation and translocation mechanism addressed in this work. (**Chapter 2**) Hypothesized receptor (gold) release facilitates PA (blue) pore formation. (**Chapter 3**) Hypothesized location of LF-N terminal domain (magenta) interactions with PA pore lumen. (**Chapter 4**) Hypothesized initial refolding of LF structural elements during translocation. Membrane bilayer shown in green.

Chapter 2: *In Vitro* evidence of Capillary Morphogenesis Protein 2 Dissociation during Anthrax Toxin Pore Formation

This chapter has previously been published and is reprinted with permission. Machen, A.J., O'Neil, P.T., Pentelute, B.L., Villar, M.T., Artigues, A. and Fisher, M.T., (2018). Analyzing Dynamic Protein Complexes Assembled On and Released From Biolayer Interferometry Biosensor Using Mass Spectrometry and Electron Microscopy. *JoVE (Journal of Visualized Experiments*), (138), p.e57902.

ABSTRACT

In vivo, proteins are often part of large macromolecular complexes where binding specificity and dynamics ultimately dictate functional outputs. In this work, the pre-endosomal anthrax toxin is assembled and transitioned into the endosomal complex. First, the N-terminal domain of a cysteine mutant lethal factor (LF_N) is attached to a biolayer interferometry (BLI) biosensor through disulfide coupling in an optimal orientation allowing protective antigen (PA) prepore to bind (K_d 1 nM). The optimally oriented LF_N-$PA_{prepore}$ complex then binds to a soluble capillary morphogenic gene-2 (CMG2) cell surface receptor (K_d 170 pM), resulting in a representative anthrax pre-endosomal complex, stable at pH 7.5. This assembled complex is then subjected to acidification (pH 5.0) representative of the late endosome environment to transition the $PA_{prepore}$ into its membrane inserted pore state. This PA_{pore} state results in a weakened binding between the CMG2 receptor and the LF_N-PA_{pore} and a substantial dissociation of CMG2 from the transition pore. The thio-attachment of LF_N to the biosensor surface is easily reversed by dithiothreitol. Reduction on the BLI biosensor surface releases the LF_N-$PA_{prepore}$-CMG2 ternary complex or the acid transitioned LF_N-PA_{pore} complexes into microliter volumes. Released complexes are then visualized and identified using electron microscopy and mass spectrometry. These experiments demonstrate how to monitor the kinetic assembly/disassembly of specific protein complexes using label-free BLI methodologies and evaluate the structure and identity of these BLI assembled complexes by electron microscopy and mass spectrometry, respectively, using easy-to-replicate sequential procedures.

INTRODUCTION

Identifying and understanding the specificity governing protein complex assembly *in vivo* is of extreme interest for biochemical researchers. Large heterogeneous protein assemblies are the norm rather than the exception. This notion is supported by spectroscopic monitoring of *in vivo* assembly, isolating complexes using gentler cell disruption techniques, evaluating products from affinity-based purification methods, and visualizing them using high resolution cryogenic electron microscopy. To understand the control of assembly specificity within the cell, researchers must routinely isolate, identify, and ultimately characterize these dynamic assembling/disassembling structures. The most heavily used molecular tool to identify the components of these assemblies frequently requires antibody-based immunoprecipitation which relies on maintaining complex stability during cell disruption. Various coupled analytical techniques were recently developed to capture complexes from cell samples using microfluidic based approaches, such as surface plasmon resonance (SPR). Following removal from the SPR surfaces, these samples were analyzed by matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) (Bellon, Buchmann et al. 2009, Kim, Yi et al. 2012). Advancing this methodology using easier protocols will allow researchers to visualize and validate predicted complexes that occur in the cellular milieu. Since SPR is a microfluidics-based system, problems often arise from aggregate formation. Circumventing this problem requires sample dilution, which in turn, can decrease the integrity of concentration liable biological complexes.

A relatively recent advancement in label-free technologies is the development of the biolayer interferometry (BLI) system (Abdiche, Malashock et al. 2008). These particular light-based reflectance systems replicate, or best emulate, SPR binding and kinetic results at a fraction of the cost (Abdiche, Malashock et al. 2008, Abdiche, Malashock et al. 2009) particularly if single

channel units are used. BLI measures changes in reflected light interference patterns between a reference layer (control) and the biolayer surface (experimental). The resulting change in phase is measured in real time as a kinetic and quantitative readout (Ercius, Alaidi et al. 2015). The biosensor surface, containing specific immobilization chemistries, is physically transferred between solutions, as opposed to buffer changes by microfluidic approaches in SPR, for measurement via wavelength phase deflections. Mass transfer effects are prevented by agitating the solution. Unlike SPR, these systems are quite useful in evaluating complexes from crude biological samples. The physical parameter measured during BLI experiments primarily depends on a change in mass or thickness at the biosensor surface which results from protein complex assembly or disassembly.

These fiber optic biosensors are easy to use and relatively inexpensive. One of the emerging useful aspects of BLI is the facile removal of newly assembled protein complexes from the surface. A recent application of this method allowed this laboratory to observe the real time kinetics of large-scale pH induced protein structure rearrangement of the anthrax toxin protective antigen (PA) component as the prepore ($PA_{prepore}$) transitions to its pore (PA_{pore}) form. This transition on the biosensor tip was verified using electron microscopy (Naik, Brock et al. 2013). Removal of complexes from biosensor surfaces avoids larger volume dilution effects frequently encountered when releasing complexes from chip surfaces when using microfluidics systems.

In the current work, anthrax toxin complexes are assembled and disassembled on the biosensor surface and then released into microliter volumes. The resulting complex components are validated orthogonally using electron microscopy and mass spectrometry.

RESULTS

The ability to monitor and validate the assembly of large macromolecular complexes is a crucial step towards understanding the specificity and functionality of large biomolecular assemblies. The results of the methods presented herein demonstrate the ease with which large protein complexes (>150 kDa mass) can be assembled using biolayer interferometry, all while monitoring the kinetics and amplitude of assembly. The unique compact nature of the biosensor surface enables assembly analysis to be extended by releasing assembled complexes into small microvolumes. These microvolumes can be used to visualize the initial physical structure of the complexes using electron microscopy, and verify the identity of the complex components using mass spectrometry. A schematic overview of this entire process is shown in **Figure 2.1**.

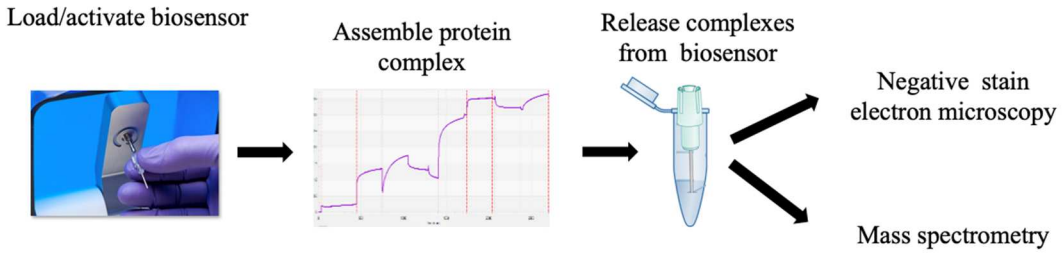


Figure 2.1. Analyzing protein complexes assembled on and isolated from BLI biosensor using EM and MS: Schematic overview for analysis of protein complexes assembled on and isolated from BLI biosensor using EM and MS.

A key element for the successful assembly and verification of macromolecular complex on the biosensor surface involves the proper orientation of the initial seed protein. This ensures the protein-protein interaction sites are accessible, not sterically blocked, and optimally positioned away from the biosensor surface. As shown in **Figure 2.2**, the proper orientation of the anthrax toxin complex is achieved by using a specifically engineered N-terminal fragment of lethal factor (LF_N) so the LF_N-$PA_{prepore}$ binding site is always positioned opposite of the biosensor covalent

attachment site (Naik, Brock et al. 2013). The subsequent buildup of the complex with binding of PA to the LF_N BLI biosensor followed by soluble CMG2 binding to PA ultimately creates a translocation competent anthrax toxin complex.

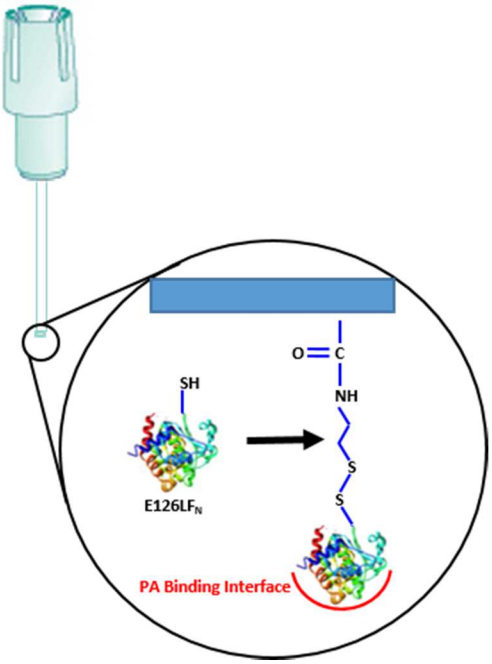


Figure 2.2. Biosensor activation: First step in orientation specific assembly on BLI biosensor. The E126C Lethal Factor N-terminal domain, (LF_N) is linked through a thiolinkage creating the properly orientated protective antigen prepore binding interface.

The BLI sensogram trace is a real-time read out of the amplitude changes due to specific addition of the anthrax toxin components as they are added onto the biosensor. **Figure 2.3** shows a representative trace paired with a model of the complex predicted to form at that step in the process. The first rise is LF_N loading onto the tip. After quenching, $PA_{prepore}$ then binds to LF_N followed by the addition of soluble CMG2 receptor resulting in the assembled pre-endosomal complex of LF_N-$PA_{prepore}$-CMG2. To progress toward the late endosome environment, the entire complex is subjected to a low pH pulse (pH 5.0) that weakens the receptor binding, allowing the

prepore to transition to its extended membrane inserted pore conformation (Naik, Brock et al. 2013). Sensogram traces of the acidification step are shown in **Figure 2.4**. The initial increase or 'spike' in amplitude is likely the pore extension (Naik, Brock et al. 2013) that must occur prior to the decrease in the CMG2 receptor binding. The larger amplitude decline is most likely substantial or complete receptor dissociation due to diminished binding affinity. Previous work in this laboratory indicated CMG2 binding to the fully extended PA_{pore} is negligible compared to the CMG2-$PA_{prepore}$ interaction (Naik, Brock et al. 2013). In addition, the sensogram kinetic trace, observed in all sensograms of LF_N-$PA_{prepore}$-CMG2 transitions to LF_N-PA_{pore} with a pH drop, is reproducible over multiple runs.

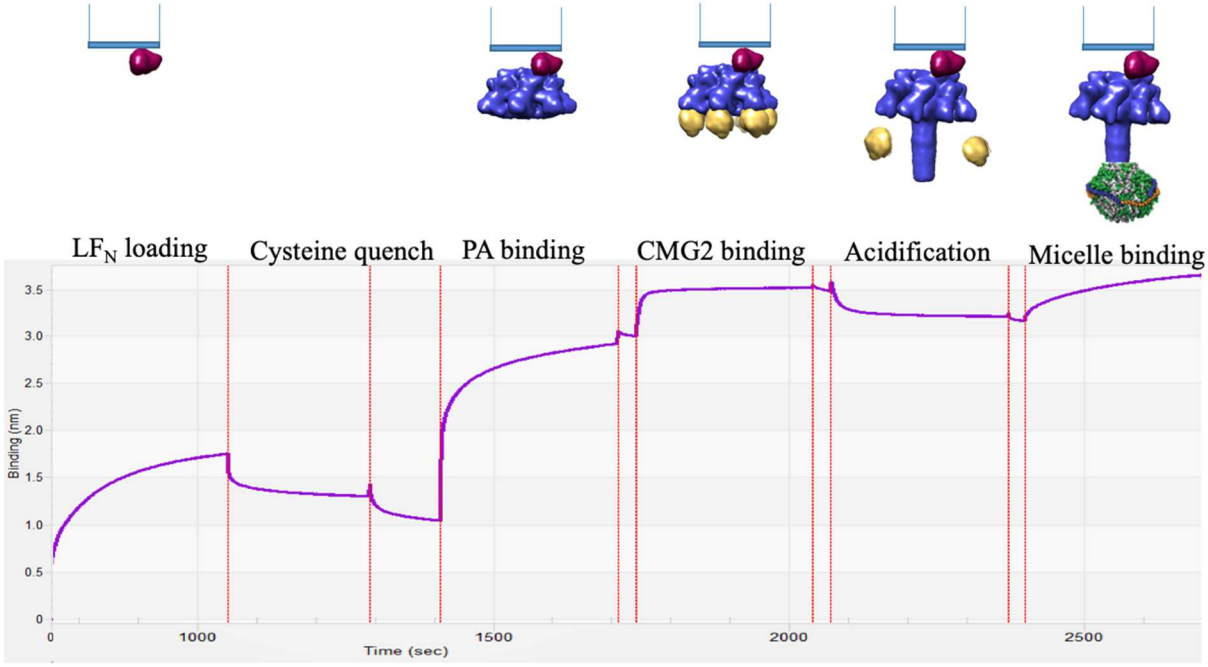


Figure 2.3. Monitoring anthrax toxin assembly and disassembly with BLI: The sensogram trace tracks the kinetics and amplitude changes due to specific addition of the anthrax toxin components as they are added onto the Biosensor starting with the LF_N Loading. $PA_{prepore}$ is then loaded onto the surface followed by the addition of soluble CMG2 receptor. The assembled pre-endosomal complex consists of an LF-$PA_{prepore}$- CMG2. To progress toward the late endosome environment, the entire assemble functional anthrax toxin is subjected to a low pH pulse (pH 5.0) that weakens the receptor binding, allowing the prepore to transition to its extended membrane inserted pore conformation. Exposure of the membrane spanning pore is confirmed and solubilized by addition of micelles.

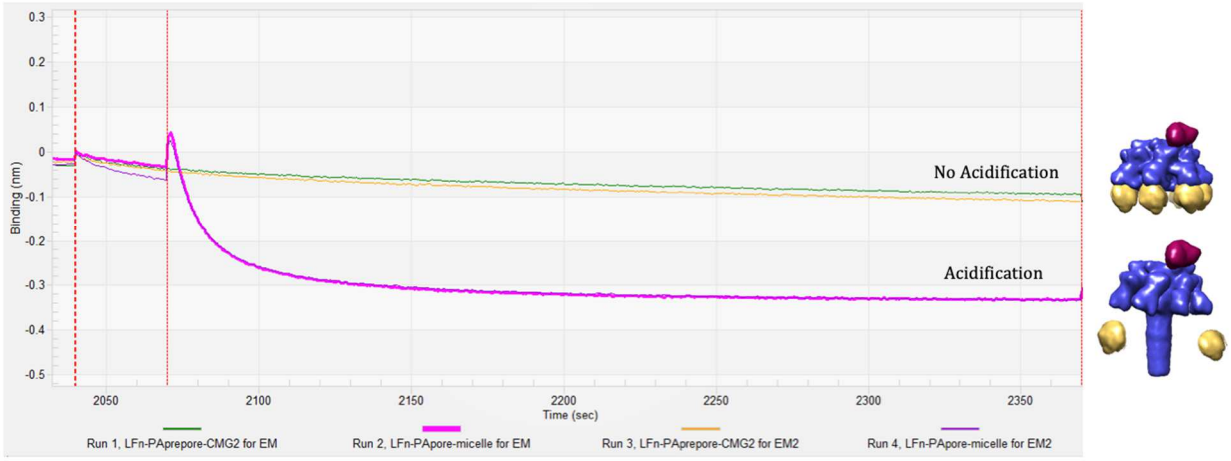


Figure 2.4. BLI sensogram of CMG2 release: Traces of the acidification step show an initial increase or 'spike' in amplitude followed by a larger amplitude decline that are likely pore formation and receptor dissociation, respectively.

Prior to and after acidification, the biosensor attached complexes are easily released for visualization by negative stain electron microscopy and identification by mass spectrometry (**Figure 2.5**). Representative complexes from the EM results are shown in **Figure 2.6**. Pre-endosomal sample grids, show densities consistent with intact ternary complexes consisting of LF_N-$PA_{prepore}$-CMG2. Post-acidification complex grids, show PA transitioned to pore and solubilized by micelle inclusion with no obvious CMG2 density.

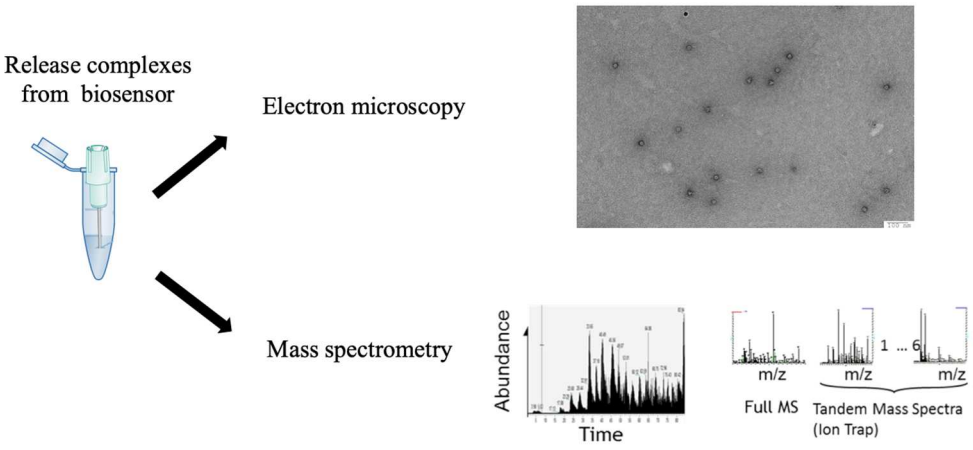


Figure 2.5. Analyzing protein complexes assembled on and isolated from BLI biosensor using EM and MS: Biosensor attached complexes are easily released into 5uL of buffer containing DTT for visualization by negative stain electron microscopy and identification by mass spectrometry.

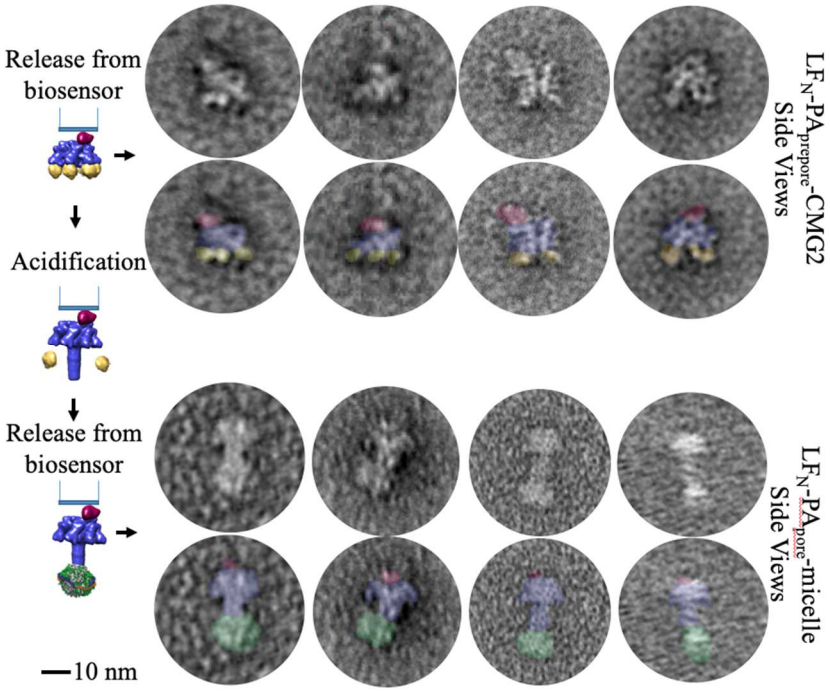


Figure 2.6. Visualization of protein complexes with electron microscopy: Pre-endosomal sample grids, show densities consistent with intact ternary complexes consisting of LF_N-$PA_{prepore}$-CMG2. Post-endosomal complex grids, show PA transitioned to pore and solubilized by micelle with no obvious CMG2 density. Models of predicted complexes (left hand side) are at the same scale as individual particles shown. Particles colorized with predicted protein (based on size of EM density) are shown below each particle.

The identities of the pre- and post-acidification complexes were verified by the mass spectrometry. A database where the sequences of PA, P13423; LF_N, P15917; and CMG2, P58335 were included in a background of a mouse proteins database derived from the NCBInr repository. Only the proteins of interest were obtained from this first database search, with the following amino acid coverage for the ternary and binary complexes: 54% and 22% for PA, 36% and 6% for LF and 43% for CMG2, respectively (CMG2 was not detected on the binary complex). In order to maximize the protein amino acid coverage, a second peptide/protein identification was performed using a protein database containing only the three protein of interest. Pre-endosomal MS samples contained peptides from all three toxin components with 60.46%, 67.97%, and 54.15% coverage for LF_N, PA, and CMG2, respectively (**Figure 2.7**). Post-endosomal results, shown in **Figure 2.8**, only contained peptides from LF_N and PA (57.41% and 67.79% coverage, respectively). The lack of CMG2 in the post-endosomal samples are consistent with the observed BLI nm decrease during pore formation.

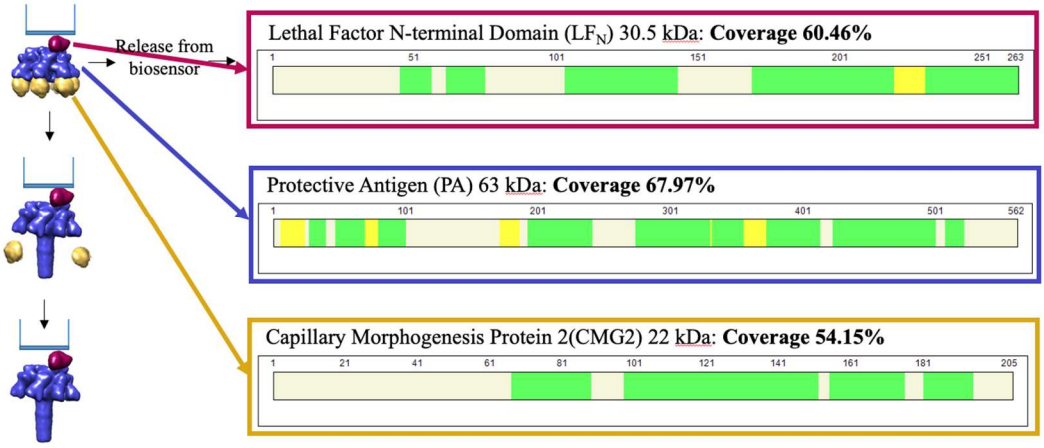


Figure 2.7. Verification of protein complexes with mass spectrometry for prepore: Pre-endosomal MS samples contained peptides from all three toxin components with 60.46%, 67.97%, and 54.15% coverage for LF_N, PA, and CMG2, respectively. Peptides detected at a false discovery rate (FDR) equal or higher than 5% shown in yellow, FDR equal or higher than 1% shown in green.

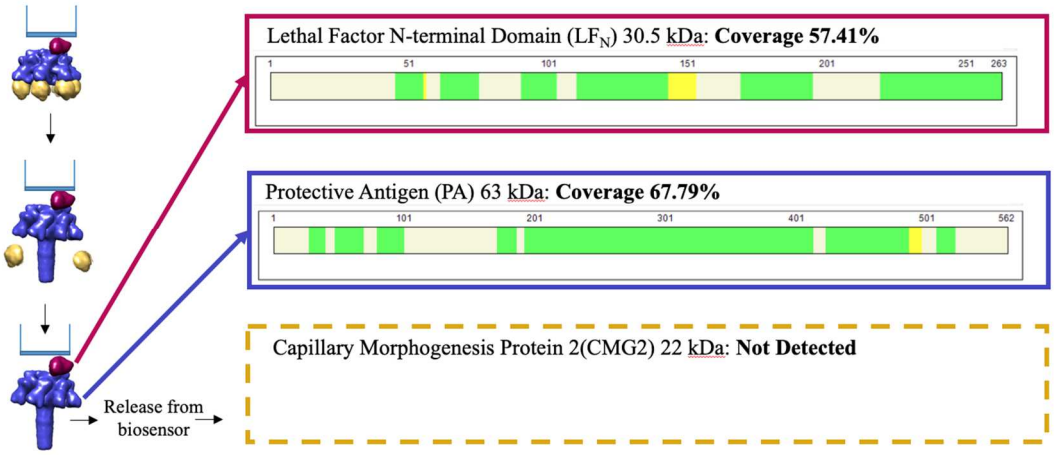


Figure 2.8. Verification of protein complexes with mass spectrometry for pore: Post-endosomal samples contained peptides from LF_N and PA (57.41% and 67.79% coverage, respectively), but not CMG2. FDR equal or higher than 5% shown in yellow, FDR equal or higher than 1% shown in green.

DISCUSSION

This demonstration illustrates how macromolecular complex formation can be easily monitored using biolayer interferometry, visualized using electron microscopy, and verified with mass spectrometry, all with microvolumes in a short time span. Structural assembly and observed complexes follow the biologically relevant predictions, further validating this combined methodology. As mentioned in the results section, the key element for assembly success requires rationally engineered cysteine mutagenesis to ensure that the complex protein-protein interfaces are properly oriented away from the biosensor surface.

Previous systems have used BLI and MS techniques to evaluate protein binding of two and three component systems as well as integrity of receptor binding of expressed protein, but, in both instances, the methods were not developed to take advantage of the tandem EM/MS approach (Jin, Cantin et al. 2011, Yamniuk, Edavettal et al. 2012).

The only other interferometry system that combined mass spectrometry analysis to help characterize interactions was a dual-polarization interferometry (Moore, Perez-Pardo et al. 2011). Unfortunately, this system is no longer available for general use. As mentioned in the introduction, there have been a number of studies completed where samples were formed on SPR-like biosensor surfaces and removed from mass spectrometry analysis. None of those examples resulted in complexes visualized using EM. We previously used BLI in combination with EM to confirm chaperone binding to denaturant destabilized proteins (Lea, O'Neil et al. 2016, O'Neil, Machen et al. 2018). See **Appendix B**. However, these were two component systems and MS was not used in parallel to EM.

Limitations

The limitations of this sequential method can be numerous but are solvable. For the initial immobilization and therefore foundation step of the assembly, a lack of knowledge of the structural interaction surfaces would certainly impede progress associated with monitoring initial assembly phases. The lack of structure information can be addressed by designing an engineered foundation construct where attachment chemistries (*e.g.*, sulfhydryl moiety for disulfide linkages/ His-tagged positioning) can be moved to various regions within the core assembly system. In the case of the anthrax toxin complex, it was fortunate that the structure of the lethal factor prepore complex is available (Feld, Thoren et al. 2010). Rational placement of the engineered cysteine was localized to regions away from the $PA_{prepore}$ binding face. With cysteine linkage chemistries, it is preferable that no other reactive cysteines are present on the protein surface.

There are various attachment chemistries which can be used to engineer specific attachment site on a protein's surfaces. One of the most popular specific attachments involves positioning a

biotin moiety at a specifically defined location on the protein surface (Fairhead and Howarth 2015). Unfortunately, biotin binding to a streptavidin or avidin coated surfaces is quite tight. Reversal of the binding interaction is not simple. The use of an engineered His-tag at the N- or C-terminus and the subsequent ease of attachment to Ni-NTA surfaces is a more universal application of affinity immobilization. Of course, one of the caveats for engineering assembly attachment sites with His-tagged systems is the requirement that the N- and C-termini of the core assembly protein remain exposed and separated so that the attachment is easy. As with all assembly processes, the interaction interface of the core assembly protein must remain available as the complex assembly progresses.

Perhaps the most common concern of using biosensor surface chemistries is non-specific binding. Streptavidin tips are often a source of significant non-specific binding effects. Specific disulfide linked biotin can be used to release very specific complexes leaving behind the reduced S-biotin linkage tightly bound to immobilized streptavidin biosensors (Naik, Kumru et al. 2014). There are other reversible chemistries becoming available such as iminoboronates and to a lesser extent ketoamide (Bandyopadhyay and Gao 2016). This field is currently underdeveloped, but there is high interest in further developing reversible covalent protocols to avoid off-target drug toxicity effects that commonly accompany the use of covalent targeted drug development.

A limitation of using EM to visualize complexes is interpretation, especially in instances where the structures of the assembled complexes are not initially known. The spatial location of components within an undefined assembled macromolecular complex can be identified using monoclonal antibodies (mAb) as specific kinetic and structural markers. For example, once a complex is formed, specific monoclonal antibodies can be added that bind to specific components. This method is frequently used in electron microscopy to identify specific

components within large assemblies (Grantham, Llorca et al. 2000). Another limitation is related to size of the complex, although there have been instances where defined symmetric assemblies as small as 70 kDa (GroES heptamer) are easily resolved using negative stain electron microscopy. Assembled complexes that are analyzed by EM are typically in the size range of ~100 Å in diameter or above. Recently however, proteins as small as 20 kDa have been resolved and low resolution structures have been obtained when using superior staining methodologies (Ercius, Alaidi et al. 2015).

For mass spectrometry analysis, the increased sensitivity of current mass spectrometry instrumentation down to the femtomolar level can in some cases increase the sensitivity of BLI detection. It is highly conceivable that protein signals that show a minimal but repeatable rise in amplitudes will result in identification of the protein in question. In addition, probing protein-protein interactions with one of the partners attached on the biosensor and the other in a cellular milieu will in effect result in a purification and subsequently easier detection of the newly formed complex. One limitation that may be observed with the current highly sensitive mass spectrometry systems is that the protein of interest may not be in the database but this observation is rare (*e.g.*, proteomes from rare species). If the sequence of the proteins of interest are know, this problem is easily solved by including the protein(s) amino acid sequence in a background protein database (as described in this work in the protocol section). Another potential limitation of the methodology results from the resistance of a protein to trypsinolysis. Trypsin digestion is typically the default method for bottom up protein identification. However, proteins may be resistant to trypsin if they lack Arg and Lys residues or access to these residues are restricted by the folded structure. These limitations are resolved, respectively, by using

alternative or a combination of proteases or including an unfolding reagent (urea or guanidine HCl) before enzymatic digestion.

Future Directions

Possible expansions of this methodology include allowing the user to follow and identify cellular assembly complexes from crude cellular lysates. It turns out testing for assembly components in concentrated cellular extracts is easy to perform using the biolayer interferometry procedures. Unlike more commonly used microfluidic based methodologies which are prone to clogging and sensitive to aggregation, the BLI approach can be used to directly immerse biolayer sensor tips into crude extracts to potentially assemble specific complexes directly from these concentrated impure samples. Once assembled, it is entirely feasible to use specific antibody probes as a follow up to the BLI system to further identify and even quantitate suspected components in cellular extracts that were identified using the microvolume MS method. Again, the key here is to use defined, properly oriented core proteins as specific affinity probes.

The ability to view the prepore to pore transition process kinetically with BLI will be highly useful in identifying potential "anti-toxin" small molecule inhibitors of the protein transitions that specifically function under late endosomal, low pH (5.0) conditions. This specific pH induced prepore to pore transition is inhibited in the presence of folding stabilizer (osmolytes) such as glycerol or sucrose and thus lends strong support for developing specific targeted folding stabilizers that load and lock the $PA_{prepore,}$ preventing PA_{pore} formation. This specific approach avoids and replaces crude aggregation-based assays where pH drops lead to protein precipitation. This latter method, although good for primary prescreening methods, often leads to false positive results where specific compounds inhibit the aggregation rather than the actual molecular transitions.

The downstream observation of the structure and identification of the individual assembled components within small microvolume samples can also be useful in validating potential lead compounds. This can be applied in instances where either specific assembly stabilization or destabilization is the target outcome. This kinetic/structure/identification parallel approach is useful for directly confirming the validity of suspected lead compound effectors of assembly and serves as a reasonable secondary confirmation screening step or medium throughput approach.

Cryogenic electron microscopy (cryoEM) is a useful technique to study the atomic details of macromolecule complexes in various states of assembly. Prior to preparing cryoEM samples, it is important to first verify a preparation contains reasonably pure homogenous complexes with negative stain EM. The work presented herein demonstrates assembly of protein complexes on BLI biosensor surfaces, release of these complexes for EM visualization, and identification of these components using mass spectrometry in microvolume. This particular methodology of controlled assembly and release can be useful in generating very specific protocols that enhance homogenous sequential sample preparation for negative stain EM, a necessary step that must be demonstrated before advancing to cryoEM. To obtain low resolution 3D structure, only 30-50 particles of complex would be needed to perform a conical tilt series (70 different 2D image views per particle) provided there is orientation diversity (multiple different views).

With respect to enhancing mass spectrometry methods, advances in sensitivity and reduction in sample volume continue to improve. Nano flows and ultra-high pressure liquid chromatography together with the development of mass spectrometers with a fast duty cycle, increased sensitivity and resolving power. Recent introduction of the orbitrap mass spectrometer, in particular the latest version (orbitrap Fusion Lumos, and its expected successor the Orbitrap Fusion Lumos 1M) as well as search algorithms greatly facilitate this process.

Conclusions

The current methodology monitors the kinetic assembly and disassembly of anthrax toxin components using label-free BLI methodologies and evaluates the structure and identity of these components using EM and MS, respectively. The use of a simple single channel BLI system coupled with routine negative staining EM analysis and elementary MS techniques are more than adequate to characterize an assembly process.

MATERIALS and METHODS

Assembly of defined macromolecular complexes on BLI biosensor surfaces

Assembly of prepore complex on PDEA-modified amine reactive biosensor surface

Hydrate amine reactive second generation (AR2G) BLI biosensor tip in 250 µL water for ten minutes. Program step times for run on BLItz instrument using BLItz software. Start BLI run by immersing biosensor tip in 250 µL water to measure initial baseline of biosensor thickness and density. Activate the biosensor by immersing tip into 250 µL 50 mM NHS (N-hydroxysuccinimide) and 200 mM EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide)) for 7 minutes. Immerse the activated biosensor in 50 mM PDEA (2-(2-pyridinyldithio)ethanamine) dissolved in 0.1 M borate buffer (pH 8.5) for 5 minutes to generate an activated thiol-reactive surface. Immerse the activated thiol-reactive biosensor into 250 µL of solution containing 100 nM E126C LF_N in 10 mM sodium acetate pH 5.0, 100 mM NaCl buffer. Immerse the LF_N tip in 50 mM L-cysteine, 1 M NaCl, 0.1 M sodium acetate pH 5.0 to quench any remaining reactive free thiol-reactive groups. Immerse the quenched LF_N tip into 0.5 µM protective antigen prepore ($PA_{prepore}$), 50 mM Tris, 50mM NaCl for 5 minutes to create LF_N-$PA_{prepore}$ complex. Once $PA_{prepore}$ is associated, remove the tip from the $PA_{prepore}$ solution and immerse the tip into 50 mM

Tris, 50 mM NaCl for 1 minute to wash away any non-specifically bound $PA_{prepore}$. Immerse the LF_N-$PA_{prepore}$ complex into 0.5 μM CMG2 receptor (without the transmembrane domain), 50 mM Tris, 50 mM NaCl for 5 minutes. Immerse the LF_N-$PA_{prepore}$-CMG2 complex into 50 mM Tris, 50 mM NaCl for 5 minutes to wash away any unbound CMG2 to form pre-endosomal complex.

For electron microscopy analysis, release the LF_N-$PA_{prepore}$-CMG2 complex from the biosensor tip by immersing the tip into a 5 μL of 50 mM DTT, 50 mM Tris, 50 mM NaCl inside a PCR tube. For tandem mass spectrometry analysis of the peptides from the complex, release the LF_N-$PA_{prepore}$-CMG2 complex from the biosensor tip by immersing the biosensor into a 5 μL volume of 50 mM DTT, 6 M GuHCl (keratin-free), 25 mM ammonium bicarbonate pH 8.0 inside a PCR tube. This is performed on a different biosensor than the one used for electron microscopy analysis.

Assembly of pore anthrax toxin complex on PDEA-modified amine reactive biosensor surface

To view the complex after pH transition, LF_N-$PA_{prepore}$-CMG2 complex were immersed into a 10mM acetate pH 5.0 to initiate the transition of the $PA_{prepore}$ to PA_{pore} transition. This transition, as indicated by increasing amplitude (approx. 0.2nm) is followed by a larger amplitude decline that is hypothesized to be substantial or complete receptor dissociation due to diminished binding affinity.

For the electron microscopy analysis sample, immerse the biosensor tip into a solution containing 1.25 mM micelles (2.5 mM MSP1D1, 25 mM Na-cholate, 162.5 mM POPC) to prevent aggregation in solution after disulfide release. For electron microscopy analysis, release the LF_N-PA_{pore}-Micelle complex from the biosensor tip by immersing the tip into a 5 μL of 50

mM DTT, 50 mM Tris, 50 mM NaCl inside a PCR tube. For tandem mass spectrometry analysis of the peptides from the complex, release the LF_N-PA_{poore} complex from the biosensor tip by immersing the biosensor into a 5 µL volume of 50 mM DTT, 6 M GuHCl (keratin-free), 25 mM ammonium bicarbonate pH 8.0 inside a PCR tube. This is performed on a different biosensor than the one used for electron microscopy analysis.

Visualizing and validating released macromolecular assemblies from BLI biosensors by negative stain electron microscopy

Glow discharge a carbon-coated Cu 300 grid. Typical glow discharge settings are 0.38 mBar stable atmosphere pressure, negative 15 mAmps, 20 seconds then vented with air. Secure grid between a pair of clean tweezers. Pipette 4 µL of released complex sample in PCR tube onto the grid and allow adsorption for 60s. Wick away remaining liquid with a filter paper wedge. Stain the grid by pipetting 5 µL of 0.75% 0.02 micron filtered uranyl formate and wicking away excess stain after 5 seconds. Allow grid to dry at room temperature. View stained sample grids using transmission electron microscope.

Identification of complete pre-endosomal anthrax toxin complex (LF_N-$PA_{prepore}$-CMG2) and transitioned complex (LF_N-PA_{pore} without CMG2) using mass spectrometry

Dilute released samples in PCR tube to 20 µL, and incubated for 1 hour. Add 2 µL of 55 mM iodoacetamide, 25 mM ammonium bicarbonate pH 8.0 and incubate for 1 hour at room temperature in dark (covered with aluminum foil). Dilute the sample with 100 µL of 25 mM ammonium bicarbonate pH 8.0 to reduce the guanidine hydrochloride concentration below 1 M. Add 5 µL of sequencing grade modified trypsin at 20 ng/ µL and incubate at 37°C overnight. Add acetic acid glacial to a final concentration of 5% to reduce the pH to <3 then reduce volume to 10 µl on speedvac. Transfer the peptide solution to the sample plate on the autosampler of the

nLC 1200 uHPLC. Load 5 µL the peptide solution onto a uHPLC reversed phase column mounted on the ionization stage of the mass spectrometry. Wash column with 15 µL of 0.1% formic acid at a maximum rate of 5 μL/min and/or maximum pressure of 800 psi. Elute peptides from the reversed-phase C18 column at a flow rate of 350 nL/min over a 90 min period using a lineal gradient of 5% to 40% of solvent B in A+B (solvent A: 0.1% formic acid; solvent B: 95% acetonitrile with 0.1% formic acid). Analyze eluting peptides on line using tandem mass spectrometry on an Orbitrap Fusion Lumos. The ionization source was operated at 2500 volts and ion transfer temperature was set to 250^{o}C. The mass spectrometer was operated under automatic control to perform continuously one MS scan followed by as many tandem MSMS as possible on a 3 seconds period, using CID and a normalized collision energy of 35. Identify peptide and protein components using standard methods. For this work, two sets of peptide and protein analysis were performed using the Sequest HT search engine included in the Protein discoverer software suite of applications.

Chapter 3: Asymmetric CryoEM Structure of Anthrax Toxin Protective Antigen Pore with Lethal Factor N-Terminal Domain

This chapter has previously been published and is reprinted with permission. Machen, AJ, Akkaladevi, N., Trecazzi, C., O'Neil, PT, Mukherjee, S., Qi, Y., Dillard, R., Im, W., Gogol, EP, White, TA and Fisher, MT (2017). Asymmetric cryoEM structure of anthrax toxin protective antigen pore with lethal factor N-terminal domain. *Toxins*, 9 (10), p.298.

ABSTRACT

The anthrax lethal toxin consists of protective antigen (PA) and lethal factor (LF). Understanding both the PA pore formation and LF translocation through the PA pore is crucial to mitigating and perhaps preventing anthrax disease. To better understand the interactions of the LF-PA engagement complex, the structure of the LF_N-bound PA pore solubilized by a lipid nanodisc was examined using cryoEM. CryoSPARC was used to rapidly sort particle populations of a heterogeneous sample preparation without imposing symmetry, resulting in a refined 17 Å PA pore structure with 3 LF_N bound. At pH 7.5, the contributions from the three unstructured LF_N lysine-rich tail regions do not occlude the Φ clamp opening. The open Φ clamp suggests that, in this translocation-compromised pH environment, the lysine-rich tails remain flexible and do not interact with the pore lumen region.

INTRODUCTION

The lethality of anthrax, a zoonotic disease and bioterrorism agent, is due to the anthrax toxin. This tripartite toxin consists of a protective antigen (PA), lethal factor (LF; a mitogen-activated protein kinase kinase protease), and edema factor (EF; an adenylate cyclase) (Young and Collier 2007). After secretion from *Bacillus anthracis*, the 83 kDa PA (PA_{83}) binds to its target host cell receptor, either capillary morphogenesis protein 2 (CMG2) or tumor endothelium marker-8 (TEM8) (Lacy, Wigelsworth et al. 2004, Santelli, Bankston et al. 2004, Young and Collier 2007, Wimalasena, Janowiak et al. 2010). PA_{83} is cleaved by proteases, resulting in 20 kDa and 63 kDa fragments. PA_{63} then self-associates to form a heptameric PA prepore that can associate with up to three molecules of LF or EF (Mogridge, Cunningham et al. 2002). Octameric PA prepores may also assemble in solution, governed by LF or EF binding to PA_{63} monomers clipped in solution (Kintzer, Thoren et al. 2009). Receptor-bound assembled complexes are endocytosed. As the endosome acidifies to pH 5.0 (late endosome), the receptor-bound PA prepore undergoes a conformational change into an extended β-barrel pore structure that penetrates the endosomal membrane. This newly-formed structure facilitates unfolding and translocation of the 90 kDa LF (or EF) enzyme across the pH gradient of the endosomal membrane through the narrow PA pore lumen in a pH-driven hypothesized Brownian ratchet mechanism (Krantz, Finkelstein et al. 2006). Translocation of α-helical regions of LF are aided by the PA α-clamp (Brown, Thoren et al. 2015, Das and Krantz 2017). LF translocation is gated by a ring of seven phenylalanine residues, termed the Φ clamp, located further down the PA pore lumen (Krantz, Melnyk et al. 2005, Melnyk and Collier 2006, Janowiak, Fischer et al. 2010, Janowiak, Jennings-Antipov et al. 2011). The directional translocation of LF depends on protonation of acidic residues, the electrostatic character of the PA pore lumen, and any residual positive charges on LF (Wynia-

Smith, Brown et al. 2012). Subsequent deprotonation of the translocating peptide after passing through the Φ clamp prevents back transfer. Translocated LF refolds on the cytosolic side of the endosomal membrane, where it disrupts cell signaling by cleaving MAP kinase kinases, resulting in cell death (Pannifer, Wong et al. 2001).

Previously, the Krantz group published work on the large-scale rearrangement of LF that occurs upon binding to the PA prepore. Specifically, the N-terminal α-helix of LF moves away from the main body of LF and is resituated into a groove in the interior surface of the PA prepore cap, termed the α-clamp region (Feld, Thoren et al. 2010). This reposition is proposed to help funnel the N-termini of LF into the PA pore lumen. The narrowest part of the pore lumen is the Φ clamp (Krantz, Melnyk et al. 2005, Jiang, Pentelute et al. 2015). PA F427, which forms the heptameric Φ clamp, is an essential residue that facilitates LF translocation (Sellman, Nassi et al. 2001, Janowiak, Fischer et al. 2010). Mutations in this residue (*e.g.*, F427A) affect the kinetics of pore formation and translocation (Sun, Lang et al. 2008). Interestingly, the loop containing F427 ($2\beta_{10}$–$2\beta_{11}$ loop) was suggested by Jiang *et al*. (Jiang, Pentelute et al. 2015) to be involved in the first unfolding step of the pore formation mechanism. This mechanism is based on a comparisons of crystal structures of the oligomeric PA prepore (Fabre, Santelli et al. 2016) and the 2.9 Å cryoEM pore structure (Jiang, Pentelute et al. 2015). The $2\beta_{10}$–$2\beta_{11}$ loop also contains D426 which forms a conserved inter-subunit salt bridge with K397 in the PA pore (Melnyk and Collier 2006). These interactions orient F427 into its constricted Φ clamp formation, which is hypothesized as pi-stacking interactions between adjacent F427 residues (Melnyk and Collier 2006, Young and Collier 2007). The first step of this pore forming mechanism is based on the increased flexibility of the $2\beta_{10}$–$2\beta_{11}$ loop in various prepore crystal structures. Early characterization of LF-PA interactions showed the N-terminal tail of LF interacts with the Φ

clamp of the PA pore at pH 5.0, which has since been verified using cysteine cross-linking (Janowiak, Fischer et al. 2010, Janowiak, Jennings-Antipov et al. 2011).

PA pore formation is necessary, but not sufficient, for lethality: LF must be translocated through the pore. Das and Krantz (Das and Krantz 2017) recently proposed the Φ clamp region is dynamic and can undergo large-scale movements to momentarily increase the pore diameter from 6 Å to 10–12 Å. These movements could resemble transient open forms due to salt bridge formation between the acidic residues in the Φ clamp loop and an adjacent monomer (Melnyk and Collier 2006). This latter conformation (open Φ clamp loop) affects translocation rates since mutations that inhibit salt bridge formation impact translocation kinetics. In this particular model, Krantz also presented single-channel evidence that α-helical structures translocate more efficiently than extended β-sheet-like structures or unstructured polypeptides containing alternating L- and D-amino acids. Pentelute *et al*. (Pentelute, Sharma et al. 2011) showed chirality is not important for translocation of the unstructured region of the N-terminal domain of LF (LF_N) either, but this does not preclude the possibility that α-helical structures could be formed upon the electrostatic interaction between LF_N and the PA pore. This would have to include α-helices that are in the d chiral form since all L- or D- amino acid α-helical structures do not slow down translocation. It would then be of value to determine if the Φ clamp loop region becomes more structurally dynamic (loss of resolution) and/or adapts a more open configuration upon interaction with the single or multiple unstructured lysine-rich tails of bound lethal factor(s).

Understanding both pore formation and LF translocation is imperative in order to develop strategies that mitigate or prevent the formation of the anthrax toxin complex or inhibit the translocation mechanism. Inhibition of circulating anthrax toxins is crucial since the toxin

components retain cell lethality even after the bacilli have been killed with antibiotics (Young and Collier 2007). To better understand the interactions between LF_N and PA, the structure of the LF_N-bound PA pore in a lipid membrane environment was examined using cryo-electron microscopy (cryoEM).

RESULTS

CryoEM Sample Preparation of PA Pore with Three LF_N Bound

With the recent publication of the cryoEM PA pore structure at pH 5.0 (Jiang, Pentelute et al. 2015), the logical, but challenging, next step in understanding anthrax toxin pore formation and translocation involves determining how bound LF influences the conformation of the PA pore. An atomic resolution structure of LF_N-bound PA pore would give molecular insight into the nuances of this interaction. In order to solve the cryoEM structure of the LF_N-PA pore, several obstacles must be overcome, including the aggregation propensity of the pore, asymmetry of the LF_N-PA complex, and orientational preferences of complexes on EM grids.

We previously published a methodology to assemble LF_N-PA pore complexes while avoiding aggregation by immobilizing PA pores before solubilizing the hydrophobic tip with lipid bilayer nanodiscs (Akkaladevi, Hinton-Chollet et al. 2013, Gogol, Akkaladevi et al. 2013, Naik, Brock et al. 2013, Akkaladevi, Mukherjee et al. 2015). After immobilization, the PA prepores were transitioned into pores using a urea/37°C pulse methodology, exposing the aggregation-prone pore tip. The nanodisc formed around the hydrophobic pore tip while the complex was immobilized (Katayama, Wang et al. 2010, Akkaladevi, Hinton-Chollet et al. 2013, Gogol, Akkaladevi et al. 2013, Akkaladevi, Mukherjee et al. 2015). A schematic of this methodology is shown in **Figure 2.1A**. Our previous low-resolution LF_N-PA-nanodisc structures were

reconstructed from samples frozen on perforated carbon containing a thin carbon layer over holes (Akkaladevi, Hinton-Chollet et al. 2013). There were a number of caveats limiting the structural analysis of that preparation. Most importantly, large diameter nanodiscs (approximately 400 Å) were generated and required the use of thicker ice. In addition, LF_N-PA-nanodisc complexes interacted with the carbon layer, resulting in complexes preferring a side view orientation which displays the long axis of the heptameric PA pore rather than allowing for more diverse conformational orientations, including top views. Although these LF_N-PA-nanodisc complexes were inherently structurally asymmetric (symmetry mismatch of seven PA subunits to a maximum of three LF_N bound), their structures were generated by imposing seven-fold symmetry, which resulted in smearing of the LF_N-bound density. This coupled with the sample-induced constraints (thin carbon backing, thick ice, and Fresnel fringe effects for the sharp nanodisc protein interface) diminished the contrast of the protein. These constraints also interfered with the visualization of the PA β-barrel in the reconstruction.

To obtain a more concise LF_N-PA-nanodisc complex structure, these sample preparation issues had to be overcome. For better contrast, samples were frozen on simple perforated carbon grids without a thin carbon layer in order to achieve greater orientational diversity and were imaged with a JEM 2200FS electron microscope (60,000× magnification) (JEOL, Peabody, MA, USA). A representative micrograph with high defocus for better contrast (for visualization, not reconstruction purposes) with individual particles highlighted with red circles is shown in **Figure 3.1B**. Low-dose, low-defocus conditions were used to collect images for 3D reconstruction. Notably, the nanodiscs for these samples were significantly smaller than the previous larger nanodisc samples. The nanodisc size was dependent on the length of time that LF_N-PA-nanodisc complexes were immobilized as well as rotation of the sample tube. Under non-ideal conditions,

the pre-nanodisc micelles may merge, generating larger nanodisc diameters. Interestingly, larger nanodiscs often resulted in multiple PA pore-inserted nanodisc complexes (*e.g.*, sometimes four PA pores inserted into one large nanodisc). These larger nanodiscs were attributed to longer dialysis times that consistently resulted in merging of pre-nanodisc micelles. Reducing the time of incubation, ensuring adequate detergent dialysis with Bio-Beads, and constant rotation during formation yielded smaller nanodiscs within the expected diameter range (100–150 Å) containing a single PA pore (**Figure 3.1B**).

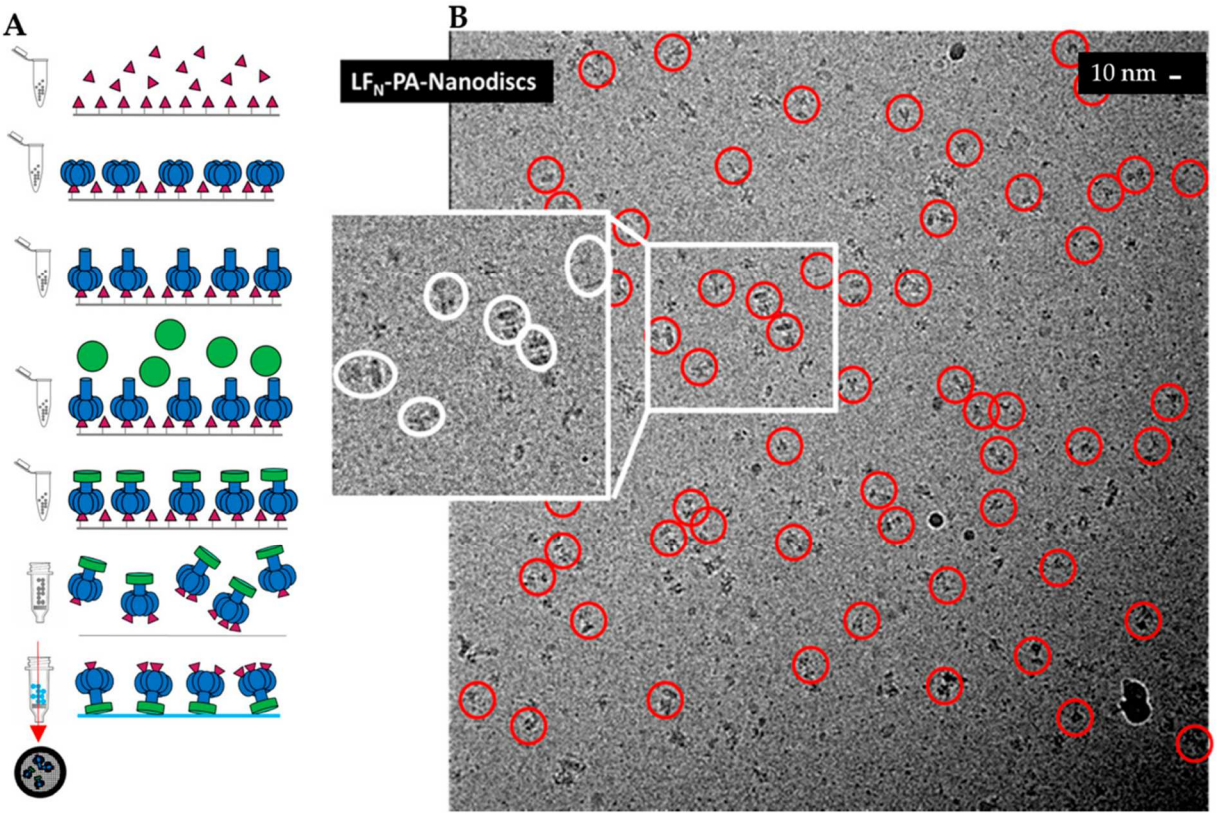


Figure 3.1. Sample preparation of 3LF$_N$-PA-nanodiscs: (**A**) schematic of LF$_N$ (magenta)-PA (blue)-nanodisc (green) complex formation with stepwise addition of LF$_N$ and PA to thiol sepharose beads; and (**B**) a higher defocus representative field for high-contrast visualization. Individual LF$_N$-PA-nanodisc complexes may be easily observed within this micrograph. Note the variable size of the nanodiscs in the insert.

Single-Particle Analysis of LF_N-PA-Nanodisc Complexes

Initial classification analysis using SPARX (Yang, Fang et al. 2012) revealed heterogeneity in the dataset with one, two, or three LF_N bound to PA pores (**Figure 3.2A–C**). The release of LF_N-PA-nanodisc complexes from the bead surface into solution also resulted in the release of non-complexed LF_N, which was then able to bind released complexes leading to particles with multiple binding events. This led to subsets of PA having one, two, or three LF_N bound. This inherent heterogeneity in LF_N binding stoichiometry made 3D reconstruction difficult. Initially, this limited particle dataset could only be used to obtain a model by imposing C7 symmetry during reconstruction using EMAN2.1 and RELION (**Figure 3.2D**).

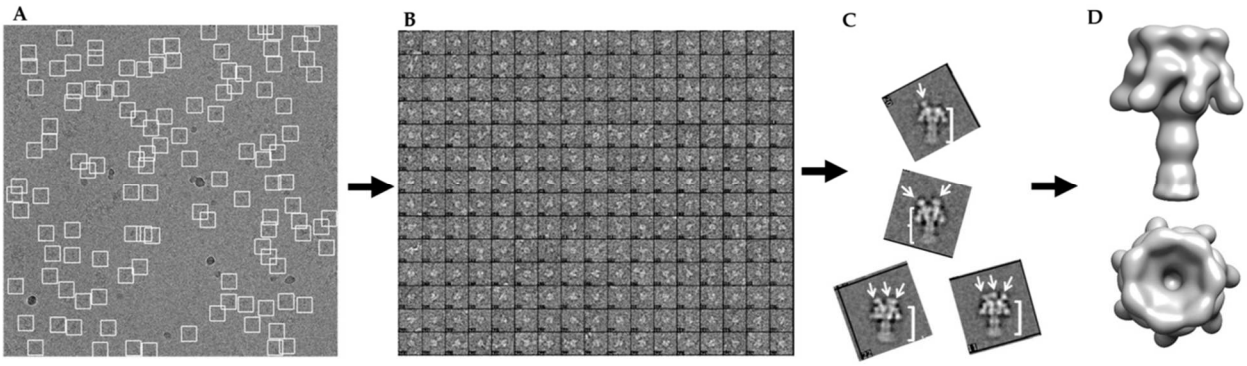


Figure 3.2. Schematic of data analysis of heterogeneous cryoEM data: (**A**) sample micrograph field showing good ice and particle distribution; (**B**) an example of individual boxed particles from micrographs with phase inversion for contrast; (**C**) SPARX 2D class averages (side views) reveal heterogeneity of sample preparation with arrows indicating LF_N binding; and (**D**) 3D model of $3LF_N$-PA pore with C7 symmetry imposed smears LF_N density into a crown around the top of the pore.

While PA alone has C7 symmetry, LF_N-bound PA in a saturated (three LF_N bound) or sub-saturated binding ratio only possesses C1 symmetry. The recent successful high-resolution reconstruction of the PA pore at pH 5.0 by Jiang *et al*. (Jiang, Pentelute et al. 2015) was accomplished using, primarily, top and side view orientations that were generated by taking

advantage of a grid adherence platform. In that sample preparation, the prepore adhered to the carbon layer through its receptor binding interface and the pore transition was accomplished by adjusting the pH of the solution to pH 5.0. Since the pore itself has an axis of seven-fold symmetry, the variable positioning of the side views of the PA pore on the carbon layer were sufficient to cover most of the conformational space to obtain the first high-resolution structure (2.9 Å) of the anthrax toxin pore translocon (Jiang, Pentelute et al. 2015). With LF_N-PA-nanodisc complexes, the nanodisc insertion procedure permits conformational diversity, which is critical for obtaining a structure without imposing sevenfold symmetry. A direction distribution map, analogous to an Euler angle map, confirmed the orientation of the LF_N-PA-nanodisc particles was conformationally diverse (**Figure 3.3**).

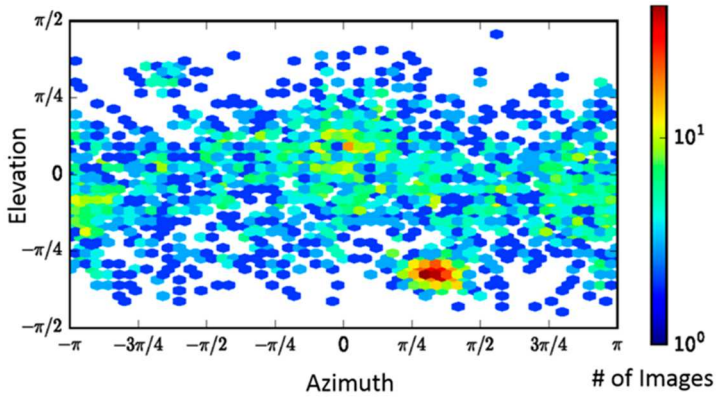


Figure 3.3. Direction distribution map of particles, analogous to an Euler angle map, showing the conformational coverage of LF_N-PA-nanodiscs.

It is important to note this diverse distribution is crucial for acquiring the asymmetric LF_N-PA-nanodisc structures since the imposition of sevenfold symmetry during 3D reconstructions distorts the density of any bound LF_N (**Figure 3.2D**). CryoSPARC is well suited to obtain unbiased, reproducible, and reliable *ab initio* 3D models rapidly even when extensive sample heterogeneity is present (Brubaker, Punjani et al. 2015, Punjani, Rubinstein et al. 2017). For

example, Ripstien *et al*. (Ripstein, Huang et al. 2017) reexamined their previous cryoEM data of the *Thermus thermophiles* V/A-ATPase using cryoSPARC and were able to determine their ATPase sample was actually populated by multiple conformations that were previously unresolved, resulting in new mechanistic insights.

To separate the heterogeneous LF_N-PA-nanodisc particles, an initial 2D classification was performed on the 30,696 particles with removal of bad classes as determined by eye (**Figure 3.4A**). An *ab initio* classification with four groups was then performed on the remaining 18,806 good particles (**Figure 3.4B**). Four groups were chosen since two LF_N can bind to PA at neighboring binding sites or with an empty binding site between them resulting in $1LF_N$, 2_ALF_N, 2_BLF_N, or $3LF_N$ bound. Group 2 was the most highly populated group identified by the cryoSPARC stochastic gradient descent (SGD) *ab initio* model generation with three distinct and equal LF_N densities (**Figure 3.4B**). Further 2D classification was performed on all four groups to assess the quality of particles within each group (**Figure 3.4C**). Group 3 contained several highly-populated classes showing sharp sevenfold symmetric top and bottom views. Group 1 and 4 particles did not result in clear classes and were discarded (**Figure 3.4C**, top and bottom panels). Since the top and bottom view classes in Group 2 were underrepresented, all particles from Group 2 (4560) and particles from the good classes in Group 3 (1159) were combined. A homogeneous refinement was run with the Group 2 *ab initio* model with the combined good particle set (**Figure 3.4D**).

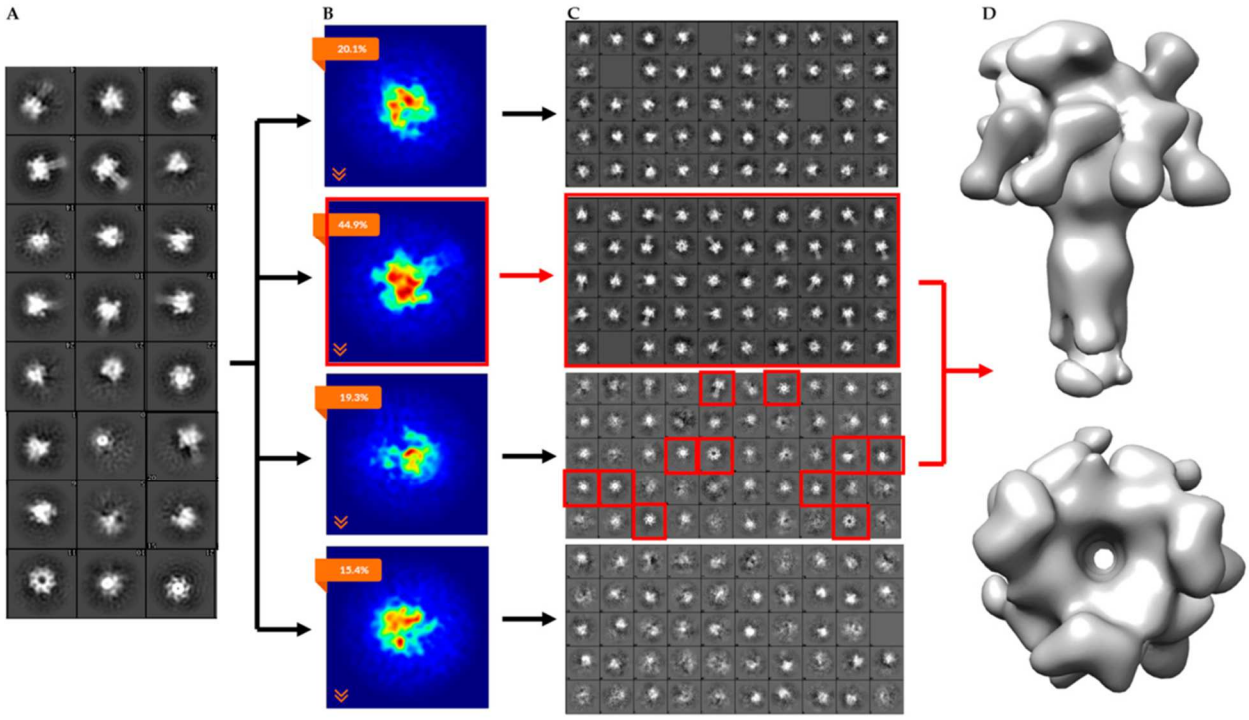


Figure 3.4. CryoSPARC data analysis flowchart of heterogeneous LF_N-PA-nanodiscs with total computational time of 3.5 h from 2D averaging to refined model: (**A**) cryoSPARC 2D class averaging of 18,806 particles; (**B**) image projection of heterogeneous *ab initio* reconstruction with four groups, the largest group, with 44.9% of the particles, corresponds to $3LF_N$; (**C**) 2D class averages of each *ab initio* particle group; and (**D**) 17 Å model of $3LF_N$-PA generated from homogeneous refinement of the Group 2 *ab initio* model with particles from top and bottom 2D class averages highlighted in red.

The homogeneous refinement resulted in a 17 Å $3LF_N$-PA pore model from 5719 particles. **Figure 3.5** shows the Fourier Shell Coefficient (FSC) used to calculate the resolution. This resulting reconstruction was not biased by outside models or symmetrization operations. The β-barrel pore of PA was not prominent in the *ab initio* model but became more apparent upon cryoSPARC refinement. The bulge in the β-barrel of the final model was also seen in the cryoEM structure of the PA pore alone where this hydrophobic region of the outer barrel bound lipids, resulting in the accumulation of additional density (Jiang, Pentelute et al. 2015). As can be seen in the 2D classification (**Figure 3.4C**, second panel), side view images reveal variation either in nanodisc size or electron density. This resulted in a lack of nanodisc structure in the

final electron density map. The irregular density at the bottom of the pore tip in the final structure can be attributed to either the presence of nanodisc or free lipid binding to exposed hydrophobic residues. As mentioned previously, the decrease in nanodisc density appears to be due to extended dialysis times during micelle to nanodisc collapse. The decreased nanodisc size did not diminish our ability to reconstruct LF_N-PA pore complexes, particularly in the PA pore cap and the initial extension of the β-barrel.

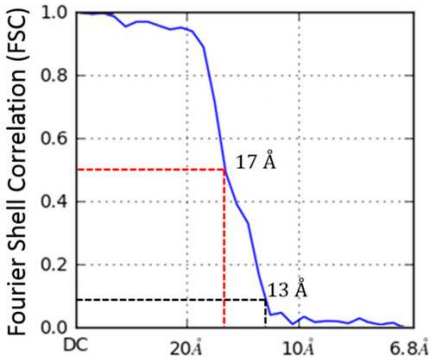


Figure 3.5. Fourier shell correlation (FSC) for 3LF_N-PA-nanodiscs. Resolution estimated to be 17 Å based on FSC with a cutoff of 0.5. This conservative cutoff agrees with filtered models shown later in Figure 10.

Constructing Samples with Highly-Populated Singly-Bound LF_N-PA for CryoEM

The heterogeneity of this sample preparation was due to the stepwise assembly of LF_N-PA complexes, shown above in **Figure 3.1A**. LF_N was immobilized onto thiol sepharose beads, then PA prepore was added, binding to the LF_N. The bulkiness of PA relative to LF_N blocked PA from binding to multiple LF_N. After LF_N-PA-nanodisc complexes were formed on the beads, they were released into solution. Any unbound LF_N was also released and, due to its high affinity for PA, bound to open binding sites of PA (**Figure 3.1A**). To obtain a larger, more homogeneous LF_N-bound PA pore particle set, the protocol was modified by pre-incubating LF_N with PA prepore in a 1:2 ratio to ensure a higher population of singly-bound LF_N-PA. A schematic of the

updated protocol is shown in **Figure 3.6A**. As proof of principle for future structure determinations, an initial cryoEM screen of complexes isolated with this new protocol was performed. **Figure 3.6B** shows a representative screening image collected on F30 twin TEM (FEI, Hillsboro, OR, USA) at 39,000 times nominal magnification and a pixel size of 3 Å on the specimen. 2D class averaging with SPARX (side views shown in **Figure 3.7**) showed the majority of the classified populations had single LF_N densities. As with all preparations using the immobilized construction of LF_N-PA pore complexes, the elution volume is easily adjusted to obtain a sufficient concentration of particles on the grid for automated screening with a high-powered microscope with a direct electron detector.

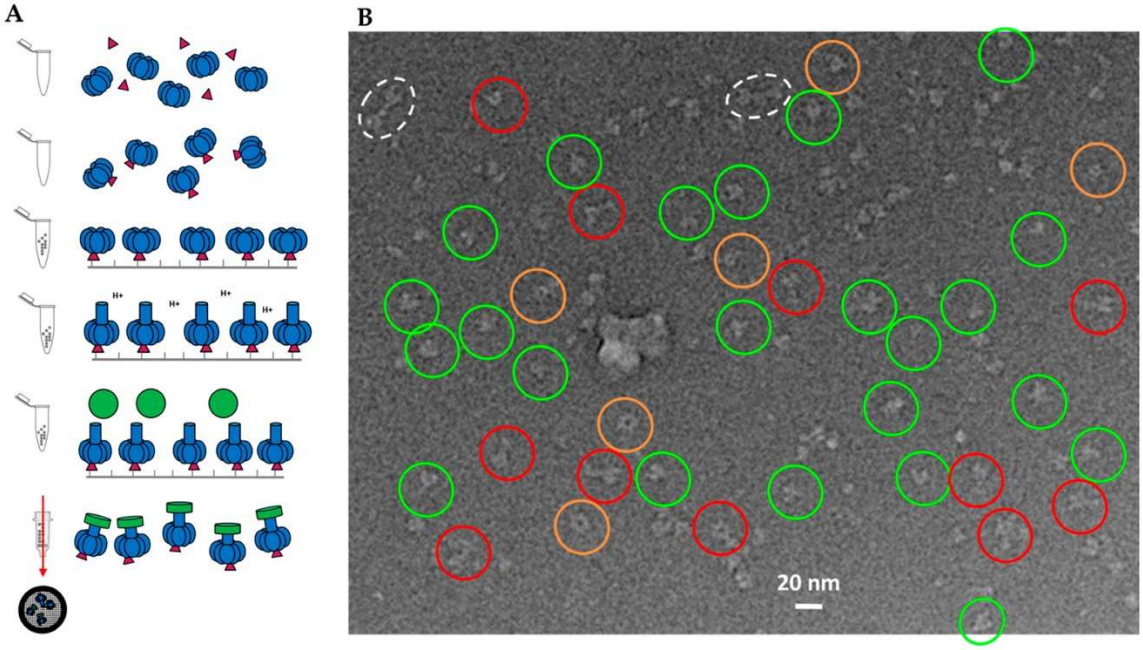


Figure 3.6. Sample preparation of 1LF_N-PA-nanodiscs: (**A**) schematic of LF_N (magenta)-PA (blue)-nanodisc (green) complex formation with LF_N and PA prepore incubated prior to immobilization; and (**B**) representative cryoEM image field of initial screening. Inverted contrast for visualization. Only select clear individual particles are noted (key: red—side views; green—various angle views; dotted white—double PA pores in a single nanodisc; orange—top and bottom views). Extra density from LF_N binding is occasionally observed, particularly in the side view orientations.

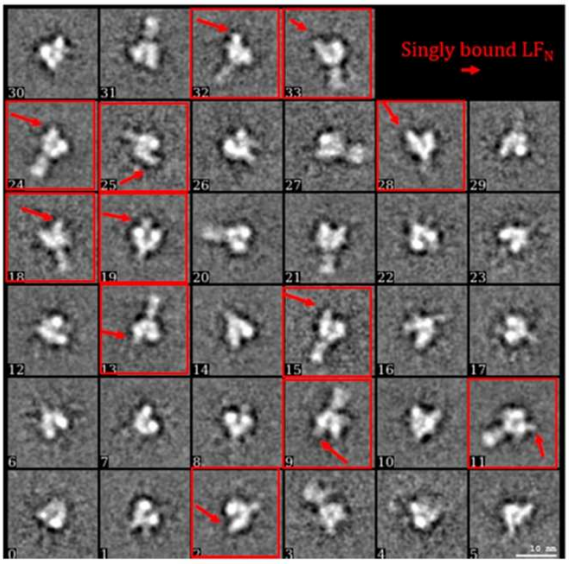


Figure 3.7. 2D classification of approximately 1200 particles using SPARX confirmed singly-bound LF$_N$ with examples of clear LF$_N$ densities highlighted in red.

Flexible fitting of 3LF$_N$-PA Pore Model into the 17 Å CryoEM Density Map

The refined 17 Å cryoEM model of 3LF$_N$-PA-nanodisc generated by cryoSPARC has several interesting asymmetric features (**Figure 3.8**). As mentioned previously, there are three LF$_N$ densities. A flexible fitting of 3LF$_N$-PA pore docked three LF$_N$, in pink, magenta, and purple, in between subunit interfaces of PA, as was seen previously in the prepore crystal structure of 4LF$_N$-8PA and confirmed by 15 Å cryoEM structures using the complete LF-PA prepore structure (Feld, Thoren et al. 2010, Fabre, Santelli et al. 2016). Previous work has shown the N-terminal tail of LF$_N$ feeds into the pore lumen and interacts with the Φ clamp. A cross-section of the model, shown in **Figure 3.9**, reveals the narrowing of the pore lumen is consistent with the positioning of the Φ clamp region in the flexible fitting model. Curiously, this pH 7.5 low-resolution triply-bound LF$_N$-PA pore structure shows an open pore region, in contrast to the closed densities observed for the previous lower-resolution, seven-fold symmetrized structures (Gogol, Akkaladevi et al. 2013).

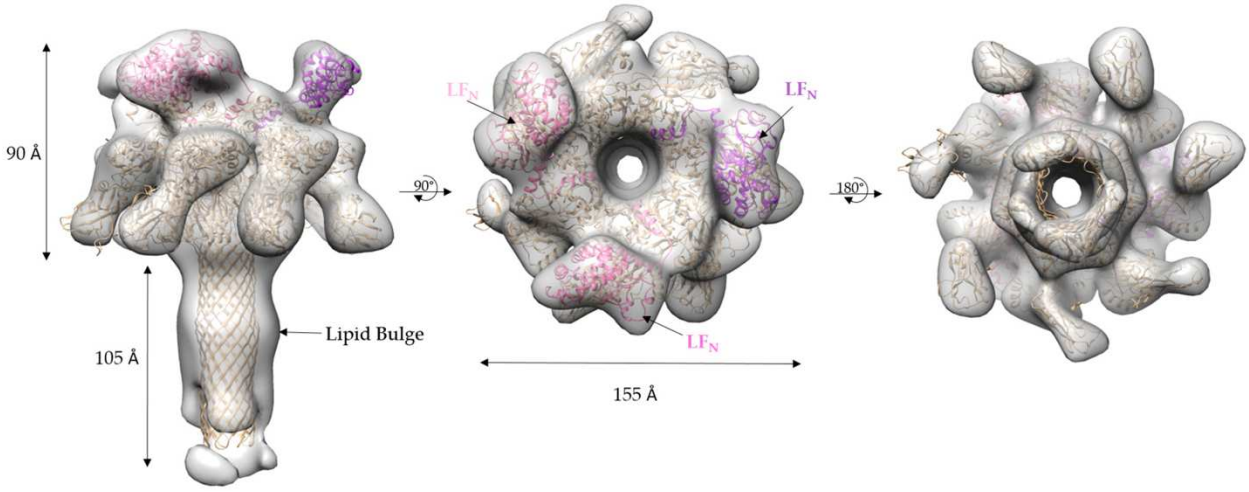


Figure 3.8. 3LF$_N$-PA cryoEM density map (grey) with the ribbon structure flexible fitting -fitted 3LF$_N$ (pink, magenta, and purple)-PA pore (gold): (left) side view; (middle) top view; and (right) bottom view.

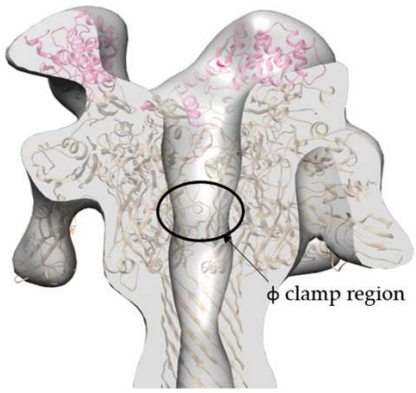


Figure 3.9. Cross-section of the side view cryoEM density map (grey) with ribbon structure flexible fitting model of LF$_N$ (pink and magenta) and PA (gold) reveals that the narrowing of the pore lumen in the density map is consistent with location of Φ clamp region.

A comparison of the flexible fitting atomic structure filtered to 17 Å with the 17 Å cryoEM-derived 3LF$_N$-PA pore structure showed surface details that were visually indistinguishable (**Figure 3.10**). For example, the top view of the cryoEM 3LF$_N$-PA structure showed LF$_N$ has a distinctive bean shape (**Figure 3.10A**). A top view of the space filled PDB structure of LF$_N$ bound to the prepore structure also had this same characteristic shape (Feld, Thoren et al. 2010). A small protrusion from the PA pore cap where LF$_N$ is absent was also present in both models.

Unlike the flexible fitting structure, the domain 4 regions of the cryoEM derived structure are not equal in density, suggesting that these regions are dynamic structures as was previously observed by Jiang *et al*. (Jiang, Pentelute et al. 2015). It is also important to note that not all surface regions in the cryoEM reconstruction are filled by flexible fitting analysis. For example, the β-barrel bulge that is due to lipid binding is not revealed in the fit structure since such a bulge in the highly-stable β-barrel is energetically restrictive.

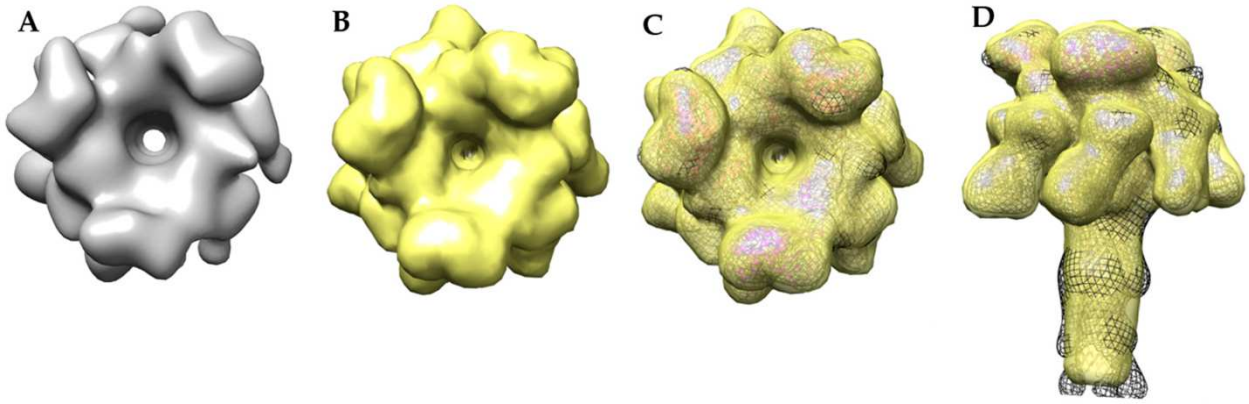


Figure 3.10. Comparison of the cryoEM and flexible fitting models show similar topology of the LF_N bean shape and PA cap protrusion: (**A**) top view of the 17 Å cryoEM map; (**B**) top view of the flexible fitting atomic resolution model filtered to 17 Å; (**C**) top view overlay of the 17 Å cryoEM model (black mesh), the flexible fitting model filtered to 17 Å (yellow), and the flexible fitting ribbon model (PA in gold, LF_N in magenta); (**D**) side view overlay of the cryoEM model (black mesh), the flexible fitting model filtered to 17 Å (yellow), and the flexible fitting ribbon model (PA in gold, LF_N in magenta).

DISCUSSION

Atomic resolution cryoEM is a rapidly evolving structural method that can be applied to examine the atomic consequences of LF_N interactions with the PA pore. The ability to generate soluble, lipid-stabilized LF_N-PA pore structures, even in this low resolution model, is the critical, important first step in demonstrating that we can obtain structural snapshots of this complex.

Sample Preparation of Highly Pure Complexes

One of the main thrusts of this work has been to demonstrate that we can routinely obtain highly-pure engagement complexes (multiply- or singly-bound LF_N) using an immobilization bead-based protocol and nanodisc technology without using columns to purify the final complexes (Akkaladevi, Hinton-Chollet et al. 2013, Akkaladevi, Mukherjee et al. 2015) and minimizing detergent influences on structure (Palazzo, Lopez et al. 2010, Shen, Yang et al. 2016). Even at 17 Å resolution, the variability of the domain 4 densities for the LF_N-PA pore indicates this region is intrinsically flexible (Jiang, Pentelute et al. 2015), ruling out the possibility that this flexibility is due to grid adherence constraints. Although it is possible the insertion of the tip region into an authentic lipid bilayer (*e.g.*, a nanodisc) may result in more ordered structures, better nanodisc resolution is required to make this assessment (Patargias, Bond et al. 2005, Cox and Sansom 2009, Eddy, Su et al. 2015). Previously, protein-bilayer interactions in nanodiscs have been noted to result in extended β-barrel protein structures (approximately two residues per strand) compared with detergent-solubilized structures (Eddy, Su et al. 2015).

Initial CryoEM Model of $3LF_N$-PA Pore

The cryoEM density map structure was created without imposing symmetry or biasing towards an initial input model using the cryoSPARC *ab initio* reconstruction and subsequent refinement procedures. This 17 Å $3LF_N$-PA pore model showed three distinct LF_N densities. In agreement with what was observed previously, the LF_N densities are positioned between two protomer interfaces of the PA pore (Feld, Thoren et al. 2010, Fabre, Santelli et al. 2016). The main contact points are on the crest of the pore and in the α-clamp. Only three LF_N are able to bind to a heptameric pore, leaving one protomer without any direct LF_N contacts.

A cross-section through the EM density map showed the location of the pore opening complete with the narrowing of the pore lumen. A flexible fitting fit starting from the atomic resolution pore structure with LF_N bound positions this narrowing region with the Φ clamp loop region and preserves the opening at the Φ clamp annulus. While the number of particles and subsequent resolution of this current cryoEM density map do not allow us to definitively define structural details of the pore lumen, it would be of interest to determine if the pore remains in a more open configuration at pH 7.5 when one or three LF monomers are bound. This further highlights the need to obtain high-resolution structures of the PA pore with one or more LF bound to determine if the Φ clamp region remains more open under these conditions. As mentioned previously, the presence of interfering electrostatic interactions appears to lead to a more open pore structure. Notably, this open pore diameter has been suggested by Das and Krantz to be necessary in order to accommodate α-helical regions during translocation at pH 5.0. These atomic resolution structures will be key to determining if varying ratios of LF bound (i.e., one vs. three) induces significant structural asymmetry (variable positioning of the Φ clamp) or concerted symmetry (all open) on the PA pore structure.

It is not uncommon to observe both small- and large-scale symmetry breakage of ordered oligomers induced by protein-protein interactions. For example, structures of protein substrate and nucleotide interactions with GroEL, a tetradecameric ring chaperonin protein, show very discernable asymmetric adjustments due to protein substrate interactions (Elad, Clare et al. 2008, Weaver, Jiang et al. 2017), as well as ATP binding and hydrolysis (Saibil and Ranson 2002). A more dramatic demonstration for ligand-induced distortion of symmetry is observed for the ATP bound vs. ADP bound ATPase unfolding machinery of the valosin-containing protein-like ATPase (VAT) recently resolved by cryoEM (Huang, Ripstein et al. 2016). In this instance, the

hexameric structure was dramatically distorted in the presence of ADP and appeared to coincide with its ATP/ADP conformational switching mechanism to provide a conformational platform that unfolds proteins prior to degradation.

It would be of great interest to compare singly bound and multiply bound LF_N-PA pore structures in different pH conditions in order to discern any distinct structural differences that may result from being in various pH environments. Observing these different states of the engagement complex (pH 5.0 vs. pH 7.5, 1 LF_N vs. 3 LF_N) would be useful in determining the position of the Φ clamp loop region and potentially defining unstructured regions of the LF_N that may become structured upon binding to the pore prior to translocation at pH 5.0. There are existing crosslinking studies by the Collier group indicating this interaction is present at pH 5.0 (Janowiak, Jennings-Antipov et al. 2011). Thus, there is precedence for this interaction and those cyro-EM structure collection experiments at pH 5.0 are currently underway. In all cases, given the intrinsic stability of the extended β-barrel at pH 5.0 and pH 7.0, it is highly unlikely that the β-barrel region will be structurally altered when LF_N binds to the PA pore cap region. Rather, the more flexible parts of the PA pore (i.e., the cap region, Φ clamp region, etc.) will be highly susceptible to LF_N-induced conformational changes. How LF structurally impacts translocation and pore formation may be manifested through long range allosteric affects.

Conclusions

Understanding both PA pore formation and LF translocation through the PA pore is crucial to mitigating, and perhaps preventing, anthrax disease. To better understand the interactions between LF_N and the PA pore, the structure of LF_N-bound PA pore was examined using cryoEM. The 17 Å structure of PA pore with 3 LF_N bound was the result of pore immobilization, nanodisc solubilization, *ab initio* modeling, and refinement. In this pH 7.5 structure, the contributions

from the three unstructured LF_N lysine-rich tail regions do not occlude the Φ clamp opening, indicating these flexible tails remain unstructured and unresolved. The next structures to examine are the LF_N-PA pore complexes at pH 5.0 to determine if the unstructured LF N-terminal tails interact with the Φ clamp.

MATERIALS and METHODS

Protein Expression and Purification

Recombinant wild-type (WT) PA was expressed in the periplasm of *Escherichia coli* BL21 (DE3) and purified by anion exchange chromatography (Miller, Elliott et al. 1999) after activation of PA with trypsin (Wigelsworth, Krantz et al. 2004). QuikChange site-directed mutagenesis (Stratagene) was used to introduce mutations into the plasmid (pET SUMO (Invitrogen)) encoding a truncated recombinant portion of lethal factor. LF_N E126C and was expressed as His_6-SUMO-LF_N, which was later cleaved by SUMO (small ubiquitin-related modifier) protease, revealing the native LF_N E126C N-terminus (Wigelsworth, Krantz et al. 2004). Membrane scaffold protein 1D1 (MSP1D1) was expressed from the pMSP1D1 plasmid (AddGene) with an N-terminal His-tag and was purified by immobilized Ni-NTA affinity chromatography as previously described (Ritchie, Grinkova et al. 2009).

Formation of LF_N-PA-Nanodisc Complexes

Heterogeneous LF_N-PA-nanodisc complexes were formed and purified as previously described (Gogol, Akkaladevi et al. 2013, Akkaladevi, Mukherjee et al. 2015). In brief, E126C LF_N was immobilized by coupling E126C LF_N to activated thiol sepharose 4B beads (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA) in Assembly Buffer (50 mM Tris, 50 mM NaCl, pH 7.5) at 4 °C for 12 h. One hundred microliters (100 μL) of 0.2 μM heptameric WT PA prepore was then

added to 50 μL of LF_N bead slurry. Beads were washed three times with Assembly Buffer to remove any unbound PA prepores. The immobilized LF_N-PA prepore complexes were then incubated in 1 M urea (Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C for 5 min to transition the PA prepores to pores. After three more washes with Assembly Buffer, pre-nanodisc micelles (2.5 μM MSP1D1, 162.5 μM 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) (Avanti, Alabaster, AL, USA) in 25 mM Na-cholate (Sigma-Aldrich, St. Louis, MO, USA), 50 mM Tris, 50 mM NaCl) were added and bound to the aggregation-prone hydrophobic transmembrane β-barrel of PA. The micelles were collapsed into nanodiscs by removing Na-cholate using dialysis with Bio-Beads (BIO RAD, Hercules, CA, USA) as previously described (Denisov, Grinkova et al. 2004). Soluble complexes were released from the thiol sepharose beads by reducing the E126C LF_N-bead disulfide bond using 50 mM dithiothreitol (DTT) (Goldbio, St. Louis, MO, USA) in Assembly Buffer. To select for LF_N-PA-nanodisc complexes, the released complexes were then incubated with Ni-NTA resin (Qiagen, Germantown, MD, USA). The His-tag on the MSP1D1 construct bound to the resin. Complexes were eluted from the Ni-NTA using 200 mM imidazole (Sigma-Aldrich, St. Louis, MO, USA) in Assembly Buffer. Assembled complexes were initially confirmed using negative-stain TEM.

Homogeneous $1LF_N$-PA pore complexes were produced using a modified protocol where E126C LF_N and PA were incubated in solution at a ratio of $1LF_N$:2PA prior to immobilization to reduce the number of complexes with multiple bound LF_N. In this particular instance, affinity purification with Ni-NTA resin was omitted to minimize sample loss and homogeneous samples were still obtained.

Other attempted protocols to LF_N-PA_{pore} assembly can be found in **Appendix C**. Briefly, these include nanodisc formation at pH 7.5, 5.5, and 5.0, complex release at pH 7.5, 5.5, and 5.0, pH

gradient introduction with liposomes, and GroEL association to unfolded, translocated LF_N. Nanodisc formation at pH 7.5 followed by LF_N-PA_{pore}-nanodisc complex release at neutral or acidic pH was successful as was PA_{pore}-liposome complex formation, and complexes of GroEL association to unfolded, translocated LF_N. LF_N-PA_{pore}-nanodisc complexes with low pH nanodisc formations and release were not successful for reason we are unsure of though LF_N aggregation is suspected to play a role as nanodiscs are able to form at low pH (**Appendix A**).

CryoEM Sample Preparation and Data Collection

CryoEM samples were prepared within 10–30 min of elution. Three to four microliters (3–4 μL) of purified LF_N-PA-nanodisc complexes were added to a glow-discharged holey carbon grid (Quantifoil R3/4 300 M Cu holey carbon) (Electron Miscroscopy, Sciences, Hatfield, PA, USA) and plunge frozen in liquid ethane using a Vitrobot (FEI, Hilsboro, OR, USA). Data were collected manually over the course of 10 sessions (8–10 h each) on a 2200FSC election microscope (JEOL, Peabody, MA, USA) at NCMI, Baylor College of Medicine. The microscope was equipped with an in-column energy filter (using a 20 eV slit) and operated at 200 kV acceleration voltage. Images were recorded on a Gatan 4k × 4k CCD camera using a 60,000× nominal magnification (1.81 Å/pixel) with an overall range of defocus values from one to three microns using a dose of approximately 20 $e^{-}/Å^2$. Approximately 650 individual micrographs were recorded. Homogeneous $1LF_N$-PA-nanodisc complexes were imaged and screened using a Tecnai F30 G2 twin transmission electron microscope (FEI, Hillsboro, OR, USA) at 200 kV at the University of Missouri Electron Microscopy Core Facility (EMC).

Image Analysis and 3D Reconstruction

The 650 raw micrographs obtained at Baylor were evaluated using EMAN2.1 (Tang, Peng et al. 2007). At the early evaluation stage, around 250 of these micrographs were rejected due to either gross contamination or charging artifacts visible in the Fourier transforms. A total of 30,696 particles were manually boxed out using the e2boxer.py routine of EMAN2.1 with a box size of 224 × 224 pixels. The data evaluated with EMAN2.1 and RELION, showed a heterogeneous population of single, double, and triple LF_N-bound PA. Due to this heterogeneity, it was difficult to use earlier versions of RELION with this smaller dataset to produce a model without imposing C7 symmetry. The approximately 30,000 particles were reevaluated using cryoSPARC (version 0.5). First, 2D class averaging was performed (**Figure 3.5A**). Bad classes were visually identified and discarded (*e.g.*, unrecognizable densities, smaller than predicted density envelopes, etc.). Using the remaining 18,806 good particles, an *ab initio* reconstruction using the cryoSPARC SGD was carried out to computationally purify the dataset into subsets containing one, two, or three bound LF_N. This computation was performed with the following settings: four groups, a group similarity factor of 0.2, and 10-fold the default iterations.

The SGD algorithm allows for *ab initio* structure determination that is insensitive to initial model inputs. An arbitrary computer-generated random initialization model improves over many noisy model iterations. Each step is based on the gradient of the approximated objective function obtained with a random selection of a small batch of initial particles. These approximate gradients do not exactly match the "overall optimization objective" (best *ab initio* model) but through multiple rounds, the derived models gradually approach this maximum. As stated by Punjani, Brubaker, and colleagues, "the success of SGD is commonly explained by the noisy sampling approximation allowing the algorithm to widely explore the space of all 3D maps to

finally arrive near the correct structure" (Brubaker, Punjani et al. 2015, Punjani, Rubinstein et al. 2017). In contrast to using the entire dataset for initial model reconstruction, cryoSPARC samples random subsets of the images during its rapid iteration processes.

The *ab initio* model with three clearly-resolved LF_N densities possessed the largest percentage of particles (44.9%). The second most populated class (20.1%) appeared to contain one prominent LF_N density with the hint of a second bound LF_N, but requires more particles in order to achieve definition (**Figure 3.11B, column 1**). After the *ab initio* model was generated, a homogeneous refinement with 100 additional passes using the branch-to-bound maximum likelihood optimization cryoSPARC algorithm. The final cryoEM map resolution was estimated to be 17 Å based on Fourier Shell Correlation (FSC) with a cut off of 0.5 (**Figure 3.5**). The *ab initio* group 1 with the second highest percentage (20.1%) had one LF_N density at a lower volume threshold. However, further processing of the potential single bound LF_N revealed added density on the PA pore cap from a mixture of one and two LF_N populations **(Figure 3.11C, column 1**). More particles are needed to populate this distribution before definitive single or double LF_N-bound structures can be obtained.

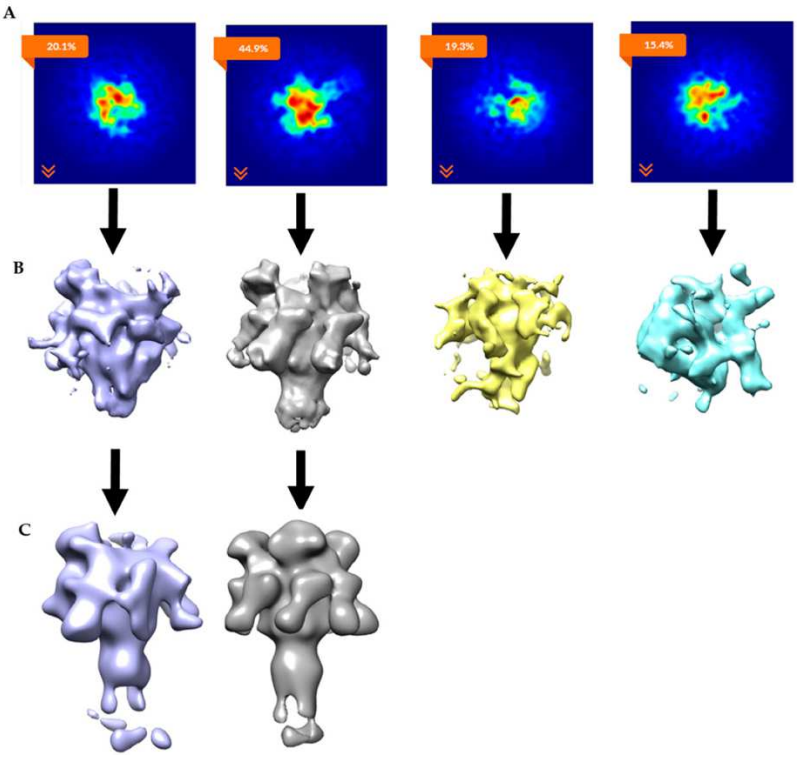


Figure 3.11. CryoSPARC data analyses parsed out heterogeneous LF_N-PA-nanodiscs: (**A**) Image projection of heterogeneous *ab initio* reconstruction with four groups, the largest group, with 44.9% of particles, corresponds to $3LF_N$; (**B**) *ab initio* 3D models (side views); and (**C**) homogeneous refinements of *ab initio* group 1 and group 2. Group 2 refined to 18 Å model of $3LF_N$-PA from 4732 particles. Group 1 clearly shows missing density in the cap region and will need more particles to determine if this structure contains sub-saturated populations (i.e., one or two LF_N bound) of LF_Nbound to the PA pore structure or that this group will split out further to separate one vs. two LF_N-bound populations.

The cryoSPARC 3D reconstruction software tool (Structura Biotechnology, Toronto, Canada) was run on a single workstation (Nova 2 Model: 2 × NVIDIA Titan Xp GPU, Intel Xeon E5-1630v4 (4-core 3.7 GHz CPU), 64 GB DDR4-2400 RAM, Intel 1.2 TB SATA solid state drive for runtime cache, and 4 × 4 TB Seagate SATA HDDs) purchased from Silicon Mechanics (Bothell, WA, USA) housed in the Fisher Laboratory. One of the main advantages of using cryoSPARC in combination with this computer system is the reduced computational time. What was once days or weeks in computational time is now only minutes or hours (Punjani, Rubinstein et al. 2017). For example, as this paper was being written, the latest version of cryoSPARC was

released (upgrade from 0.41 to 0.5). All Baylor collected data was reanalyzed with the newer version as a test for reproducibility in the span of 4 h (from reevaluating 2D classification, removing poor particles, etc.) where the final output *ab initio* models, reevaluated 2D class averages from separated populations and refined structures were reproduced using the single workstation described above. The use of SGD algorithms to generate *ab initio* models are now being β tested or implemented in other software packages.

Flexible fitting of 3LF$_N$-PA

A molecular model was fit into the cryoEM density map using flexible fitting methods (Qi, Lee et al. 2016) which apply an additional potential derived from the density map to the molecules. The starting molecular model was built by rigid docking three LF_N (PDB 3KWV) onto the PA pore cap (PDB 3J9C). The cryoEM density map and initial molecular model were spatially aligned in Sculptor (Birmanns, Rusu et al. 2011, Wahle and Wriggers 2015). The density map was then converted from mrc to a situs file extension for compatibility with the Visual Molecular Dynamics (VMD) software suite. The atomic model and density map files were prepared for flexible fitting in VMD by the typical flexible fitting tutorial progression [47,48]. The model was minimized for 2000 steps simulated for 50 ps at 300 K in vacuum. The grid-scaling factor, which controls the relative strength of the flexible fitting potential was set to 0.3. **Figure 3.10** compares the 17 Å filtered flexible fitting structure with the 17 Å cryoEM derived structure to show distinct similarities in surface topologies (Humphrey, Dalke et al. 1996, Trabuco, Villa et al. 2008, Birmanns, Rusu et al. 2011).

Chapter 4: Caught in the Act: Anthrax Toxin Translocation Complex Reveals insight into the Lethal Factor Unfolding and Refolding mechanism

This chapter is a preprint available online. Machen, A.J., Fisher, M.T. and Freudenthal, B.D., (2020). Caught in the Act: Anthrax Toxin Translocation Complex Reveals insight into the Lethal Factor Unfolding and Refolding mechanism. *bioRxiv*.

ABSTRACT

Translocation is essential to the anthrax toxin mechanism. Protective antigen, the translocon component of this AB toxin, forms an oligomeric pore with three key clamp sites that aid in the efficient entry of lethal factor or edema factor, the enzymatic components of the toxin, into the cell. LF and EF translocate through the protective antigen pore with the pH gradient between the endosome and the cytosol facilitating rapid translocation *in vivo*. Structural details of the translocation process have remained elusive despite their biological importance. To overcome the technical challenges of studying translocation intermediates, we developed a novel method to immobilize, transition, and stabilize anthrax toxin to mimic important physiological steps in the intoxication process. Here, we report a cryoEM snapshot of what we predict is PA_{pore} translocating the N-terminal domain of LF. The resulting 3.1 and 3.2 Å structures of the complex trace density consistent with LF_N as it unfolds near the α clamp, translocates through the Φ clamp, and begins to refold in the charge clamp. In addition, density consistent with an α helix is seen inside the β barrel channel suggests LF secondary structural elements begin to refold at the charge clamp site. We conclude the anthrax toxin uses an extended β barrel to effeciently fold its enzymatic payload prior to channel exit. This hypothesized refolding mechanism has broader implications for pore length of other protein translocating toxins.

INTRODUCTION

The anthrax toxin is not only a deadly *Bacillus anthracis* virulence factor, but also serves as a model system of protein translocation and as a peptide therapeutic delivery platform (Young and Collier 2007, Thoren and Krantz 2011). It's biological importance and biotechnology utility have spurred significant biochemical and biophysical advances in understanding the anthrax intoxication mechanism. In order to gain entry into the cell, this archetypical AB toxin must cross the endosomal membrane. This is accomplished by the B component of anthrax toxin, termed protective antigen (PA). PA forms a translocon pore through which lethal factor (LF) or edema factor (EF), the A component, translocate. Here, we developed an approach to elucidate the structural and mechanistic details of the anthrax toxin during translocation in an effort to understand how LF unfolds in the endosome, translocates through PA, and refolds in the cytosol. An overview of the anthrax toxin mechanism has been reviewed by the Collier lab (Young and Collier 2007) and is briefly summarized here. The first step in intoxication is the 85 kDa monomeric PA binding to host cell receptors. Then the pro-domain of PA is cleaved leaving the 63 kDa PA to oligomerize into heptameric or octameric prepore ($PA_{prepore}$) (Santelli, Bankston et al. 2004, Kintzer, Thoren et al. 2009). Up to three LF and/or EF components can bind to the $PA_{prepore}$ heptamer (Mogridge, Cunningham et al. 2002, Kintzer, Thoren et al. 2009, Antoni, Quentin et al. 2020). The AB toxin complex (**Figure 4.1Aa**) is endocytosed through clatherin mediated endocytosis (**Figure 4.1Ab**) (Abrami, Liu et al. 2003). As the endosome acidifies, $PA_{prepore}$ undergoes a conformational change to a pore (PA_{pore}) (**Figure 4.1Ac-d**) (Miller, Elliott et al. 1999). This pore inserts into the endosomal membrane to form a channel. The low pH of the endosome and the pH gradient between the endosome and the cytosol facilitate LF or EF to unfold and rapidly translocate into the cytosol in a hypothesized Brownian ratchet mechanism

(**Figure 4.1Ae**) (Krantz, Finkelstein et al. 2006). Natively refolded LF and EF in the cytosol are then able to perform their virulent enzymatic functions (**Figure 4.1Af**) (Duesbery, Webb et al. 1998).

The overall structure of the PA_{pore} translocon can be divided into two regions: the funnel and the channel (**Figure 4.1B**). The first region, the funnel, facilitates binding and unfolding of LF. LF binds to the rim of the PA_{pore} funnel and is guided down the narrowing structure. The second region of PA_{pore} is the channel, a β barrel that extends from the funnel and spans the endosomal membrane. Three nonspecific PA_{pore} clamp sites aid in the translocation of LF (**Figure 4.1B**). The α clamp is located at the PA funnel rim, is formed by adjacent PA protomers, and binds helical portions of LF to position them towards the pore lumen. A crystal structure of the N-terminal domain of LF (LF_N) bound to the $PA_{prepore}$ revealed the α clamp binding site, but was unable to resolve the 28 amino acids of LF_N passed the α clamp (Feld, Thoren et al. 2010). The second clamp site is the Φ clamp, a ring of seven phenylalanine residues that maintain the pH gradient between the endosome and the cytosol (Krantz, Melnyk et al. 2005). The 2.9 Å structure of apo PA_{pore} revealed the Φ clamp forms a narrow 6 Å diameter ring (Jiang, Pentelute et al. 2015). Secondary structural elements, such as α helices, are too wide to fit through this narrow seal. Therefore, it is hypothesized that peptide substrates must completely unfold and refold in order to translocate through the PA_{pore} and enter the cytosol of the cell (Jiang, Pentelute et al. 2015). The Φ clamp also assists in the unfolding of LF as an unfolding chaperone (Thoren and Krantz 2011). The third clamp site, the charge clamp, is located within the β barrel of PA_{pore}. The charge clamp deprotonates acidic side chains of LF and ensures unidirectional movement of the polypeptide (Pentelute, Sharma et al. 2011, Brown, Thoren et al. 2015). Interestingly, the diameter of the PA_{pore} β barrel is large enough to accomidate an α helix, which would allow for

initial refolding to occur inside the pore prior to LF entering the cytosol. However, it remains unclear what structural state LF is in when interacting with the charge clamp and within the β barrel channel.

One of the many challenges in studying the anthrax toxin is that it is a dynamic membrane protein that functions under acidic conditions. Thus many questions remain, such as what path LF travels down the endosomal pore lumen from the α clamp to the Φ clamp, whether the Φ clamp adopts multiple states during translocation, and whether LF forms α helices inside the β barrel pore. To address these questions, we developed a novel toxin immobilization, translocation, and nanodisc stabilization (TITaNS) method in combination with cryoEM in an effort to characterize PA_{pore} translocating the N-terminal domain of LF (LF_N). This approach provides unique mechanistic insight into how LF_N interacts with the three clamp sites of PA_{pore}. We observed density consistent with LF_N unfolding prior to the α clamp, translocating through the dynamic Φ clamp, and beginning to refold in the charge clamp of the PA_{pore}.

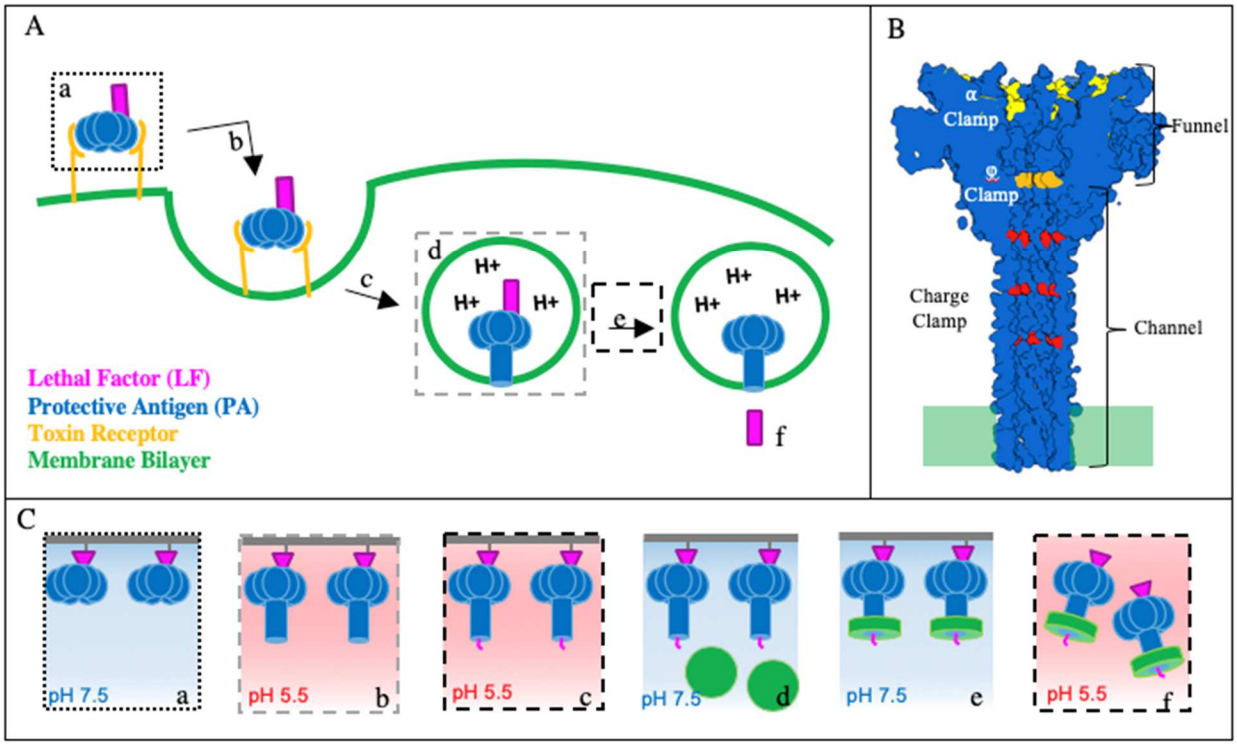


Figure 4.1. Anthrax toxin mechanism, structure, and complex assembly. (A) Important physiolocal steps in the anthrax intoxication mechanism (a) starting with assembled complex of LF-$PA_{prepore}$ on host cell surface. (b) Complex is endocytosed and (c) trafficked to the late endosome (d) where the acidic environment causes a PA conformational change from prepore to pore. (e) LF is translocated through PA_{pore} and (f) refolds in the cytosol. **(B)** PA_{pore} side view slice highlighting important pore lumen features with funnel shape from α clamp (yellow) to Φ clamp (orange) and charge clamp (red) inside pore β barrel. **(C)** Anthrax toxin immobilization, translocation, and nanodisc stabilization (TITaNS) method beginning with (a) immobilization of LF_N-$PA_{prepore}$ complex on thiol sepharose beads (grey surface) to prevent aggregation followed by (b) $PA_{prepore}$ transitioned to PA_{pore} under acidic conditions. (c) Predicted translocation complex of LF_N -PA_{pore} at low pH. (d) Addition of pre-nanodisc micelle (green) at pH 7.5 to solubilize complex. (e) Detergent dialyzed to form lipid bilayer nanodisc. (f) LF_N-PA_{pore}-Nanodisc complexes at pH 5.5 on cryoEM grid. Physiological states mimicked by TITaNS highlighted in dot and dashed boxes.

RESULTS

Assembly of anthrax translocation complexes

Previous approaches to assemble PA_{pore} complexes have generally used urea to transition from $PA_{prepore}$ to PA_{pore} to avoid aggregation (Katayama, Wang et al. 2010, Akkaladevi, Hinton-Chollet et al. 2013, Gogol, Akkaladevi et al. 2013, Akkaladevi, Mukherjee et al. 2015, Machen, Akkaladevi et al. 2017, Hardenbrook, Liu et al. 2020). These approaches have limitations in that they do not account for the low pH electrostatic microenvironment in the pore lumen predicted to be important for LF-PA interactions (Ma, Cardenas et al. 2017) and they assume similar outcomes for chaotrope and acid induced unfolding. In order to overcome these limitations we have developed a novel assembly method for toxin immobilization, translocation, and nanodisc stabilization (called TITaNS, **Figure 4.1C**). TITaNS was designed to mimic important low pH physiological states during the anthrax intoxication mechanisms. This approach allows for endosomal pH pore formation and imaging of individual complexes in a lipid bilayer in the biologically relevant low pH environment (Machen, Akkaladevi et al. 2017, Fisher and Naik 2019). TITaNS can be used in combination with techniques other than cryoEM, including mass spectrometry, nuclear magnetic resonance, surface plasmon resonance, and biolayer interferometry. TITaNS also has the potential to be adapted to screen potential pharmaceuticals that arrest or prevent endosomal membrane insertion (Fisher and Naik 2019). Reversible immobilization is key to the TITaNS methodology, because it allows the stabilized complexes to be released from the beads. We began with recombinantly purified, soluble forms of LF_N and $PA_{prepore}$. The toxin components were incubated in solution. The binary complex of LF_N bound to $PA_{prepore}$ was then immobilized onto thiol sepharose beads by covalently coupling E126C LF_N to the bead surface (**Figure 4.1Ca**). The LF_N-$PA_{prepore}$ complex was immobilized on the beads and

oriented such that when the bead slurry was washed in pH 5.5 buffer to transition PA from prepore to pore, the pore extended away from the bead surface (**Figure 4.1Cb**). We predict this low pH environment initiates translocation of LF_N through PA_{pore} *in vitro* (**Figure 4.1Cc**). We base this predication on computational and experimental evidence of early translocation events induced by low pH. Specifically, molecular simulations of anthrax early translation events predict the events are strongly influenced by the protonation state of LF and are highly favorable at low pH (Ma, Cardenas et al. 2017). Low pH induced PA_{pore} channel blockage and partial translocation has also been observed in planar lipid bilayers (Krantz, Finkelstein et al. 2006). After pore formation and predicted translocation, the next step in TITaNS was solublization of LF_N-PA_{pore} translocation complexes using nanodisc technology (Denisov, Grinkova et al. 2004, Ritchie, Grinkova et al. 2009, Bayburt and Sligar 2010). Pre-nanodisc micelles were added to the bead slurry and associated with the transmembrane portion of PA_{pore} (**Figure 4.1Cd**). To promote lipid bilayer formation we dialyzed away excess detergent (**Figure 4.1Ce**). The soluble LF_N-PA_{pore}-nanodisc complexes were then eluted off the thiol sepharose beads using a reducing agent. Eluted complexes at pH 7.5 were added to the cryoEM grid. Then the pH of solution on the grid was dropped to pH 5.5 to capture the native complex at low pH prior to blotting and plunge freezing (**Figure 4.1Cf**). Acidification of complexes was time sensitive as unfolded, potentially partially translocated LF_N would rapidly aggregate in solution at low pH and nanodiscs tended to migrate to the air-water interface given enough time (**Appendix D**). Therefore, we waited to acidify complexes until they were on the cryoEM grid and plunge freeze grids within 30 seconds of sample application.

Using our TITaNS methodology, we were able to reconstruct a 3.1 Å and a 3.2 Å cryoEM map of what we predict is LF_N translocating through PA_{pore} (**Figure 4.2A)**. We based this prediction

on a loss of LF_N density in the canonical binding site above the PA_{pore} and an addition of density inside the pore lumen. We traced density predicted to be unfolded LF_N peptide in the PA_{pore} translocon using local refinement, volume subtraction, and 3D variability analysis. These densities were most notable near the three clamp sites within PA_{pore}, suggesting the α, Φ, and charge clamps stabilize translocation intermediates. Using the cryoEM densities inside the pore, molecular models of LF_N translocating through the PA_{pore} were built (**Figure 4.2B**). These models are our interpretation of the cryoEM density maps.

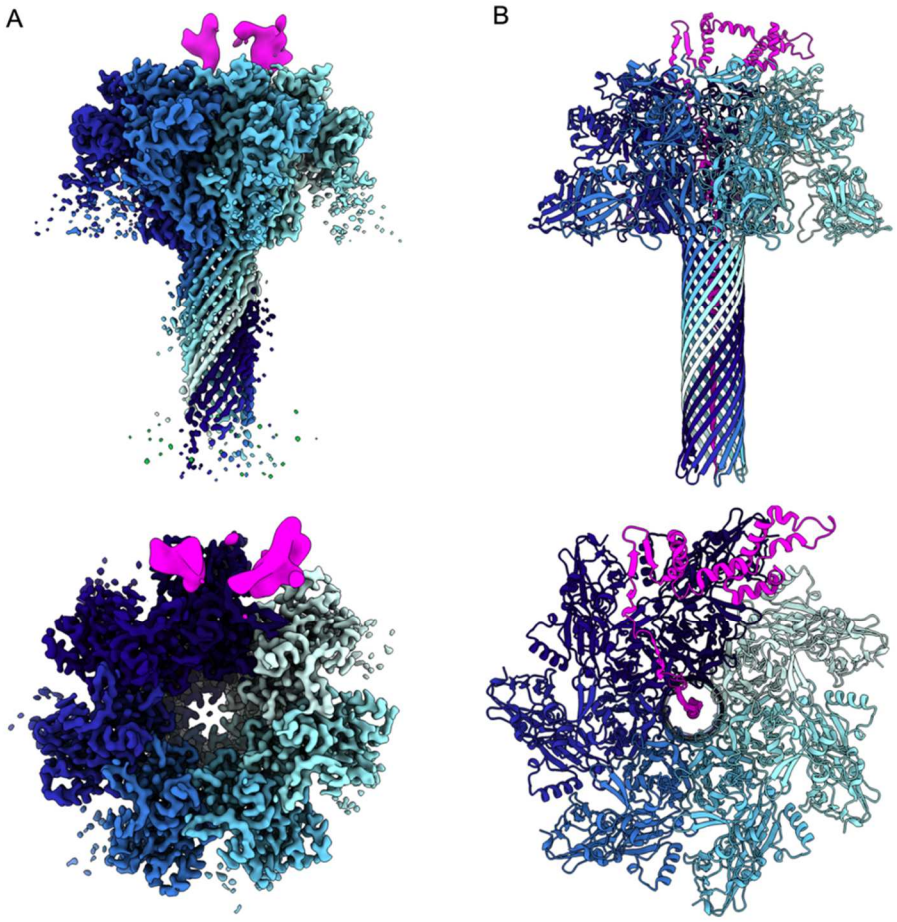


Figure 4.2. Overview of cryoEM anthrax toxin complex with the N-terminal domain of lethal factor (magenta) translocating through the protective antigen pore (blue gradient). (A) 3.06 Å cryoEM density map of LF_N and PA_{pore} with side view (top) and top down view (bottom). (B) Molecular model of PA_{pore} with side view (top) and top down view (bottom).

Unfolding intermediates of LF_N during translocation

Prior to translocation, LF_N is bound to the cap of PA at the interface of two PA protomers with helix **α**1 bound to the α clamp (Feld, Thoren et al. 2010, Hardenbrook, Liu et al. 2020). In our complex, we observe density consistent with LF_N at the interface of the two PA protomers and at the α clamp. In order to explore potential heterogeneity of the data, we used symmetry expansion in combination with 3D variability analysis. The resulting vectors revealed a second density that we predict is a second unfolded LF_N intermediate bound to the rim of PA_{pore} (**Figure 4.3**). The trajectory of these densities suggests steps in LF_N unfolding during translocation. Prior to translocation, the density for LF_N is a bean shaped structure with 2 lobes (Feld, Thoren et al. 2010, Machen, Akkaladevi et al. 2017). The first unfolded state shows density consistent with these two lobes separating. The second unfolded state shows density that suggests Lobe 1 translocating down the α clamp and Lobe 2 shifting counter clockwise towards the α clamp, primed for translocation. Interestingly, this counterclockwise shift would align full length LF, which extends from Lobe 2, directly above the α clamp for efficient unfolding and translocation (**Figure 4.3**). Together, these predicted intermediates of LF_N unfolding indicate LF travels through the α clamp during translocation.

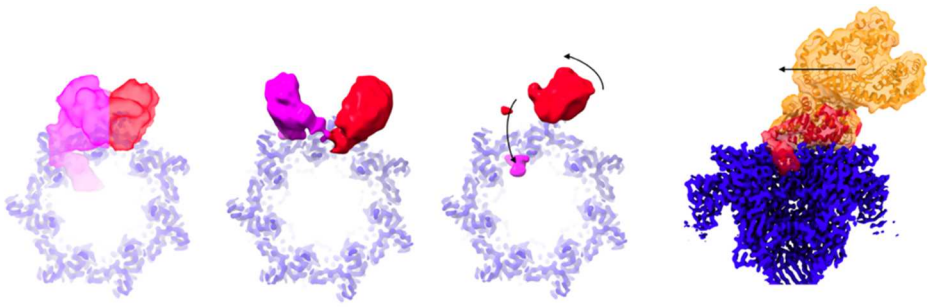


Figure 4.3. LF_N unfolding intermediates prior to translocation through PA_{pore} (A) Top down view of LF_N bound to PA_{pore} prior to translocation. Bean shaped LF_N density (pdb 3KWV) with lobe 1 shown in magenta and lobe 2 shown in red. (B) Hypothesized translocating LF_N density 1 from 3D variability analysis suggests lobe 1 and lobe 2 separating under acidic conditions. (C) Hypothesized LF_N density 2 interpreted as Lobe 1 unfolding and translocating through PA_{pore} while Lobe 2 shifts counterclockwise towards the α clamp. (D) Side view slice of full length LF (pdb 1J7N, orange) and LF_N (3KWV, red) docked on PA_{pore} (blue) with arrow indicating shift of LF towards α clamp consistent with unfolded intermediates

Local refinement of the densities above the pore showed density consistent with the two unfolded states of LF_N in greater detail. To gain insight into the unfolding of LF_N during translocation, we modeled secondary structural elements in the tubular densities to allowed us to visualize potential unfolding of LF_N (**Figure 4.4A**). In our theoretical model, immobilized residue E126C has not yet translocated suggesting translocation may have occurred while the residue was coupled to the bead surface (**Figure 4.4B-C**). Previous comparisons of folded LF_N bound to $PA_{prepore}$ vs PA_{pore} noted LF_N moves up and away from the PA binding interface (Hardenbrook, Liu et al. 2020). Our density that we interpret as unfolding LF_N suggest this motion trajectory continues and is important for unfolding prior to translocation. These models could potentially be of use for molecular simulations of LF unfolding during translocation.

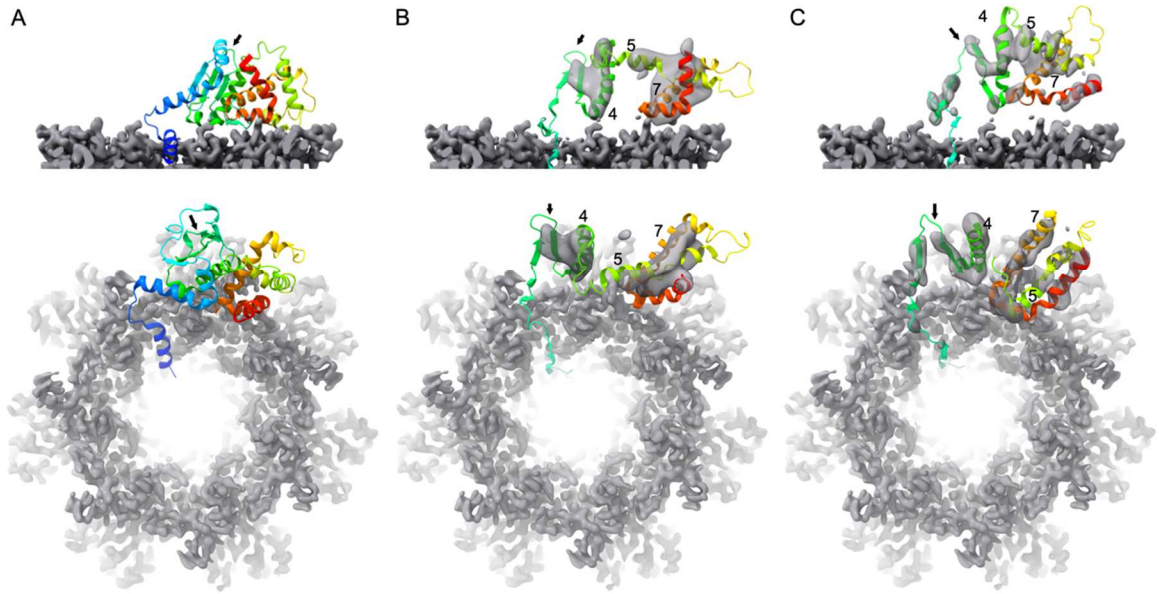


Figure 4.4. Predicted unfolding intermediates of translocating LF_N. (A) Folded LF_N (6PSN) (rainbow model) docked on PA_{pore} (grey map) for comparison. (B) Predicted unfolded LF_N 1 shows separation of Lobe 1 and Lobe 2 via helix 4 (green) and 7 (orange) with helix 5 (green-yellow) stretched between the 2 lobes. (C) Unfolded intermediate 2 showing further unfolding, vertical separation from PA, and β hairpin separation. Black arrow indicates immobilized residue E126C.

Tracing translocating LF_N through the clamp sites of the PA_{pore}

To trace LF_N from the α clamp to the Φ clamp, we subtracted PA_{pore} model density from our unsharpened map to reveal electron density in the pore lumen not accountable for the PA_{pore}. Density consistent with translocating LF_N started in the α clamp and traveled down the pore lumen, interacting with β sheet 39 of PA_{pore}. Density near the top of the PA_{pore} cap was in proximity to several hydrophobic residues of PA_{pore} (**Figure 4.5A**). Some of these residues, including Phe202, Phe236 and Phe464 were in the α clamp and have previously been proposed to aid in unfolding LF and stabilizing unfolded intermediates (Feld, Thoren et al. 2010). Additional hydrophobic residues further down the pore lumen, such as Trp226, Tyr456, and Tyr375 were also near the predicted LF_N density and may help to stabilize the unfolded peptide.

The density of predicted unfolded peptide is also in proximity to acidic residues Glu465, Asp426, and Glu393 closer to the Φ clamp. Previously, molecular modeling using milestoning noted these residues are important for early translocation events (Ma, Cardenas et al. 2017). Our results are consistent with these residues facilitating translocation of LF_N into the pore in an unfolded state.

Added asymmetric density in and around the Φ clamp (**Figure 4.5C**) was seen in the final refined map without further processing (*e.g.*, local refinement, map subtraction, sharpening). Compared to the previously published apo PA_{pore} cryoEM structure (Jiang, Pentelute et al. 2015), the Φ clamp region in our translocating structure has added density. Specifically, density for each of the Phe427 was smeared in plane with the benzyl ring suggesting rotameric states moving up and down (**Figure 4.5D)**. There is also density in the center of the Φ clamp (**Figure 4.5C**). We attribute this density to the unfolded LF_N interacting with PA_{pore} Φ clamp loop as LF_N is translocating through the pore.

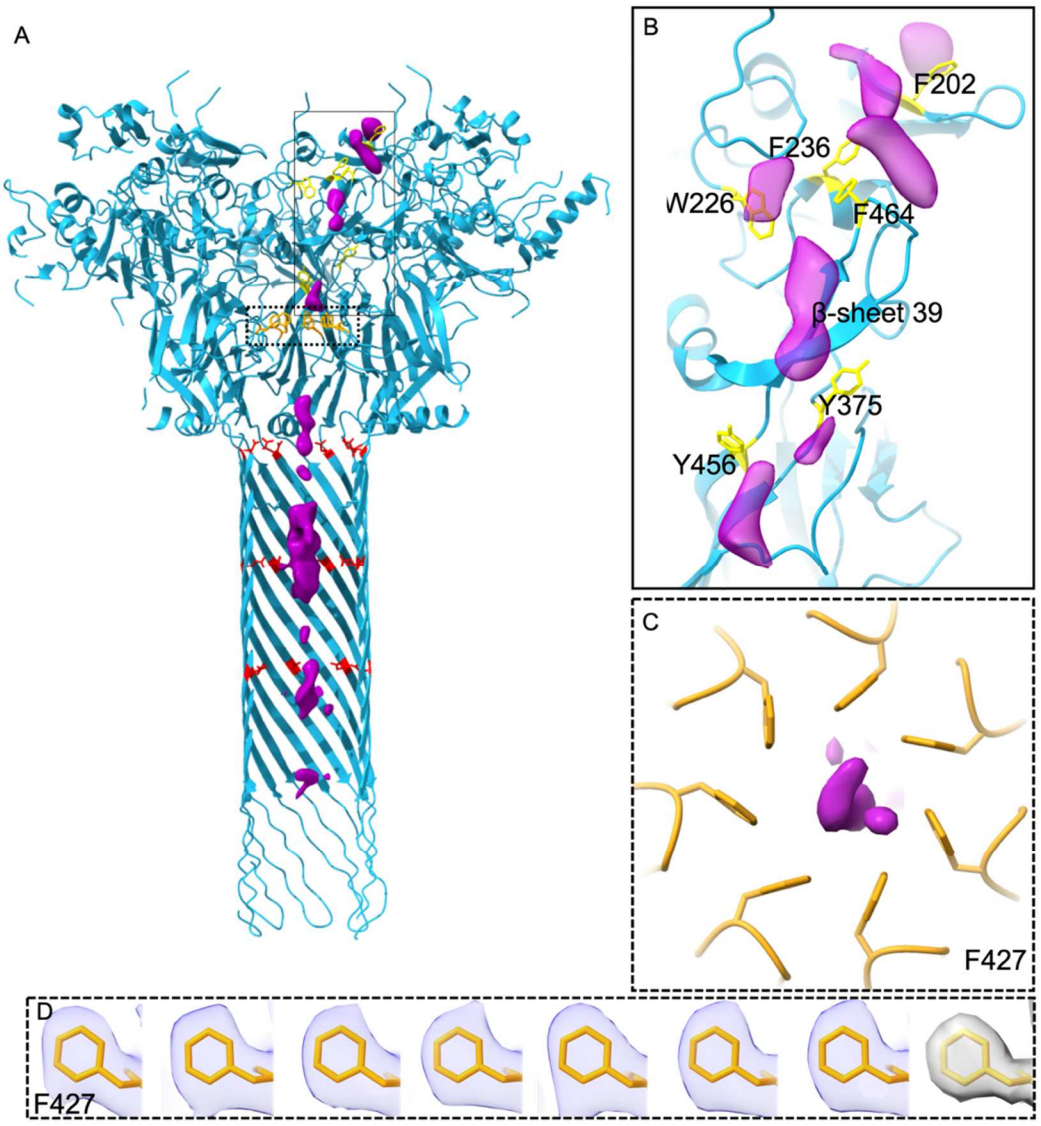


Figure 4.5. Translocating LF_N visible at key clamp sites. (A) Overview of the anthrax pore (cyan) with density of tranlsocaitng LF_N (magenta). (B) close up of key hydrophobic residues (yellow) in early translocation from α clamp to φ clamp. (C) Top view of Φ clamp (orange) with LF_N density passing through (D) Individual F427 residues (orange) with associated cryoEM density for each subunit compared to modelled density (grey, 3 Å).

Density was also observed in the β barrel of the PA_{pore}. Focused refinement of the β barrel interior revealed density consistent with the density of an α helix along with a portion of unfolded peptide above and below the helix (**Figure 4.6)**. Notably, the density consistent with an α helix starts in the PA_{pore} charge clamp suggesting the deprotonated state favors helix formation

(Jas, Childs et al. 2019). Canonical charge clamp residues Asp276, Glu434, and Glu335 are shown in **Figure 4.6** with each of the seven PA_{pore} subunits contributing one residue and the predicted LF_N density translocating through the center of the channel. This density consistent with an α helix in the charge clamp provides evidence for initial refolding of LF secondary structure inside the PA_{pore}**.** The hypothesized newly formed helix likley stabilizes unfolded intermediates of LF as well as contributes to native folding of the enzyme upon exit of the PA_{pore}.

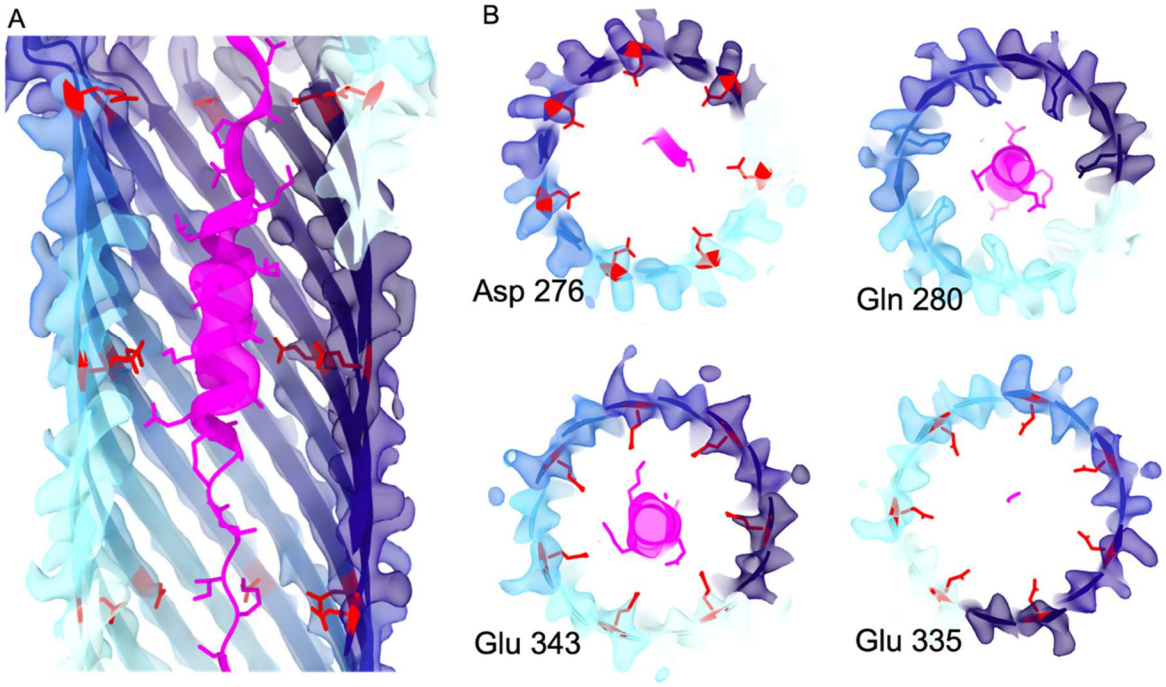


Figure 4.6. Focused refinement of β barrel interior reveals refolding of LF_N (A) Side slice of PA_{pore} (blue) charge clamp (red) with LF_N density (pink) (B) Top view slices of the PA_{pore} charge clamp at key acidic residues.

DISCUSSION

The anthrax toxin PA_{pore} is a nano trojan horse that unfolds, translocates, and refolds it's enzymatic subtrate. Three clamp sites aid in peptide translocation: the α clamp, the Φ clamp, and

the charge clamp. We report here, for the first time, a cryoEM density consistent with nascent polypeptide chain translocating the length of PA_{pore}. In our model, LF_N can be seen unfolding prior to the α clamp, passing through the dynamic φ clamp, and refolding in the charge clamp. In order to translocate, LF must completely unfold. However, if the entire 90 kDa enzyme were to unfold at once, deleterious folded intermediates or aggregates would likely block the PA_{pore} translocon, especially when multiple LF are bound to PA. Therefore, in order to efficiently translocate and refold, LF unfolds from the N to C terminus (Zhang, Finkelstein et al. 2004). While the low pH of the endosome destabilizes the enzyme, it does not completely unfold into its primary sequence (Gupta, Singh et al. 2001, Krantz, Trivedi et al. 2004). We were able to capture two densities we interpret as unfolded intermediate structures of LF_N with density for secondary strucural elements. Our results suggest that molten globular translocation intermediates of LF rotate towards the α clamp. The α clamp would then able to apply additional unfolding force on the protein (Thoren and Krantz 2011) and funnel LF towards the Φ clamp. Translocation requires unfolding and stabilization of the unfolded intermediates. When LF binds to PA_{pore}, the first helix of LF moves away from the main body of LF and binds to the α clamp of PA_{pore} (Feld, Thoren et al. 2010). From here, LF has multiple paths it could take through the PA_{pore} funnel, gated by the Φ clamp. Our results suggest a favorable path from the α to Φ clamp involves a series of hydrophobic residues that are amenable to unfolded translocation intermediates and likely serve as checkpoints to verify the unfolded state of LF prior to the Φ clamp.

The Φ clamp plays a crucial roll in translocation by acting as a hydrophobic seal between the endosome and cytosol (Krantz, Finkelstein et al. 2006). Multiple Φ clamp states have been hypothesized at pH 5.5 (Das and Krantz 2016). While our analysis did not reveal multiple

distinct states, the smeared density is indicative of a dynamic clamp. The density does not imply dilation of the clamp (Das and Krantz 2016), so much as a up and down motion along the pore axis. This motion could be conserted or individual F427 may move to accommodate various side chains. The compressive and tensile forces generated by the unfolding LF in the PA_{pore} funnel above and hypthesized refolding LF in the β barrel channel below may also contribute to this movement. Too much flexibility or dialation would cause the seal to be lost. However, this dynamic motion maintains the pH gradient between the endosome and cytosol, while accomidating any side chain, ensuring efficient translocation.

Helix formation inside the PA_{pore} β barrel has been hypothesized but, to our knowledge, never observed (Krantz, Trivedi et al. 2004, Jiang, Pentelute et al. 2015). We report here, preliminary evidence of α helix formation and translocation inside the β barrel of the PA_{pore} (**Figure 4.6)**. We hypothesized that, along with changing the charge state of the peptide substrate, the charge clamp allows for a local folding environment within the PA_{pore}. Helical portions of LF have previously been shown to dock into the α clamp,with the periodicity of these helices aiding in efficient unfolding of LF (Das and Krantz 2017). We predict this periodicity is also important for hypothesized refolding of LF, beginning in the charge clamp. Our hypothesis is consistent with other anthrax toxin subtrates, such as LF_N fused to the catalytic chain of diphtheria toxin (LF_N-DTA), that do not have the same helical periodicity and did not evolve to fold in the PA_{pore} channel. Interestingly, these non-native substrates require chaperones for enzymatic activity (Dmochewitz, Lillich et al. 2011) indicating the DTA portion of these proteins do not form helices in the PA channel at optimal intervals. Thus LF_N-DTA is less likely to emerge as folded intermediates therefore requiring chaperones to fold or refold properly. This model is reminesant of the ribosome, where helix folding in the exit tunnel aids in co-translational folding of native

proteins (Wilson and Beckmann 2011). We predict helix folding in the PA_{pore} β barrel aids in co-translocational folding by temporally altering LF emersion from the tunnel allowing regions to fold into tertiary structures.

A proposed unfolding-refolding translocation model for the anthrax toxin is shown in **Figure 4.7** starting with LF_N bound to the funnel rim of $PA_{pore.}$ LF_N lobes 1 and 2 separate and rotate counter clockwise towards the α clamp as LF_N is unfolded and funnelled towards the Φ clamp, aided by hydrophobic residues along the funnel slope. LF_N acidic residues are protonated in the acidic environment of the funnel. Completely unfolded LF_N then passes the Φ clamp. This ring of F427 remains restrictive enough to maintain a seal while cushioning the mechanical forces of translocation and chemical interactions of side chains. As the channel widens in the charge clamp, acidic residues are deprotonated. Folding of α helical portions would place mechanical force on the translocating peptide, contributing to efficient translocation, and possibly overcoming local energy minimum that would otherwise stall the complex. The newly formed, secondary structure favors unidirectional translocation by discouraging retrograde transfer through the narrow Φ clamp.

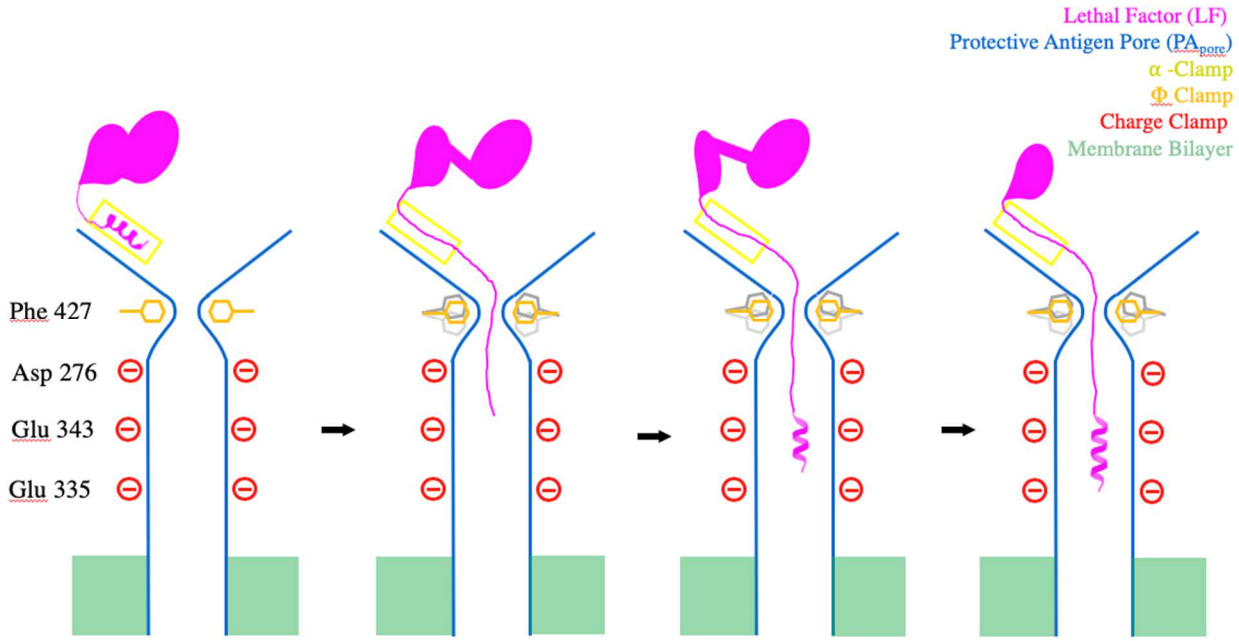


Figure 4.7. Proposed anthrax toxin unfolding-refolding translocation mechanism. Unfolding-refolding translocation model of anthrax toxin. LF_N (magenta) bound to PA_{pore} (blue) with helix bound to α clamp (yellow). Under acidic conditions of the endosome, LF_N lobes separate and shift toward α clamp during translocation. Unfolded peptide passes the dynamic Φ clamp (orange and grey) and acidic residues are deprotonated by three PA charge clamp residues (red) in the channel. Helical portions of LF_N begin to refold in the channel, disfavoring retrograde transfer, overcoming local energy barriers, and ensuring proper teritiary folding. Membrane bilayer represented in green.

In order to capture anthrax toxin for cryoEM, we developed a novel toxin immobilization, translocation and nanodisc stabilization (TITaNS) method. This method can be adapted to attain intermediates of a wide range of host cell interactions under endosomal conditions. Currently, the method involves immobilization on a bead based substrate with bulk solution washes in an eppendorf tube which results in an ensemble of translocation complexes. To obtain more descrete complexes, the method is being expanded to include immoblization on a column, allowing for low pH pulse chase. E126C was selected as the immobilized residue for its location with respect to the PA binding site. We also predict it may stall the complex early in translocation. Moving the immobilization residue further from the N-terminus has the potential to capture mid and late translocation complexes. In the future, TITaNS could be adapted to time-lapse cryoEM which has previously been used to characterize ribosome processivity (Frank 2017).

Aromatic gates are found in other toxin and translocation systems. PA Phe427 is equivalent to Phe454 of *Clostridium perfringens* iota toxin (Knapp, Maier et al. 2015); Phe428 of *Clostridium botulinum* C2II binary toxin (Neumeyer, Schiffler et al. 2008); and Trp318 of *Vibrio cholerae* cytolysin (De and Olson 2011). It is not just bacterial toxins that use aromatic gates. Other transport systems with rings of hydrophobic residues in the pore lumen include mammalian cell entry (MCE) complexes (Ekiert, Bhabha et al. 2017) and proteasome-activating nucleotidase (PAN) (Zhang, Hu et al. 2009). We hypothesize a dynamic hydrophobic seal model is a common translocon mechanism, applicable to these other systems. TITaNS could be adapted to these systems to investigate other hydrophobic seals.

Our results suggest the anthrax toxin has an extended β barrel channel to effeciently fold its enzymatic payload. The longer pore is also necessary to interact with a receptor binding site

further removed from the membrane surface. Initial refolding in the pore channel is likely not unique to the anthrax toxin. Indeed, other translocon shown in **Figure 4.8**, such as the iota toxin and TcA, have extended pores (Piper, Brillault et al. 2019, Yamada, Yoshida et al. 2020). Not all pore forming toxins translocate proteins. Some, like *Vibrio cholerae* cytolysin and *Staphylococcus aureus* α hemolysin, form pores to distrupt ion concentrations (De and Olson 2011, Sugawara, Yamashita et al. 2015). The pore length of these toxins is noticeably shorter (**Figure 4.8**). We predict translocon pores have evolved extended pores to faciliate substrate refolding inside the translocon for effective intoxication.

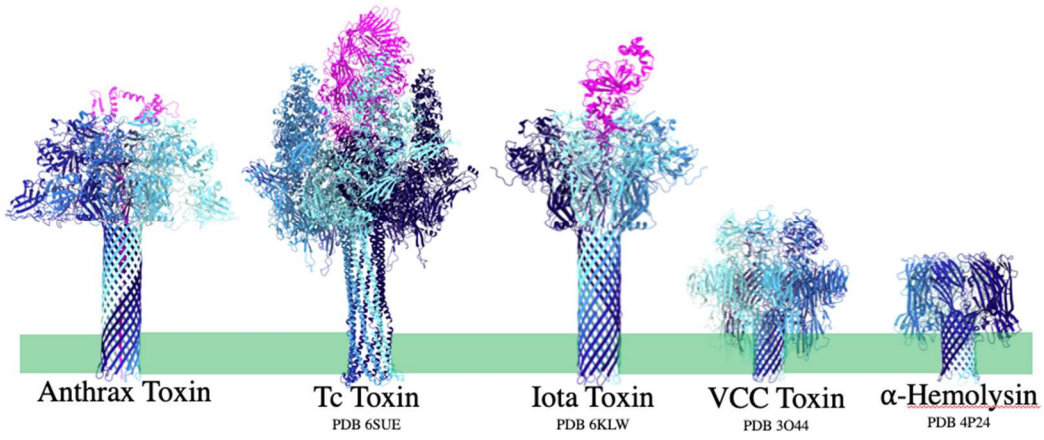


Figure 4.8. Comparison of toxin pore length. Comparison of toxin pore length between toxins that translocate proteins (anthrax, Tc, iota) vs toxins that disrupt ion gradients (VCC and α-Hemolysin). Membrane bilayer represented in green. Note only the N-terminal domain of anthrax toxin lethal factor is shown.

MATERIALS and METHODS

Protein expression and purification

Proteins were purified as previously described (Machen, Akkaladevi et al. 2017). Briefly, His_6-SUMO-LF_N E126C was expressed in BL21 cells, purified using anion exchange, and cleaved by small ubiquitin-related modifier protease (Wigelsworth, Krantz et al. 2004). Recombinant wild-type PA_{83} was expressed in the periplasm of *Escherichia coli* BL21 (DE3) and purified by

ammonium precipitation and anion exchange chromatography (Miller, Elliott et al. 1999). After trypsin activation (Wigelsworth, Krantz et al. 2004), PA_{63} heptameric prepores were formed using anion exchange and size exclusion chromatography. Membrane scaffold protein 1D1 (MSP1D1) was expressed from the pMSP1D1 plasmid (AddGene) with an N-terminal His-tag and was purified by affinity chromatography(Ritchie, Grinkova et al. 2009).

LF_N-PA-Nanodisc complex formation for cryoEM with TITaNS

E126C LF_N and PA were incubated in solution at a ratio of $4LF_N$:1PA. Complexes were then immobilized by coupling E126C LF_N to activated thiol sepharose 4B beads (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA) in Assembly Buffer (50 mM Tris, 50 mM NaCl, 10mM $CaCl_2$ pH 7.5) at 4 °C for 12 hr. Beads were washed three times with Assembly Buffer to remove any unbound PA prepores. The immobilized LF_N-PA prepore complexes were then incubated in low pH buffer to transition the $PA_{prepore}$ to PA_{pore} and are predicted to initiate translocation of LF_N. The beads were washed in Assembly Buffer at neutral pH three times. Then, pre-nanodisc micelles (2.5 μM MSP1D1, 97.5 μM 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) (Avanti, Alabaster, AL, USA), 65 (POPG) in 25 mM Na-cholate (Sigma-Aldrich, St. Louis, MO, USA), 50 mM Tris, and 50 mM NaCl) were added and bound to the aggregation-prone hydrophobic transmembrane β-barrel of PA_{pore}. The micelles were collapsed into nanodiscs by removing Na-cholate using dialysis with Bio-Beads (BIO RAD, Hercules, CA, USA). Soluble complexes were released from the thiol sepharose beads by reducing the E126C LF_N-bead disulfide bond using 50 mM dithiothreitol (DTT) (Goldbio, St. Louis, MO, USA) in Assembly Buffer. Assembled complexes were initially confirmed using negative-stain TEM. Complexes were stored at -80C prior to cryoEM grid preparation.

Grid preparation for cryoEM

Complexes stored at -80C were thawed on ice. A glow discharged Quantifoil R1.2/1.3 300M Cu holey carbon grid was placed inside the FEI Vitrobot Mark IV humidity chamber at 100% humidity. Then, 2ul of thawed sample was applied to the grid followed by 0.5uL of 1M acetate pH 5.5. The grids were then blotted and plunge frozen in liquid ethane. Frozen grids were stored in liquid nitrogen prior to use.

CryoEM data collection and image processing

CryoEM grids were loaded into a FEI Titan Krios electron microscope operated at 300 kV for automated image acquisition with serialEM (Mastronarde 2005). cryoEM micrographs were recorded as movies on a Gatan K2 Summit direct electron detection camera using the electron counting mode in super resolution mode at ×130K nominal magnification, a pixel size of 0.535 A per pixel, and defocus ranging between −1 and −3 μm. Total dose was 50.76 e-/A^2. Total exposure time was 9s and fractionated into 45 frames with 200 ms exposure time for each frame. In total, 3701 micrographs were taken in a continuous session. Frames in each movie were aligned and averaged for correction of beam-induced drift using MotionCor2 and cryoSPARC patch motion correction to generate a micrograph. Micrographs generated by averaging all frames of each movie were used for defocus determination and particle picking. Micrographs obtained by averaging frames 2-23 (corresponding to ~30 electrons per square ångström) were used for two- and three-dimensional image classifications. The best 3,223 micrographs were selected for the following in-depth data processing.

Single particle analysis and structure determination

Single particle analysis was performed using cryoSPARC v2.14.2 (Punjani, Rubinstein et al. 2017) (**Figure 4.9**). A random subset of micrographs was selected for blob particle picking.

These particles were subjected to 2D classification in order to obtain a set of five particle templates. Using these templates, 1,772,616 particles were selected from 3,223 micrographs. After multiple rounds of 2D classification, the remaining 209,513 ‘good’ particles were used to create a C7 symmetric 3D model. C7 symmetry expansion was performed on these particles followed by 3D variability analysis. Three orthogonal principle modes (i.e. eigenvectors of the 3D covariance) were solved with a filter resolution of 6Å. 124,919 particle from four resulting clusters with potential LF_N density were selected for further processing. 3D classification, consisting of two rounds of heterogenous refinements were performed. 3D classes without LF_N density were discarded leaving two classes. A non-uniform refinement on the per particle motion and CTF corrected particles was performed resulting in a 3.06 Å and 3.24 Å model. Resolution was determined using gold standard Fourier shell correlation with a cut off of 0.143. Resolution and particles distribution are shown in **Figure 4.10**.

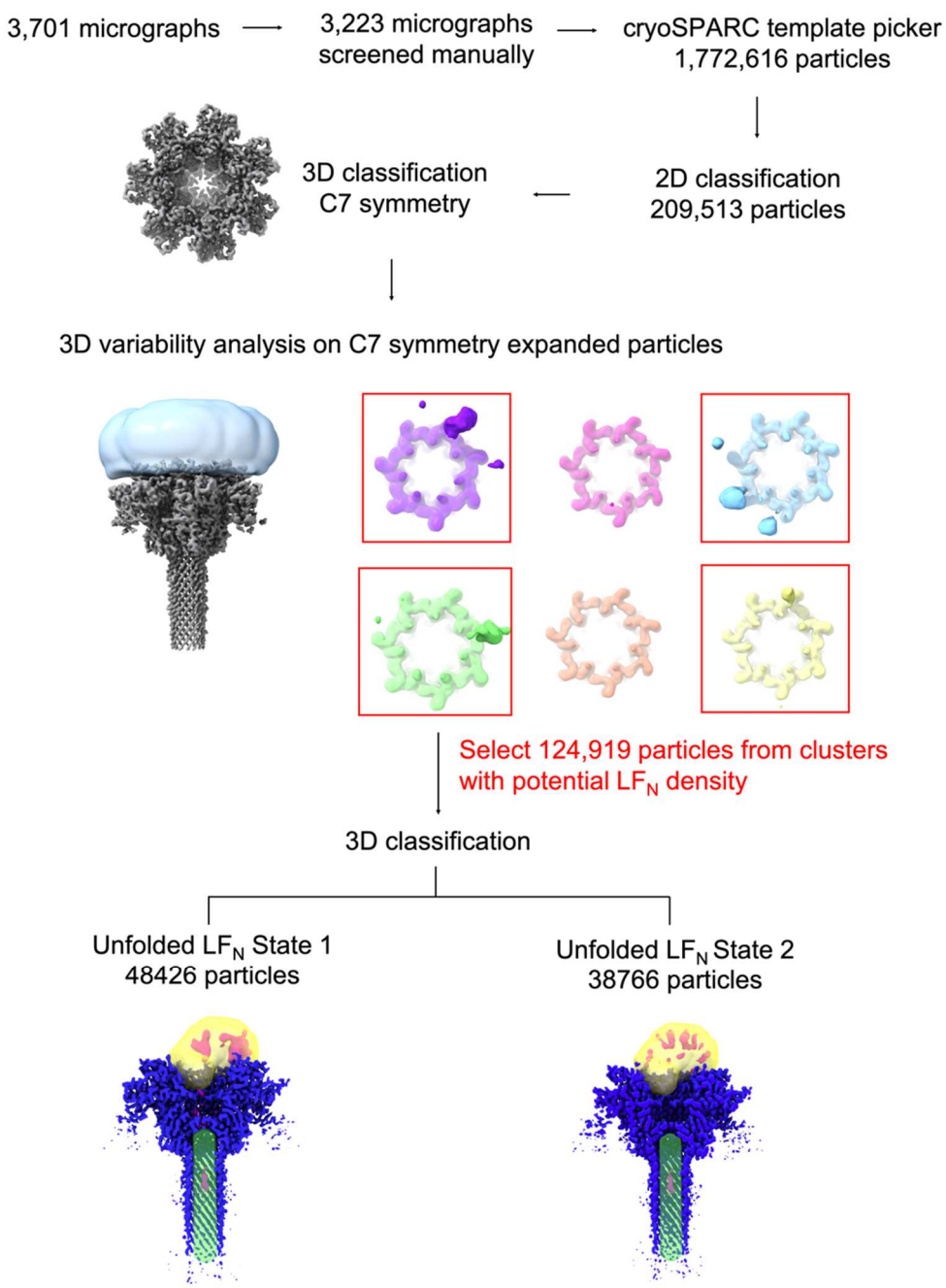


Figure 4.9. Single particle analysis of anthrax toxin translocating complexes. C7 PA_{pore} in grey with light blue mask used for 3D variabiltiy analysis. Results of 3D variability analysis shown in pastels (purple, pink, blue, green, orange, yellow) with potential LF_N density containing maps boxed in red. Refined PA_{pore} maps shown in blue with LF_N in magenta, yellow mask for unfolding LF_N local refinement, and green mask for refolding LF_N local refinement.

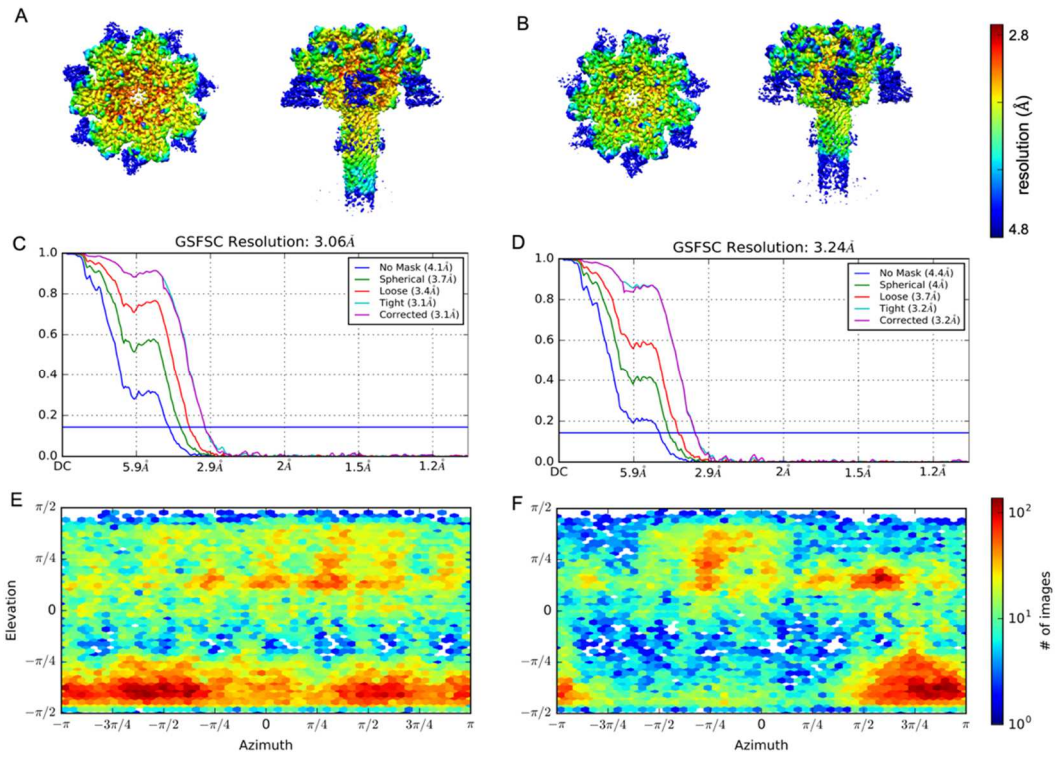


Figure 4.10. Resolution and particle distribution of CryoEM maps. Local resolution of translocation complex (A) state 1 and (B) state 2. Gold Standard FSC curves for translocation complex (C) state 1 and (D) state 2. Euler distribution of particle poses for translocation complex (E) state 1 and (F) state 2.

Local refinement was performed on the cap of the PA pore to further characterize bound LF_N using a mask of LF_N (PDB 3KWV) low pass filtered to 30Å. To clearly visualize density in the PA_{pore} lumen not attributable to the PA_{pore}, a 3.0Å volume of PA_{pore} was created using Chimera molmap (Pettersen, Goddard et al. 2004). This volume was subtracted from the cryoEM map. The resulting density was sharpened in PHENIX (Adams, Afonine et al. 2010) to a resolution of 6 Å to obtain density of LF_N from the α to the Φ clamp. Local refinement was performed on the β barrel interior using a cylindrical mask.

Model Building and Refinement

An initial model using PDB 6PSN for the PA_{pore} and PDB 3KWV for LF_N was docked into the cryoEM map and corresponding LF_N local refinement map, respectively, using Chimera

map to model (Pettersen, Goddard et al. 2004). The LF_N coarse model was adjusted manually using Coot (Emsley and Cowtan 2004) to fit the density starting at the C-terminus. Model α helical assignments were based on helicity in original model (PDB 3KWV), cryoEM density diameter, and consistency with previously published helical density (**Figure 4.11**). The PA_{pore} coarse model was refined using PHENIX real space refine (Adams, Afonine et al. 2010). Individual atomic model side chains were manually adjusted to fit the density map using Coot (Emsley and Cowtan 2004). This process was repeated iteratively until an optimal model was obtained. Ramachandran plots and MolProbity (Chen, Arendall et al. 2010) were used to assess model quality. Supplementary **Table 4.1** is a summary of cryoEM data collection and processing as well as model building and validation.

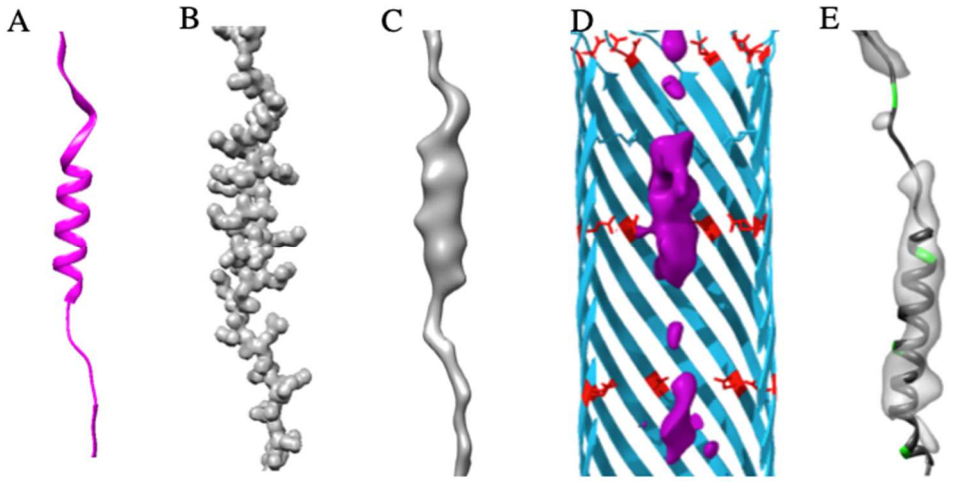


Figure 4.11. Examples of α helix models and electron density maps. (A) LF_N α helix model in our complex (magenta ribbon). (B) 2.0 Å density map of model LF_N. (C) Low pass filter density of map from model. (D) CryoEM data of beta barrel pore consistent with α helix. (E) Example of refolding α helix (in a ribosome exit tunnel) in the literature (Wilson and Beckmann 2011)

Table 4.1. CryoEM data collection, data processing, model building, and model validation

	LF_N-PA_{pore} Unfolded State 1	LF_N-PA_{pore} Unfolded State 2
Magnification	130,000X	130,000X
Voltage (kV)	300	300
Electron exposure (e^{-}/Å^2)	1.79 per frame	1.79 per frame
Defocus range (μm)	-1 to -3 uM	-1 to -3 uM
Pixel size (Å)	0.535	0.535
Symmetry imposed	C1	C1
Initial particle images (no.)	1,772,616	1,772,616
Final particle images (no.)	48,426	38,766
Map resolution (Å)	3.06	3.24
FSC threshold	0.143	0.143
Refinement		
Initial model used (PDB ID)	PA_{pore} (6PSN)	PA_{pore} (6PSN)
Model resolution (Å)	3.1	3.2
FSC threshold	0.143	0.143
Model composition		
Nonhydrogen atoms	31,031	31,031
Protein residues	3,934	3,934
Ligands	CA:14	CA:14
***B* factors (Å ^{2})**		
Protein	115.75	118.01
Ligand	76.41	74.69
r.m.s. deviations		
Bond lengths (Å)	0.007	0.006
Bond angles (°)	0.74	0.76
Validation		
MolProbity score	2.99	3.05
Clashscore	11.16	13.06
Ramachandran plot		
Favored (%)	92.68	92.60
Allowed (%)	7.32	7.40
Disallowed (%)	0	0

Chapter 5: Discussion

Summary Insight into Anthrax Toxin Pore Formation and Translocation

The morbidity and mortality of anthrax disease is associated with the anthrax toxin. The anthrax toxin is a two-component system, termed AB toxin, with an active (A) moiety and a binding (B) moiety. The anthrax lethal toxin A component is lethal factor (LF), a mitogen-activated protein kinase kinase protease. LF must be in the cytosol of the host cell in order to perform its virulent physiological function (Pannifer, Wong et al. 2001). Cellular entry is accomplished by the B component of the toxin termed protective antigen (PA). PA binds a target host cell receptor and forms a pore (PA_{pore}) to translocate LF into the cytosol (Santelli, Bankston et al. 2004). In this work, we explore the interplay between toxin components during pore formation and translocation in an effort to provide a more complete picture of the anthrax intoxication mechanism at these crucial steps. Specific questions we sought to address include: Does PA release from its cell receptor during pore formation? How does the LF N-terminal tail interact with the PA_{pore} lumen? Does LF form structural elements inside the PA_{pore} β barrel during translocation? To address receptor binding during pore formation, we investigated the anthrax toxin receptor CMG2 binding capabilities to the PA_{pore} under endosomal conditions. Our results provide evidence for receptor release prior to pore formation consistent with the hypothesized receptor clamp mechanism. This mechanism involves the receptor preventing premature pore formation by binding to both the receptor binding domain of PA as well as the pore forming domain of PA (**Chapter 2**). We then characterized the structure of three LF_N bound to PA_{pore} at neutral pH. Our results indicated the N-terminal tail of LF remains flexible in the translocation incompetent neutral pH environment and underscores the necessity of using physiologically relevant conditions when studying this system (**Chapter 3**). To investigate whether LF forms structural elements inside the PA_{pore} during translocation, we captured intermediates of LF

translocation using cryoEM. Our results support the hypothesis that initial refolding of LF structural elements occurs in the PA_{pore} β barrel during translocation. (**Chapter 4**). Cumulatively we have made significant contributions to our understanding of the anthrax intoxication mechanism at multiple biologically relevant steps.

Receptor Dissociation During Pore Formation

In **Chapter 2**, we present data suggesting CMG2 does not remain bound to the pore complex and propose this disassociation is necessary for pore formation in the endosome. Receptor binding is important for pore formation with CMG2 stabilizing the prepore complex at pH values greater than 5.2 (Scobie, Marlett et al. 2007). This stabilization occurs through the burial of 2000 $Å^2$ of PA surface involving domain 4 and domain 2. In the receptor clamp mechanism, CMG2 is hypothesized to act as a clamp or latch preventing premature pore formation by disallowing the hinge motion of domain 4 that is needed to provide the space for domain 2 rearrangement and membrane insertion of the PA pore (Lacy, Wigelsworth et al. 2004). Partial or complete receptor disassociation would be consistent with the receptor latch mechanism. However, whether or not CMG2 remained bound to the pore form of PA was not known. We hypothesized the receptor would disassociate during pore formation. To test this hypothesis, we characterized the prepore and pore complex *in vitro* using three complementary techniques: biolayer interferometry (BLI), mass spectrometry (MS), and negative stain electron microscopy (EM). We monitored the assembly of LF_N-$PA_{prepore}$-CMG2 complexes using BLI and observed sensogram changes consistent with the dissociation of CMG2 upon pore formation, thus supporting our hypothesis that the receptor disassociates during pore formation. To provide further evidence of CMG2 receptor release during PA_{pore} formation, we then adapted EM and MS to determine the composition of the components on the BLI biosensor before and after pore formation. Notably,

CMG2 was not detected in the samples after pore formation providing further evidence for receptor release during pore formation. It remains to be seen whether there is a physiological benefit of PA-receptor dissociation or if it is a bi-product of the pore formation process.

Combining the Receptor Clamp and Pore Formation Mechanisms

The anthrax toxin receptor clamp mechanism is not part of the current model of PA pore formation, and our results fall short of bridging these two key steps in the anthrax toxin mechanism. Moreover, while our results do suggest receptor release, they do not determine when during pore formation the receptor release occurs. The pore formation mechanism is based on the observed differences between the structures of the prepore and pore complexes. Specifically, the flipping of the 2β10–2β11 loop was suggested as 'the first step' in the pore formation followed by Domain 2 rotation and β hairpin insertion (Jiang, Pentelute et al. 2015). The authors note other parts of PA may regulate pH sensing, but the question remains: When during pore formation does the receptor release? The receptor is hypothesized to prevent premature β hairpin insertion. Does it release at this final step? Or does it release earlier in the process, perhaps even before the 2β10–2β11 loop flipping? Other groups have started to investigate intermediates of pore formation using 2β10–2β11 loop mutants (*e.g.*, D425A) to arrest pore formation, presumably at the loop flipping step (Scott III 2018). This work continues to focus on structural changes of PA_{pore} and excludes consideration of the effect of receptors. However, the results and approaches from **Chapter 2** could be used in combination with the pore forming loop mutant to determine when during pore formation the receptor is released, creating a more complete picture of the pore formation process. Specifically, using our BLI, MS, and EM methodology, the D425A mutants could be tested to see whether CMG2 remains bound during the loop flipping step of pore formation. If CMG2 does remain bound, a prepore-pore intermediate structure with

CMG2 bound could be obtained using cryoEM. These results would clarify the structural detail from receptor bound $PA_{prepore}$ to receptor released PA_{pore}, filling in important gaps between the receptor clamp and pore formation mechanisms.

Structural Studies of LF bound to PA_{pore} and Sample Heterogeneity

Due to our results indicating receptor release during pore formation, we chose to exclude CMG2 from our subsequent structural biology work of PA_{pore}. The first structure we determined was a structure of the PA_{pore} with 3 LF_N bound at pH 7.5. This cryoEM data was collected on a charge-coupled device, not the direct electron detectors that spurred the 'cryoEM revolution' (Kühlbrandt 2014) (**Appendix E**). One of our main challenges in getting a 3D reconstruction of $3LF_N$ bound to PA_{pore} was the amount of heterogeneity in our data. Indeed, this is often an issue for cryoEM data analysis with classification schemes unable to separate small conformational or compositional differences due to coarse angular samplings (Scheres 2016). Our heterogeneity manifested in the number of LF_N bound to PA_{pore} with 2D classes containing one, two, or three LF_N. Three LF_N bound was the predominant species in our data. However, a C1 symmetric model could not be obtained without the use of stochastic gradient descent (SGD) modelling, which is capable of sampling a large conformational space without the use of an input model. To this end, CryoSPARC *ab initio,* which uses SGD modelling, allowed us to perform 3D classification to sort out three bound LF_N complexes (Punjani, Rubinstein et al. 2017). While 3D classification is now standard practice for single particle analysis, it was a new concept at the time of the data analysis in **Chapter 3**. 3D classification allowed us to obtain a 17 Å density map of 3 LF_N bound to PA_{pore} at neutral pH. We used previously published X-ray crystal and cryoEM structures of LF_N and PA_{pore}, respectively, to create a flexible fitted model in our cryoEM density. The model showed LF_N bound to PA_{pore} in the same canonical binding conformation as

LF_N bound to $PA_{prepore}$ with the first helix bound in the α clamp. We had predicted that the LF N-terminal tail would become more structured and interact with the Φ clamp. However, we did not see the LF N-terminal tail density in the pore lumen. We attribute this result to the electrostatic environment of the PA_{pore} lumen at neutral pH, and therefore it would be of interest to the field to perform a follow up study at low pH.

Comparison of Chaotropic and pH Induced Pore Formation

The conformational changes that occur going from $PA_{prepore}$ to the PA_{pore} can be described using a protein folding landscape. The PA_{pore} would be the free energy minimum of the landscape with $PA_{prepore}$ existing as a local minimum along the folding pathway. Importantly, there is an energy barrier between these two states. *In vivo*, this barrier is overcome by the protonation of specific histidine residues and the destabilization of receptor-PA interactions. *In vitro*, urea has been used to induce $PA_{prepore}$ to PA_{pore} conformational changes. Chaoptropes, like urea, introduce entropy to the system. Likely, this entropy overcomes the energy barrier between prepore and pore. The urea induced refolding path and the acid induced refolding path may or may not be the same along the energy landscape. However, the end point, the structure of PA_{pore}, appears to be the highly similar. Specifically, we compared our molecular model of PA_{pore} at pH 5.5 to a published molecular model of PA_{pore} at pH 7.5 (1M urea induced pore formation, PDB 6PSN). The models were highly similar with 99.2% of the 3919 residues within 2 Å of each other (Phenix PDB chain comparison). The most variable region between the two models was domain 4, which was also the lowest resolution regions for both cryoEM maps, likely due to flexibility of domain 4 with respect to the rest of PA after pore formation. While the structure of PA_{pore} at neutral and acidic pH are similar, the electrostatic microenvironment of the pore lumen would likely be different.

Importantly, this pH difference would have an effect on the interaction between the PA_{pore} lumen and the N-terminal tail of LF.

Lethal Factor N-Terminal Tail Flexibility and Translocation Order

Further evidence in support of our hypothesis that the LF N-terminal tail is flexible at neutral pH, **Chapter 3**, has recently been published. A *bioRxiv* preprint structure of the $PA_{prepore}$ with three bound LFs suggesting LF N-terminal tail flexibility as well as LF structural heterogeneity (Antoni, Quentin et al. 2020). In the 3LF-$PA_{prepore}$ structure, only the middle LF, termed LF_2, is engaged with the canonical α clamp binding site, with α helix 1 bound to the amphipathic cleft between $PA_{prepore}$ protomers (**Figure 5.1A-B**). LF_2 was said to be in an 'open state' due to the movement of α helix 1 away from LF and into the clamp site (**Figure 5.1C**). Its C-terminal domain was also highly flexible, thus suggesting it was primed for unfolding. Interestingly, the LFs on either side of the LF_2 are not engaged with the α clamp, but in an 'intermediate state' with their N terminal tails unresolved, presumably flexible above the prepore lumen (**Figure 5.1D**). Unlike LF_2, their C-terminal domains are stabilized.

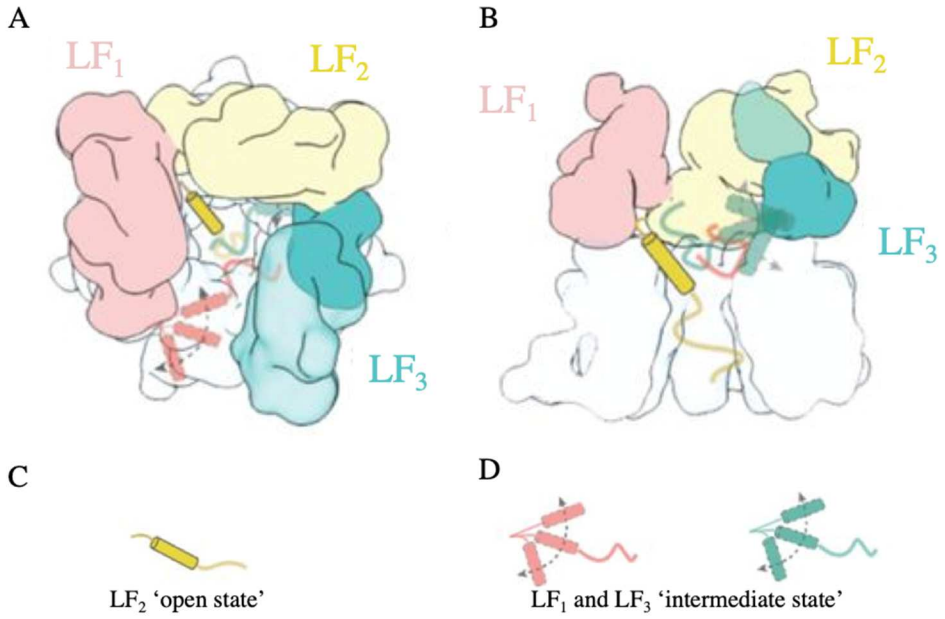


Figure 5.1. Proposed structural state of LF N-terminal tails in structure of three LF bound to $PA_{prepore}$. (A) cartoon top and (B) side view of $PA_{prepore}$ with three LF bound. (C) Cartoon depiction of LF_2 open state bound to the α clamp of $PA_{prepore}$. (D) Cartoon depiction of LF_1 and LF_3 'intermediate state' flexible above the pore lumen. Adapted from (Antoni, Quentin et al. 2020).

What determines which LF translocates first, second, and third through the pore when all three binding sites are occupied is a major question in the field (Fabre, Santelli et al. 2016, Krantz 2016). As shown in **Figure 5.1**, this recent study provided insight into the order of LF translocation. The authors hypothesize that translocation order is dictated by LF binding order, at least for the first translocating LF. In their model, LF_2 is predicted to bind $PA_{prepore}$ first and would be the first factor to translocate followed by either LF_1 or LF_3. In support of the model, their prepore structure shows only one LF N-terminal tail funneled down the prepore lumen. Because only one tail is shown funneled toward the Φ clamp, these results may also clarify why the complex does not stall, as one would expect, from multiple tails trying to translocate at once.

Whether the model remains true for multiple LFs bound to the pore form a PA at low pH remains to be seen.

Proposed Future Structural Studies of Translocation

In **Chapter 4**, we propose a translocation unfolding-refolding mechanism whereby the extended pore of PA facilitates initial refolding of α helical elements prior to channel exit. This initial refolding would disfavor retrograde transfer, overcome local energy barriers, and ensure proper tertiary folding. Our cryoEM translocation complexes offer preliminary evidence of helical formation in the pore in the form of electron density consistent with a helix in the PA_{pore} β barrel. However, more evidence is needed to support this hypothesis. To confirm translocation, a biotin bobber could be placed on the LF N-termini. Streptavidin binding, visualized by EM density or label free technology (SPR or BLI), would allow for confirmation that the biotin bobber had translocated through the PA_{pore}. As a negative control for helix formation, substrates unable to form helical components, such as alternative L and D amino acids or proline inserted sequences could be used. LF_{1-50} peptide sequences with modified stereochemical properties have previously been developed to investigate the effect of helical substrates on the rate of translocation (Brown, Thoren et al. 2015, Das and Krantz 2016{Brown, 2015 #22). We predict these substrates would not form a helix-like density in the β barrel. This negative control would capture non-helical portions of translocating peptide, allowing for characterization of such intermediates. The resulting density (or lack thereof) of non-helical peptide in the charge clamp would be a useful comparison to our hypothesized helical density. Alternatively, a different immobilization residue on full length LF could also be used that is further away from the N-terminus to allow for the capture of translocation events at different steps in the process. Processivity of other systems has been studied using time-lapse cryoEM. For example, time-lapse cryoEM of ribosome

processivity was performed using spray technology coupled to a stop flow mixer (Lu, Shaikh et al. 2009). Owing to its potential in capturing processive structural snapshots, time-lapse cryoEM would also be useful for future structural studies of translocation complexes. In fact, the thiol sepharose bead slurry, used in **Chapter 3 and 4** to immobilize complexes and prevent aggregation, could be poured into a column allowing for a low pH pulse chase potentially resulting in more discrete intermediates. Moreover, instead of using beads, a plug flow reactor design could be used with the immobilization chemistry on the interior wall of the reactor. In this case, the column or reactor could then be coupled to a quick spray nozzle capable of rapidly depositing translocation complexes onto a grid as they are eluted. This combination of pulse chase column and spray deposition has the potential to carry out the entire sample preparation process at low pH resulting in a physiological complex.

Translocational Refolding in the Context of AB Toxins

We hypothesize that our refolding mechanism is used by other AB toxins with extended pores and that the pore length has evolved to facilitate this refolding (**Chapter 4**). Other toxins with extended pores include *C. botulinum* C2 toxin*, C. perfringens* Iota toxin*, C. Difficile* TcdA, TcdB, and transferase toxin (Tsuge, Nagahama et al. 2003, Schleberger, Hochmann et al. 2006, D'Urzo, Malito et al. 2012, Krakauer 2016, Chen, Lam et al. 2019). However, there are also several examples of AB toxins that have pores that do not extend above the membrane bilayer, *e.g*., cholera toxin and shiga toxin (Fraser, Chernaia et al. 1994, Zhang, Scott et al. 1995), or do not use translocational pores, *e.g*., diphtheria toxin (Ladokhin, Vargas-Uribe et al. 2017). It is not surprising that bacteria have evolved multiple ways to intoxicate host cells. However, one cannot help but wonder: What physiological benefit is there for an AB toxin to have an extended pore? One possibility may have to do with the molecular weight of the translocated A component. In

support of this idea, AB toxins with extended pores tend to have higher molecular weight A components (30-90 kDa) compared to other AB toxins (less than 50 kDa). The molecular weight of the active A component of several AB toxins is shown in **Table 5.1**. We predict the larger or more complex proteins use extended pores to efficiently refold structural elements leading to proper tertiary fold of toxin A components. Importantly, without the extended pore, there would be an increased probability of misfolded A components leading to a decrease in toxicity.

Table 5.1. Molecular weight of the A component of AB toxins

Toxin	Bacteria	A component MW (kDa)	
Membrane Spanning Pore			
Cholera toxin	*Vibrio cholerae*	28	(Zhang, Scott et al. 1995)
Heat labile toxin	*Escherichia coli (ETEC)*	29	(Merritt, Sarfaty et al. 1995)
Pertussis toxin	*Bordetella pertussis*	26	(Stein, Boodhoo et al. 1994)
Shiga toxin	*Shigella dysenteriae*	35	(Fraser, Chernaia et al. 1994)
Extended Pore			
Anthrax edema toxin	*Bacillus anthracis*	89	(Drum, Yan et al. 2002)
Anthrax lethal toxin	*Bacillus anthracis*	90	(Pannifer, Wong et al. 2001)
C2	*Clostridium botulinum*	49	(Schleberger, Hochmann et al. 2006)
CST	*Clostridium spiroforme*	47	(Papatheodorou, Wilczek et al. 2012)
Iota	*Clostridium perfringens*	48	(Tsuge, Nagahama et al. 2003)
Toxin A (TcdA)	*Clostridium difficile*	34	(D'Urzo, Malito et al. 2012)
Toxin B (TcdB)	*Clostridium difficile*	63	(Chen, Lam et al. 2019)
Toxin complex (Tc)	*Photorhabdus luminescens*	30	(Busby, Panjikar et al. 2013)
Transferase toxin (CDT)	*Clostridium difficile*	48	(Krakauer 2016)
Other*			
Botulinum neurotoxin (BoNT), A-G	*Clostridium botulinum*	50	(Lacy, Tepp et al. 1998)
Cytotoxic necrotizing factor 1 (CNF)	*Escherichia coli*	33	(Buetow, Flatau et al. 2001)
Exotoxin A	*Pseudomonas aeruginosa*	37	(Wedekind, Trame et al. 2001)
Diphtheria toxin	*Corynebacterium diphtheriae*	21	(Choe, Bennett et al. 1992)
Tetanus neurotoxin (TeNT)	*Clostridium tetani*	50	(Masuyer, Conrad et al. 2017)

*AB toxins that have not been shown to form stable pores for the purpose of translocation

Comparison of EF and LF Toxin Complexes

While much of this work focused on LF, edema factor (EF) is also a key component of the anthrax toxin. EF has been structurally characterized by other groups, and crystal structures of EF bound to calmodulin were solved in 2002 (Drum, Yan et al. 2002). During the course of our work on LF_N bound to PA, three structures of EF bound to PA_{pore} were published (Hardenbrook, Liu et al. 2020). Interestingly, EF undergoes a large-scale conformational rearrangement upon

binding to PA. Specifically, the EF helical domain (domain 4) swings down and bridges EF_N (domain 1) and the adenylate cyclase domains (domains 2 and 3) to form a more compact complex (Hardenbrook, Liu et al. 2020). The EF pore studies also characterized LF bound to PA_{pore} at neutral pH. However, unlike EF, only the N-terminal domain of LF was resolvable in the structure suggesting the LF C-terminus is destabilized when bound to the pore. Of note, this destabilization would have to be the result of structural changes and not the endosomal environment as the pH was neutral. To our knowledge, no structural work has been done on hetero-enzyme complexes with both LF and EF bound to PA, despite the ability of LF-EF-PA complexes to assemble *in vivo* with a single PA_{pore} capable of delivering both enzymes to the cytosol of a host cell (Pimental, Christensen et al. 2004). Additional questions that will need to be addressed in the future include: Is there a preferential arrangement of enzymes bound to $PA_{prepore}$, *e.g.*, LF-EF-LF or LF-EF-EF?; and Does the binding of one enzyme to PA affect the binding of subsequent enzymes? Several single-molecule techniques would be amenable to probing these questions (Shashkova and Leake 2017). For example, potential cooperativity during the assembly of LF-$PA_{prepore}$ complexes could be investigated using negative stain EM and 2D classification of $PA_{prepores}$ with increasing concentrations of the A components. For hetero-enzyme complexes, single-molecule total internal reflection fluorescence, or TIRF, with differentially labelled LF and EF could be used to observe processive binding events.

References

Abdiche, Y., D. Malashock, A. Pinkerton and J. Pons (2008). "Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet." Anal Biochem **377**(2): 209-217.

Abdiche, Y. N., D. S. Malashock, A. Pinkerton and J. Pons (2009). "Exploring blocking assays using Octet, ProteOn, and Biacore biosensors." Anal Biochem **386**(2): 172-180.

Abrami, L., S. Liu, P. Cosson, S. H. Leppla and F. G. van der Goot (2003). "Anthrax toxin triggers endocytosis of its receptor via a lipid raft–mediated clathrin-dependent process." J Cell Biol **160**(3): 321-328.

Abrami, L., S. Liu, P. Cosson, S. H. Leppla and F. G. van der Goot (2003). "Anthrax toxin triggers endocytosis of its receptor via a lipid raft–mediated clathrin-dependent process." The Journal of cell biology **160**(3): 321-328.

Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral and R. W. Grosse-Kunstleve (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D: Biological Crystallography **66**(2): 213-221.

Akkaladevi, N., L. Hinton-Chollet, H. Katayama, J. Mitchell, L. Szerszen, S. Mukherjee, E. Gogol, B. Pentelute, R. Collier and M. Fisher (2013). "Assembly of anthrax toxin pore: Lethal-factor complexes into lipid nanodiscs." Protein Science **22**(4): 492-501.

Akkaladevi, N., S. Mukherjee, H. Katayama, B. Janowiak, D. Patel, E. P. Gogol, B. L. Pentelute, R. J. Collier and M. T. Fisher (2015). "Following natures lead: on the construction of membrane-inserted toxins in lipid bilayer nanodiscs." The Journal of membrane biology **248**(3): 595-607.

Anderson, D. M., M. J. Sheedlo, J. L. Jensen and D. B. Lacy (2020). "Structural insights into the transition of Clostridioides difficile binary toxin from prepore to pore." Nature microbiology **5**(1): 102-107.

Andrews, R. K., J. López and M. C. Berndt (1997). "Molecular mechanisms of platelet adhesion and activation." The international journal of biochemistry & cell biology **29**(1): 91-105.

Antoni, C., D. Quentin, A. E. Lang, K. Aktories, C. Gatsogiannis and S. Raunser (2020). "Cryo-EM structure of the fully-loaded asymmetric anthrax lethal toxin in its heptameric pre-pore state." bioRxiv.

Auton, M., M. A. Cruz and J. Moake (2007). "Conformational stability and domain unfolding of the Von Willebrand factor A domains." Journal of molecular biology **366**(3): 986-1000.

Auton, M., L. M. F. Holthauzen and D. W. Bolen (2007). "Anatomy of energetic changes accompanying urea-induced protein denaturation." Proceedings of the National Academy of Sciences **104**(39): 15317-15322.

Auton, M., E. Sedlák, J. Marek, T. Wu, C. Zhu and M. A. Cruz (2009). "Changes in thermodynamic stability of von Willebrand factor differentially affect the force-dependent binding to platelet GPIbα." Biophysical journal **97**(2): 618-627.

Auton, M., K. E. Sowa, S. M. Smith, E. Sedlák, K. V. Vijayan and M. A. Cruz (2010). "Destabilization of the A1 domain in von Willebrand factor dissociates the A1A2A3 tri-domain and provokes spontaneous binding to glycoprotein Ibα and platelet activation under shear stress." Journal of Biological Chemistry **285**(30): 22831-22839.

Bandyopadhyay, A. and J. Gao (2016). "Iminoboronate-based peptide cyclization that responds to pH, oxidation, and small molecule modulators." Journal of the American Chemical Society **138**(7): 2098-2101.
Bayburt, T. H. and S. G. Sligar (2010). "Membrane protein assembly into Nanodiscs." FEBS letters **584**(9): 1721-1727.
Bellon, S., W. Buchmann, F. Gonnet, N. Jarroux, M. Anger-Leroy, F. Guillonneau and R. Daniel (2009). "Hyphenation of surface plasmon resonance imaging to matrix-assisted laser desorption ionization mass spectrometry by on-chip mass spectrometry and tandem mass spectrometry analysis." Anal Chem **81**(18): 7695-7702.
Birmanns, S., M. Rusu and W. Wriggers (2011). "Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes." Journal of structural biology **173**(3): 428-435.
Blaustein, R. O., T. M. Koehler, R. J. Collier and A. Finkelstein (1989). "Anthrax toxin: channel-forming activity of protective antigen in planar phospholipid bilayers." Proceedings of the National Academy of Sciences **86**(7): 2209-2213.
Bonuccelli, G., F. Sotgia, P. G. Frank, T. M. Williams, C. J. De Almeida, H. B. Tanowitz, P. E. Scherer, K. A. Hotchkiss, B. I. Terman and B. Rollman (2005). "ATR/TEM8 is highly expressed in epithelial cells lining Bacillus anthracis' three sites of entry: implications for the pathogenesis of anthrax infection." American Journal of Physiology-Cell Physiology **288**(6): C1402-C1410.
Bradley, K. A., J. Mogridge, M. Mourez, R. J. Collier and J. A. Young (2001). "Identification of the cellular receptor for anthrax toxin." Nature **414**(6860): 225-229.
Brillault, L. and M. J. Landsberg (2020). Preparation of Proteins and Macromolecular Assemblies for Cryo-electron Microscopy. Protein Nanotechnology, Springer**:** 221-246.
Brown, M. J., K. L. Thoren and B. A. Krantz (2015). "Role of the α clamp in the protein translocation mechanism of anthrax toxin." Journal of molecular biology **427**(20): 3340-3349.
Brubaker, M. A., A. Punjani and D. J. Fleet (2015). Building proteins in a day: Efficient 3D molecular reconstruction. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
Buetow, L., G. Flatau, K. Chiu, P. Boquet and P. Ghosh (2001). "Structure of the Rho-activating domain of Escherichia coli cytotoxic necrotizing factor 1." Nature structural biology **8**(7): 584-588.
Busby, J. N., S. Panjikar, M. J. Landsberg, M. R. Hurst and J. S. Lott (2013). "The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device." Nature **501**(7468): 547-550.
Bush, L. M., B. H. Abrams, A. Beall and C. C. Johnson (2001). "Index case of fatal inhalational anthrax due to bioterrorism in the United States." New England Journal of Medicine **345**(22): 1607-1610.
Chen, P., K.-h. Lam, Z. Liu, F. A. Mindlin, B. Chen, C. B. Gutierrez, L. Huang, Y. Zhang, T. Hamza and H. Feng (2019). "Structure of the full-length Clostridium difficile toxin B." Nature structural & molecular biology **26**(8): 712-719.
Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallographica Section D: Biological Crystallography **66**(1): 12-21.
Choe, S., M. J. Bennett, G. Fujii, P. M. Curmi, K. A. Kantardjieff, R. J. Collier and D. Eisenberg (1992). "The crystal structure of diphtheria toxin." Nature **357**(6375): 216-222.

Chopra, A. P., S. A. Boone, X. Liang and N. S. Duesbery (2003). "Anthrax lethal factor proteolysis and inactivation of MAPK kinase." Journal of Biological Chemistry **278**(11): 9402-9406.
Cox, K. and M. S. Sansom (2009). "One membrane protein, two structures and six environments: a comparative molecular dynamics simulation study of the bacterial outer membrane protein PagP." Molecular membrane biology **26**(4): 205-214.
Croix, B. S., C. Rago, V. Velculescu, G. Traverso, K. E. Romans, E. Montgomery, A. Lal, G. J. Riggins, C. Lengauer and B. Vogelstein (2000). "Genes expressed in human tumor endothelium." Science **289**(5482): 1197-1202.
D'Urzo, N., E. Malito, M. Biancucci, M. J. Bottomley, D. Maione, M. Scarselli and M. Martinelli (2012). "The structure of Clostridium difficile toxin A glucosyltransferase domain bound to Mn2+ and UDP provides insights into glucosyltransferase activity and product release." The FEBS journal **279**(17): 3085-3097.
Dandey, V. P., W. C. Budell, H. Wei, D. Bobe, K. Maruthi, M. Kopylov, E. T. Eng, P. A. Kahn, J. E. Hinshaw and N. Kundu (2020). "Time-resolved cryoEM using Spotiton." bioRxiv.
Das, D. and B. A. Krantz (2016). "Peptide-and proton-driven allosteric clamps catalyze anthrax toxin translocation across membranes." Proceedings of the National Academy of Sciences **113**(34): 9611-9616.
Das, D. and B. A. Krantz (2017). "Secondary Structure Preferences of the Anthrax Toxin Protective Antigen Translocase." Journal of molecular biology **429**(5): 753-762.
De, S. and R. Olson (2011). "Crystal structure of the Vibrio cholerae cytolysin heptamer reveals common features among disparate pore-forming toxins." Proceedings of the National Academy of Sciences **108**(18): 7385-7390.
Denisov, I., Y. Grinkova, A. Lazarides and S. Sligar (2004). "Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size." Journal of the American Chemical Society **126**(11): 3477-3487.
Dmochewitz, L., M. Lillich, E. Kaiser, L. D. Jennings, A. E. Lang, J. Buchner, G. Fischer, K. Aktories, R. J. Collier and H. Barth (2011). "Role of CypA and Hsp90 in membrane translocation mediated by anthrax protective antigen." Cellular microbiology **13**(3): 359-373.
Drum, C. L., S.-Z. Yan, J. Bard, Y.-Q. Shen, D. Lu, S. Soelaiman, Z. Grabarek, A. Bohm and W.-J. Tang (2002). "Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin." Nature **415**(6870): 396-402.
Duesbery, N. S., C. P. Webb, S. H. Leppla, V. M. Gordon, K. R. Klimpel, T. D. Copeland, N. G. Ahn, M. K. Oskarsson, K. Fukasawa and K. D. Paull (1998). "Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor." Science **280**(5364): 734-737.
Eddy, M. T., Y. Su, R. Silvers, L. Andreas, L. Clark, G. Wagner, G. Pintacuda, L. Emsley and R. G. Griffin (2015). "Lipid bilayer-bound conformation of an integral membrane beta barrel protein by multidimensional MAS NMR." Journal of biomolecular NMR **61**(3-4): 299-310.
Ekiert, D. C., G. Bhabha, G. L. Isom, G. Greenan, S. Ovchinnikov, I. R. Henderson, J. S. Cox and R. D. Vale (2017). "Architectures of lipid transport systems for the bacterial outer membrane." Cell **169**(2): 273-285. e217.
Elad, N., D. K. Clare, H. R. Saibil and E. V. Orlova (2008). "Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections." Journal of structural biology **162**(1): 108-120.

Elliott, J. L., J. Mogridge and R. J. Collier (2000). "A quantitative study of the interactions of Bacillus anthracis edema factor and lethal factor with activated protective antigen." Biochemistry **39**(22): 6706-6713.
Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D: Biological Crystallography **60**(12): 2126-2132.
Ercius, P., O. Alaidi, M. J. Rames and G. Ren (2015). "Electron Tomography: A Three-Dimensional Analytic Tool for Hard and Soft Materials Research." Advanced Materials **27**(38): 5638-5663.
Ezzell, J. W. and T. G. Abshire (1992). "Serum protease cleavage of Bacillus anthracis protective antigen." Microbiology **138**(3): 543-549.
Fabre, L., E. Santelli, D. Mountassif, A. Donoghue, A. Biswas, R. Blunck, D. Hanein, N. Volkmann, R. Liddington and I. Rouiller (2016). "Structure of anthrax lethal toxin prepore complex suggests a pathway for efficient cell entry." The Journal of general physiology **148**(4): 313-324.
Fairhead, M. and M. Howarth (2015). "Site-specific biotinylation of purified proteins using BirA." Site-Specific Protein Labeling: Methods and Protocols: 171-184.
Fasanella, A., D. Galante, G. Garofolo and M. H. Jones (2010). "Anthrax undervalued zoonosis." Veterinary microbiology **140**(3-4): 318-331.
Favaloro, E. J. (2011). Von Willebrand disease: local diagnosis and management of a globally distributed bleeding disorder. Seminars in thrombosis and hemostasis, © Thieme Medical Publishers.
Feld, G. K., K. L. Thoren, A. F. Kintzer, H. J. Sterling, Tang, II, S. G. Greenberg, E. R. Williams and B. A. Krantz (2010). "Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers." Nat Struct Mol Biol **17**(11): 1383-1390.
Fisher, M. T. (1992). "Promotion of the in vitro renaturation of dodecameric glutamine synthetase from Escherichia coli in the presence of GroEL (chaperonin-60) and ATP." Biochemistry **31**(16): 3955-3963.
Fisher, M. T. and S. Naik (2019). Systems and methods for identifying protein stabilizers, Google Patents.
Frank, J. (2017). "Time-resolved cryo-electron microscopy: Recent progress." Journal of structural biology **200**(3): 303-306.
Fraser, M. E., M. M. Chernaia, Y. V. Kozlov and M. N. James (1994). "Crystal structure of the holotoxino from shigella dysenteriae at 2.5 Å resolution." Nature structural biology **1**(1): 59-64.
Gogol, E., N. Akkaladevi, L. Szerszen, S. Mukherjee, L. Chollet-Hinton, H. Katayama, B. Pentelute, R. Collier and M. Fisher (2013). "Three dimensional structure of the anthrax toxin translocon–lethal factor complex by cryo-electron microscopy." Protein Science **22**(5): 586-594.
Grantham, J., O. Llorca, J. M. Valpuesta and K. R. Willison (2000). "Partial occlusion of both cavities of the eukaryotic chaperonin with antibody has no effect upon the rates of beta-actin or alpha-tubulin folding." J Biol Chem **275**(7): 4587-4591.
Gupta, P., S. Singh, A. Tiwari, R. Bhat and R. Bhatnagar (2001). "Effect of pH on stability of anthrax lethal factor: correlation between denaturation and activity." Biochemical and biophysical research communications **284**(3): 568-573.
Gupta, P. K., H. Chandra, R. Gaur, R. K. Kurupati, S. Chowdhury, V. Tandon, Y. Singh and K. Maithal (2003). "Conformational fluctuations in anthrax protective antigen: a possible role of calcium in the folding pathway of the protein." FEBS letters **554**(3): 505-510.
Halvorson, H. (1997). Two generations of spore research: from father to son.

Hanna, P. C., D. Acosta and R. J. Collier (1993). "On the role of macrophages in anthrax." Proceedings of the National Academy of Sciences **90**(21): 10198-10201.
Hardenbrook, N. J., S. Liu, K. Zhou, K. Ghosal, Z. H. Zhou and B. A. Krantz (2020). "Atomic structures of anthrax toxin protective antigen channels bound to partially unfolded lethal and edema factors." Nature Communications **11**(1): 1-10.
Huang, R., Z. A. Ripstein, R. Augustyniak, M. Lazniewski, K. Ginalski, L. E. Kay and J. L. Rubinstein (2016). "Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study." Proceedings of the National Academy of Sciences: 201603980.
Humphrey, W., A. Dalke and K. Schulten (1996). "VMD: visual molecular dynamics." Journal of molecular graphics **14**(1): 33-38.
Janowiak, B. E., A. Fischer and R. J. Collier (2010). "Effects of introducing a single charged residue into the phenylalanine clamp of multimeric anthrax protective antigen." Journal of Biological Chemistry **285**(11): 8130-8137.
Janowiak, B. E., L. D. Jennings-Antipov and R. J. Collier (2011). "Cys− Cys Cross-Linking Shows Contact between the N-Terminus of Lethal Factor and Phe427 of the Anthrax Toxin Pore." Biochemistry **50**(17): 3512-3516.
Jas, G. S., E. W. Childs and K. Kuczera (2019). "Kinetic pathway analysis of an α-helix in two protonation states: Direct observation and optimal dimensionality reduction." The Journal of chemical physics **150**(7): 074902.
Jiang, J., B. L. Pentelute, R. J. Collier and Z. H. Zhou (2015). "Atomic structure of anthrax PA pore elucidates toxin translocation." Nature **521**(7553): 545.
Jin, H., G. T. Cantin, S. Maki, L. C. Chew, S. M. Resnick, J. Ngai and D. M. Retallack (2011). "Soluble periplasmic production of human granulocyte colony-stimulating factor (G-CSF) in Pseudomonas fluorescens." Protein Expr Purif **78**(1): 69-77.
Katayama, H., B. E. Janowiak, M. Brzozowski, J. Juryck, S. Falke, E. P. Gogol, R. J. Collier and M. T. Fisher (2008). "GroEL as a molecular scaffold for structural analysis of the anthrax toxin pore." Nature structural & molecular biology **15**(7): 754.
Katayama, H., J. Wang, F. Tama, L. Chollet, E. Gogol, R. Collier and M. Fisher (2010). "Three-dimensional structure of the anthrax toxin pore inserted into lipid nanodiscs and lipid vesicles." Proceedings of the National Academy of Sciences **107**(8): 3453-3457.
Keeney, S. and A. Cumming (2001). "The molecular biology of von Willebrand disease." Clinical & Laboratory Haematology **23**(4): 209-230.
Khan, A. S. (2011). "Public health preparedness and response in the USA since 9/11: a national health security imperative." The Lancet **378**(9794): 953-956.
Kim, Y. E., S. Y. Yi, C. S. Lee, Y. Jung and B. H. Chung (2012). "Gold patterned biochips for on-chip immuno-MALDI-TOF MS: SPR imaging coupled multi-protein MS analysis." Analyst **137**(2): 386-392.
Kintzer, A. F., I. I. Tang, A. K. Schawel, M. J. Brown and B. A. Krantz (2012). "Anthrax toxin protective antigen integrates poly-γ-d-glutamate and pH signals to sense the optimal environment for channel formation." Proceedings of the National Academy of Sciences **109**(45): 18378-18383.
Kintzer, A. F., K. L. Thoren, H. J. Sterling, K. C. Dong, G. K. Feld, I. I. Tang, T. T. Zhang, E. R. Williams, J. M. Berger and B. A. Krantz (2009). "The protective antigen component of anthrax toxin forms functional octameric complexes." Journal of molecular biology **392**(3): 614-629.

Knapp, O., E. Maier, E. Waltenberger, C. Mazuet, R. Benz and M. R. Popoff (2015). "Residues involved in the pore-forming activity of the C lostridium perfringens iota toxin." Cellular microbiology **17**(2): 288-302.
Krakauer, T. (2016). Enterotoxins: Microbial Proteins and Host Cell Dysregulation, Multidisciplinary Digital Publishing Institute.
Krantz, B. A. (2016). "Anthrax lethal toxin co-complexes are stabilized by contacts between adjacent lethal factors." Journal of General Physiology **148**(4): 273-275.
Krantz, B. A., A. Finkelstein and R. J. Collier (2006). "Protein translocation through the anthrax toxin transmembrane pore is driven by a proton gradient." Journal of molecular biology **355**(5): 968-979.
Krantz, B. A., R. A. Melnyk, S. Zhang, S. J. Juris, D. B. Lacy, Z. Wu, A. Finkelstein and R. J. Collier (2005). "A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore." Science **309**(5735): 777-781.
Krantz, B. A., A. D. Trivedi, K. Cunningham, K. A. Christensen and R. J. Collier (2004). "Acid-induced unfolding of the amino-terminal domains of the lethal and edema factors of anthrax toxin." Journal of molecular biology **344**(3): 739-756.
Kühlbrandt, W. (2014). "The resolution revolution." Science **343**(6178): 1443-1444.
Lacy, D. B., H. C. Lin, R. A. Melnyk, O. Schueler-Furman, L. Reither, K. Cunningham, D. Baker and R. J. Collier (2005). "A model of anthrax toxin lethal factor bound to protective antigen." Proceedings of the National Academy of Sciences **102**(45): 16409-16414.
Lacy, D. B., M. Mourez, A. Fouassier and R. J. Collier (2002). "Mapping the anthrax protective antigen binding site on the lethal and edema factors." Journal of biological chemistry **277**(4): 3006-3010.
Lacy, D. B., W. Tepp, A. C. Cohen, B. R. DasGupta and R. C. Stevens (1998). "Crystal structure of botulinum neurotoxin type A and implications for toxicity." Nature structural biology **5**(10): 898-902.
Lacy, D. B., D. J. Wigelsworth, R. A. Melnyk, S. C. Harrison and R. J. Collier (2004). "Structure of heptameric protective antigen bound to an anthrax toxin receptor: a role for receptor in pH-dependent pore formation." Proceedings of the National Academy of Sciences of the United States of America **101**(36): 13147-13151.
Ladokhin, A. S., S. Jayasinghe and S. H. White (2000). "How to measure and analyze tryptophan fluorescence in membranes properly, and why bother?" Analytical biochemistry **285**(2): 235-245.
Ladokhin, A. S., M. Vargas-Uribe, M. V. Rodnin, C. Ghatak and O. Sharma (2017). "Cellular entry of the diphtheria toxin does not require the formation of the open-channel state by its translocation domain." Toxins **9**(10): 299.
Lea, W. A., P. T. O'Neil, A. J. Machen, S. Naik, T. Chaudhri, W. McGinn-Straub, A. Tischer, M. T. Auton, J. R. Burns and M. R. Baldwin (2016). "Chaperonin-based biolayer interferometry to assess the kinetic stability of metastable, aggregation-prone proteins." Biochemistry **55**(35): 4885-4908.
Liu, S. and S. H. Leppla (2003). "Cell surface tumor endothelium marker 8 cytoplasmic tail-independent anthrax toxin binding, proteolytic processing, oligomer formation, and internalization." Journal of Biological Chemistry **278**(7): 5227-5234.
Lu, Z., T. R. Shaikh, D. Barnard, X. Meng, H. Mohamed, A. Yassin, C. A. Mannella, R. K. Agrawal, T.-M. Lu and T. Wagenknecht (2009). "Monolithic microfluidic mixing–spraying

devices for time-resolved cryo-electron microscopy." Journal of structural biology **168**(3): 388-395.
M Beierlein, J. and A. C Anderson (2011). "New developments in vaccines, inhibitors of anthrax toxins, and antibiotic therapeutics for Bacillus anthracis." Current medicinal chemistry **18**(33): 5083-5094.
Ma, P., A. E. Cardenas, M. I. Chaudhari, R. Elber and S. B. Rempe (2017). "The impact of protonation on early translocation of anthrax lethal factor: Kinetics from molecular dynamics simulations and milestoning theory." Journal of the American Chemical Society **139**(42): 14837-14840.
Machen, A., N. Akkaladevi, C. Trecazzi, P. O'Neil, S. Mukherjee, Y. Qi, R. Dillard, W. Im, E. Gogol and T. White (2017). "Asymmetric cryo-EM structure of anthrax toxin protective antigen pore with lethal factor N-terminal domain." Toxins **9**(10): 298.
Machha, V. R., A. Tischer, L. Moon-Tasson and M. Auton (2017). "The von willebrand factor a1–collagen III interaction is independent of conformation and type 2 von willebrand disease phenotype." Journal of molecular biology **429**(1): 32-47.
Maldonado-Arocho, F. J., J. A. Fulcher, B. Lee and K. A. Bradley (2006). "Anthrax oedema toxin induces anthrax toxin receptor expression in monocyte-derived cells." Molecular microbiology **61**(2): 324-337.
Martin, G. J. and A. M. Friedlander (2010). "Bacillus anthracis (anthrax)." Mandell, Douglas, and Bennett's principles and practice of infectious diseases. 7th ed. Philadelphia: Churchill Livingstone: 2715-2725.
Mastronarde, D. N. (2005). "Automated electron microscope tomography using robust prediction of specimen movements." Journal of structural biology **152**(1): 36-51.
Masuyer, G., J. Conrad and P. Stenmark (2017). "The structure of the tetanus toxin reveals pH-mediated domain dynamics." EMBO reports **18**(8): 1306-1317.
Maxfield, F. R. and T. E. McGraw (2004). "Endocytic recycling." Nature reviews Molecular cell biology **5**(2): 121-132.
Melnyk, R. A. and R. J. Collier (2006). "A loop network within the anthrax toxin pore positions the phenylalanine clamp in an active conformation." Proceedings of the National Academy of Sciences **103**(26): 9802-9807.
Merritt, E. A., S. Sarfaty, M. Pizza, M. Domenighini, R. Rappuoli and W. G. Hol (1995). "Mutation of a buried residue causes loss of activity but no conformational change in the heat-labile enterotoxin of Escherichia coli." Nature structural biology **2**(4): 269-272.
Meselson, M., J. Guillemin, M. Hugh-Jones, A. Langmuir, I. Popova, A. Shelokov and O. Yampolskaya (1994). "The Sverdlovsk anthrax outbreak of 1979." Science **266**(5188): 1202-1208.
Miller, C. J., J. L. Elliott and R. J. Collier (1999). "Anthrax protective antigen: prepore-to-pore conversion." Biochemistry **38**(32): 10432-10441.
Mogridge, J., K. Cunningham and R. J. Collier (2002). "Stoichiometry of anthrax toxin complexes." Biochemistry **41**(3): 1079-1082.
Moore, J. D., M. A. Perez-Pardo, J. F. Popplewell, S. J. Spencer, S. Ray, M. J. Swann, A. G. Shard, W. Jones, A. Hills and D. G. Bracewell (2011). "Chemical and biological characterisation of a sensor surface for bioprocess monitoring." Biosens Bioelectron **26**(6): 2940-2947.
Mourez, M., M. Yan, D. B. Lacy, L. Dillon, L. Bentsen, A. Marpoe, C. Maurin, E. Hotze, D. Wigelsworth and R.-A. Pimental (2003). "Mapping dominant-negative mutations of anthrax

protective antigen by scanning mutagenesis." Proceedings of the National Academy of Sciences **100**(24): 13803-13808.
Naik, S., S. Brock, N. Akkaladevi, J. Tally, W. Mcginn-Straub, N. Zhang, P. Gao, E. Gogol, B. Pentelute and R. J. Collier (2013). "Monitoring the kinetics of the pH-driven transition of the anthrax toxin prepore to the pore by biolayer interferometry and surface plasmon resonance." Biochemistry **52**(37): 6335-6347.
Naik, S., S. Brock, N. Akkaladevi, J. Tally, W. McGinn-Straub, N. Zhang, P. Gao, E. P. Gogol, B. L. Pentelute, R. J. Collier and M. T. Fisher (2013). "Monitoring the Kinetics of the pH-Driven Transition of the Anthrax Toxin Prepore to the Pore by Biolayer Interferometry and Surface Plasmon Resonance." Biochemistry.
Naik, S., O. S. Kumru, M. Cullom, S. N. Telikepalli, E. Lindboe, T. L. Roop, S. B. Joshi, D. Amin, P. Gao and C. R. Middaugh (2014). "Probing structurally altered and aggregated states of therapeutically relevant proteins using GroEL coupled to bio-layer interferometry." Protein Science **23**(10): 1461-1478.
Neumeyer, T., B. Schiffler, E. Maier, A. E. Lang, K. Aktories and R. Benz (2008). "Clostridium botulinum C2 Toxin IDENTIFICATION OF THE BINDING SITE FOR CHLOROQUINE AND RELATED COMPOUNDS AND INFLUENCE OF THE BINDING SITE ON PROPERTIES OF THE C2II CHANNEL." Journal of Biological Chemistry **283**(7): 3904-3914.
Nguyen, T. H. D., W. P. Galej, X.-c. Bai, C. Oubridge, A. J. Newman, S. H. Scheres and K. Nagai (2016). "Cryo-EM structure of the yeast U4/U6. U5 tri-snRNP at 3.7 Å resolution." Nature **530**(7590): 298-302.
Nogales, E. and S. H. Scheres (2015). "Cryo-EM: a unique tool for the visualization of macromolecular complexity." Molecular cell **58**(4): 677-689.
O'Neil, P. T., A. J. Machen, B. C. Deatherage, C. Trecazzi, A. Tischer, V. R. Machha, M. T. Auton, M. R. Baldwin, T. A. White and M. T. Fisher (2018). "The chaperonin GroEL: a versatile tool for applied biotechnology platforms." Frontiers in Molecular Biosciences **5**: 46.
Palazzo, G., F. Lopez and A. Mallardi (2010). "Effect of detergent concentration on the thermal stability of a membrane protein: The case study of bacterial reaction center solubilized by N, N-dimethyldodecylamine-N-oxide." Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics **1804**(1): 137-146.
Pannifer, A. D., T. Y. Wong, R. Schwarzenbacher, M. Renatus, C. Petosa, J. Bienkowska, D. B. Lacy, R. J. Collier, S. Park and S. H. Leppla (2001). "Crystal structure of the anthrax lethal factor." Nature **414**(6860): 229-233.
Papatheodorou, P., C. Wilczek, T. Nölke, G. Guttenberg, D. Hornuss, C. Schwan and K. Aktories (2012). "Identification of the cellular receptor of Clostridium spiroforme toxin." Infection and immunity **80**(4): 1418-1423.
Park, J. M., F. R. Greten, Z.-W. Li and M. Karin (2002). "Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition." Science **297**(5589): 2048-2051.
Patargias, G., P. J. Bond, S. S. Deol and M. S. Sansom (2005). "Molecular dynamics simulations of GlpF in a micelle vs in a bilayer: conformational dynamics of a membrane protein as a function of environment." The Journal of Physical Chemistry B **109**(1): 575-582.
Pentelute, B. L., O. Sharma and R. J. Collier (2011). "Chemical dissection of protein translocation through the anthrax toxin pore." Angewandte Chemie International Edition **50**(10): 2294-2296.
Petosa, C., R. J. Collier, K. R. Klimpel, S. H. Leppla and R. C. Liddington (1997). "Crystal structure of the anthrax toxin protective antigen." Nature **385**(6619): 833-838.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera—a visualization system for exploratory research and analysis." Journal of computational chemistry **25**(13): 1605-1612.
Pimental, R.-A. L., K. A. Christensen, B. A. Krantz and R. J. Collier (2004). "Anthrax toxin complexes: heptameric protective antigen can bind lethal factor and edema factor simultaneously." Biochemical and biophysical research communications **322**(1): 258-262.
Piper, S. J., L. Brillault, R. Rothnagel, T. I. Croll, J. K. Box, I. Chassagnon, S. Scherer, K. N. Goldie, S. A. Jones and F. Schepers (2019). "Cryo-EM structures of the pore-forming A subunit from the Yersinia entomophaga ABC toxin." Nature communications **10**(1): 1952.
Punjani, A. and D. J. Fleet (2020). "3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images." bioRxiv.
Punjani, A., J. L. Rubinstein, D. J. Fleet and M. A. Brubaker (2017). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nature Methods **14**(3): 290-296.
Qi, Y., J. Lee, A. Singharoy, R. McGreevy, K. Schulten and W. Im (2016). "CHARMM-GUI MDFF/xMDFF Utilizer for Molecular Dynamics Flexible Fitting Simulations in Various Environments." The Journal of Physical Chemistry B **121**(15): 3718-3723.
Rainey, G. J. A., D. J. Wigelsworth, P. L. Ryan, H. M. Scobie, R. J. Collier and J. A. Young (2005). "Receptor-specific requirements for anthrax toxin delivery into cells." Proceedings of the National Academy of Sciences **102**(37): 13278-13283.
Ravelli, R. B., F. J. Nijpels, R. J. Henderikx, G. Weissenberger, S. Thewessem, A. Gijsbers, B. W. Beulen, C. Lopez-Iglesias and P. Peters (2019). "Automated cryo-EM sample preparation by pin-printing and jet vitrification." bioRxiv: 651208.
Ripstein, Z. A., R. Huang, R. Augustyniak, L. E. Kay and J. L. Rubinstein (2017). "Structure of a AAA+ unfoldase in the process of unfolding substrate." Elife **6**: e25754.
Ritchie, T., Y. Grinkova, T. Bayburt, I. Denisov, J. Zolnerciks, W. Atkins and S. Sligar (2009). "Chapter eleven-reconstitution of membrane proteins in phospholipid bilayer nanodiscs." Methods in enzymology **464**: 211-231.
Rmali, K., M. Puntis and W. Jiang (2005). "TEM-8 and tubule formation in endothelial cells, its potential role of its vW/TM domains." Biochemical and biophysical research communications **334**(1): 231-238.
Russo, C. J. and L. A. Passmore (2014). "Ultrastable gold substrates for electron cryomicroscopy." Science **346**(6215): 1377-1380.
Saibil, H. R. and N. A. Ranson (2002). "The chaperonin folding machine." Trends in biochemical sciences **27**(12): 627-632.
Santelli, E., L. A. Bankston, S. H. Leppia and R. C. Liddington (2004). "Crystal structure of a complex between anthrax toxin and its host cell receptor." Nature **430**(7002): 905.
Scheres, S. H. (2016). Processing of structurally heterogeneous cryo-EM data in RELION. Methods in enzymology, Elsevier. **579:** 125-157.
Schleberger, C., H. Hochmann, H. Barth, K. Aktories and G. E. Schulz (2006). "Structure and action of the binary C2 toxin from Clostridium botulinum." Journal of molecular biology **364**(4): 705-715.
Schuler, M. A., I. G. Denisov and S. G. Sligar (2013). Nanodiscs as a new tool to examine lipid–protein interactions. Lipid-Protein Interactions, Springer**:** 415-433.
Scobie, H. M., J. M. Marlett, G. J. A. Rainey, D. B. Lacy, R. J. Collier and J. A. Young (2007). "Anthrax toxin receptor 2 determinants that dictate the pH threshold of toxin pore formation." PloS one **2**(3).

Scobie, H. M., G. J. A. Rainey, K. A. Bradley and J. A. Young (2003). "Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor." Proceedings of the National Academy of Sciences **100**(9): 5170-5174.
Scobie, H. M., D. Thomas, J. M. Marlett, G. Destito, D. J. Wigelsworth, R. J. Collier, J. A. Young and M. Manchester (2005). "A soluble receptor decoy protects rats against anthrax lethal toxin challenge." The Journal of infectious diseases **192**(6): 1047-1051.
Scott III, H. W. (2018). USING ELECTRON MICROSCOPY TO GAIN STRUCTURAL INSIGHT INTO BIOLOGICALLY RELEVANT, LABILE OR DESTABILIZED PROTEIN COMPLEXES, Case Western Reserve University.
Sellman, B. R., M. Mourez and R. J. Collier (2001). "Dominant-negative mutants of a toxin subunit: an approach to therapy of anthrax." Science **292**(5517): 695-697.
Sellman, B. R., S. Nassi and R. J. Collier (2001). "Point mutations in anthrax protective antigen that block translocation." Journal of biological chemistry **276**(11): 8371-8376.
Serna, M. (2019). "Hands on methods for high resolution cryo-Electron Microscopy structures of heterogeneous macromolecular complexes." Frontiers in molecular biosciences **6**: 33.
Shashkova, S. and M. C. Leake (2017). "Single-molecule fluorescence microscopy review: shedding new light on old problems." Bioscience reports **37**(4): BSR20170031.
Shen, P. S., X. Yang, P. G. DeCaen, X. Liu, D. Bulkley, D. E. Clapham and E. Cao (2016). "The structure of the polycystic kidney disease channel PKD2 in lipid nanodiscs." Cell **167**(3): 763-773. e711.
Stein, P. E., A. Boodhoo, G. D. Armstrong, S. A. Cockle, M. H. Klein and R. J. Read (1994). "The crystal structure of pertussis toxin." Structure **2**(1): 45-57.
Sugawara, T., D. Yamashita, K. Kato, Z. Peng, J. Ueda, J. Kaneko, Y. Kamio, Y. Tanaka and M. Yao (2015). "Structural basis for pore-forming mechanism of staphylococcal α-hemolysin." Toxicon **108**: 226-231.
Sun, J., A. E. Lang, K. Aktories and R. J. Collier (2008). "Phenylalanine-427 of anthrax protective antigen functions in both pore formation and protein translocation." Proceedings of the National Academy of Sciences **105**(11): 4346-4351.
Tachibana, C. (2020). Democratizing cryo-EM: Broadening access to an expanding field, AMER ASSOC ADVANCEMENT SCIENCE 1200 NEW YORK AVE, NW, WASHINGTON, DC 20005 USA.
Tang, G., L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees and S. J. Ludtke (2007). "EMAN2: an extensible image processing suite for electron microscopy." Journal of structural biology **157**(1): 38-46.
Thoren, K. L. and B. A. Krantz (2011). "The unfolding story of anthrax toxin translocation." Molecular microbiology **80**(3): 588-595.
Timofeev, V., I. Bahtejeva, R. Mironova, G. Titareva, I. Lev, D. Christiany, A. Borzilov, A. Bogun and G. Vergnaud (2019). "Insights from Bacillus anthracis strains isolated from permafrost in the tundra zone of Russia." PloS one **14**(5).
Tischer, A., P. Madde, L. Moon-Tasson and M. Auton (2014). "Misfolding of vWF to pathologically disordered conformations impacts the severity of von Willebrand disease." Biophysical journal **107**(5): 1185-1195.
Trabuco, L. G., E. Villa, K. Mitra, J. Frank and K. Schulten (2008). "Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics." Structure **16**(5): 673-683.

Tsuge, H., M. Nagahama, H. Nishimura, J. Hisatsune, Y. Sakaguchi, Y. Itogawa, N. Katunuma and J. Sakurai (2003). "Crystal structure and site-directed mutagenesis of enzymatic components from Clostridium perfringens iota-toxin." Journal of molecular biology **325**(3): 471-483.
Turnbull, P. C., J. Kramer and J. Melling (1991). "Bacillus." Manual of clinical microbiology **5**: 296-303.
Voziyan, P. A. and M. T. Fisher (2000). "Chaperonin-assisted folding of glutamine synthetase under nonpermissive conditions: off-pathway aggregation propensity does not determine the co-chaperonin requirement." Protein Science **9**(12): 2405-2412.
Wahle, M. and W. Wriggers (2015). "Multi-scale visualization of molecular architecture using real-time ambient occlusion in sculptor." PLoS computational biology **11**(10): e1004516.
Wang, J. Y. and M. H. Roehrl (2005). "Anthrax vaccine design: strategies to achieve comprehensive protection against spore, bacillus, and toxin." Medical Immunology **4**(1): 4.
Weaver, J., M. Jiang, A. Roth, J. Puchalla, J. Zhang and H. S. Rye (2017). "GroEL actively stimulates folding of the endogenous substrate protein PepQ." Nature Communications **8**.
Wedekind, J. E., C. B. Trame, M. Dorywalska, P. Koehl, T. M. Raschke, M. McKee, D. FitzGerald, R. J. Collier and D. B. McKay (2001). "Refined crystallographic structure of Pseudomonas aeruginosa exotoxin A and its implications for the molecular mechanism of toxicity." Journal of molecular biology **314**(4): 823-837.
Wei, H., V. P. Dandey, Z. Zhang, A. Raczkowski, W. J. Rice, B. Carragher and C. S. Potter (2018). "Optimizing "self-wicking" nanowire grids." Journal of structural biology **202**(2): 170-174.
Wigelsworth, D. J., B. A. Krantz, K. A. Christensen, D. B. Lacy, S. J. Juris and R. J. Collier (2004). "Binding stoichiometry and kinetics of the interaction of a human anthrax toxin receptor, CMG2, with protective antigen." Journal of Biological Chemistry **279**(22): 23349-23356.
Wilson, D. N. and R. Beckmann (2011). "The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling." Current opinion in structural biology **21**(2): 274-282.
Wimalasena, D. S., B. E. Janowiak, S. Lovell, M. Miyagi, J. Sun, H. Zhou, J. Hajduch, C. Pooput, K. L. Kirk and K. P. Battaile (2010). "Evidence that histidine protonation of receptor-bound anthrax protective antigen is a trigger for pore formation." Biochemistry **49**(33): 6973-6983.
Wolfe, J. T., B. A. Krantz, G. J. A. Rainey, J. A. Young and R. J. Collier (2005). "Whole-cell voltage clamp measurements of anthrax toxin pore current." Journal of biological chemistry **280**(47): 39417-39422.
Wrapp, D., N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham and J. S. McLellan (2020). "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation." Science **367**(6483): 1260-1263.
Wynia-Smith, S. L., M. J. Brown, G. Chirichella, G. Kemalyan and B. A. Krantz (2012). "Electrostatic ratchet in the protective antigen channel promotes anthrax toxin translocation." Journal of biological chemistry **287**(52): 43753-43764.
Yamada, T., T. Yoshida, A. Kawamoto, K. Mitsuoka, K. Iwasaki and H. Tsuge (2020). "Cryo-EM structures reveal translocational unfolding in the clostridial binary iota toxin complex." Nature Structural & Molecular Biology **27**(3): 288-296.
Yamniuk, A. P., S. C. Edavettal, S. Bergqvist, S. P. Yadav, M. L. Doyle, K. Calabrese, J. F. Parsons and E. Eisenstein (2012). "ABRF-MIRG benchmark study: molecular interactions in a three-component system." J Biomol Tech **23**(3): 101-114.

Yang, Z., J. Fang, J. Chittuluru, F. J. Asturias and P. A. Penczek (2012). "Iterative stable alignment and clustering of 2D transmission electron microscope images." Structure **20**(2): 237-247.
Young, J. A. and R. J. Collier (2007). "Anthrax toxin: receptor binding, internalization, pore formation, and translocation." Annu. Rev. Biochem. **76**: 243-265.
Zhang, F., M. Hu, G. Tian, P. Zhang, D. Finley, P. D. Jeffrey and Y. Shi (2009). "Structural insights into the regulatory particle of the proteasome from Methanocaldococcus jannaschii." Molecular cell **34**(4): 473-484.
Zhang, R.-G., D. L. Scott, M. L. Westbrook, S. Nance, B. D. Spangler, G. G. Shipley and E. M. Westbrook (1995). "The three-dimensional crystal structure of cholera toxin." Journal of molecular biology **251**(4): 563-573.
Zhang, S., A. Finkelstein and R. J. Collier (2004). "Evidence that translocation of anthrax toxin's lethal factor is initiated by entry of its N terminus into the protective antigen channel." Proceedings of the National Academy of Sciences **101**(48): 16756-16761.
Zhong, E. D., T. Bepler, B. Berger and J. H. Davis (2020). "CryoDRGN: Reconstruction of heterogeneous structures from cryo-electron micrographs using neural networks." bioRxiv.
Zhou, M., Y. Li, Q. Hu, X.-c. Bai, W. Huang, C. Yan, S. H. Scheres and Y. Shi (2015). "Atomic structure of the apoptosome: mechanism of cytochrome c-and dATP-mediated activation of Apaf-1." Genes & development **29**(22): 2349-2361.
Zimmermann, M. T., A. Tischer, S. T. Whitten and M. Auton (2015). "Structural origins of misfolding propensity in the platelet adhesive von Willebrand factor A1 domain." Biophysical journal **109**(2): 398-406.
Zivanov, J., T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl and S. H. Scheres (2018). "New tools for automated high-resolution cryo-EM structure determination in RELION-3." Elife **7**: e42166.

Appendix A: Lipid Bilayer Nanodisc Stability under Neutral and Acidic Conditions

INTRODUCTION

Nanodiscs, discoidal lipid bilayers stabilized by two amphipathic helical membrane scaffold proteins (MSP), are a highly useful model system for biophysical and structural studies of membrane associated and membrane inserted proteins (Denisov, Grinkova et al. 2004, Bayburt and Sligar 2010). Nanodisc size is adjustable with different combinations of apolipoprotein A-I derived helical segments resulting in MSPs of varying lengths (Schuler, Denisov et al. 2013). MSP1D1, H1(1-11)-H2-H3-H4-H5-H6-H7-H8-H9-H10, forms ~ 10 nm diameter nanodiscs (Denisov, Grinkova et al. 2004). There are two tryptophan residues in the MSP1D1 peptide (H2 and H4) which hinder internal fluorescence measurements often used to characterize membrane proteins (Ladokhin, Jayasinghe et al. 2000). Removal of the MSP tryptophan residues would allow for these biophysical studies to be carried out in nanodiscs. We present here microscopic assessment of nanodiscs with wild-type MSP1D1 and mutant-type MSP1D1 (2W/2Y). Our results show tryptophan-less MSP is capable of forming discoidal nanodiscs, expanding the biophysical repertoire of these membrane mimetics.

RESULTS

Wild-type MSP forms functional nanodisc under acidic conditions

During the course of this work, aggregation of protein-nanodiscs complexes at low pH was a common challenge. The individual particles needed for structural studies could not be seen by negative stain electron microscopy due to this aggregation. To ensure aggregation was not the result of nanodisc instability at low pH, nanodisc formation in the absence of protein was carried out at low pH and compared to neutral pH nanodiscs. Nanodisc formation was confirmed by negative stain electron microscopy (**Figure A.1**). We conclude micelle to nanodisc conversion is

possible under acidic conditions and formed nanodisc are stable at low pH. We attributed aggregation of protein-nanodisc complexes in our work to issues with the protein component of the complex.

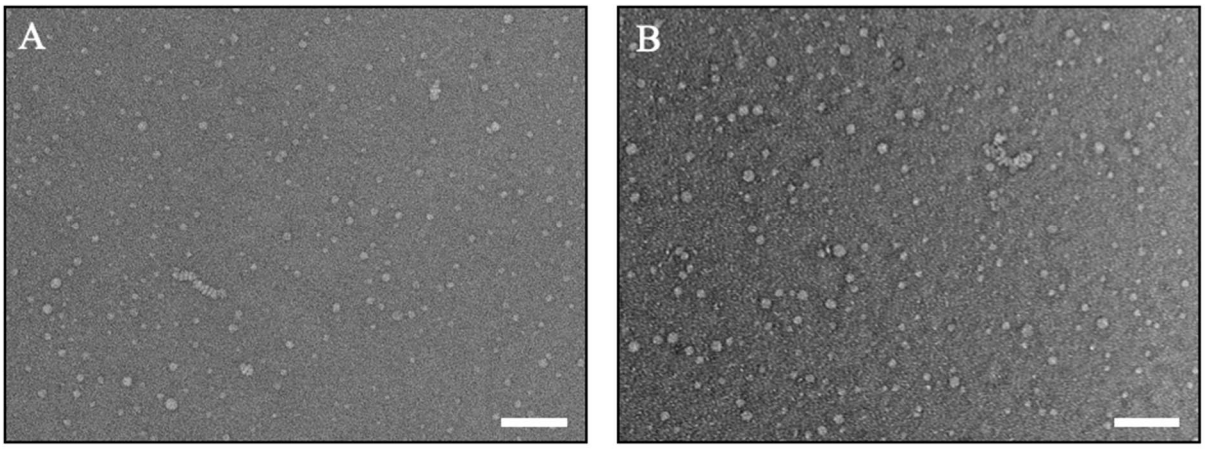


Figure A.1. Negative stain micrographs of wild-type MSP1D1 nanodiscs at neutral and acidic pH. POPC nanodiscs formed at (A) pH 7.5 and (B) pH 5.5 with discoidal size and homogeneity confirmed by electron microscopy. Scale bar: 100 nm.

Tryptophan-less MSP forms functional nanodisc under neutral and acidic conditions

We visually confirm mutant nanodisc formation at pH 8 with negative stain EM. (**Figure A.3.A**). To ensure mutant MSP nanodiscs were amenable lipid composition and pH studies, we also formed nanodiscs with a lipid ratio of PC75:PG25 (**Figure A.3.B**) and repeated the experiments at pH 5.0 (**Figure A.3.C-D**). Our results show mutant MSP is functional when interacting with zwitterionic or charged lipids at neutral and acidic pH.

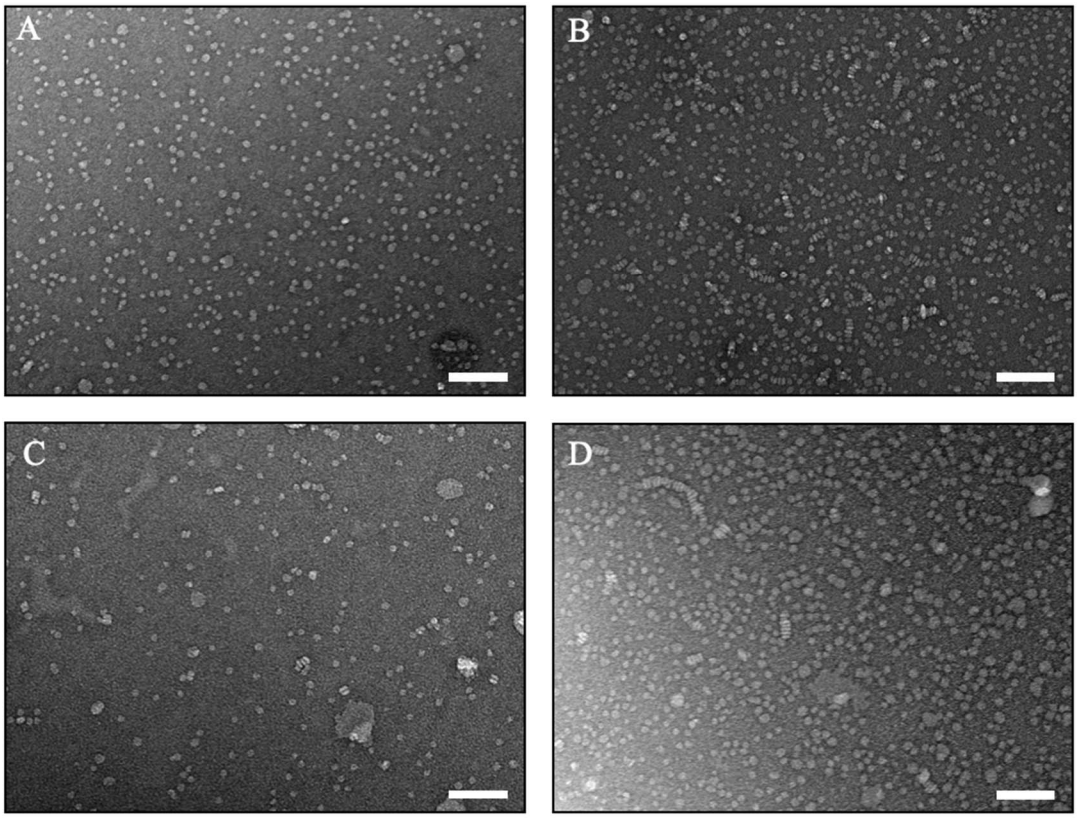


Figure A.2. Negative stain EM micrographs confirm mutant-type MSP1D1 nanodisc formation at neutral and acidic pH. (A) MSP1D1 2W/2Y, POPC nanodiscs at pH 8. (B) MSP1D1 2W/2, PC75:PG25 nanodiscs at pH 8. (C) MSP1D1 2W/2Y, POPC nanodiscs at pH 5. (D) MSP1D1 2W/2Y, PC75:PG25 nanodiscs at pH 5. Scale bar:100 nm

METHODS

Materials

Wild-type membrane scaffold protein 1D1 (MSP1D1) was expressed from the pMSP1D1 plasmid (AddGene) with an N-terminal His-tag and was purified by immobilized Ni-NTA affinity chromatography as previously described (Ritchie, Grinkova et al. 2009). MSP1D1 (H1(1-11)-H2-H3-H4-H5-H6-H7-H8-H9-H10) contains two tryptophans: STFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEM ELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDEL RQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSF LSALEEYTKKLNTQ.

Mutant type MSP1D1 was developed by the Ladohkin lab using standard molecular biology protocols. 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG) were purchased from Avanti (Alabaster, AL, USA). Uranyl formate was purchased from Electron Microscopy Sciences (22451). Sodium cholate was purchased from Sigma-Aldrich, (St. Louis, MO, USA). Bio-Beads were purchased from BIO RAD (Hercules, CA, USA).

Nanodisc Assembly

Nanodiscs were assembled using standard protocol as previously described (Denisov, Grinkova et al. 2004). Briefly, stock solutions of PC and PC75:PG25 in chloroform were evaporated under nitrogen gas. Dried lipids were solubilized in sodium phosphate buffer. MSPs were reconstituted in sodium phosphate buffer with 25 mM Na-cholate and mixed with the lipid solution to a final pre-nanodisc micelle mix of 2.5 μM MSP1D1, 162.5 μM POPC, 25mM Na-cholate in phosphate buffer at desired pH. Pre-nanodisc micelles were incubated at 4°C for 10 minutes followed by 4 hrs of dialysis with bio beads to remove Na-cholate and form lipid bilayer.

Negative Stain Electron Microscopy

Nanodisc samples were diluted to 10 nM and 4 μL of diluted sample were deposited onto a 200 mesh carbon-coated copper (Electron Microscopy Sciences CF200-Cu) glow discharged (20 s at −15 mA in 39 mBar atmosphere) grid. After one minute, excess sample was wicked away with P8 Grade filter paper. Grids were washed three times in ultrapure water. Then, grids were stained for 5 seconds using 5 μL 0.75% w/v uranyl formate, 0.022 μm filtered, followed by wicking of excess stain with filter paper. Stained grids were imaged using JEOL JEM 1400 transmission electron microscope (Peabody, MA, USA) operated at 100 keV at magnification of 30-50,000x.

Appendix B: Kinetic Stability of Wild- and Mutant-Type von Willebrand Factor

This appendix is adapted from the works cited below and is reprinted with permission.

Lea, W.A., O'Neil, P.T., Machen, A.J., Naik, S., Chaudhri, T., McGinn-Straub, W., Tischer, A., Auton, M.T., Burns, J.R., Baldwin, M.R. and Khar, K.R., 2016. Chaperonin-based biolayer interferometry to assess the kinetic stability of metastable, aggregation-prone proteins. Biochemistry, 55(35), pp.4885-4908.

O'Neil, P.T., Machen, A.J., Deatherage, B.C., Trecazzi, C., Tischer, A., Machha, V.R., Auton, M.T., Baldwin, M.R., White, T.A. and Fisher, M.T., 2018. The chaperonin GroEL: a versatile tool for applied biotechnology platforms. Frontiers in Molecular Biosciences, 5, p.46.

INTRODUCTION

von Willebrand Factor (vWF)

vWF is a multimeric plasma glycoprotein which initiates platelet adhesion at sites of vascular injury (Andrews, López et al. 1997). Under high shear stress, vWF unravels and binds to platelets and collagen to create a plug to stop bleeding. von Willebrand Disease (vWD) is a bleeding disorder affecting approximately 1% of the world population (Favaloro 2011). This hereditary disease is caused by mutations that cause quantitative deficiencies of vWF or qualitatively alter vWF function (Keeney and Cumming 2001). Mutations in A1 of the triple A domain of vWF alter its specificity for the platelet receptor GP1bα. Mutations in A3 can affect its collagen binding affinity. Finally, mutations in A2 cause defective intracellular transport or enhance proteolysis of a scissile bond recognized by the soluble blood metalloprotease (ADAMTS13), which helps regulate the multimeric size of vWF (Keeney and Cumming 2001). Some vWD mutations that change A1-GP1bα binding specificity result in local misfolding of the A1 domain (Tischer, Madde et al. 2014, Zimmermann, Tischer et al. 2015, Machha, Tischer et al. 2017) causing both gain and loss of function phenotypes. Two such mutations are V1314D, a gain of function mutation that causes increased platelet adhesion, and F1369I, a loss of function mutation that does not adhere to platelets at all. The GroEL-based BLI denaturant pulse assay was used to assess the kinetic stability of vWF A1-A2-A3 for both wild-type and partially disordered vWD point mutants.

Denaturant Pulse Assay

The kinetic stability of aggregation-prone proteins can be determined using a unique chaperonin dependent denaturant pulse assay. This technique assesses the stability of proteins immobilized on BLI biosensor surfaces after a time-controlled pulse in various denaturant solutions. GroEL

binding to hydrophobic patches on the unfolded protein amplifies the unfolded protein signal. To illustrate the expanding utility of this method, denaturation isotherms of wild- and mutant-type von Willebrand Factor (vWF) triple A domain were collected and compared. Many missense mutants tend to aggregate in solution during conventional stability analysis. Therefore, the immobilization of missense mutants prior to performing the denaturant pulse assay avoids this common difficulty.

Release of GroEL-vWF Complexes From Ni-NTA Biosensors

BLI output is in nanometer shift in the interference spectrum and therefor does not carry structural data. it requires the assumption that the expected protein complexes are being constructed on the biosensor surface. To validate complex assembly, the purported GroEL-protein complexes can be orthogonally confirmed using negative-stain EM. Using this technique, the GroEL-protein complexes can be directly visualized. For other techniques that reveal high and low resolution election density envelops of protein complexes such as crystallography or small angle X-ray scattering respectively, it is ideal that the sample consists of homogeneous complexes. If the target protein is inherently varied with respect to conformational heterogeneity (multiple conformations), analyzing the structural outputs by these methods becomes problematic. With EM analysis, each complex can be examined individually and therefore the binding heterogeneity can potentially be revealed in each case, particularly if random tilt series methods are applied. While mass spectroscopy can also identify target protein composition upon ATP induced release from the GroEL chaperonin, the maintenance of solubility during EM analysis is critical. Including the natural anti-aggregation chaperonin protein allows one to obtain low resolution structures of both free and bound substrate protein chaperonin complexes. Direct

EM visualization allows one to broadly identify which folded yet hydrophobic region(s) of the vWF interact with the GroEL chaperonin at its promiscuous protein substrate binding site.

RESULTS

Kinetically controlled isotherms from Denaturant Pulse Assay

We assessed the kinetic stability of the vWF triple A domain using chaperonin BLI denaturation pulse procedures. The vWF A3 domain contains a His tag that can be attached to the Ni-NTA biosensor surface and mimics the attachment that naturally occurs when the triple A domain protein binds collagen. This same attachment scheme was used to attach the His-tagged triple A domain onto Cu^{2+}-coated surfaces for rheodynamic analysis of platelet adherence and pause time measurements (Tischer, Madde et al. 2014). A sample complete run for von Willebrand Factor is represented in **Figure B.1**.

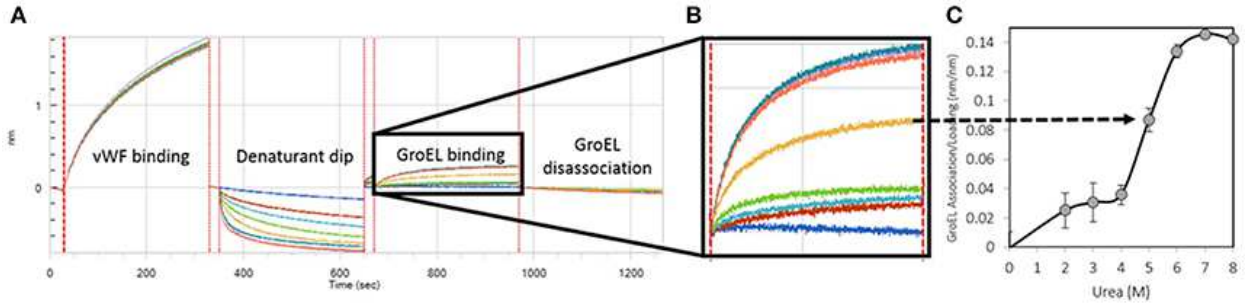


Figure B.1. Generation of a Kinetically Controlled Denaturant Pulse Isotherm for the Wild-type von Willebrand Factor A1-A2-A3 Triple Domain Fragment. (A) Protein is partially denatured during automated urea pulse steps, and (B) the GroEL binding amplitude was (C) plotted as a function of urea concentration to generate a kinetically controlled denaturation isotherm

The GroEL binding signal, plotted as a function of the increasing denaturant pulse concentration, increased with an increasing denaturant concentration and resulted in a kinetically controlled denaturation isotherm. BLI runs were performed in triplicate with tip regeneration performed

after runs 1 and 2 only. The GroEL binding signal was plotted as a function of denaturant concentration to create kinetically controlled denaturation isotherms for wild-type, gain of function (V1314D), and loss of function (F1369I) mutant vWF.

Comparison of vWF Kinetically Controlled Denaturation Profile to Equilibrium Denaturation

Figure B.2 shows the denaturation profile for WT vWF with increasing denaturant pulse time. As the denaturation pulse times increase, the kinetically controlled denaturation isotherm derived from the GroEL binding amplitude shifts toward the circular dichroism (CD)-derived equilibrium isotherms (Auton, Cruz et al. 2007). The equilibrium profiles show transitions between 2 and 3 M urea known to be associated with the A1 and A2 domains unfolding and a transition at ~5M urea that is attributed to domain A3 unfolding (**Figure B.2.B**). The A1 and A2 unfolding events were not readily observed by the chaperonin BLI detection platform even after the longest (10 min) denaturant pulse. Only at higher denaturant concentrations (~5 M urea) does the GroEL binding amplitudes approach the equilibrium denaturation profiles that are attributed to domain A3 unfolding.

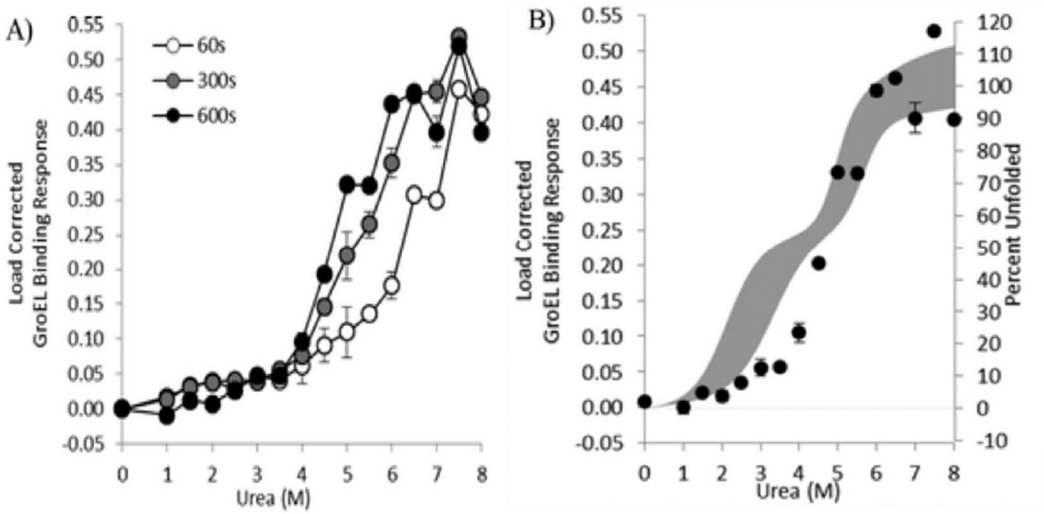


Figure B.2. Kinetic stability isotherm of von Willebrand factor triple A domain generated with the BLI denaturant pulse assay. (A) Time-dependent left shift of the load-corrected GroEL binding response as a function of denaturant concentration. The automated kinetically controlled denaturation isotherm consisted of two separate runs with three runs per data point and error bars of one standard deviation. To record points every 0.5 M urea, two separate automated runs were conducted with eight biosensors followed by regeneration. Increasing the denaturant pulse time from 1 to 10 min results in a leftward shift in the denaturation isotherms. (B) An overlay of the kinetically controlled denaturation profile generated under a 10 min kinetic denaturant pulse (●) compared with the previously generated equilibrium denaturation profile with a 95% confidence interval (gray shaded region) shows a better fit at higher denaturant concentrations (Auton, Cruz et al. 2007).

vWF Wild- & Mutant-Type Denaturant Isotherms

The kinetically controlled denaturant isotherms for WT, F1369I, and V1314D vWF triple A domains are shown in **Figure B.3**. Even under little to no denaturant conditions, The gain of function mutant V1314D had exposed hydrophobic patches as evidenced by GroEL association from 0-3M urea. The denaturant pulse profiles for wild-type and V1314D were similar at high urea concentrations. By contrast, F1369I and wild-type had similar GroEL binding at low urea concentrations, but at high denaturant conditions the F1369I exhibited significantly higher binding by GroEL.

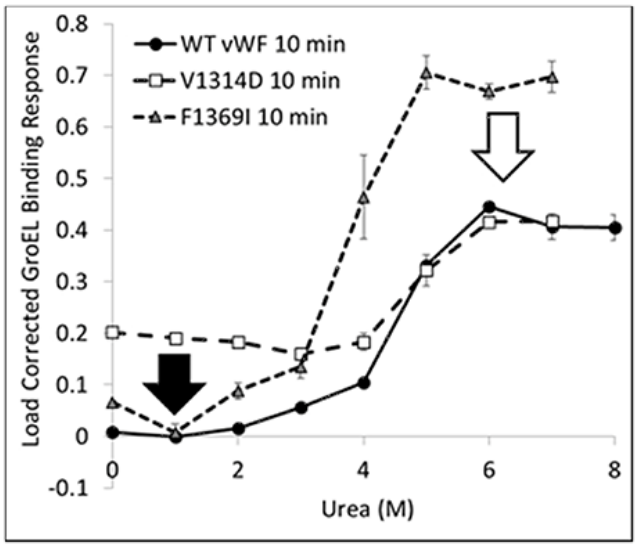


Figure B.3. vWF Kinetically Controlled Denaturant Pulse Isotherms for Wild- and Mutant-Types. Kinetic denaturation isotherms for the wild-type (circles), gain of function mutant (squares), and loss of function mutant (triangle) overlaid for comparison. The closed and open arrows denote the potential corrective ligand induced responses that may restore wild-type denaturant pulse profile and protein function.

Negative stain EM of GroEL-vWF complexes

It is possible to visually confirm of the denaturant induced GroEL association of the denaturant pulse assay by reversing the immobilization. For experiments where the protein of interest has a His_6 tag, the protein can be immobilized using Ni-NTA biosensors. The coordination between the His_6 tag and the Ni^{++} ion can be gently reversed using either imidazole competition or EDTA chelation. It is imperative to optimize the eluent concentration and elution time using the BLI. By releasing captured proteins into a microvolume drop (3–4 μL), the concentration of any released protein will be relatively high and appropriate for EM. As an example of this release and visualization, the Ni-NTA tip used for the 2 M urea pulse on V1314D was released and stained for negative-stain EM (**Figure B.4**). In this micrograph, it is possible to discern the distinct A1-A2-A3 domain extension, especially in the top view (**Figure B.4, red box**). As the His_6 tag is on the C-terminus of vWF, this orients the A3 domain closest to the biosensor surface and, therefore, sterically hindered against GroEL binding. Additionally, the previous solution

equilibrium data indicates the A1 domain unfolds first (Auton et al., 2007a). With these two facts, it is most likely the A1 domain captured by GroEL. Although it is not possible to definitively identify the interacting domain with this sample, additional experiments could identify the interacting domains. For example, the addition of an anti-A1 antibody and its distinct density on the GroEL-vWF complex would help identify the GroEL interacting domain via negative-stain EM. In this particular field, the free GroEL observed is a consequence of not washing the biosensor before release. Extensive dissociation of the GroEL from vWF is not observed.

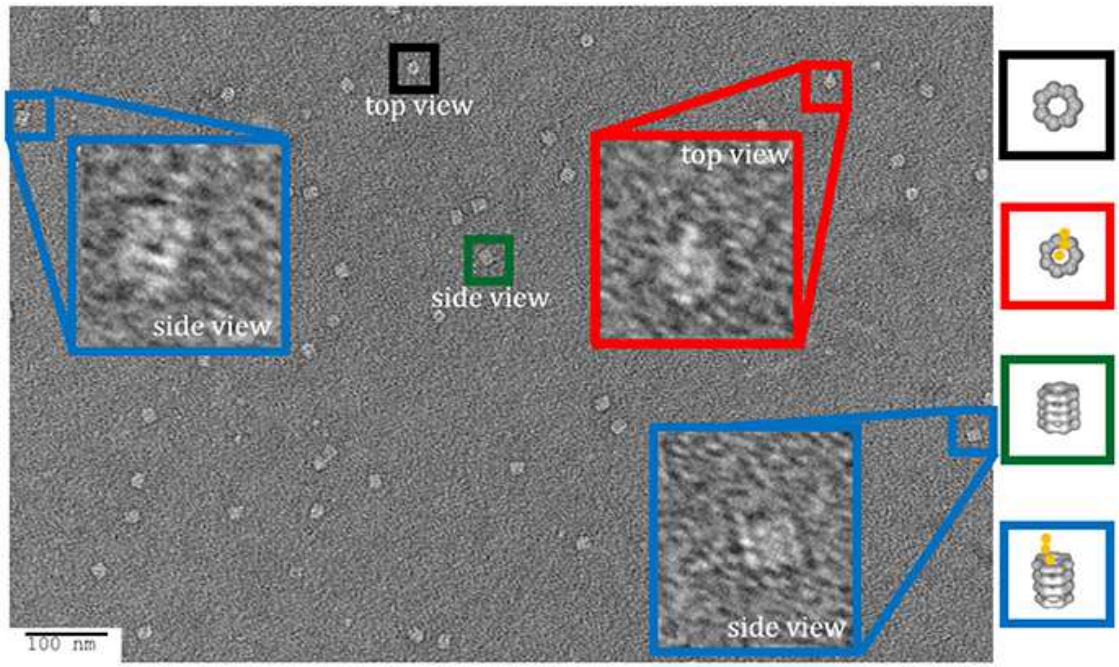


Figure B.4. Imidazole Release of GroEL-vWF Complexes from Ni-NTA Biosensor. Representative field of gridded and stained V1314D vWF-GroEL complexes formed after a 2 M urea denaturant pulse. Multiple complexes can be seen as white on dark background. Top view of the complex is boxed in red, in which it is possible to observe the three domains of vWF. Side view of the complex is boxed in blue. Non-complexed GroEL top and side views are boxed in black and green, respectively.

DISCUSSION

In equilibrium denaturation profiles using urea, the thermodynamics of vWF triple A domain unfolding has been described as three domains linked in a linear fashion in which unfolding of each domain proceeds orderly with the simultaneous unfolding of A1 and A2 at low urea followed by A3 at high urea (Auton, Holthauzen et al. 2007). Evidence suggests that in the absence of urea, the A1, A2, and A3 domains interact with each other and mutations can disrupt these interactions as a result of their intrinsic effects on the single domain (Auton, Sowa et al. 2010). It is possible that the interactions between domains observed in the wild-type protein are differentially altered depending on the structural location of a mutation (Zimmermann, Tischer et al. 2015), whether the mutation occurs at a domain interface, its intrinsic effect on thermodynamic stability of a domain (Auton, Sedlák et al. 2009), and/or its propensity for local disorder (Zimmermann, Tischer et al. 2015). These intrinsic properties of the vWF triple A domain would therefore lead to different GroEL binding efficacies thus altering the urea denaturant pulse dependence of GroEL binding.

For V1369D, the disordered structure of the A1 domain (Tischer, Madde et al. 2014) is such that GroEL is able to bind to the vWF triple A domains even with low urea concentrations. By contrast, F1369I requires a much higher urea concentration to yield extensive GroEL binding. These observations imply that structural disorder induced in the A1 domain by these mutations result in altered quaternary A1-A2-A3 domain interactions that are differentially recognized by GroEL. The structural disruption of A1 by V1314D is so severe that GroEL readily recognizes exposed hydrophobic regions without urea denaturation. Conversely, F1369I, which also misfolds the A1 domain, may cause A1-A2-A3 domain to reorganize its quaternary structure forming unnatural domain interfaces which are stabilized against urea denaturation, thereby

requiring higher urea to achieve similar levels of GroEL binding. This differential binding by GroEL depending on the mutation may be reduced by post-translational glycosylation, which normally decorates the vWF surface but are lacking when bacterially expressed.

The BLI denaturation pulse assay for the wild- and mutant-type proteins may have potential to be used as a rapid drug discovery platform. Performed with candidate small molecule stabilizers generated using in silico selection algorithms, this assay could be used to determine if the test compounds rectify the structural origins of misfolding. Any compound which returns proper folding to the mutant-type protein would return the mutant denaturation isotherm to match that for wild-type. Although these mutations presented herein are both in the A1 domain, a different stabilizing compound may be required to correct each specific mutation as the two mutants do not have the same effect on the denaturation isotherm and likely represent two different misfolding events which need to be stabilized.

The data presented herein illustrates the broad utility of using the promiscuous chaperonin to (1) capture kinetic transients, (2) distinguish various mutant-type folds, and (3) enhance structure assessment of large proteins using electron microscopy. All of these applications arose from the simple observation that the binding affinity of some folding proteins leads to folding arrest and long-term sequestration of protein substrates. It will be interesting to expand the role of chaperonin capture and release strategies to examine initial structural stages of protein aggregation, a truly elusive reaction time regime that may provide enormous benefits in understanding the molecular basis of some human protein folding diseases.

MATERIALS and METHODS

GroEL Chaperonin Purification and Storage

Highly purified GroEL was obtained using the purification scheme outlined in a previous work (Voziyan and Fisher 2000). Because GroEL does not contain tryptophan, the purity was determined by following the diminishing contributions from tryptophan (proteins or peptides) as assessed by contributions from contaminant tryptophan fluorescence (excitation at 297 nm) or by noting the indole contributions using second-derivative UV analysis (Fisher 1992). It is crucial to obtain highly pure GroEL because small amounts of contaminating proteins and peptides diminish the effectiveness of protein capture and binding. In addition, highly purified GroEL has a tendency to slowly dissociate into heptamers and monomers, where monomers can bind to remaining oligomeric GroEL, further compromising the GroEL preparation. To avoid dissociation of the GroEL tetradecamer, the purified GroEL lots are stored in 50% glycerol at 4 °C. This storage solution is removed prior to immediate GroEL use and replaced with a GroEL buffer (preferred in protein refolding assays) containing 50 mM Tris, 50 mM KCl, 10 mM MgCl2, and 0.5 mM EDTA (pH 7.5) at 25 °C. For elevated temperatures (37 °C), the Tris-HCl buffer can be replaced with 50 mM HEPES, but the pH decline is only to 7.2 for the Tris buffer.

vWF A1-A2-A3 Purification

Wild- and mutant-type vWF A1-A2-A3 tridomains with von Willebrand Disease point mutations (V1314D and F1369) engineered into the A1 domain (Tischer, Madde et al. 2014)(Tischer et al., 2014) were expressed with a C-terminal His_6 tag on the A3 domain and purified as previously described (Auton et al., 2007a). Purified protein was stored in vWF buffer (25 mM TRIS, 150 mM NaCl, pH 7.5) at 4°C and used within 2 weeks.

Denaturant Pulse Assay for vWFA1-A2-A3

A sample complete run for von Willebrand Factor is represented in **Figure B.1**. This triple A domain protein (A1-A2-A3-His_6 tag) was attached to a Ni-NTA BLI biosensor in an orientation where the A3 domain is closest to the biosensor surface. The vWF denaturant pulse assays were performed on an automated eight-channel Octet RED96 instrument (*forté*BIO) shaking at 1,000 rpm, 25°C. The programmed steps were as follows with the urea range from 0 to 7 M by 1 M step:

Table B1. Denaturant pulse assay steps for vWF with the urea range from 0 to 7 M by 1 M step

Step	Time (s)	Event	Composition
1	30	Initial Baseline	vWF Buffer (vWFB)
2	300	Loading	0.6 μM vWF Protein in vWFB
3	30	Baseline	vWFB
4	600	Custom	Urea Range
5	10	Baseline	GroEL Buffer (GB)
6	300	Association	0.5 μM GroEL in GB
7	300	Dissociation	GB
8	5	Regeneration	10 mM glycine, pH 1.7
9	5	Regeneration	GB
10	5	Regeneration	10 mM glycine, pH 1.7
11	5	Regeneration	GB
12	5	Regeneration	10 mM glycine, pH 1.7
13	5	Regeneration	GB
14	60	Regeneration	10 mM $NiCl_2$

Runs were performed in triplicate with tip regeneration performed after runs 1 and 2 only. The GroEL binding signal was plotted as a function of denaturant concentration to create kinetically controlled denaturation isotherms for wild-type, gain of function (V1314D), and loss of function (F1369I) mutant vWF.

Transmission Electron Microscope Sample Preparation

Two hundred mesh carbon-coated copper grids (Electron Microscopy Sciences CF200-Cu) were glow discharged for 20 s at −15 mA in 39 mBar atmosphere. After glow discharge, grids were rested for approximately 20 min before sample application. 4 μL of sample was applied to the rested grid for 1 min and then wicked off using Fisherbrand™ P8 Grade filter paper. Grids were washed once with ultrapure water and wicked off as fast as possible. The grids were then stained for 5 s with 0.022 μm filtered 0.75% uranyl formate (Electron Microscopy Sciences 22451) in ultrapure water and then wicked dry. Grids were completely dried overnight on filter paper inside a 100 mm Petri plate. EM image was acquired using 100 keV JEOL-JEM 1400 transmission electron microscope.

Appendix C: Alternative approaches to LF_N-PA_{pore} complex assembly

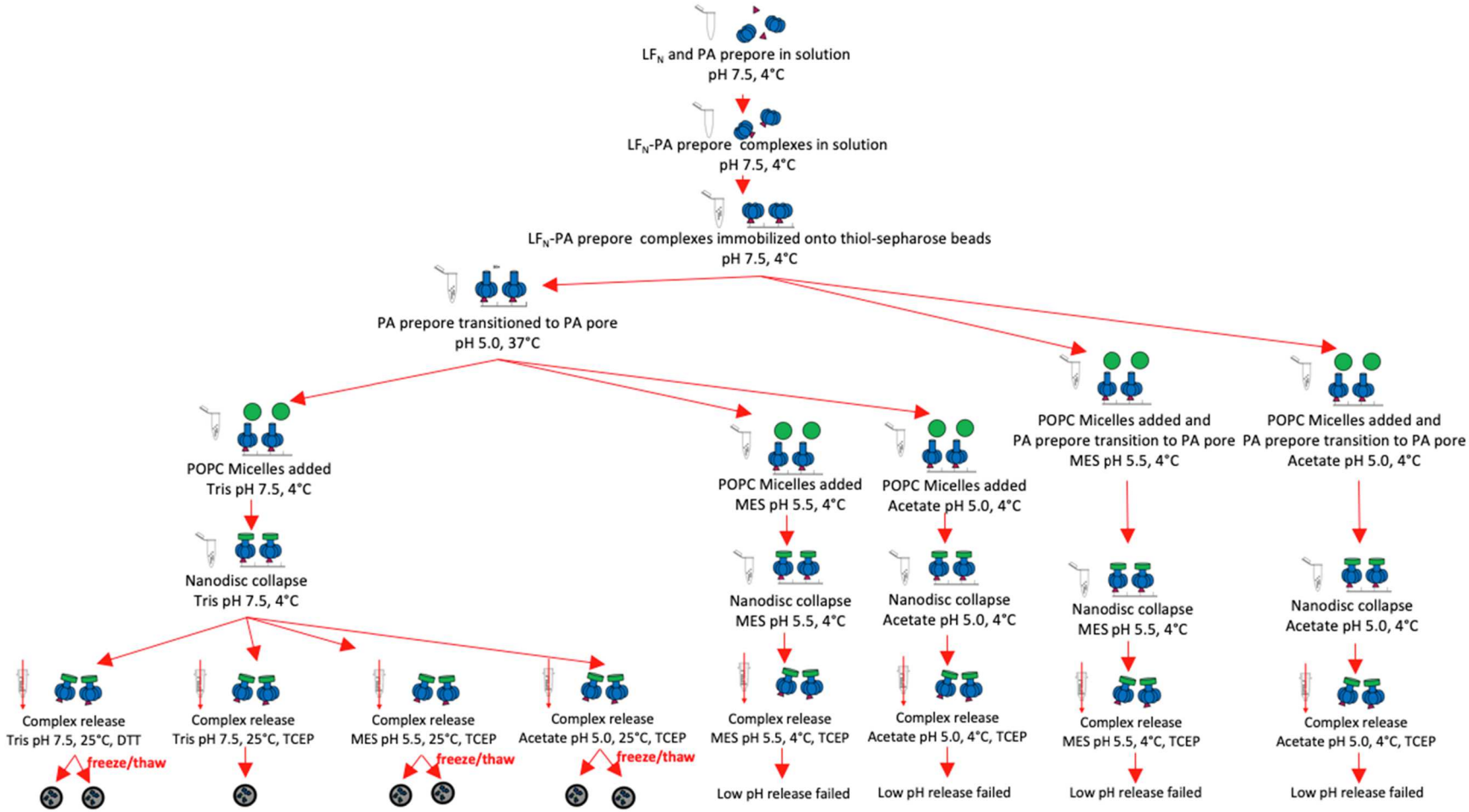


Figure C.1. LF$_N$-PA$_{pore}$-Nanodics complex assembly protocols with varying pH and temperature of nanodisc assembly, complex release. Complex formation confirmed by negative stain EM.

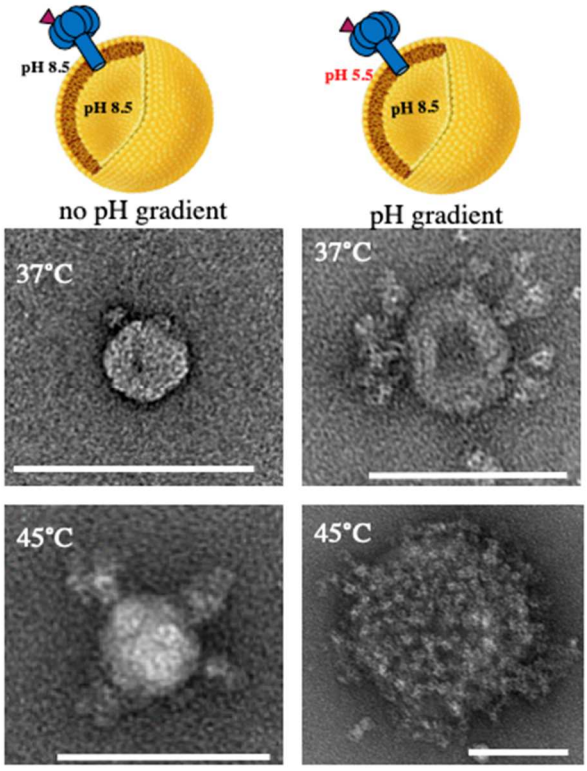
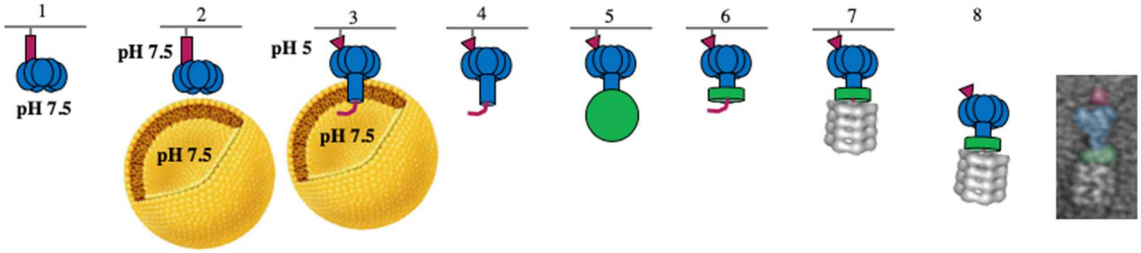


Figure C.2. Temperature and pH dependence of LF_N-PA_{pore}-liposome complex formation starting with $3LF_N$-$PA_{prepore}$ complexes and POPC liposomes.



(1) Immobilize LF_N-PA prepore
(2) Add LUV mix to immobilized LF_N-PA
(3) Drop pH to transition PA pore, insert into LUV, and translocate LF_N
(4) Remove vesicles with complex still attached to beads
(5) Add pre-nanodisc micelles
(6) Collapse micelles into nanodiscs
(7) Add GroEL to bind N-terminal tail of LF_N, confirming translocation arrested complex
(8) Release (soluble) complex of LF_N partially translocated through the PA pore in nanodisc with GroEL bound

Figure C.3. Protocol for translocation arrested co-toxin complexes using POPC liposome to induce pH gradient driven translocation and GroEL to capture emerging LF_N. Colorized negative stain EM particle (far right) from micrograph shown in **Figure B.4** shows proof of concept.

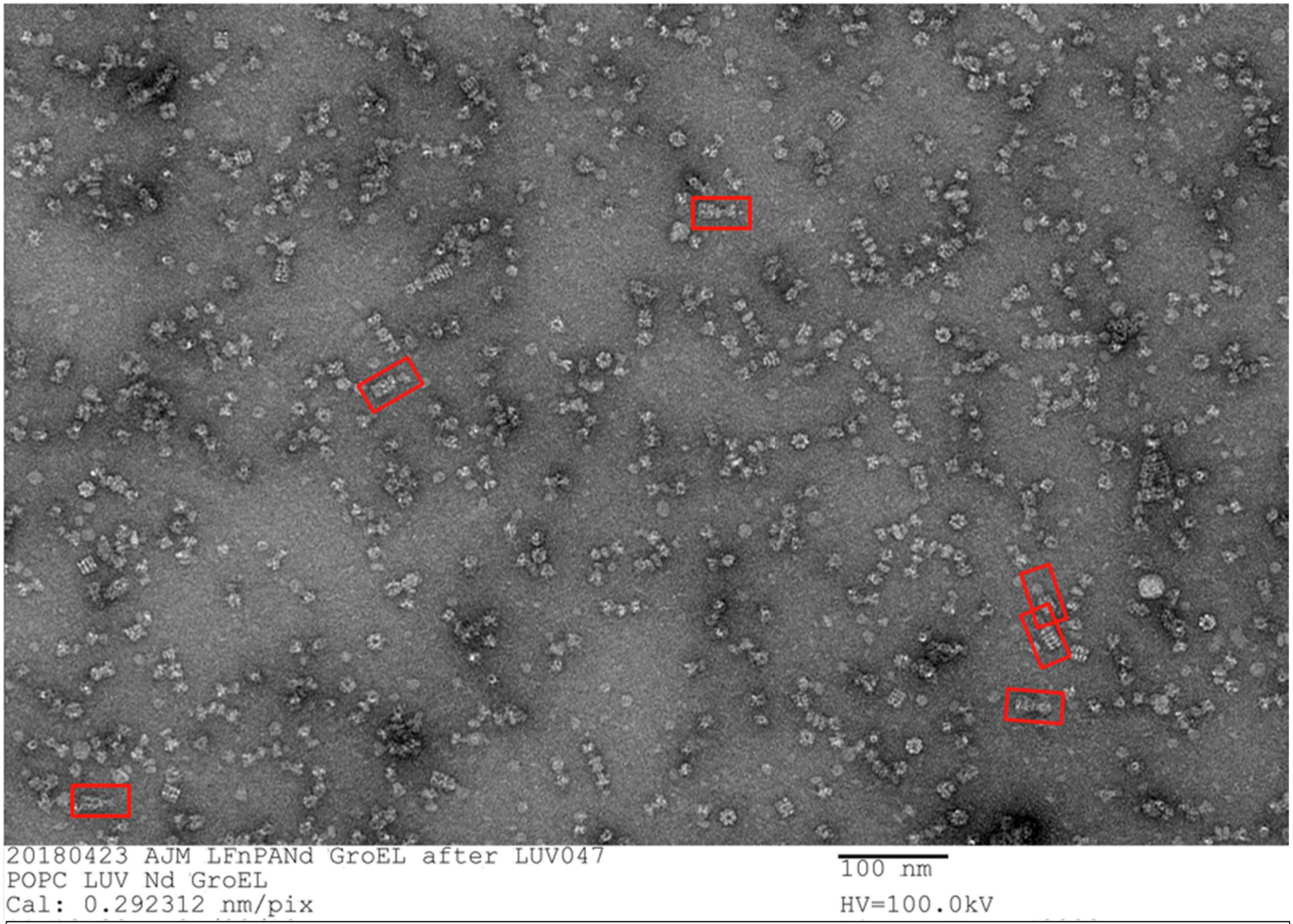


Figure C.4. Negative stain EM micrograph of LF_N-PA_{pore}-Nanodisc-GroEL complexes from protocol in **Figure B.3** show heterogenous mix of complexes with and without GroEL. Boxed particles show clear examples of LF_N-PA_{pore}-nanodisc-GroEL complexes

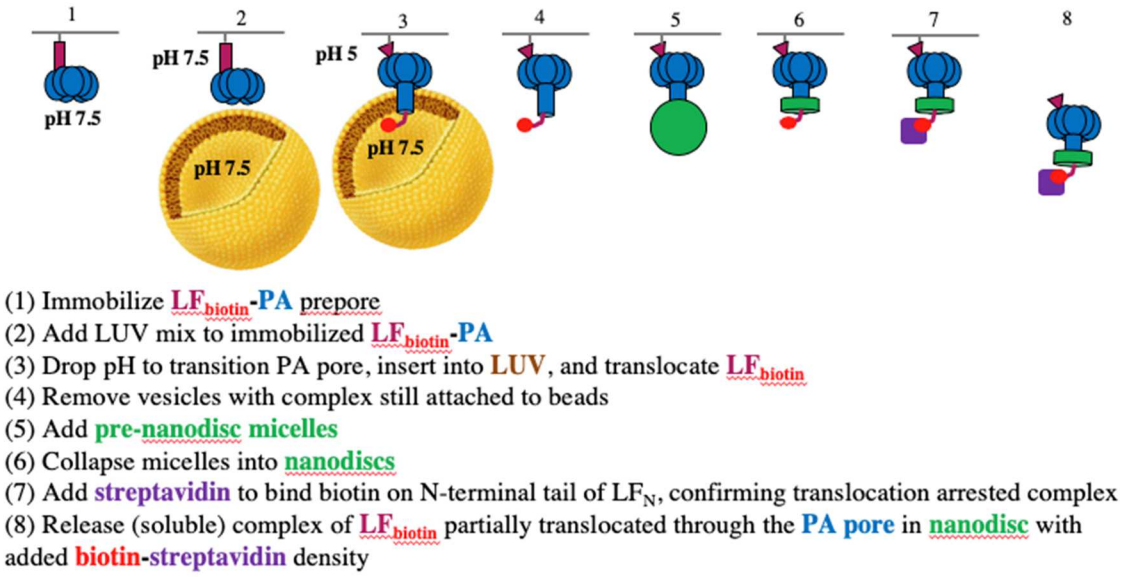


Figure C.5. Proposed protocol for translocation arrested co-toxin complexes. using liposomes to induce pH gradient driven translocation and biotin bobber to confirm translocation.

Appendix D: Particle orientation of cryoEM data collections

Our first cryoEM data collection suffered from severe orientation issues. **Figure D.1.A** shows a representative motion corrected micrograph and 2D class averages with most particles pointing the PA_{pore} β barrel perpendicular to the page. Our second data collection had diverse particle orientation. **Figure D.1.B** shows a more densely populated micrograph with various tilt and side views seen in the 2D class averages. To attain diverse particle orientation, we decreased the time from sample application on grid to blot and plunge freeze to avoid nanodisc and/or unfolded LF_N preference at the air water interface. CryoEM grids were imaged using a Titan Krios with energy filter and K2 direct electron detector. Movies were acquired in super resolution mode at 0.535 Å per pixel. 2D classification was performed in cryoSPARC.

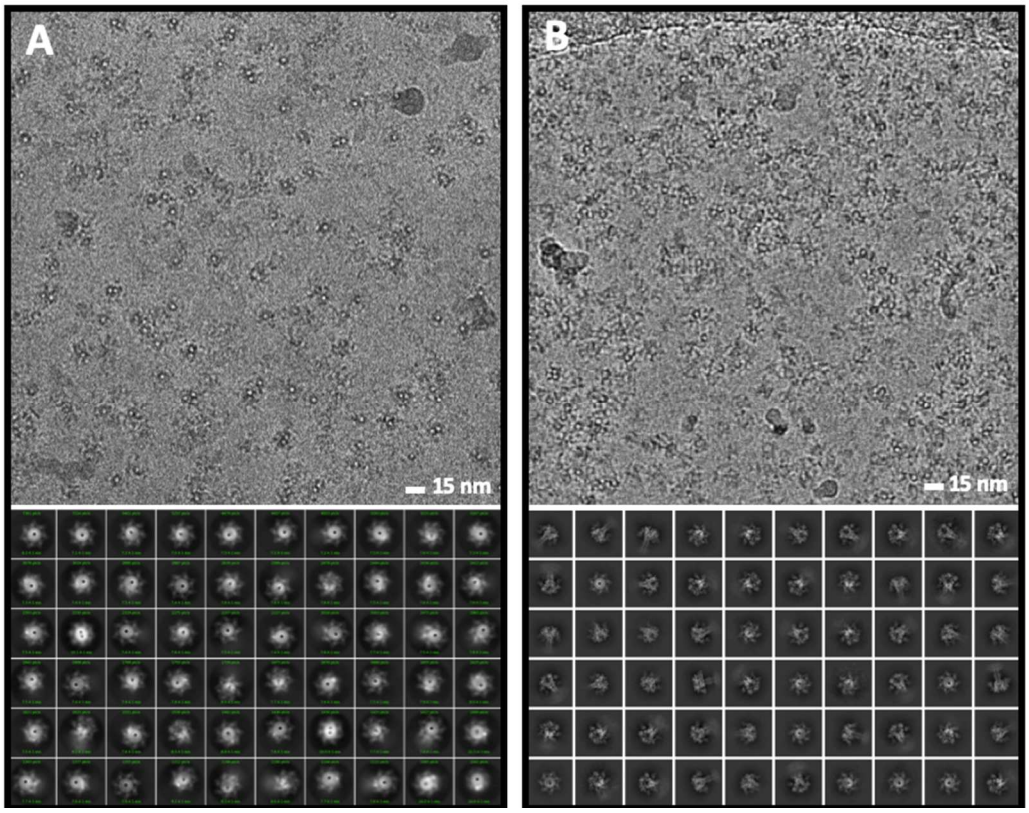


Figure D.1. Motion corrected CryoEM micrographs (top) and representative 2D classification (bottom) of LF_N-PA_{pore}-Nd complexes at pH 5.5 with (A) preferential orientation of nanodiscs to the air water interface and (B) diverse orientation of particles

Appendix E: Concluding Comments on the cryoEM Resolution Revolution

While this work was driven by the biological questions related to anthrax toxin intoxication, it is important that we contextualize this PhD dissertation and my training with the cryoEM 'resolution revolution' that occurred at the same time (Kühlbrandt 2014). In 2014, when I matriculated into the Interdisciplinary Graduate Program in Biomedical Sciences (IGPBS), there were only 36 cryoEM maps deposited in the EMDB at a resolution of 4 Å or better. In 2019, that number had grown to an astonishing 1,195. Figure 5.2 shows the exponential growth of cryoEM over the last seven years. The anthrax toxin has been on the front lines of this revolution, thus was a model system used to address challenges related to data heterogeneity and sample preparation. Our work has contributed to this revolution and we implemented many of the cutting-edge approaches as they were released (e.g., stochastic gradient descent analysis, principle component analysis (PCA), nanodisc insertion of membrane proteins, and translocation arrested complexes).

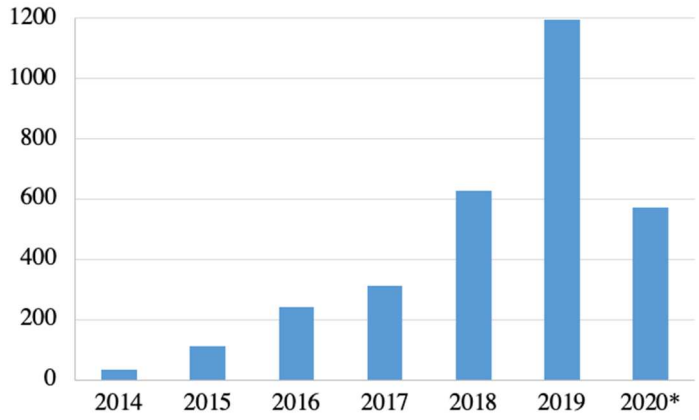


Figure E.1. Exponential growth of cryoEM. Number of cryoEM maps at a resolution of 4 Å or better deposited in the electron microscopy data bank (EMDB). *2020 map number is as of May 1, 2020.

Computational advancements have continued to emerge since stochastic gradient descent. Not only is discrete heterogeneity being addressed using cryosparc's *ab* initio (Punjani, Rubinstein et al. 2017) and relion's *de novo* (Zivanov, Nakane et al. 2018), continuous heterogeneity is also

being explored in programs like 3D variability analysis (Punjani and Fleet 2020) and cryoDRGN (Zhong, Bepler et al. 2020) using PCA. CryoEM data heterogeneity, when handled properly, can reveal motion or 'breathing' of complexes (Nguyen, Galej et al. 2016, Wrapp, Wang et al. 2020). This type of PCA has proved invaluable during data analysis of our mid-translocation complexes, resulting in, to our knowledge, the first structure of an actively translocating toxin. In the last six months, there has been a boom of toxin cryoEM structures (Piper, Brillault et al. 2019, Anderson, Sheedlo et al. 2020, Antoni, Quentin et al. 2020, Hardenbrook, Liu et al. 2020, Yamada, Yoshida et al. 2020). Many of these toxins are formed by the assembly of multiple identical subunits. This is a common theme in biology, as it allows for efficient storage of genetic information and protein production. While it is tempting to assume the subunits are always symmetric (it certainly improves resolution!), many biologically relevant answers may be lost by imposing symmetry. In our work, LF_N density is smeared or lost if we impose symmetry operations. However, we were able to deal with the pseudo symmetry of the PA_{pore} by using symmetry expansion techniques which separate each symmetric unit of a particle into a new particle allowing for masked 3D classification and refinement with local angular searches (Zhou, Li et al. 2015, Serna 2019). Thus, both symmetry expansion and PCA were instrumental in acquiring complexes we interpret as LF_N translocating through the PA_{pore}.

With advancements in cryoEM data collection, *e.g.*, direct electron detectors (DEDs) and phase plates, as well as cryoEM data processing, *e.g.*, graphics processing units (GPUs) and machine learning, the next wave in cryoEM advancements appears to be in sample preparation (Brillault and Landsberg 2020). For example, quick spray applications, such as chameleon (Dandey, Budell et al. 2020) and vitrojet (Ravelli, Nijpels et al. 2019), are under development to improve consistency of thin ice and diverse particle orientations. Grids are improving as well, with gold

support structures to reduce z-directional vibration in the microscopes (Russo and Passmore 2014) and grid capillaries for ice thickness consistency (Wei, Dandey et al. 2018). For membrane proteins, strategies to paint membranes across grids are under development. Toxin immobilization, translocation, and nanodisc stabilization (TITaNS) was our contribution to cryoEM sample preparation advancements. This method facilitates the capture and release of macromolecular machines in action. Overall, cryoEM has proven to be a well suited technique to solve structures of large macromolecular assemblies (Nogales and Scheres 2015). As the technique becomes more widely available (Tachibana 2020), structures of intermediates will become more prolific and necessary to answer new, challenging structural biology questions.