Title
Molecular Mechanisms of Anthrax Toxin Assembly and Transport

Author
Feld, Geoffrey Keith

Publication Date
2012

Peer reviewed|Thesis/dissertation

Molecular Mechanisms of Anthrax Toxin Assembly and Transport

by

Geoffrey Keith Feld

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Bryan Krantz, Chair
Professor David Wemmer
Professor Susan Marqusee

Spring 2012

Molecular Mechanisms of Anthrax Toxin Assembly and Transport

Geoffrey Keith Feld

Abstract

Molecular mechanisms of anthrax toxin assembly and transport

by

Geoffrey Keith Feld

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Bryan Krantz, Chair

Because proteins carry out their cellular functions in specific locations, protein transport represents a crucial step in the central dogma of biochemistry. Transport across a membrane barrier requires energy; therefore, nature has provided a class of molecular machines known as translocases that catalyze protein translocation. Often, these machines contain channels that are too narrow to transport a folded protein; thus, substrate unfolding adds another layer of complexity to the poorly understood transport mechanism. The anthrax toxin system represents a useful model for elucidating the mechanisms of translocation-coupled protein unfolding, both in terms of structure and function.

In order to understand the mechanism by which an unfolding machine interacts with its substrate, the X-ray structure of an anthrax protective antigen (PA) oligomer prechannel was solved in complex with the amino-terminal PA-binding domain of the substrate lethal factor (LF_N). The structure revealed how PA interacts with unfolded polypeptide *via* a hydrophobic cleft we call the α clamp. Working with my colleague, Katie Thoren, we show how the α clamp is critical for toxin assembly, substrate binding, unfolding, and translocation through a nonspecific binding mechanism.

The recent discovery that PA can form two oligomeric states exemplifies the complexity of multimeric protein nanomachine assembly. In order to better understand the molecular mechanism of heterogenous assembly, a series of PA constructs with varying length crosslinks was produced, their structures were solved to high resolution, and their assembly products were analyzed by electron microscopy. The flexibility of PA's receptor-binding domain (D4) relative to the main body of the protein provides a mechanism for controlling oligomeric stoichiometry, whereby D4 can adopt a pro-PA_7 or pro-PA_8 conformation.

Finally, an overall model for protein translocation is presented, based on data discussed herein and other recent studies involving anthrax toxin transport.

Dedicated to my family.

Table of Contents

List of Figures..... iv
List of Tables.....v
List of Abbreviations.....vi
Acknowledgements.....vii

Chapter 1
Ratcheting up protein translocation with anthrax toxin.....1
1.1 Nanomachines as ratchets.....1
The physical environment of the cell.....5
1.2 Molecular mechanisms of biological nanomachines.....5
1.3 Anthrax toxin as a protein translocation model system.....7
1.4 The PA channel.....11
The LF/EF binding sites.....12
The ϕ clamp.....12
The β barrel.....14
1.5 Translocation powered by a proton gradient.....15
1.6 Translocation-coupled protein unfolding.....16
1.7 Processing unfolded protein.....17

Chapter 2
Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers.....19
2.1 Abstract.....19
2.2 Introduction.....19
2.3 Results.....20
Crystal structure of the $PA_8(LF_N)_4$ complex.....20
The carboxy-terminal binding subsite.....22
The α-clamp binding subsite.....22
Both LF-binding subsites are critical for cytotoxicity activity.....31
The role of the α clamp in LT assembly.....31
Mapping the LF_N-binding interaction with the PA channel.....31
The α clamp possesses nonspecific binding activity.....34

LF_N must unfold to bind the α-clamp subsite..... 36
Binding to PA induces strain and disorder into LF_N..... 36
The role of the α clamp in protein translocation..... 42
2.4 Discussion..... 42
2.5 Materials and Methods..... 49

Chapter 3
Domain flexibility modulates the heterogenous assembly mechanism of anthrax protective antigen..... 56
3.1 Abstract..... 56
3.2 Introduction..... 56
3.3 Results..... 59
Crystal structure of the $PA^{\Delta MIL}$ monomer..... 59
Ordered loop regions..... 61
Receptor-binding loop-helix $2\alpha_1$..... 61
Changes in orientation of PA D4..... 61
Crystal structures of cross-linked PA monomers..... 63
Crystal structure of $PA^{\Delta MIL}$ at pH 6.5..... 63
The D2-D4 interface modulates the propensity of PA to form octamers..... 63
Correlation of D4 conformation and oligomeric heterogeneity..... 66
The furin-cleavage site..... 66
Furin-dependent PA processing is not ANTXR2-dependent..... 69
3.4 Discussion..... 69
Protoxin maturation by furin..... 70
Molecular determinants of PA's heterogenous oligomerization mechanism..... 71
3.5 Materials and methods..... 72

Chapter 4
Concluding remarks..... 76
4.1 Conformational entropy, diffusion, and substrate orientation..... 78
4.2 A proton-engine mechanism for translocation..... 79
4.3 An entropic force-generation mechanism..... 81

References..... 84

List of Figures

Figure		Page
1.1	Real and imagined heat engines and pumps	2
1.2	Flashing Brownian ratchet model	4
1.3	Molecular mechanisms of nanomachines	6
1.4	Components of anthrax toxin	9
1.5	Anthrax toxin assembly and transport	10
1.6	Peptide clamps in the PA channel	13
2.1	Mass spectrometry analysis of the PA_8LF_4 and $PA_8(LF_N)4$ complexes	21
2.2	Structure of LF's PA-binding domain in complex with the PA octamer	24
2.3	LF_N electron density in the $PA_8(LF_N)_4$ complex	25
2.4	Stereo image for LF_N electron density in the $PA_8(LF_N)_4$ complex	26
2.5	Stereo images for α-clamp electron density	27
2.6	The PA octamer binds LF_N in two distinct substites	28
2.7	Changes in equilibrium binding free energy ($\Delta\Delta G$) for PA channel complexes	29
2.8	Probing the role of the PA-LF_N interaction in toxin assembly and cytotoxicity	30
2.9	Role of LF_N α1 and β1 in PA assembly	32
2.10	Mass spectrometry analysis of Δ47-LF_N-driven PA assembly	33
2.11	Equilibrium and kinetic measurements of LF_N binding to the PA channel	35
2.12	Sequence-alignment and helical-wheel analysis of α1/β1-replacement sequences	37
2.13	The first 20 or 60 amino-terminal residues of LF_N are sufficient to bind the PA channel at the α-clamp subsite	38
2.14	Dynamics and thermodynamics of the pre-translocation unfolding of LF_N	39
2.15	LF_N unfolding increases its affinity for PA but reduces its stability	40
2.16	Fluorescence anisotropy changes upon PA-LF_N complex formation	43
2.17	The role of the α clamp in LF_N and LF translocation	44
2.18	LF_N translocates similarly via PA_7 and PA_8 channels	45
2.19	Translocation efficiencies for LF_N and mutants on PA channels	46
2.20	LF translocation records	47
2.21	The α clamp catalyzes translocation by interacting with unfolded structure	48
3.1	Anthrax toxin assembly is heterogeneous and requires proteolyic processing	58
3.2	The structure of $PA^{\Delta MIL}$	60
3.3	Cysteine cross-linking of PA D2 and D4	64
3.4	The interface of PA D2 and D4 controls oligomeric stoichiometry	65
3.5	Molecular basis for PA oligomeric stoichiometry	67
3.6	Structural basis for furin-dependent cleavage of PA	68
4.1	The α clamp	77
4.2	Loop flexibility in PA oligomers	80
4.3	Translocation by a proton-driven machine	82

List of Tables

Table		Page
2.1	Data collection and refinement statistics	23
2.2	Thermodynamic stability free-energy parameters for Δn LF_N truncations	41
3.1	Data collection and refinement statistics	62

List of Abbreviations

2Fo-Fc	weighted electron density map
ΔΨ	membrane potential
ΔpH	proton gradient
ANTXR1	anthrax toxin receptor 1
ANTXR2	anthrax toxin receptor 2
ADP	atomic displacement parameter
ATP	adenosine-5'-triphosphate
Atx	anthrax toxin
B-factor	crystallographic temperature factor
CD	circular dichroism
CD_{222}	CD signal at 222 nm
COL	center of oligomer lumen
COM	center of mass
DBM	n-dodecyl-β-D-maltoside
DCA	1,3-dichloroacetone
DTA	diphtheria toxin A chain
DTT	dithiothreitol
EDTA	ethylenediaminetetra acetic acid
EF	edema factor
EF_N	amino-terminal domain of edema factor
FA	fluorescence anisotropy
Fo-Fc	difference electron density map
GdmCl	guanidinium chloride
LT	lethal factor + PA
LF	lethal factor
LF_N	amino-terminal domain of lethal factor
LSQ	least-squares quotient
MIL	membrane insertion loop
MR	molecular replacement
nanoESI-MS	nanoelectrospray ionization mass spectrometry
NMR	nuclear magnetic resonance
PA	protective antigen
$PA^{\Delta MIL}$	protective antigen Δ303-324, V303P and H304G
PA_{20}	~20 kDa fragment of PA
PA_{63}	~63 kDa fragment of PA
PA_{83}	~83 kDa full-length PA
PDB	protein data bank
PMSF	phenylmethylsulfonyl fluoride
RMS	root-mean square
SASA	solvent-accessible surface area
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TCEP	tris(2-carboxyethyl)phosphine
TOP	torsion optimization procedure
UBB	universal bilayer buffer
WT	wild type

Acknowledgements

This work would not have been possible without the combined efforts of a fantastic group of people. I owe a debt of gratitude to the individuals who have dedicated both time and effort in mentoring me during my graduate career. I would first like to thank my advisor, Professor Bryan Krantz, who has been a close mentor for me, and from whom I have learned invaluable lessons in science, especially in the forms of presentation and communication. I would also like to thank Professors David Wemmer, Susan Marqusee, and John Kuriyan, who have been immensely helpful in shaping my scientific acumen and career.

As science is best conducted as part of a team, I would especially like to thank former and current members of the Krantz Lab, who are the embodiment of the word 'colleague': Katie Thoren, Alex Kintzer, Mike Brown, Jen Colby, Yoki Tang, Sarah Wynia-Smith, Melinda Hanes, Shon Greenberg, Adam Schawel, Sam Stephenson, Seth Garwin, and George Shan. I would also like to thank Harry Sterling and Catherine Cassou of the Williams Lab for a special and fruitful collaboration.

Being at UC-Berkeley has afforded me the expertise and companionship of a number of individuals and labs. I would like to thank the Advanced Light Source beamline 8.3.1 team James Holton, George Meigs, and Jane Tamanachi; Nat Echols of Paul Adams's PHENIX team; and members of the Berger and Alber labs, especially Art Lyubimov, Teri Lang, and Nathan Thomsen, for helpful discussions in crystallography. I would also like to thank my friends in the Long Group and the College of Chemistry Fighters for successful softball campaigns, as well as Julian Bigi, John Weinstein, and Tammy Vinson for their friendship and support. Finally, I would like to extend a very special thank you to my family, my girlfriend, Colleen Kellenberger, and my dog, Patton, for their infinite support and encouraging me to do my best work possible.

This work was supported by the National Institutes of Health Research Grant R01-AI077703 (to Bryan Krantz).

Chapter 1

Ratcheting up protein translocation with anthrax toxin

Proteins, organelles, metabolites, and various cargos are continuously transported into, out of, and around cellular compartments. Generally, these movements are catalyzed by macromolecular machines, which consume energy and directionally translocate along specific polymeric tracks. These polymer tracks come in various forms, including microtubules, nucleic acids, and polypeptides. While it is consistently observed that chemical energy is consumed to do mechanical work, the underlying molecular mechanisms of energy transduction and directed movement at the nanoscale level are poorly understood and remain intense areas of research (Matouschek 2003; Prakash & Matouschek 2004; Sauer *et al.* 2004; Wickner & Schekman 2005; Gennerich & Vale 2009; Sauer & Baker 2011; Thoren & Krantz 2011).

Biological processes occur on the nanoscale under ambient chemical and temperature conditions. Scientists have long mused about the ability of nanoscopic power devices to harness energy from their environments. The most famous thermodynamic trick of the 19th century is credited to Maxwell, who proposed in *Theory of Heat* how energy is statistically distributed in a population of gas molecules (Maxwell 1871). When Maxwell began to imagine the molecular world, he expressed a most remarkable "contradiction" by the conclusion of his manuscript, which he subtitled, "Limitation of the Second Law of Thermodynamics." To illustrate, he proposed a *Gedankenexperiment* (thought experiment) (Figure 1.1a) involving a vessel divided into two sections, A and B. Each side contains a similar proportion of gas molecules with a similar distribution of velocities and temperature. The division between the two sections contains a hole with a trap door operated by a "being" who only permits the movement of the faster molecules from A to B, and concurrently the movement of slower molecules from B to A. The being, coined "Maxwell's demon" by Lord Kelvin (Thomson 1874) would "thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the Second Law of Thermodynamics," (Maxwell 1871). Later, it was reasoned that because the demon would need precise information on each gas molecule to sort them by velocity, the entropy change of the system and surroundings upon creating a thermal gradient would not be less than zero. Obtaining and consuming that information came at a price, and Maxwell's *Gedankenexperiment*, in fact, obeyed the Second Law of Thermodynamics. The device is, therefore, an information-powered heat pump.

The drinking bird novelty device (Figure 1.1b) patented by Sullivan in 1946 is now a famous staple demonstration in science classrooms all over the world (Sullivan 1946). Deliberately flirting with one's disbelief in perpetual-motion machines, the happy-go-lucky device teeters relentlessly, dipping its beak into a glass of water. At first glance, one may think the bird is a "free" energy device. Of course, there is a man behind the curtain: the drinking bird, like a Stirling heat engine, functions by means of a *required* thermal gradient (ΔT). (Evaporative cooling on the wetted bird beak creates the ΔT; see Figure 1.1b for details.) As noted by Carnot, "the motive power undoubtedly increases with difference in temperature between the warm and cold bodies," (Carnot 1824). Thus a nonzero ΔT is the *driving force* that powers the heat engine to produce spontaneous motion.

1.1 Nanomachines as ratchets

Further adoption of Maxwell's demon led Smoluchowski (Smoluchowski 1912) and later

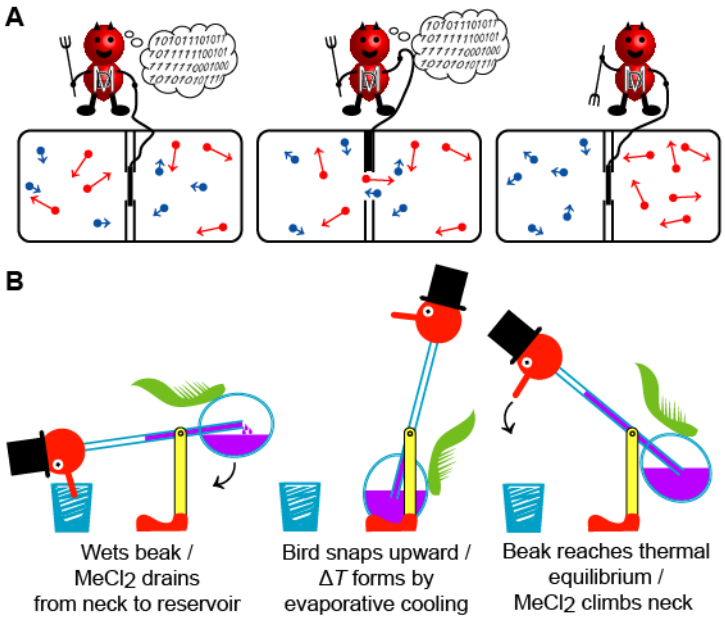

Figure 1.1. Real and imagined heat engines and pumps. (A) Maxwell's demon is an example of an information-powered heat pump. (left) Two compartments are divided by a partition with a trap door operated by the demon. Hot (red) and cold (blue) gas atoms are shown evenly divided into each compartment. (middle and right) In order to obey the Second Law, the demon must consume stored information ('*101101011*...') on the individual atoms velocities and positions, to properly sort the hot ones to the right and the cool ones to the left. **(B)** Thermodynamic cycle of Sullivan's drinking bird exemplifies a basic heat engine. Inertial movements of the bird are linked to the establishment and dissipation of ΔT gradients across the glass vessel body, which contains $MeCl_2$ (purple fluid). (right) When the bird dips its beak into the shot glass, $MeCl_2$ leaks from the opposite end of the tube into the reservoir, acting to ratchet the bird upward. (middle) Evaporative cooling leads to $MeCl_2$ condensation. (left) $MeCl_2$ in the reservoir travels up the neck, shifting the center of mass. The bird lurches forward to complete the thermodynamic cycle.

Feynman (Feynman *et al.* 1963) to propose a nanoscale heat-engine device in related *Gedankenexperiments* (see (Miller *et al.* 1999; Reimann *et al.* 2002) for a comprehensive treatment of Feynman's nanoscale device). Imagine a system with two thermal reservoirs; across the two reservoirs is a drive shaft containing a pulley, a rope, and a weight. Gas molecules in the hotter reservoir transfer random thermal energy to large vanes, albeit the applied forces are random in direction. An anisotropic sawtooth-patterned ratchet wheel resides within the cooler reservoir. The ratchet wheel engages a pawl that effectively biases rotation of the wheel in one direction. Therefore, while Brownian motion of particles contained within the hotter reservoir pushes the vaned wheel randomly in either direction, the pawl-and-ratchet device in the colder reservoir rectifies these random fluctuations, biasing net motion in the allowed direction; and hence the pulley and rope turn, lifting the weight.

The ΔT energy gradient across the reservoirs is an essential feature of the ratcheting nanodevice (Feynman *et al.* 1963; Blaustein *et al.* 1989). If stochastic motion within the pawl in the colder reservoir were on the order of that on the hotter vaned-wheel reservoir (i.e., the condition of thermal equilibrium), then the system could not possibly do useful work, and the weight could not be lifted. Thus the Smoluchowski-Feynman ratchet is a nanomechanical heat engine (analogous to Sullivan's drinking bird). In addition to a nonzero thermal gradient, the machine operates on the principle that the mean displacement of random motion is on the order of the period, L, of the spatially periodic potential function during the time in which the barrier is switched to its minimized state (Figure 1.2). A diode represents an analogous ratchet-like component for an electrical circuit. A four-diode bridge device can rectify an oscillating electrical current produced by a magnet displaced at random through a coil of wire; however, the diode bridge cannot use Nyquist noise (random *equilibrium* thermal fluctuations of electrons in the electrical circuit) to do work. Thus the pawl in Feynman's ratchet, the one-way valve in Smoluchowski's ratchet, and the demon-operated trap door in Maxwell's *Gedankenexperiment* can be thought of as mechanical diodes; however, each is only capable of using *non-equilibrium* energy fluctuations, energy gradients (e.g. ΔT), or stores of ordered information to do useful work.

We can simulate net motion in the Smoluchowski-Feynman ratchet when we use a spatially anisotropic potential but not an isotropic one (Figure 1.2). Here we depict the ratchet wheel as a linearized repeating anisotropic potential and the pawl as a movable particle. To keep the model simple, the pawl (particle) can move relative to a stationary ratchet wheel in effect. When the potential is momentarily *flashed off*, the ratchet particle can freely diffuse. In principle, these energy barriers can be modulated *on* or *off* by any number of means, as long as the barriers are approximately $k_B T$, when switched *off*. (T is the temperature, and k_B is the Boltzmann constant.) When the energy well is offset and anisotropic, the particle is more likely to cross a closer peak, rather than a more distant peak. When the potential is *flashed on*, the particle will have a significantly higher probability of progressing rather than regressing. Directional flux is, therefore, supported by anisotropy. A perfectly symmetrical energy potential cannot support net movement in any particular direction (Figure 1.2). Energy input into such a system is required in order to flash, rock, or alter the potential function. Of course, the energy input may also be used to more directly push or drive the particle (via a power stroke). However, because of the significance of diffusion and random motion at the molecular level, the ratchet-like property of the flashing asymmetrical energy potential is critical to net directional translocation. Therefore, a minimal structural feature for net motion is a physically periodic substrate and a modulated clamping interaction with the nanomachine. Because ΔT at the required nanometer length scale

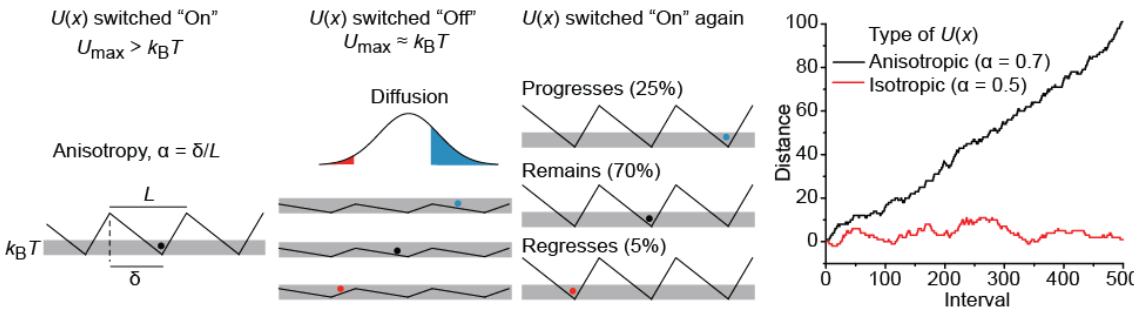

Figure 1.2. Flashing Brownian ratchet model. (far left) A saw-tooth potential function, $U(x)$, with respect to distance, x, is depicted, where the positional anisotropy, α, of $U(x)$ is related to the distance between the maximum and minimum, δ, of each period, L, such that $\alpha = \delta/L$. When $U(x)$ is switched *on*, the particle is trapped in a well, since $U_{max} > k_BT$. (middle left) When the potential is switched *off*, the particles diffuse freely according to Einstein's relation. (middle right) When $U(x)$ is switched back *on*, one interval is completed and the particle is trapped again. Thus the probability of progressing is greater than regressing. (far right) A Monte Carlo simulation of the flashing BR model plotting x versus the number of switching intervals for an anisotropic (black, $\alpha = 0.7$) and an isotropic (red, $\alpha = 0.5$) $U(x)$.

is impossible to achieve inside of cells, biological nanomachines must use other kinds of energy gradients to do work. Hence the interconversion of chemical- or electrical-potential energy to mechanical kinetic energy is ubiquitous in biological molecular motors, switches, pumps, and transporters. How do nature's molecular machines perform their necessary work? As we learned from Sullivan's drinking bird (Figure 1.1b), defining the problem facing a nanomechanical system, its environment, parts, and energy source are essential to elucidating its mechanism.

The physical environment of the cell. The environment of the cell is extraordinarily violent, where molecular movements carried out by nanomachines would be akin to sailing in a hurricane. Robert Brown initially observed the chaotic motion of pollen grains in solution; and these motions are referred to as Brownian motion. For the erratic path an individual pollen grain travels, its Brownian motion emerges statistically from the unequal number of collisions between the larger particle with many more numerous solvent molecules (having a large distribution of velocities) (Smoluchowski 1906; Brown *et al.* 2011). Einstein described the average diffusion of particles of radius, r, in a solution of viscosity, η, as the mean square displacement, $\langle \Delta x^2 \rangle$, where $\langle \Delta x^2 \rangle = 2Dt$: t is the time; the diffusion constant, D, is $k_B T/\beta$; and β, the frictional coefficient, is $6\pi\eta r$ (Einstein 1905). Thus given equal thermal energy inputs, larger objects diffuse slower than smaller ones. While macroscopic forces are largely inertial and depend heavily upon mass, m, and acceleration, a, (i.e. $F = ma$), Langevin noted that microscopic Brownian particles experience a noisy, fluctuating force, ζ, which he called the "complementary force" (Langevin 1908; Lemons & Gythiel 1997). The fluctuating force, ζ, unpredictably arises from a barrage of solvent molecules colliding with the Brownian particle, and it allows an instantaneous force to be assessed for a single particle. Langevin hypothesized forces are related to velocity, v, the frictional coefficient, and fluctuating force by $F = -\beta v + \zeta$. For nanomachines, forces scale to surface area and not volume. For any single molecule, ζ can be quite significant, hence the random, unpredictable nature of any individual Brownian walk.

The behavior of polymers in solution is especially relevant for nanomachines that unfold and translocate proteins. From statistical mechanics, an unstructured protein has many more degrees of freedom than a simple, rigid Brownian particle. A peptide backbone in the unfolded state can be assumed to have 3 degrees of freedom per amino acid, considering relevant ϕ/ψ-angle conformations. The multiplicity, W, of this peptide system, given as $W = 3^N$, translates into a configurational entropy, S, of $Nk_B \ln 3$. Therefore, a polypeptide with $N = 50$ residues has a thermal free energy of ~33 kcal mol^{-1} (~140 kJ mol^{-1}). These energies are much more significant for the polymer relative to a simple, rigid Brownian particle, which experiences only three translational degrees of freedom.

1.2 Molecular mechanisms of biological nanomachines

Generally, two competing mechanisms describing the functionality of nanoscale molecular machines have been presented (Tomkiewicz *et al.* 2007). On one hand, nanomachines falling into the Brownian-ratchet (BR) classification do work by using external energy sources to harness Brownian thermal energy (Figure 1.3a) (Feynman *et al.* 1963; Simon *et al.* 1992; Astumian 1997; Krantz *et al.* 2006). A Brownian ratchet on the biological scale is analogous to a Smoluchowski-Feynman ratchet and requires an external gradient as an energy source. Imagine a nanomachine acting on an unfolded polypeptide chain with sites that can be modified to be in permissive and non-permissive states. The ratchet can switch between the permissive *on* and non-permissive *off* states through the modification of the substrate, for example by protonation. Analogously, a gate in the transporter switches to a permissive *open* or non-permissive *closed*

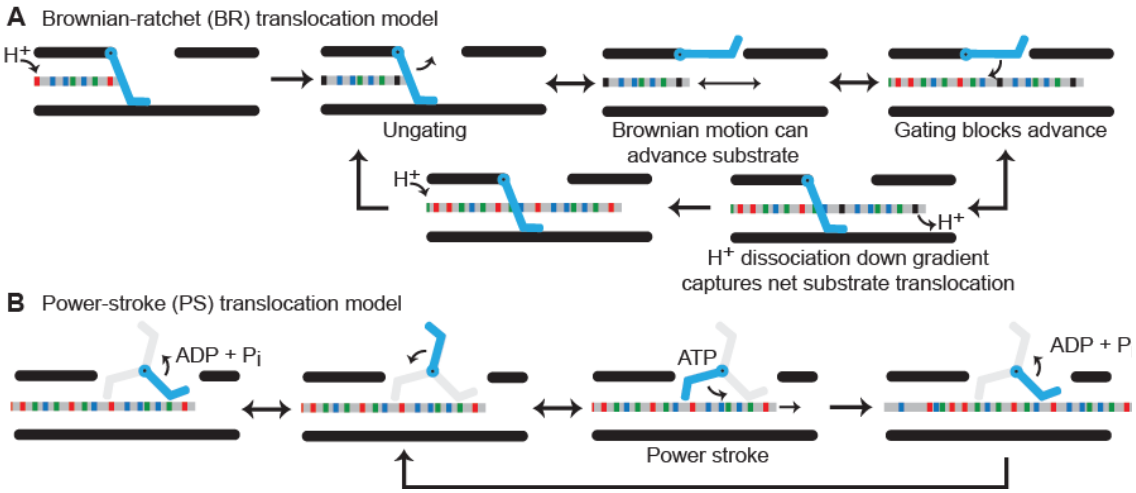

Figure 1.3. Molecular mechanisms of nanomachines. (A) A ΔpH-dependent BR mechanism for protein translocation. The substrate polypeptide chain is depicted as a simplistic gray rod with different functional groups colored as follows acidic (red), basic (blue), and nonpolar (green). A gate (cyan) electrostatically excludes anionic charges on deprotonated acidic residues. In this cyclical mechanism, substrate acidic residues are protonated (black); the gate opens, allowing for Brownian motion to take place. The peptide can only advance up to the point where deprotonated acidic residues enter the channel. Closing of the gate traps the peptide in the channel, as the dissipation of H^+ ions down the gradient upon deprotonation prevents retrograde movements. Further protonation then allows the cycle to repeat. **(B)** An ATP-dependent PS mechanism for protein translocation. The substrate chain is colored as in (A) while in this case, the gate acts more like a paddle with active (cyan) and inactive (gray) states. In the ADP-bound state, the paddle has low affinity for peptide; the paddle exchanges ADP for ATP, and subsequent ATP hydrolysis triggers a conformational change, allowing the paddle to push the polypeptide chain forward (power stroke). Cycles of ADP release and ATP hydrolysis allows the mechanism to continue.

state, depending on the protonation state of the polymer. Under the influence of a chemical gradient (e.g. a proton gradient, ΔpH) residues may be more likely to be in a permissive state on one side of the membrane than on the other. Thus translocation can proceed in the direction of non-permissive modification, since non-permissively modified substrate will be unable to retro-translocate. A number of chemical modifications have been suggested to follow such a mechanism. Protonation (Krantz *et al.* 2006; Brown *et al.* 2011), chaperone binding (Matlack *et al.* 1999; Liebermeister *et al.* 2001), glycosylation, disufide bonding, and conditions that promote protein folding (Simon *et al.* 1992) are all capable of biasing movement through a translocase channel.

On the other hand, the power-stroke (PS) classification suggests that molecular machines use external energy sources to directly drive motion without the need for harnessing Brownian thermal fluctuations (Figure 1.3b) (Glick 1995). The binding of ATP, for example, induces a conformational change in the protein machine that performs useful work such as pulling or pushing a substrate through the translocase. The machine engages with the translocating polypeptide chain via a polypeptide clamp or other active site loop structure. ATP hydrolysis and release of ADP and inorganic phosphate (P_i) allows the machine to return to its original conformation. However, the resetting of the polypeptide clamp or other active-site structure must occur via a different path back to the initial state, else it would essentially undo the work done in the ATP binding step.

The major difference between the BR and PS models is that the latter does not consider Brownian thermal fluctuations (or Langevin's random force, ζ) to be part of the mechanism. However, BR and PS mechanisms should not be held as mutually exclusive models. Instead, these two types of mechanisms may occur at distinct junctures in the transport cycle. For example, the two-headed motor, kinesin, uses an ATP-driven PS to nudge the trailing leg forward; however, this push provided by ATP binding is far too small to drive the leg the 8 nm span between binding sites on the microtubule, and a BR phase is required to bridge the remaining distance by allowing the head to diffuse and seek its next binding site, thus completing the transport cycle (Bier 2003). Furthermore, as operating on polymers in a cell is akin to sailing in a hurricane, consider two readily available types of machinery for hoisting up a sail: a simple hand crank and a hand crank with a ratchet. Under ideal conditions, the crank driven by a PS may suffice in performing the task. However, under tropical cyclone conditions, a system that includes both a PS-driven crank and a ratchet may become necessary. While extremely windy and diffusive weather likely makes it difficult for the sailor to turn the simple hand crank in the proper direction, the ratchet ensures that sub-steps of a crank cycle are not lost to diffusive backsliding.

1.3 Anthrax toxin as a protein translocation model system

The focus of this thesis is on protein pumps or translocases that move proteins within the cell. These nanomachines consume energy in order to disassemble and translocate folded polypeptide substrates. They are found both in solution and in lipid bilayer membranes. The former serve the cell as disaggregases, protein complex disassembly machines, and unfoldases that aid in intracellular protein degradation (Rapoport 2007; Horwich & Fenton 2009; Merdanovic *et al.* 2011). The latter allow proteins to be transported either across or into lipid bilayers (Matouschek 2003; Wickner & Schekman 2005; Thoren & Krantz 2011). The binary A/B bacterial toxins (Drazin *et al.* 1971)—including diphtheria, botulinum, and anthrax toxin—are a widely studied class of membrane transport systems, which use their own transport

machinery to enter cells (Falnes & Sandvig 2000; Barth *et al.* 2004; Collier 2009). These A/B toxins assemble into complexes on cell surfaces and then utilize host cell chemical potentials to unfold and translocate enzymatic factors into the host cell.

The binary A_2B toxin (Barth *et al.* 2004) system secreted by *Bacillus anthracis* (Young & Collier 2007; Thoren & Krantz 2011), anthrax toxin (Figure 1.4), represents an attractive model system for probing the molecular mechanism of protein translocation across membrane bilayers for a variety of reasons. First, its three protein components can be expressed recombinantly and studied independently. Second, the PA translocase itself and individual steps of substrate translocation can be readily reconstituted from purified proteins and studied using planar bilayer electrophysiology at the ensemble (Zhang *et al.* 2004; Krantz *et al.* 2005; Krantz *et al.* 2006; Basilio *et al.* 2009; Thoren *et al.* 2009; Janowiak *et al.* 2010; Pentelute, B. L. *et al.* 2010; Basilio *et al.* 2011; Basilio *et al.* 2011; Brown *et al.* 2011; Fischer *et al.* 2011; Pentelute, B.L., Sharma, O., Collier, R.J. 2011) and single-molecule level (Blaustein & Finkelstein 1990; Kintzer *et al.* 2009; Thoren *et al.* 2009; Basilio *et al.* 2011; Brown *et al.* 2011). Critical to this electrophysiological approach (also used in other systems (Hoch *et al.* 1985; Henry *et al.* 1996; Senzel *et al.* 1998; Grigoriev *et al.* 2004; Movileanu *et al.* 2005; Montal 2009; Harsman *et al.* 2011; Mahendran *et al.* 2012)) is the ability to precisely control the driving force and solution conditions on either side of the membrane. Third, for a membrane-protein system, structural studies using X-ray crystallography are tractable, because the PA translocase also exists in a soluble state (Petosa *et al.* 1997; Lacy *et al.* 2004; Kintzer *et al.* 2009). In this manner, researchers have been able to obtain structural information (Petosa *et al.* 1997; Lacy *et al.* 2004; Katayama *et al.* 2008; Kintzer *et al.* 2009; Katayama *et al.* 2010) (Pannifer *et al.* 2001; Drum *et al.* 2002) and distinguish possible translocation models using a wide variety of functional assays (Zhang *et al.* 2004; Krantz *et al.* 2005; Krantz *et al.* 2006; Basilio *et al.* 2009; Thoren *et al.* 2009; Janowiak *et al.* 2010; Pentelute, B. L. *et al.* 2010; Basilio *et al.* 2011; Basilio *et al.* 2011; Brown *et al.* 2011; Fischer *et al.* 2011; Jennings-Antipov *et al.* 2011; Pentelute, B.L., Sharma, O., Collier, R.J. 2011).

The structures of the three anthrax toxin components are known (Figure 1.4). The protective antigen (PA) component, which forms the translocase channel, is secreted as an 83-kDa proprotein (Petosa *et al.* 1997). The other two components that are transported by PA are ~90-kDa enzymes, lethal factor (LF) (Pannifer *et al.*) and edema factor (EF) (Drum *et al.*). In order for the toxin to function, PA must first be proteolytically nicked by a furin-family protease (Klimpel *et al.* 1992; Molloy *et al.* 1992), releasing an amino-terminal 20-kDa fragment, PA_{20}. The resulting 63-kDa fragment can assemble into an active holotoxin complex comprised of multiple copies of LF and EF bound to a ring-shaped PA oligomer (Figure 1.5). There are two known PA "prechannel" oligomeric architectures, a homoheptamer (PA_7) (Milne *et al.* 1994; Petosa *et al.* 1997; Lacy *et al.* 2004) and a homooctamer (PA_8) (Kintzer *et al.* 2009; Kintzer *et al.* 2010). The assembly and binding interactions between the PA oligomer and its substrates are well characterized (Chauhan & Bhatnagar 2002; Cunningham *et al.* 2002; Mourez *et al.* 2003; Lacy *et al.* 2004; Schueler-Furman *et al.* 2005; Melnyk *et al.* 2006), and is a primary focus of Chapter 2 of this thesis. These prechannel complexes either assemble on cells by binding a specific protein receptor (Bradley *et al.* 2001; Scobie *et al.* 2003) or localize to cells after assembling in the plasma (Kintzer *et al.* 2010). Internalization and subsequent acidification within the endosomal compartment converts PA prechannels into membrane-embedded channels (Koehler & Collier 1991; Miller *et al.* 1999; Qa'dan *et al.* 2005), which are strongly cation-selective (Blaustein *et al.* 1989).

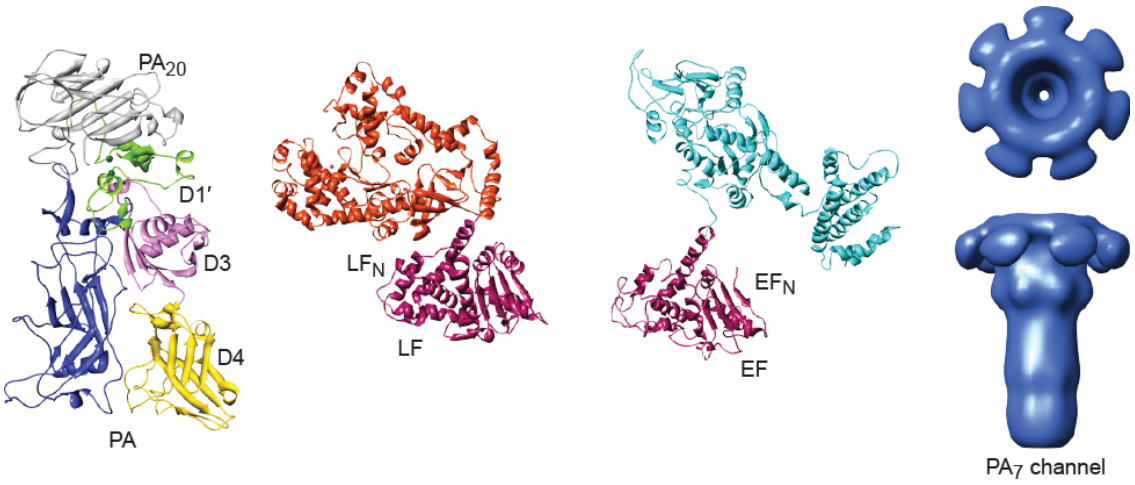

Figure 1.4. Components of anthrax toxin. (left to right) Ribbons depiction of PA (3TEW (Feld *et al.* 2012)) colored by domain: PA_{20} (gray), D1' (green), Ca^{2+} ions (dark green), D2 (blue), D3 (magenta) and D4 (yellow). The enzymes, LF (1JZN (Pannifer *et al.* 2001)) and EF (1YOV (Shen *et al.* 2005)); their amino-terminal PA binding domains (LF_N and EF_N, respectively) are colored red-violet and their catalytic domains colored orange and cyan, respectively. Axial view (above) and side view (below) of a three-dimensional EM reconstruction of the PA_7 channel (Katayama *et al.* 2008) (colored denim). (Prof. Mark Fisher graciously provided the EM density map).

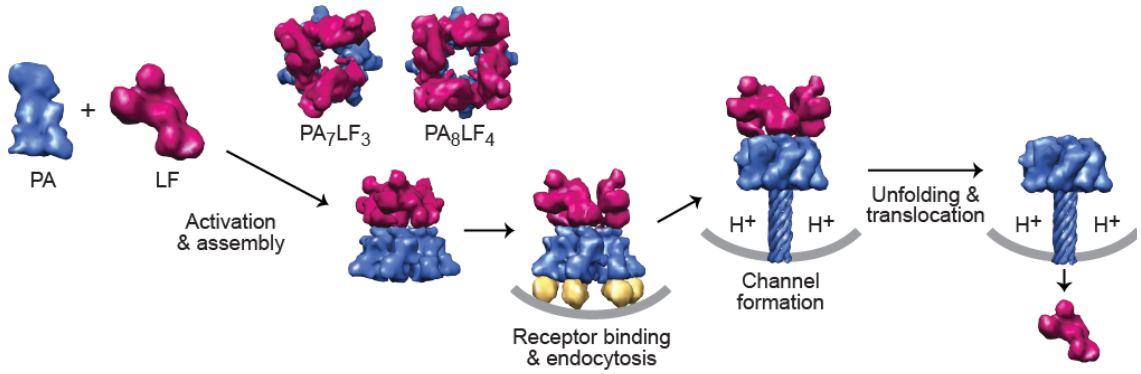

Fig. 1.5. Anthrax toxin assembly and transport. PA (denim) is proteolytically nicked and assembles with LF (red-violet) and forms PA_8LF_4 and PA_7LF_3 prechannel complexes (Mogridge *et al.* 2002; Kintzer *et al.* 2009) (based on 3KWV (Feld *et al.* 2010) and 1TZO, (Lacy *et al.* 2004) respectively). Prechannel complexes bind cellular receptors (gold; 1T6B (Santelli *et al.* 2004) and 1TZN (Lacy *et al.* 2005)) triggering endocytosis; acidic pH conditions in the endosome induce PA to form a transmembrane channel (Katayama *et al.* 2008) (atomic model 1V36 (Nguyen 2004)); the pH gradient that develops across the endosomal membrane destabilizes LF (Krantz *et al.* 2004), drives LF unfolding (Thoren *et al.* 2009) and translocation (Krantz *et al.* 2006; Brown *et al.* 2011) through the PA channel.

The acidic environment of the endosome is required for the action of the toxin (Friedlander 1986), and it generates a proton motive force (PMF) comprised of both a chemical potential (ΔpH) and an electrical potential (Δψ) capable of driving LF/EF unfolding (Krantz *et al.* 2004; Thoren *et al.* 2009; Brown *et al.* 2011) and translocation (Krantz *et al.* 2005; Krantz *et al.* 2006; Thoren *et al.* 2009; Brown *et al.* 2011). Translocation initiates from the amino-terminus of LF (Zhang *et al.* 2004), albeit a cationic unstructured amino-terminus is sufficient for initiating translocation (Blanke *et al.* 1996). The ΔpH component of the PMF is sufficient to unfold and translocate proteins (Brown *et al.* 2011). Moreover, while either the ΔpH or Δψ is sufficient to translocate both LF_N and the amino-terminal PA binding domain of EF (EF_N) the translocation of the physiological full-length LF and EF substrates requires the ΔpH more so than the Δψ (Krantz *et al.* 2006). The molecular mechanism of ΔpH-driven translocation is believed to be consistent with a charge-state BR model (Krantz *et al.* 2006; Finkelstein 2009; Pentelute, B. L. *et al.* 2010; Brown *et al.* 2011; Pentelute, B.L., Sharma, O., Collier, R.J. 2011), where differential protonation of anionic charges in the substrate polypeptide are essential for transport (Krantz *et al.* 2006; Brown *et al.* 2011) (Figure 1.3a). In support of the charge-state BR model, ΔpH-driven translocation depends upon charged residues (Pentelute, B. L. *et al.* 2010; Pentelute, B.L., Sharma, O., Collier, R.J. 2011) and, more precisely, acidic residues (Brown *et al.* 2011) in the translocating protein sequence.

1.4 The PA channel

Recent reports have provided insights into the structure of the PA translocase both in its soluble prechannel state and its membrane-inserted channel state. Many studies have shown that PA forms PA_7 oligomers, mostly due to the nature of the preparation used (Milne *et al.* 1994; Petosa *et al.* 1997; Lacy *et al.* 2004). When PA co-assembles with either of its substrate proteins, PA oligomerization turns out to be much more complex and heterogeneous. Under these conditions, PA forms mixtures of PA_7 and PA_8 oligomers in solution and on cell surfaces (Kintzer *et al.* 2009). Additional molecular determinats deciding the oligomeric fate of PA are presented and discussed in Chapter 3. The PA_8 oligomer can carry a maximum payload of four LF or EF molecules, while the PA_7 oligomer can carry a maximum payload of three substrates (Figure 1.5) (Mogridge *et al.* 2002; Kintzer *et al.* 2009). Despite a slight increase in the internal diameter of the oligomeric ring and a corresponding increase in the single-channel ion conductance of the PA_8 oligomer relative to the PA_7 oligomer, the change in channel diameter appears to have a very subtle impact on the translocation mechanism (Kintzer *et al.* 2009). The PA_8 oligomer, however, has the advantage of increased thermostability relative to the PA_7 oligomer, where the added interdomain contacts and surface burial in PA_8 oligomer provide additional stabilizing contacts (Kintzer *et al.* 2009; Kintzer *et al.* 2010). Thus the more stable PA_8 oligomer is less sensitive to premature channel formation under physiological temperature and pH conditions (Kintzer *et al.* 2010). This stability difference, however, is not observed when either PA oligomer is bound to the extracellular domain of its host cell receptor, and the receptor domain effectively stabilizes PA_7 and PA_8 oligomers through a similar mechanism (Miller *et al.* 1999; Lacy *et al.* 2004; Kintzer *et al.* 2010b). The octamer also represents a useful structural tool for crystallographic studies due in part to the architecture's increased thermostability and internal symmetry (Kintzer *et al.* 2009).

Using electron microscopy (EM), the structure of the PA_7 channel has been recently imaged (Figure 1.4) (Katayama *et al.* 2008). The PA channel is mushroom-shaped and approximately 170 Å tall × 125 Å wide at its maximum dimensions. The wider, cap-shaped part

of the structure is about 70 Å long and likely contains the LF/EF binding sites. Beneath the cap is a 100 Å long stem, which is likely a 14-strand β-barrel structure; the stem ultimately inserts into and spans the membrane bilayer (Katayama *et al.* 2010). Earlier electrophysiology studies suggested the stem is an extended β-barrel structure (Nassi *et al.* 2002). From basic modeling studies (Krantz *et al.* 2004), the β-barrel stem is likely only able to accommodate structures as wide as an α helix (10-15 Å-wide); therefore, LF and EF must unfold during translocation.

The PA translocase channel can be divided into three sections: the substrate docking surface in the cap, a critical hydrophobic constriction point about one-third of the distance into the translocase (called the ϕ-clamp site), and the highly charged solvophilic β-barrel stem portion, which comprises the bottom two-thirds of the translocase (Figure 1.6). Interestingly, there appear to be separate polypeptide interaction sites, or clamps, associated with these different points in the PA translocase channel. These clamps also allow the channel to interact with the substrate nonspecifically. Here, the notion of a clamp is considered to more closely resemble a dynamic binding site for polypeptide, where, for example, a clamp site can be modulated from a higher-affinity binding mode to a lower-affinity binding mode. This dynamic is critical to allowing the translocase to both favor unfolding and translocation but preventing tight binding interactions from occurring that would otherwise impede transport. In the following sections, these clamp structures and their associated activities are summarized.

The LF/EF binding sites. Because LF and EF unfold prior to translocation through the PA channel, much attention has been focused on understanding the interaction between PA and its substrates. These substrates translocate *via* their N-termini first; (Zhang *et al.* 2004), therefore, LF_N represents a useful model protein for studying PA-dependent unfolding and translocation. Residues implicated in binding and assembly both on the PA prechannel (Chauhan & Bhatnagar 2002; Cunningham *et al.* 2002) and on the substrates LF and EF (Lacy *et al.* 2002) were identified using biochemical and cellular assays (Mourez *et al.* 2003). Hydrogen/deuterium exchange mass spectrometry experiments determined the regions of LF_N that become solvent inaccessible upon formation of PA-LF_N complexes (Melnyk *et al.* 2006). Furthermore, using Rosetta-Dock (Schueler-Furman *et al.* 2005), a static computational dock between LF_N and the D1' region of a PA dimer model was generated. While the model was consistent with most available biochemical data (Lacy *et al.* 2005), it failed to account for the interaction of PA R178 with LF_N, which has strong biochemical support (Cunningham *et al.* 2002). A crystal structure describing the interaction between the PA oligomer and LF_N was recently solved (Feld *et al.* 2010), which is the focus of Chapter 2 of this thesis.

The ϕ clamp. The acidified endosomal trafficking route produces conditions that thermodynamically destabilize the substrates, LF and EF (Krantz *et al.* 2004), and causes them to populate molten-globule intermediates. To determine how LF and EF are linearized and fully unfolded during translocation, a series of experiments were performed to identify potential active-site residues lining the PA channel (Krantz *et al.* 2005). An interesting conserved residue, F427, was identified in an otherwise hydrophilic PA loop, N^{422}AQDD**F**SSTP. As most of the channel-lining residues are hydrophilic and/or anionic (Nassi *et al.* 2002), this aromatic site is unusual in terms of its chemistry and prominence in the interior of the PA channel. Structurally, each PA subunit contributes one F427 residue, and the collective ring of aromatic rings forms a hydrophobic bottleneck in the PA channel (Figure 1.6). The site, dubbed the "ϕ clamp," is critical for translocase function (Krantz *et al.* 2005). Analogous ϕ-clamp sites have been identified in numerous protein translocases, including soluble ones that hydrolyze ATP (Martin *et al.* 2008), revealing the general importance of the site to the mechanism of translocation.

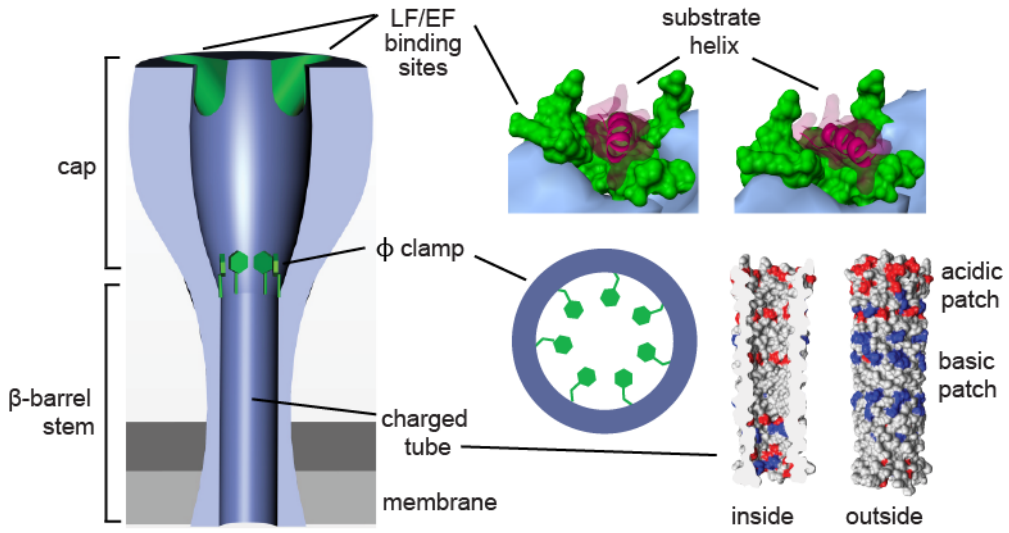

Figure 1.6. Peptide clamps in the PA channel. The PA channel (denim) contains clamping sites (green) and a β-barrel tube with positively (blue) and negatively (red) charged patches. Partially unfolded substrates (red-violet) first bind to the LF/EF sites (see Chapter 2). Subsequently, the peptide would thread into the ϕ clamp, which is comprised of a ring of F427 residues (Krantz *et al.* 2005). Finally, the substrate polypeptide chain encounters the β-barrel tube (Benson *et al.* 1998; Nassi *et al.* 2002), which may act as a charge clamp, attracting cationic sequences and preventing the retrotranslocation of deprotonated acidic residues.

PA's ϕ clamp has been characterized extensively by numerous biophysical techniques to better understand the molecular mechanism of translocation. The F427 residues radially converge within the channel, forming a symmetric ring of rings, which is very narrow. Single-channel electrophysiology studies confirmed that the ϕ clamp forms a conductance bottleneck (or constriction point) within the channel. As the ion conductance path and the protein translocation path are the same, it is unsurprising that defects arising from mutations to the ϕ clamp manifest in increased substrate diffusion, backsliding, and retrotranslocation. The retrotranslocation phenotype of these ϕ-clamp mutants likely explains their >1000-fold losses in translocation activity (Krantz *et al.* 2005); moreover, the ratchet-like behavior of the ϕ clamp indicates it is a dynamic clamping site, allowing translocation to proceed in one direction but opposing backsliding in the opposite.

The ϕ-clamp has broad substrate specificity with preference for hydrophobic aromatic compounds (Krantz *et al.* 2005). The broad specificity is consistent with the π-cloud electrons of the Phe residues making cation-π, π-π, and π-dipole interactions. Recent reports (Janowiak *et al.* 2010) indicate substitution of a single Phe residue in the ϕ clamp with a charged residue fully inactivates the translocase, which is consistent with a cooperative site. Protein folding studies have revealed that protein folding is highly cooperative and driven by the hydrophobic effect (Kauzmann 1959; Tanford 1968; Dill *et al.* 1993); therefore, it is reasonable to hypothesize that the ϕ clamp is a hydrophobic interaction site, operating similar to a protein chaperone, where the site interacts with a broad spectrum of peptide chemistries during translocation.

An analysis of known prechannel oligomeric crystal structures reveals that the ϕ-clamp loop is quite flexible and exists in two different conformations (Lacy *et al.* 2004; Kintzer *et al.* 2009; Feld *et al.* 2010). In the ligand-free structure of PA_8 (Protein Data Bank 3HVD) (Kintzer *et al.* 2009), both the ϕ-clamp loop and the hydrophilic loop containing K397, located directly above the ϕ-clamp loop, are tilted upward, relative to the $PA_8(LF_N)_4$ structure (3KWV, see Chapter 2) (Feld *et al.* 2010) and the PA_7 structure (1TZO) (Lacy *et al.* 2004). While these structures are of PA prechannels, some evidence supports the functional implications of coordinated loop movement (Melnyk & Collier 2006). A mutagenesis study of K397 and D426 found that these residues may make a salt bridge upon channel formation (Melnyk & Collier 2006). This salt bridge linking these two flexible loops may position the ϕ clamp in an active conformation (Melnyk & Collier 2006; Bann 2012), or the different configurations observed may be alternative binding modes of the ϕ clamp, i.e., a high affinity and a lower affinity state of the site. Proton binding may be able to modulate the salt-bridge interaction, altering the substrate affinity of the ϕ clamp site.

The β barrel. The length, diameter and electrostatic composition of the PA channel and its extended β-barrel stem likely play a key role in the molecular mechanism of translocation. While an atomic-resolution structure of the PA channel is currently unavailable, a β-barrel model (Nguyen 2004) has been proposed (Benson *et al.* 1998; Nassi *et al.* 2002). EM studies also show a tube-like stem consistent with a β-barrel architecture (Figure 1.4) (Katayama *et al.* 2008; Katayama *et al.* 2010). The β barrel's inner diameter is no wider than the width of an α helix (Krantz *et al.* 2004) and may stabilize helical structure due to favorable Van der Waals contacts and backbone desolvation. A number of charged residues populate both the inside and outside of the β barrel. These charges create a sawtooth electrostatic potential within the barrel (Figure 1.6). The precise role of the charged groups in the barrel is unknown; however, the charge-state BR mechanism (Figure 1.3a) proposes that differential electrostatic repulsion between the channel

and substrate polypeptide is critical to the overall mechanism of ΔpH-driven translocation.

1.5 Translocation powered by a proton gradient

The PA translocase machine is powered by the chemical potential component, ΔpH, of the PMF (Krantz *et al.* 2006; Brown *et al.* 2011). While a pure Δψ can drive the translocation of the LF's amino-terminal domain, LF_N (Zhang *et al.* 2004), the ΔpH is required for the translocation of the full-length substrates, LF and EF (Krantz *et al.* 2006), and a ΔpH alone is sufficient for translocation (Brown *et al.* 2011). A charged-state BR model was initially proposed for the molecular mechanism of ΔpH-driven translocation (Figure 1.3a) (Krantz *et al.* 2006). The model is based on the fact that the PA channel itself is cation-selective (or anion-repulsive) (Blaustein *et al.* 1989), and yet LF and EF are anionic substrates with isoelectric points of 5.4 and 6.6, respectively. (More importantly, their amino-terminal domains, LF_N and EF_N, have pI values of 4.9.) Krantz *et al*. proposed this paradox is resolved if acidic residues can be protonated during their translocation through the PA channel, thereby allowing their anionic charges to be neutralized (Krantz *et al.* 2006). The ΔpH naturally favors substrate protonation on the endosomal side of the membrane, where the pH is ~5.5, but once the substrate reaches the higher pH side of the membrane (neutral cytosolic pH), these acidic residues would spontaneously deprotonate. Thus, while these protonated acidic residues could then pass through the channel's cation-selective filter, they would be trapped on the opposite side of the membrane upon substrate deprotonation. In this mechanism, substrate motion is largely explained by Brownian motion, and the ratchet is an electrostatic trap created via cycles of acidic residue protonation and deprotonation on either side of the channel's charge-selectivity filter.

Several recent studies support the charged-state BR model of translocation. In one study, it was shown that sulfate groups attached to LF_N via cysteine linkage inhibited Δψ-mediated translocation (Basilio *et al.* 2009). Since the sulfate moiety essentially cannot be protonated under the experimental conditions, the authors concluded that only titratable negative charges could pass through the cation-selective channel. Another study reached a similar conclusion by incorporating cysteic acid residues in a semisynthetic LF_N construct (Pentelute, B. L. *et al.* 2010). Furthermore, semisynthetic LF_N constructs lacking titratable acidic residues in the amino-terminal presequence display significant translocation defects (Pentelute, B. L. *et al.* 2010). A recent report extensively probed the role of both positively and negatively charged residues in LF_N's presequence (Brown *et al.* 2011). Interestingly, substrates lacking acidic residues could be translocated by a Δψ alone, yet they do not display ΔpH-dependent translocation. This evidence strongly supports the model that the ΔpH drives translocation by protonating acidic residues and is consistent with the charge-state BR model. Furthermore, it was demonstrated that basic residues likely act as chaperones for deprotonated acidic residues. That is, segments of sequence containing high densities of negative charge will not enter the PA channel as efficiently as those containing positively-charged residues in the vacinity.

The ΔpH plays a role not only in translocation but also in the unfolding of the substrate. As discussed above, the thermal energy of even a small polypeptide is quite large (over 100 kJ mol^{-1}) and, if properly rectified, can lead to significant force generation (tens of piconewtons), which is sufficient to accelerate the protein unfolding process. The most ΔpH-dependent step of translocation is associated with the protein folding barrier (Thoren *et al.* 2009), which is consistent with the observation that a ΔpH alone is sufficient to unfold and translocate a fully folded substrate (Brown *et al.* 2011). Furthermore, acidic residues in LF_N are naturally concentrated in an optimal position immediately prior to the substrate's folding domain and

separating this region of optimal charge density from the folded domain greatly inhibits translocation (Brown *et al.* 2011). Taken together, these results imply a model for ΔpH-dependent unfolding. As the substrate presequence is translocated and captured on the high pH side of the membrane, the remainder of the polypeptide is caught in an extended state with more limited conformational options. This results in entropic tension that is only relieved via substrate unfolding. The details of this mechanism are presented in further detail in Chapter 4.

1.6 Translocation-coupled unfolding

The narrow architecture of the PA translocase channel necessitates that substrates must unfold prior to translocation. Critical to understanding protein translocation is determining how driving forces are coupled to substrate unfolding. Thoren *et al*. probed the folding step by site-directed mutagenesis, thermodynamic stability studies, and planar lipid bilayer electrophysiology (Thoren *et al.* 2009). The ΔpH- or Δψ-driving-force dependencies of the activation energy of translocation ($\Delta G^{\ddagger}$) revealed a boomerang-shaped curve with two distinct limiting slopes. At low driving force, the $\Delta G^{\ddagger}$ is 10-fold more force dependent than at high driving forces. Thus two major barriers are crossed during translocation: one that is strongly force-dependent and one that is largely force-independent. Thoren *et al*. were then able to show that substrate unfolding occurs at the more force-dependent barrier and little unfolding occurs at the more force-independent barrier (Thoren *et al.* 2009).

By examining destabilizing mutations throughout substrate protein's structure, the location of the folded substructure that is rate-limiting to the unfolding step of translocation was identified. For LF_N, the substrate's β-sheet subdomain corresponds to this "mechanical breakpoint" region. β-sheet regions often represent the rate-limiting mechanical breakpoint in a force-dependent unfolding mechanism, as probed by optical force microscopy; and the orientation and topology of β-sheet substructures are key determinants of the forces required for unfolding (Brockwell *et al.* 2003) .

Finally, because so little unfolding appears to occur in the less force-dependent step, it has been proposed that this step is limited by the translocation of the unfolded chain (Thoren *et al.* 2009). The translocation barrier by its nature of having a shallow force dependence cannot be reduced significantly and places an overall speed limit on the rate of protein translocation for anthrax toxin (Krantz *et al.* 2006; Thoren *et al.* 2009) and other systems (Huang *et al.* 1999; Burton *et al.* 2001; Kenniston *et al.* 2003). Despite the fact that other translocases use ATP as their source of free energy, it appears the translocation step is bottlenecked in the range of ~1-10 s for protein substrates that range in size from 100 to 800 residues in length. Most likely, this bottlenecking limits the ultimate rate of translocation and reflects that the overall force-independence of the post-unfolding event is a general feature in many systems.

A different model proposed by Basilio *et al*. proposes that translocation is not a barrier-limited process but rather an electrodiffusive movement of a "charged rod" through the channel (Basilio *et al.* 2011). By definition, this particular model does not consider unfolding during LF translocation to be rate-limiting. It is difficult to reconcile this model with the majority of published studies on protein unfolding, which show, in fact, that unfolding is barrier limited (Krantz *et al.* 2002) and force-dependent (Brockwell *et al.* 2003; Cecconi *et al.* 2005; Borgia *et al.* 2008; Crampton & Brockwell 2010). Forces along the translocation axis of a membrane-embedded translocase are readily achieved for charged ions, q, in an electric field, E, by $F = Eq$. Similarly, optical force-microscopy measurements of substrate proteins being pulled through ATP-dependent unfolding machines show significant barriers to unfolding, where the most

force-dependent step is related to unfolding (Maillard *et al.* 2011). Thus, studies of soluble ATP-dependent unfolding machines agree more closely with the conclusions of Thoren *et al*. (Thoren *et al.* 2009) and are less consistent with the charged-rod hypothesis proposed by Basilio *et al*. (Basilio *et al.* 2011).

1.7 Processing unfolded protein

The nature of unfolded polypeptide presents significant complexities and challenges for the nanomachine, including: (i) chemical complexity (Krantz *et al.* 2005; Tian *et al.* 2005) and (ii) configurational complexity, diffusion, and orientation (Thoren *et al.* 2009). By contrast, the nanomachine kinesin moves along a very specific polymer track, a microtubule. The regular chemical periodicity of microtubule tracks provides convenient, highly organized highways for kinesin to transverse (Gennerich & Vale 2009). Dealing with the chemical and configurational complexity of an unfolded protein polymer implies that the fundamental challenge the translocase faces is entropic in nature. In order to do work on a reasonable timescale, protein translocases must contend with entropy.

The catalytic power of orienting substrates in an enzyme reaction was initially described by Page and Jencks, where enzyme catalyzed rates are accelerated by the appropriate orientation of reactive chemical groups (Page & Jencks 1971). This orientation effect, referred to as an "entropy trap," was considered to be critical to the observed rate accelerations of enzyme-catalyzed reactions. The entropy trap results from the exchange of enthalpic interactions in an enzyme's active site binding pockets with the substrates; these binding interactions can reduce the overall rotational and translational degrees of freedom and accelerate the chemical reaction. The individual chemical steps of anthrax toxin translocation are protonation, deprotonation, and translocation (Figure 1.3a). It is likely these steps require oriented and somewhat ordered substrates, thereby reducing the conformational search and subsequently lowering the $\Delta G^{\ddagger}$ for each reaction. Efficient, well-designed protein translocases are enzyme catalysts, and as such, they would tend to maximize ordered motion, and minimize heat dissipation, to the greatest extent possible and within the limits of the Second Law. As concluded by Jencks, the more an enzyme minimizes counterproductive motions of its substrate, the more significant acceleration in catalysis is realized.

In Chapter 2, the crystal structure of a PA prechannel complex containing PA_8 and LF_N is presented. In the structure, the N-terminal α1 helix and β1 strand of LF_N unfold from their native conformations and dock into a novel protein-binding site, called the "α clamp," on the surface of the PA oligomer. Along with my colleague, Dr. Katie Thoren, and co-workers we characterized the functionality of the α clamp. Our results describe a binding site that is particularly equipped to handle the challenges of dealing with unfolded polypeptide mentioned above. Much like how kinesin's motor functionality is facilitated by the regularity of the microtubule track it transverses, we believe that translocase channels simplify the task of protein unfolding and translocation by ordering their substrate chains into regular structure, such as an α helix.

Of course, before a multimeric translocase channel can catalyze the movement of polypeptide substrates, it must first assemble into an active nanomachine. In the case of anthrax toxin, my co-workers and I have shown that PA forms two oligomeric states: the previously known heptamer and a novel octamer. In Chapter 3, a molecular mechanism for stoichiometric preference is described, whereby the positioning of the flexible D4 relative to the main body of PA dictates the final oligomeric state. Moreover, the models derived from high-resolution (~1.5 Å) structures contain the furin-cleavage site, providing a structural basis for furin-dependent

activation of proproteins, including the protoxin PA.

Finally, in Chapter 4 I present conclusions of this thesis in regards to protein translocation in the anthrax toxin system. We believe that the mechanisms we have deduced from the anthrax toxin translocation system are universal and should apply broadly to other translocation systems that require the nanomachine to handle unfolded polypeptide.

Chapter 2

Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers

2.1 Abstract

The protein transporter, anthrax lethal toxin, is comprised of protective antigen (PA), a transmembrane translocase, and lethal factor (LF), a cytotoxic enzyme. Following assembly into holotoxin complexes, PA forms an oligomeric channel that unfolds LF and translocates it into the host cell. We report the crystal structure of the core of a lethal toxin complex to 3.1-Å resolution; the structure contains a PA octamer bound to four LF PA-binding domains (LF_N). The first α helix and β strand of each LF_N unfold and dock into a deep amphipathic cleft on the surface of the PA octamer, which we call the α clamp. The α clamp possesses nonspecific polypeptide binding activity and is functionally relevant to efficient holotoxin assembly, PA octamer formation, and LF unfolding and translocation. This structure provides insight on the mechanism of translocation-coupled protein unfolding.

2.2 Introduction

Protein secretion and degradation are essential cellular processes that allow for protein trafficking, organelle biogenesis, protein quality control, and cell-cycle regulation (Sauer *et al.* 2004; Wickner & Schekman 2005; Navon & Ciechanover 2009). Since folded proteins are thermodynamically stable under typical conditions, these processes often require complex, energy-consuming molecular machines (Matouschek 2003; Sauer *et al.* 2004; Young & Collier 2007; Cheng 2009), which catalyze a series of unfolding and translocation reactions (Huang *et al.* 1999; Huang *et al.* 2002; Kenniston *et al.* 2003; Krantz *et al.* 2005; Krantz *et al.* 2006; Martin *et al.* 2008; Thoren *et al.* 2009). Anthrax toxin (Smith & Keppie 1954; Young & Collier 2007), a three-protein virulence factor secreted by *Bacillus anthracis*, is an example of such a transmembrane protein delivery system. This bacterial toxin follows the classical two-component *AB* paradigm, where the *A* component is an active enzyme that localizes to and enters cells by forming complexes with the cell-binding, or *B* component. Anthrax toxin is composed of two *A* components, LF (91 kDa) and edema factor (EF, 89 kDa), and one *B* component, PA (83 kDa). Therefore, two different toxic complexes can form: lethal toxin (LT, PA plus LF) and edema toxin (ET, PA plus EF). LT (which we focus on herein) causes macrophage lysis (Friedlander 1986), immune system suppression (Agrawal & Pulendran 2004), and death (Smith & Keppie 1954).

For LT to inflict its cytotoxic effects, PA and LF must assemble into active holotoxin complexes, which can translocate LF into host cells. Proteases present either on host-cell surfaces or in blood serum potentiate LT assembly by proteolytically nicking PA, yielding $_nPA$ (Ezzell & Abshire 1992; Milne *et al.* 1994; Kintzer *et al.* 2009). Dissociation of a 20-kDa amino-terminal fragment from $_nPA$ exposes LF-binding sites, permitting assembly. The resulting LT complex contains multiple copies of LF bound to either a ring-shaped PA homoheptamer, PA_7 (Milne *et al.* 1994; Petosa *et al.* 1997; Katayama *et al.* 2008; Kintzer *et al.* 2009), or homooctamer, PA_8 (Kintzer *et al.* 2009). Octameric PA forms more robust LT complexes than heptameric PA under physiological conditions (Kintzer *et al.* 2010). The crystal structures of the individual PA and LF monomers (Petosa *et al.* 1997; Pannifer *et al.* 2001) and the assembled PA

heptamer (Lacy *et al.* 2004) and octamer (Kintzer *et al.* 2009) are known. However, an atomic-resolution X-ray crystal structure of a lethal toxin co-complex has not been described.

After the LT complex is endocytosed, the PA oligomer transforms into a transmembrane, β-barrel channel (Benson *et al.* 1998) through which LF translocates to enter the cytosol. Due to the narrowness of the channel, LF unfolds during translocation. The acidic endosomal pH conditions required for toxin action (Friedlander 1986) not only aid in the destabilization of LF (Krantz *et al.* 2004) but also drive further LF unfolding (Thoren *et al.* 2009) and translocation by means of a proton motive driving force (Krantz *et al.* 2006). This driving force is comprised of a proton gradient (ΔpH) and membrane potential (ΔΨ). Efficient coupling of the ΔpH requires a catalytic active site in the channel, called the ϕ clamp, composed of a narrow ring of phenylalanine residues (Krantz *et al.* 2005; Krantz *et al.* 2006). The ϕ clamp forms a narrowly apposed substrate clamping site in the central lumen of the PA channel (Krantz *et al.* 2005), and it allows the channel to catalyze unfolding (Thoren *et al.* 2009) and translocation (Krantz *et al.* 2005) presumably by forming transient interactions with the unfolded translocating chain (Krantz *et al.* 2005).

Many, but not all, protein processing machines that translocate, unfold and/or refold proteins utilize analogous polypeptide clamping features to denature a protein and engage with its unfolded structure. The features that bind to unstructured or unfolded polypeptides include hydrophobic/aromatic pore loops (Wang, J. *et al.* 2001; Van den Berg *et al.* 2004; Krantz *et al.* 2005; Lum *et al.* 2008; Martin *et al.* 2008), polypeptide clamping sites (Krantz *et al.* 2005; Zimmer *et al.* 2008), and other substrate binding clefts or adapters (Levchenko *et al.* 2003; Hinnerwisch *et al.* 2005; Levchenko *et al.* 2005). Some of these machines utilize tandem polypeptide binding sites (Hinnerwisch *et al.* 2005; Krantz *et al.* 2005; Thoren *et al.* 2009): one site is a substrate docking site, which feeds into a second hydrophobic site found deeper within the pore. Questions surround the mechanism of these clamping sites and their interactions with unfolded substrates. How do these sites unfold proteins? How do they process the wide chemical complexity and configurational flexibility contained in an unfolding substrate? These questions have remained unanswered, in part because atomic resolution structures of unfolding intermediates in complex with these clamps have not been described. Here we report a structure of a partially unfolded substrate, the PA-binding domain of LF, in complex with its unfolding machine, the PA oligomer.

2.3 Results

Crystal structure of the $PA_8(LF_N)_4$ complex. For these crystallographic studies, we focused on the PA_8 oligomer, considering its enhanced thermostability as well as its advantageous fourfold, square-planar symmetry (Kintzer *et al.* 2009). All of the crystallography work was done by Geoff Feld, and the mass spectrometry data was collected by Harry Sterling in the Williams lab. By mass spectrometry, we find that the PA_8LF_4 complex is physiologically relevant, as it assembles from the full-length, wild-type (WT) PA and LF subunits (Figure 2.1a). Our best diffracting crystals contain LF_N (LF residues 1-263) and a PA construct lacking its membrane-insertion loop (Kintzer *et al.* 2009), which is superfluous to the known PA-LF_N interaction (Cunningham *et al.* 2002). LF_N, the minimal portion of LF that specifically binds PA (Arora & Leppla 1993), can translocate heterologous domains as amino- or carboxy-terminal fusions into cells (Arora & Leppla 1994; Milne *et al.* 1995). EF contains a homologous PA-binding domain, and the PA-LF_N interaction is likely general to LT and ET complexes (Lacy *et al.* 2002). Homogenous $PA_8(LF_N)_4$ complexes (Figure 2.1b) form crystals in the $P42_12$ space

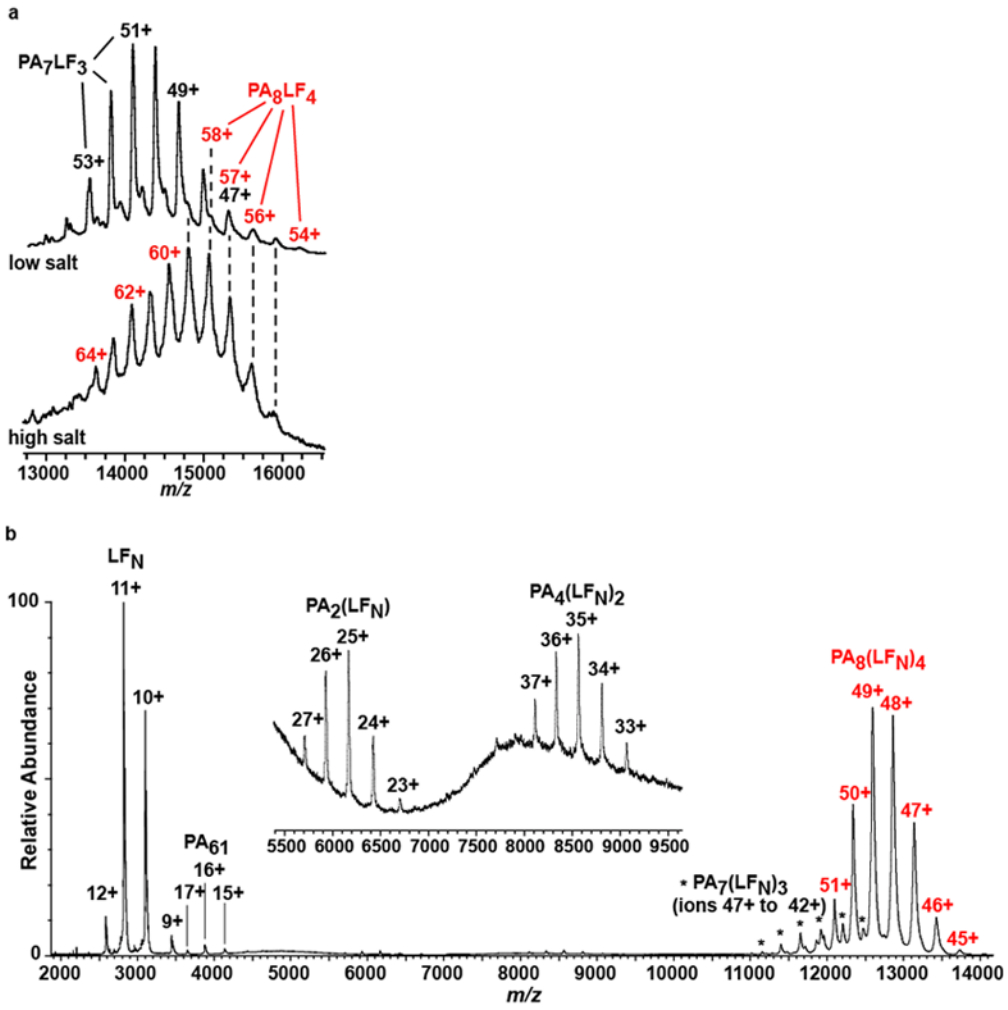

Figure 2.1. Mass spectrometry analysis of the PA_8LF_4 and $PA_8(LF_N)_4$ complexes. (a) LT complexes were assembled from WT $_n$PA and WT LF for 1 hour at room temperature at pH 7.8 in either low salt (150 mM ammonium acetate, upper panel) or high salt (300 mM ammonium acetate, lower panel). NanoESI-MS spectra of the LT assembly reactions were obtained. Molecular masses of 718,726 (±491) and 873,881 (±288) Da are obtained from the charge-state distributions (or *m/z* peaks), which correspond to the PA_7LF_3 (black) and PA_8LF_4 (red) complexes, respectively. **(b)** The PA construct used in the crystallographic experiments, $PA^{\Delta MIL}$, is a membrane-insertion-loop-deleted construct. This $PA^{\Delta MIL}$ construct was assembled, and the octameric assembly state was purified (Kintzer *et al.* 2009). PA octamers were liganded with LF_N at a 1:1 stoichiometric ratio (LF_N:PA monomer) to make $PA_8(LF_N)_4$ complexes. The $PA_8(LF_N)_4$ complexes were analyzed by mass spectrometry. The mass spectrum revealed the largest relative abundances for free LF_N and the octameric complexes, $PA_8(LF_N)_4$. The octameric complex's *m/z* peaks are numbered in red. Low relative abundances were observed for PA_{20}-dissociated $PA^{\Delta MIL}$ monomer (PA_{61}) and $PA_7(LF_N)_3$ as well as $PA_2(LF_N)$ and $PA_4(LF_N)_2$ (magnified inset). The latter two species are likely intermediates in the assembly of the $PA_8(LF_N)_4$ complex (Kintzer *et al.* 2009).

group that diffract X-rays to 3.1 Å (Table 2.1). Molecular replacement solutions indentify two PA_2 complexes and significant (2.7σ) unassigned electron density (F_o-F_c) for α helices located proximal to the domain 1′ (D1′) surface of each PA_2 complex. Rounds of polyalanine-helix modeling and refinement reveal that the novel helical density aligns well with α2, α4, α9, and α10 of LF_N. The two occurrences of the PA_2LF_N ternary complex (Figure 2.2a) in the asymmetric unit are structurally identical; its PA subunits are structurally similar to the full-length PA monomer (Petosa *et al.* 1997) and the PA subunits observed in the PA_7 and PA_8 prechannel oligomers (Lacy *et al.* 2004; Kintzer *et al.* 2009). Thus the biological unit—the $PA_8(LF_N)_4$ prechannel complex (Figure 2.2b)—is comprised of four PA_2LF_N ternary complexes (Figure 2.2c).

Interestingly, LF_N α1/β1 (residues 29-50) unfolds and adopts a novel conformation relative to free LF (1J7N (Pannifer *et al.* 2001)). LF_N α1/β1 docks in the cleft formed between adjacent PA subunits and aligns well with the experimental electron density (Figure 2.3a,b). We can assign this unique conformation of α1/β1 since it extends from LF_N α2 as a contiguous stretch of electron density contoured at σ=1 (Figure 2.4). LF_N's carboxy terminus also reveals well-defined electron density (Figure 2.3c). Overall, LF_N excludes 1900 $Å^2$ of solvent accessible surface area (SASA) on the PA dimer. This surface is comprised of two discontinuous LF_N-binding subsites (Figure 2.2a) formed by adjacent PA subunits, termed PA_N and PA_C (to reflect whether the PA subunit interacts primarily with the amino terminus or carboxy terminus of LF_N, respectively). The details of these respective subsites, called the α-clamp binding subsite and the carboxy-terminal binding subsite, are depicted in Figure 2.6. Thus upon binding the PA oligomer, LF_N partially unfolds, whereby its first α helix and β strand (i) separate from the main body of the protein, (ii) dock into the cleft between two adjacent PA subunits (Figure 2.2a), and (iii) orient toward the center of the PA oligomer lumen (Figure 2.2b).

The carboxy-terminal binding subsite. At the carboxy-terminal subsite, LF_N's carboxy-terminal subdomain excludes ~900 $Å^2$ on PA_C (Figure 2.6b). The structure reveals a hydrophobic interface, involving PA_C Phe202, Pro205, Ile207, and Ile210 and LF Val232, Leu235, His229, Tyr223, Leu188, and Tyr236. In particular, LF Tyr236 is well packed against PA_C Ile210 (Figure 2.3c) and its phenol hydroxyl forms a hydrogen-bonding network with PA_C His211 and Asp195 near the center of the hydrophobic interface (Figure 2.6b). Additional electrostatic interactions surround this hydrophobic core. The carboxyl side chain of PA_C Glu190 forms a pair of hydrogen bonds with both the γ hydroxyl and amide nitrogen of LF Thr141; PA_C Lys197, Lys213, Lys214 and Lys218 form salt bridges with LF Asp182, Asp187, Asp184, and Glu142, respectively; and PA_N Arg200 forms a salt bridge with LF Glu139. PA and LF residues localized in this binding subsite are corroborated by mutagenesis studies, probing binding (Figure 2.7a,b), assembly/binding (Figure 2.8a), (Chauhan & Bhatnagar 2002; Cunningham *et al.* 2002; Lacy *et al.* 2002; Lacy *et al.* 2005; Melnyk *et al.* 2006) and cytotoxicity (Chauhan & Bhatnagar 2002) (Figure 2.8b,c).

The α-clamp binding subsite. At the α-clamp subsite, PA_N and PA_C interact with LF_N's unfolded α1 and β1 structures (Figure 2.6a). Remarkably, hydrogen bonds lost upon LF_N unfolding are reformed on the surface of PA: LF_N α1 maintains a similar helical conformation; and LF_N β1 (Ile43 and Lys45) forms parallel β-sheet hydrogen bonds with Leu203 in PA_N β13 (Figure 2.3b). PA_N Pro205, which is positioned at the end of PA_N β13, terminates the parallel-sheet interactions with LF_N β1. Overall, LF_N α1/β1 excludes 1000 $Å^2$ of SASA on PA. LF_N α1 is docked deep into the α-clamp cleft at the interface of adjacent PA subunits (Figures 2.2a and 2.6a). Reminiscent of calmodulin complexes with peptide helices

Table 2.1. **Data collection and refinement statistics**

	$PA_8(LF_N)_4$ [a]
Data collection	
Space group	$P42_12$
Cell dimensions	
a, *b*, *c* (Å)	178.38, 178.38, 240.36
α, , β,, γ (°)	90, 90, 90
Resolution (Å)	49.8-3.1(3.2-3.1)[b]
$R_{p.i.m.}$	6.9(46.0)
$I / \sigma I$	11.4(2.2)
Completeness (%)	92.0(78.0)
Redundancy	7.9(8.0)
Refinement	
Resolution (Å)	49.8-3.1
No. reflections	65,165
R_{work} / R_{free}	24.9/ 28.1
No. atoms	
Protein	20,397
Ligand/ion	8
Water	4
B-factors	
Protein	100.7
Ligand/ion	53.3
Water	56.7
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.610

[a]Data for this complex were collected from a single crystal.
[b]Values in parentheses are for the highest-resolution shell.

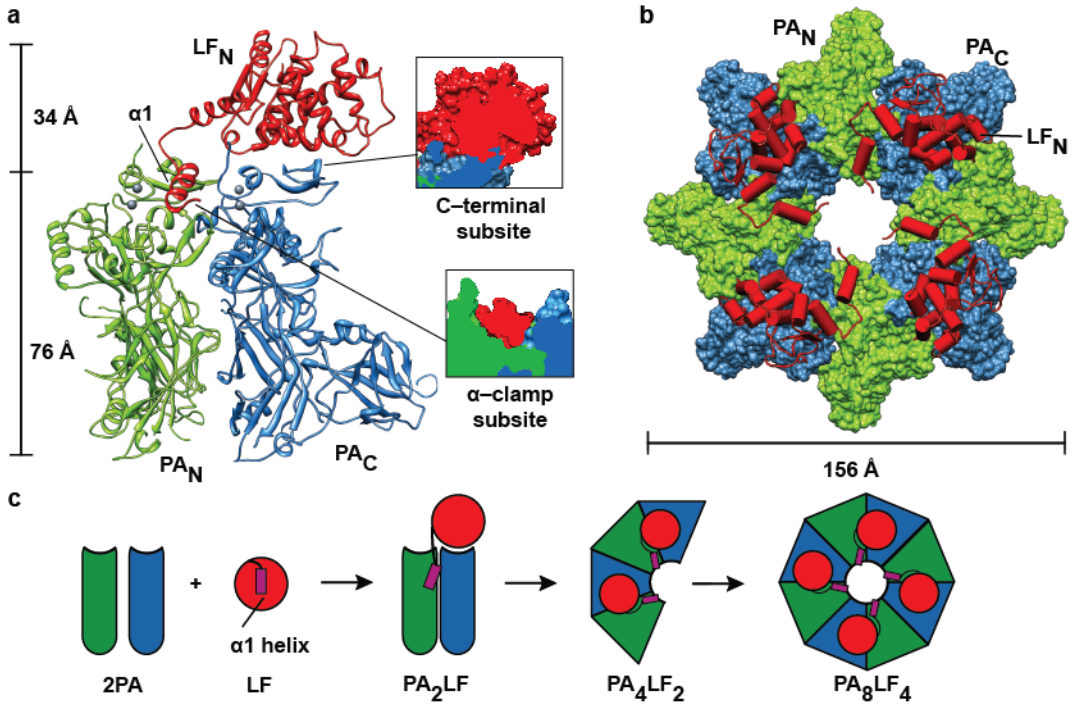

Figure 2.2. Structure of LF's PA-binding domain in complex with the PA octamer. **(a)** (Left) Ribbon depiction of the PA_2LF_N ternary complex. LF_N (red), PA_N (green), PA_C (blue), and calcium ions (gray spheres). (Right) Slices through a surface rendering of the two LF_N-binding subsites, with the carboxy-terminal binding subsite at the top and the α-clamp subsite at bottom. **(b)** Axial rendering of the biological unit, the $PA_8(LF_N)_4$ complex, colored as in (a). The PA octamer is shown as a molecular surface, and LF_N's helices and strands are cylinders and planks, respectively. **(c)** LF_N α1-β1 binds the α-clamp subsite formed at the interface of two PA subunits, driving the assembly of dimeric and tetrameric PA intermediates (Kintzer *et al.* 2009), which in turn form PA_8 complexes.

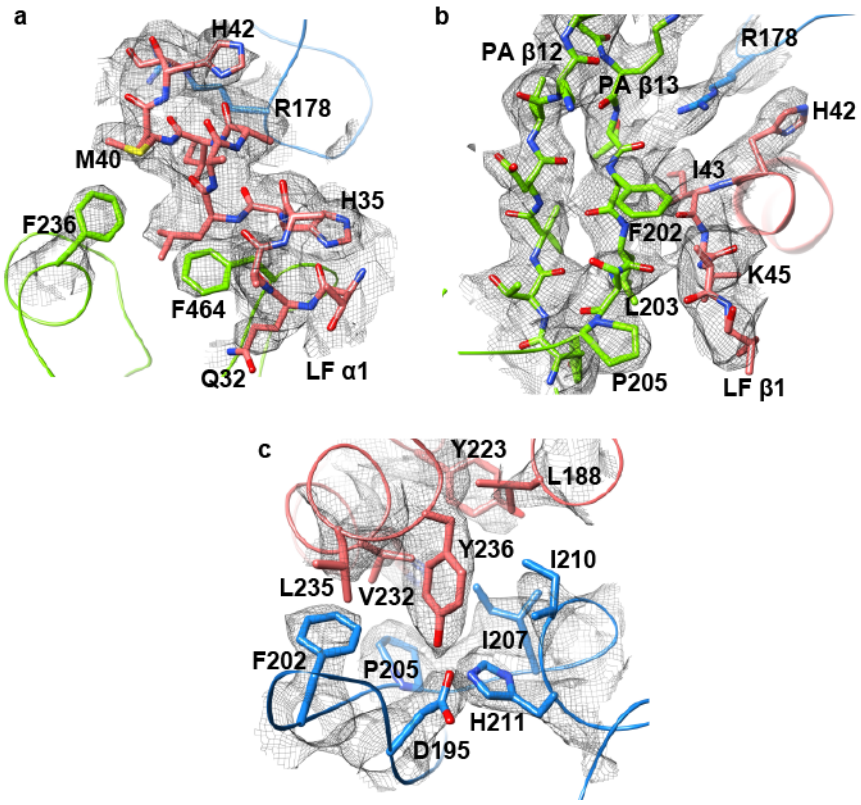

Figure 2.3. LF_N electron density in the $PA_8(LF_N)_4$ complex. A composite simulated-annealing (SA) omit map calculated in PHENIX (Adams *et al.* 2004) to 3.1 Å contoured at σ=1 (gray mesh). The models of PA_N, PA_C and LF_N are rendered in green, blue, and red, respectively. Secondary structure elements and individual residues are labeled. Nitrogen, oxygen and sulfur atoms are colored blue, red and yellow, respectively. **(a)** LF_N α1 (residues 31-42) in complex with PA_N. Lysine and glutamate residues are truncated to Cβ for clarity. **(b)** LF_N β1 in complex with PA_N β12-13. LF_N Lys45 is truncated to Cβ for clarity. **(c)** LF_N's carboxy-terminal binding subsite interaction with PA_C. Additional stereo-pair images of LF_N omit maps following SA refinement are depicted in Figure 2.5.

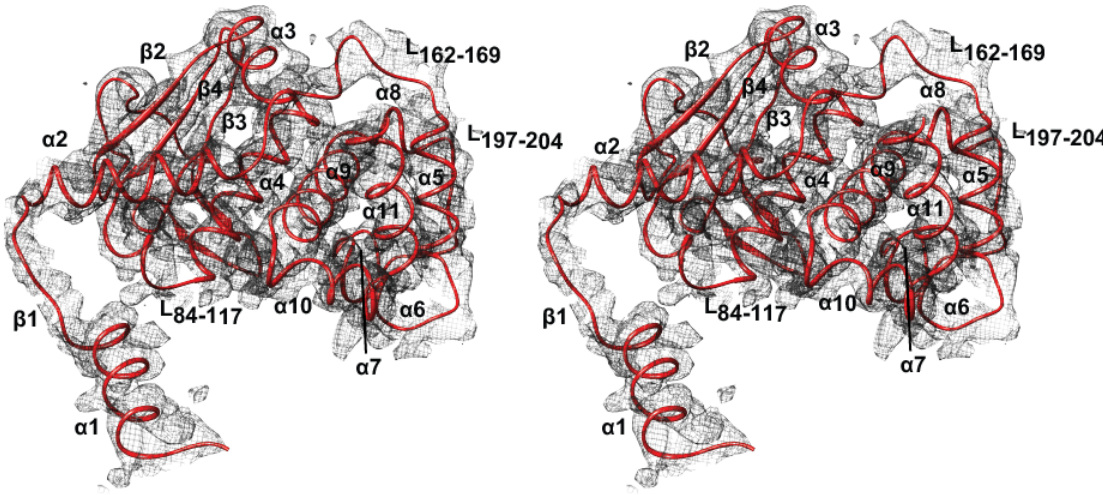

Figure 2.4. Stereo image for LF_N electron density in the $PA_8(LF_N)_4$ complex. A composite simulated-annealing omit map was calculated in PHENIX (Adams *et al.* 2004) (1 σ contour, gray mesh) for LF_N (chain C, residues 29-250), which is rendered as a red ribbon. Its secondary structure elements and loop regions (designated L_{x-y} where the indices are residue numbers) are labeled. Note the other LF_N in the asymmetric unit (chain F) has an identical structure and corresponding electron density.

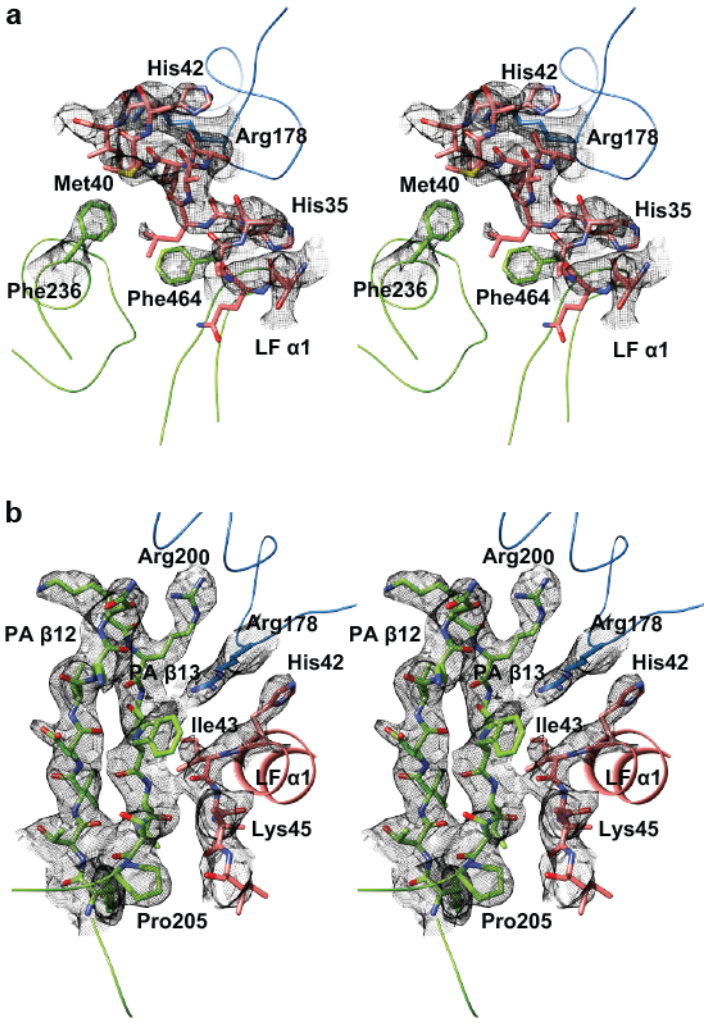

Figure 2.5. Stereo images for α-clamp electron density. (a) LF's α1 interaction with α clamp phenylalanine residues and **(b)** LF's β1 in a parallel β-sheet interaction with PA β13. The electron density (gray mesh) is a 2*F*o-*F*c simulated-annealing omit map calculated in PHENIX to 3.1 Å and contoured at σ=1 with LF_N residues 31-45 omitted. PA_N and PA_C are rendered in green and blue, respectively.

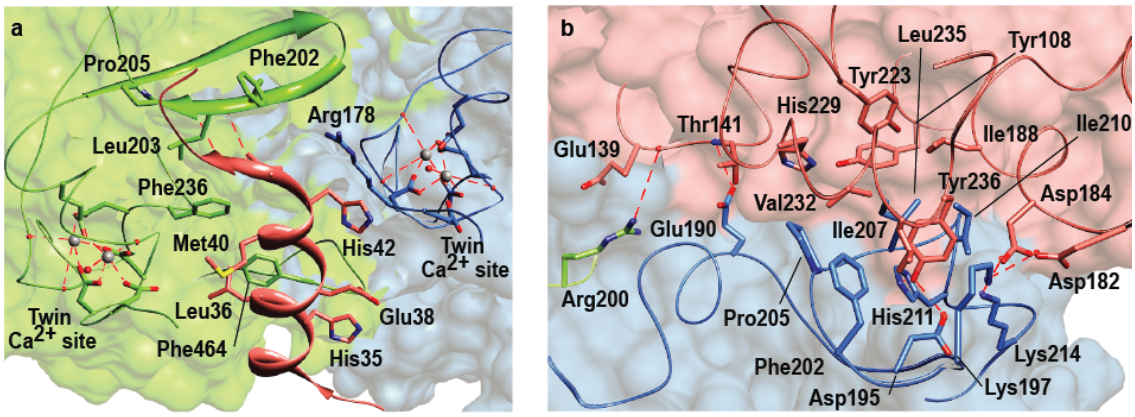

Figure 2.6. The PA octamer binds LF_N in two distinct subsites. Detailed views of **(a)** the α-clamp binding subsite and **(b)** the carboxy-terminal binding subsite are depicted. Highlighted non-covalent interactions are indicated with red dashed lines. Chains and Ca^{2+} ions are colored as in Figure 2.2a; nitrogen, oxygen and sulfur atoms are colored blue, red, and yellow, respectively.

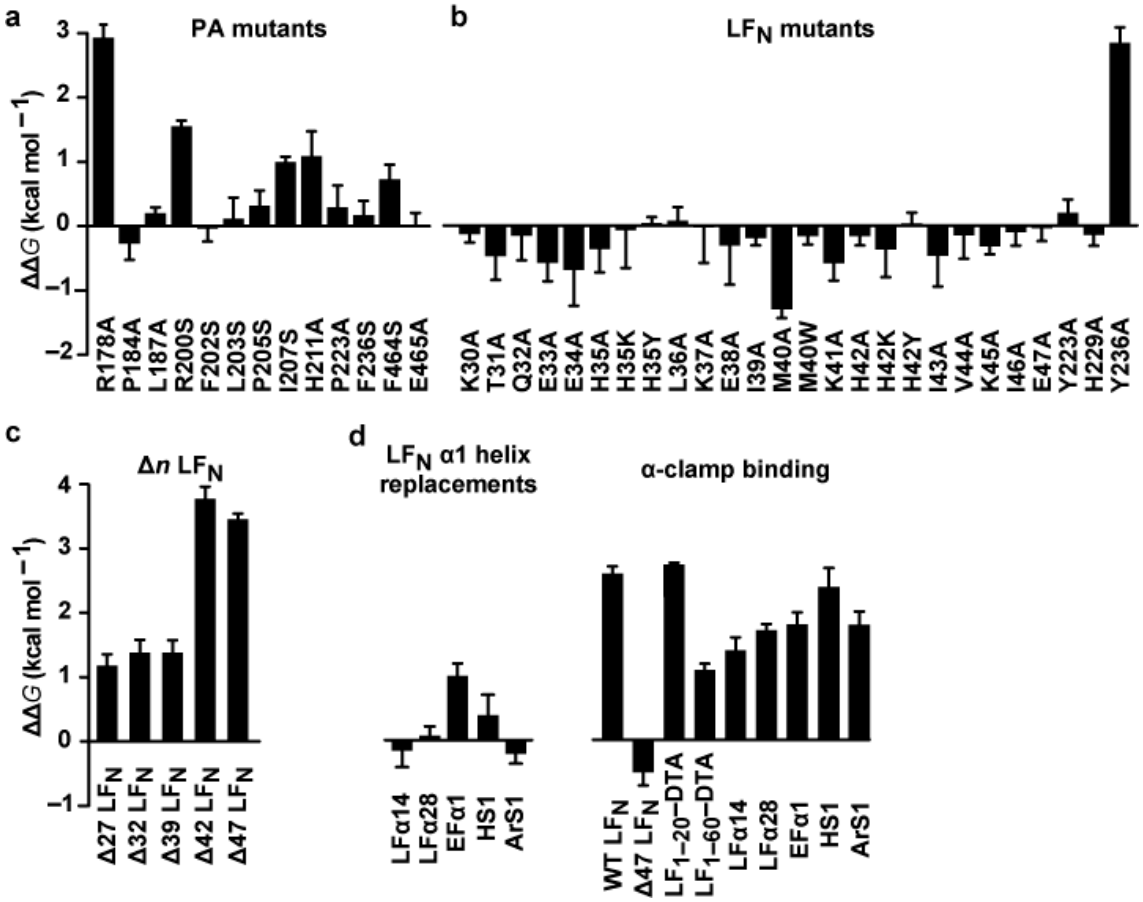

Figure 2.7. Changes in equilibrium binding free energy (ΔΔ*G*) for PA channel complexes, comparing **(a)** site-directed mutants of PA, **(b)** site-directed mutants of LF_N, and **(c)** Δ*n* LF_N amino-terminal truncation mutants. In (a-c), the reference state is WT LF_N:WT PA. **(d)** (Left) LF_N α1/β1-replacement mutant binding to WT PA; ΔΔ*G* values are referenced to WT LF_N. (Right) $LF_{1\text{-}20}$-DTA, $LF_{1\text{-}60}$-DTA, Δ47 LF_N and LF_N α1/β1-replacement mutant binding to PA R178A; ΔΔ*G* values are referenced to WT PA. LF_N α1/β1-replacement mutants either include multiple point mutations in the α1/β1 sequence (^{32}QEEHLKEIMKHIVK^{46}I) or replacements of the α1/β1 sequence with other sequences from LF or EF. The name, replacement sequence, and sequence identity (%) are listed for each: LFα14, SEEGRGLLKKLQI (23%); LFα28, NSKKFIDIFKEEG (23%); EFα1, EKEKFKDSINNLV (31%); hydrophilic sequence 1 (HS1), QEEHSKEISKHSVKS (73%); aromatic sequence 1 (ArS1), QEEHFKEIFKHFVKF (73%). See Figure 2.12 for alignments and helical-wheel depictions of the α1/β1-replacement sequences. In (a-d), $\Delta\Delta G = RT \ln K_d^{MUT} / K_d^{WT}$, where the equilibrium dissociation constants (K_d) were measured for the mutant (MUT) and WT proteins at pH 7.4, ΔΨ = 0 mV (Figure 2.11); *R* is the gas constant; and *T* is the temperature. The error bars are the mean ±s.d. (*n* = 2-6).

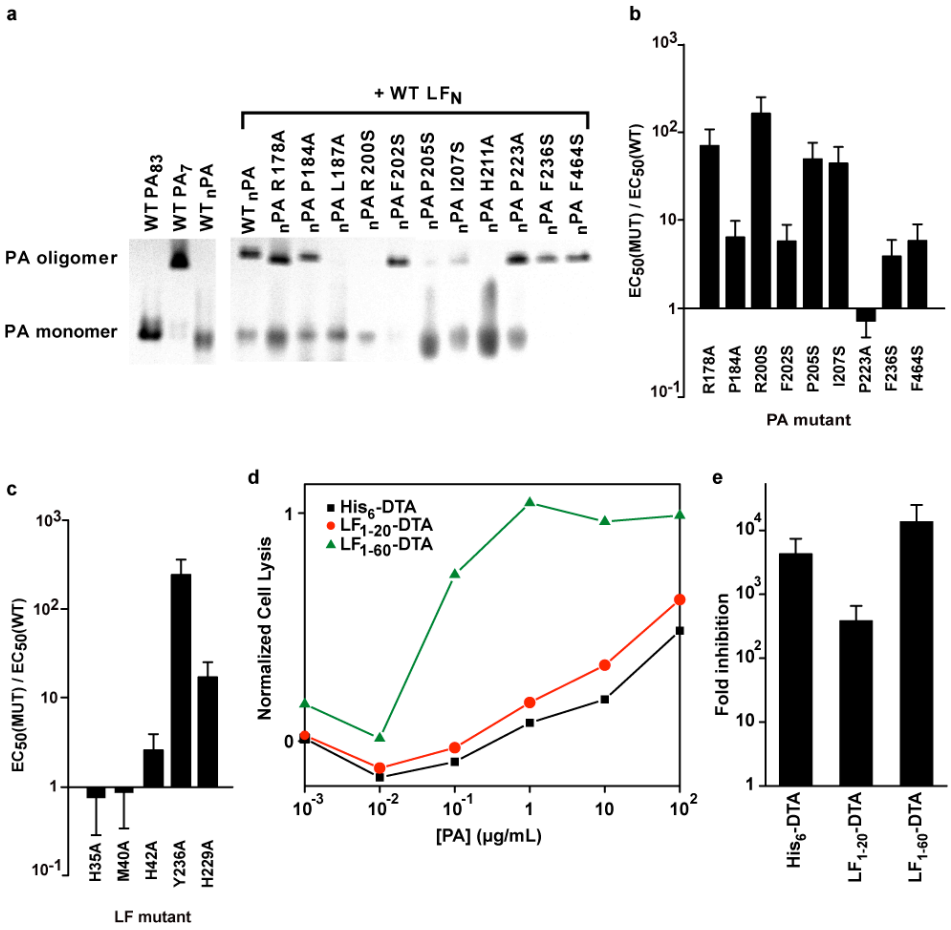

Figure 2.8. Probing the role of the PA-LF$_N$ interaction in toxin assembly and cytotoxicity. (a) Native PAGE analysis of LF$_N$-driven PA assembly. (Right) WT $_n$PA and site-directed $_n$PA mutants at 1 mg/ml were mixed with a stoichiometric equivalent of WT LF$_N$ and incubated for 3 hours at 25 °C, pH 8.0. Protein complexes were separated on a 4-20% gradient gel under native conditions and stained with coomassie blue. Individual lanes show the assembly reactions for either nicked WT PA or nicked PA ($_n$PA) point mutants and WT LF$_N$. The bands corresponding to assembled PA oligomer and unassembled full-length PA monomer (PA$_{83}$) are indicated based upon controls (shown on the Left). WT PA$_7$ refers to $_n$PA assembled over a Q-sepharose column (Blaustein *et al.* 1989); the resulting oligomers are predominantly heptameric (Milne *et al.* 1994; Kintzer *et al.* 2009). **(b)** LT cytotoxicity was monitored by an enzyme-coupled LDH release assay (Decker & Lohmann-Matthes 1988) using immortalized bone marrow macrophage cells from 129 mice. Various PA and LF mutants were assayed in two types of experiments. Relative cytotoxicity of site-directed mutants of PA. Each PA mutant and WT PA were held at a constant concentration, and the concentration of WT LF was varied. Each bar represents the ratio of the effective-concentration-for-50% lysis (EC$_{50}$) when comparing the mutant PA to WT PA (MUT:WT). **(c)** Relative cytotoxicity of site-directed mutants of LF. WT PA was held at a constant concentration, and the concentration of WT LF or a LF mutant was varied. Each bar represents the ratio in EC$_{50}$ (MUT:WT) for the LF mutants. EC$_{50}$ values used to calculate the ratios in each panel were determined by curve fits to the normalized-cell-lysis curves as a function of the LF concentration. Errors in (a) and (b) are the mean ±s.d. (n = 3). **(d)** The substrate, His$_6$-DTA, LF$_{1\text{-}20}$-DTA, or LF$_{1\text{-}60}$-DTA, was added to the macrophage cell suspensions at 10 µg/ml, and the WT PA concentration was varied. Cell lysis was monitored by LDH release after 20-24 hours of exposure to the mixture of PA and the various DTA substrates. Normalized cell lysis curves are plotted for His$_6$-DTA (black ■), LF$_{1\text{-}20}$-DTA (red ●) and LF$_{1\text{-}60}$-DTA (green ▲). Plotted values are the mean of three trials. **(e)** Fold inhibition of DTA cytotoxicty was determined by measuring the individual EC$_{50}$s for each DTA construct in the presence of PA R178A or WT PA. Fold inhibition = EC$_{50}$(PA R178A) / EC$_{50}$(WT PA). The error bars are the mean ±s.d. (n = 3).

(Meador *et al.* 1992; Meador *et al.* 1993), PA's twin Ca^{2+}-binding sites scaffold the cleft and define its distinct shape and chemical character, including: (i) a delocalized anionic potential created by the excess of negatively-charged PA residues chelating the two Ca^{2+} ions and (ii) a large proportion of SASA contributed by PA backbone atoms. LF_N's side chains are not well-packed with side chains in the α-clamp cleft, in contrast to the carboxy-terminal binding subsite (Figure 2.6). Interestingly, PA contacts the side chains of LF Met40 and His35 through backbone interactions. PA_C Arg178 contacts the hydrophilic face of α1 at LF His42 while maintaining a hydrogen bond with the backbone carbonyl of PA_N Thr201. Aromatic residues, PA_N Phe236 and Phe464, and aliphatic residues, PA_N Leu187 and Leu203, line the cleft face opposite of PA_C Arg178. Upon binding LF_N, PA_N Phe202 repositions its phenyl group toward LF_N β1, shielding β1's backbone hydrogen bonds with PA_N Leu203. The chemical nature of the α-clamp cleft suggests that it is well-suited to bind an unfolded β strand and an amphipathic helix with a positively-charged face.

Both LF-binding subsites are critical for cytotoxicity activity. We initially characterized the PA-LF binding interaction using cytotoxicity assays. Site-directed mutagenesis studies on PA and LF residues involved in either binding subsite reveal defects in LT-induced macrophage cytolysis (Figure 2.8b,c). To further address the interaction between LF_N's α1/β1 sequence and the α clamp, we created fusions of the first 20 or 60 residues of LF and the A fragment from diphtheria toxin (DTA), called LF_{1-20}-DTA and LF_{1-60}-DTA, respectively. When co-administered with PA, we find LF_{1-60}-DTA is 100-fold more cytotoxic than LF_{1-20}-DTA or hexahistadine-tagged DTA (His_6-DTA, DTA with an amino-terminal, 18-residue leader containing the hexahistidine sequence, Figure 2.8d). Interestingly, despite lacking the α1/β1 sequence, His_6-DTA (Blanke *et al.* 1996) and LF_{1-20}-DTA are cytotoxic when co-treated with WT PA (Figure 2.8d); however, all of these DTA constructs are ~1000-fold less cytotoxic when co-treated with the α-clamp mutant, PA R178A (Figure 2.8e). Thus the α clamp has broad substrate specificity. However, the role of the α1/β1-α-clamp interaction in toxin function is difficult to surmise from cytotoxicity assays alone, since toxin uptake involves multiple steps (e.g., PA assembly, LF binding, unfolding and translocation).

The role of the α clamp in LT assembly. To determine the role of the α clamp in LT assembly, we performed multiple in vitro PA-LF_N assembly assays. By native PAGE, we find that PA mutations introduced into the LF_N-PA-binding interface disrupt PA co-assembly with LF_N (Figure 2.8a). To focus on the role of LF_N α1/β1 in PA assembly, we labeled PA K563C with two different fluorescent probes. A 1:1 ratiometric mixture of these labeled $_n$PA K563C constructs ($_n$PA*) produces an increase in fluorescence resonance energy transfer (FRET) upon assembly with LF_N (Christensen *et al.* 2006). Using this FRET assay, we find that 5-fold more $_n$PA* assembles with WT LF_N than with the Δ47 LF_N amino-terminal truncation (which lacks both α1 and β1, Figure 2.9a). The circular dichroism (CD) spectra of Δ47 and WT LF_N are comparable, demonstrating that the assembly defect is not due to the misfolding of Δ47 LF_N (Figure 2.9b). Using electron microscopy (EM) (Figure 2.9c), native PAGE (Figure 2.9d), and mass spectrometry (Figure 2.10), we find that the percentage of octameric PA oligomers is greatly reduced for Δ47 LF_N relative to WT LF_N. By EM, we estimate that ~3% of the PA oligomers produced with Δ47 LF_N are octameric (10-fold less than that observed with WT LF_N, Figure 2.9d). Thus LF_N's α1 and β1 structures not only drive PA oligomerization, but also they are critical to the mechanism of PA octamer formation (Figure 2.2c).

Mapping the LF_N-binding interaction with the PA channel. Using electrophysiology, we measure LF_N binding by observing kinetic and equilibrium changes in channel conductance

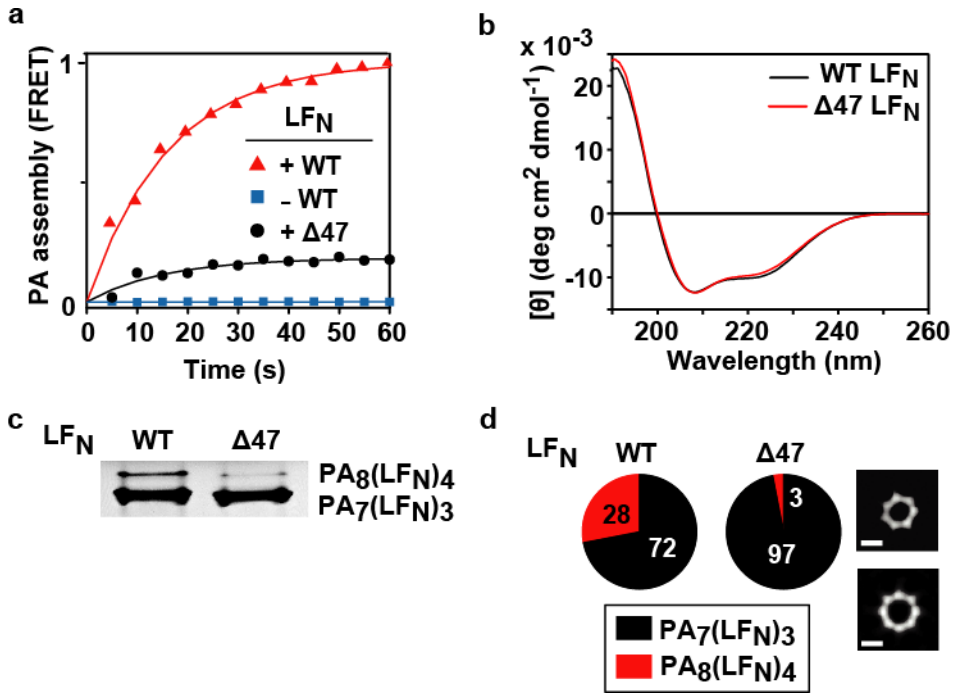

Figure 2.9. Role of LF_N α1 and β1 in PA assembly. (a) PA assembly was monitored by fluorescence resonance energy transfer (FRET) using a mixture of fluorescently-labeled $_nPA$ ($_nPA$*). Shown are FRET-probed assembly records for $_nPA$* in the presence of WT LF_N (red ▲) and Δ47 LF_N (black ●) as well as a control record for assembly in the absence of LF_N (blue ■) at room temperature, pH 7.4. **(b)** Circular dichroism spectra (average of six scans, 2 nm bandwidth, 0.1 cm pathlength) of WT (black) and Δ47 LF_N (red) in 20 mM sodium phosphate, pH 7.5. **(c)** Native PAGE and **(d)** EM analysis of the oligomeric distribution of PA_7 and PA_8 complexes produced after one hour of $_nPA$ co-assembly with either WT LF_N or Δ47 LF_N at room temperature, pH 8. In the assembly reaction, 1 mg/ml WT $_nPA$ was incubated with either WT LF_N or Δ47 LF_N added at 1:1 molar stoichiometry (with respect to $_nPA$). (Left) Pie charts report the distribution for WT-LF_N-assembled and Δ47-LF_N-assembled PA observed by EM. (Right) Representative class-average EM images of the PA_7 and PA_8 complexes are shown with a 5-nm scale bar (white).

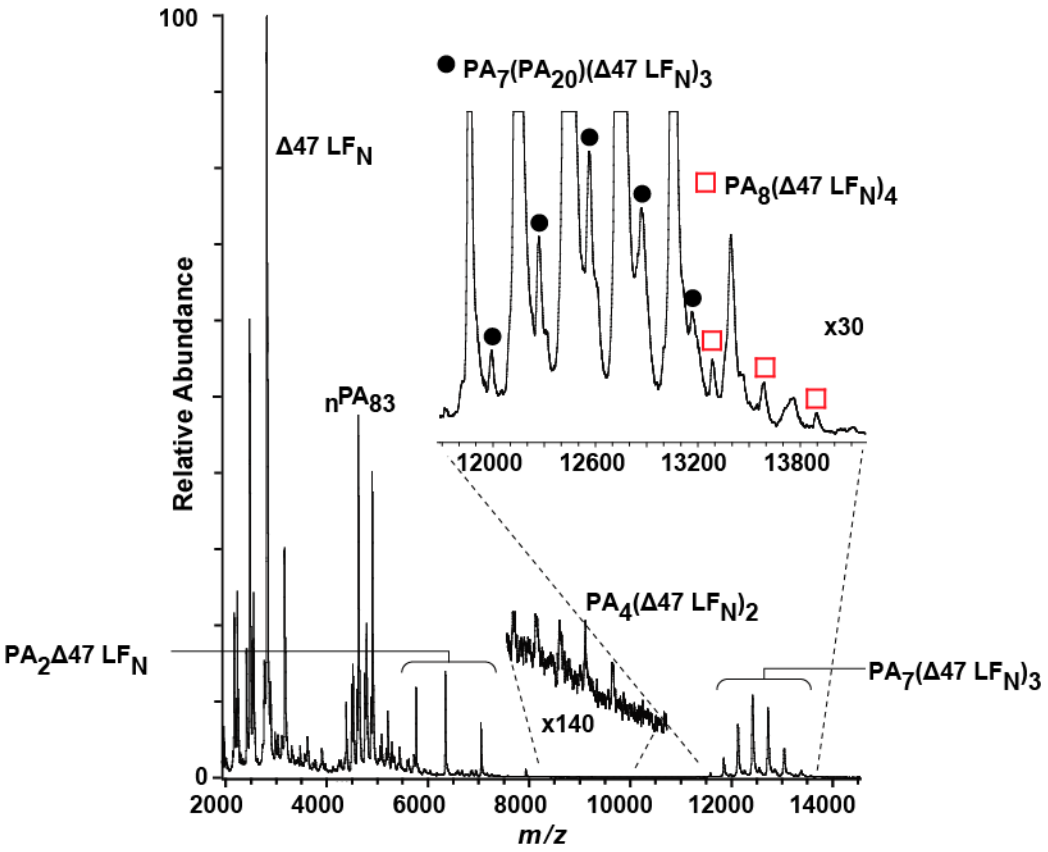

Figure 2.10. Mass spectrometry analysis of Δ47-LF$_N$-driven PA assembly. WT $_n$PA (1 mg/ml) was mixed with Δ47 LF$_N$ at a 1:1 stoichiometric ratio (Δ47 LF$_N$:PA monomer) in ammonium acetate buffer at room temperature, pH 7.8. The products of the assembly reaction were analyzed using nanoESI-MS. The mass spectrum revealed the largest relative abundances for Δ47 LF$_N$ and unassembled WT $_n$PA. The PA$_7$(Δ47 LF$_N$)$_3$ complex represented >90% of the observed oligomers. Low relative abundances were observed for PA$_8$(Δ47 LF$_N$)$_4$ complexes (red □) and other dimeric and tetrameric PA intermediates. Of note, an aberrant oligomeric state, containing PA$_7$(PA$_{20}$)(Δ47 LF$_N$)$_3$ (black ●), was also observed at low relative abundance, albeit at higher levels than octameric PA$_8$(Δ47 LF$_N$)$_4$ complexes. The aberrant complex contains the PA$_{20}$ moiety, and it can be modeled free of steric hindrance, allowing 3 LF$_N$ and 1 PA$_{20}$ to simultaneously bind to the PA heptamer.

(Krantz *et al.* 2005) (Figure 2.11a-c); i.e., when LF_N binds to the PA channel, it inserts its amino-terminal end into the channel and blocks conductance. We monitor binding in the absence of an applied $\Delta\Psi$ to eliminate its influence on the channel-substrate interaction. Since PA_7 and PA_8 have similar translocation (Kintzer *et al.* 2009) and cell cytotoxicity (Kintzer *et al.* 2010) activities, we use the PA_7 oligomer to maintain consistency with prior reports. (Krantz *et al.* 2005; Krantz *et al.* 2006; Thoren *et al.* 2009) To determine the overall thermodynamic contribution of LF_N α1/β1, we made a series of additional Δ*n* LF_N amino-terminal truncations (where *n* is the number of deleted residues). These Δ*n* LF_N do not block PA channel conductance, as they lack sufficient unfolded/unstructured sequence on their amino termini. We use a competition assay to measure Δ*n* LF_N binding: first we block PA channel conductance with WT LF_N (~100 pM); then we add the competitor Δ*n* LF_N and monitor the restoration of the conductance (Figure 2.11d,e). We find that Δ42 and Δ47 LF_N reduce WT PA-channel-binding affinity by 3.6-3.8 kcal mol^{-1} relative to WT LF_N (Figure 2.7c). However, since Δ27, Δ32, and Δ39 LF_N destabilize the complex by about 1.2-1.4 kcal mol^{-1}, the α1/β1 interaction is worth ~2.5 kcal mol^{-1}. We assume that downstream interactions within the channel provide the additional ~1 kcal mol^{-1} of stabilization. We conclude that LF_N α1/β1 binds to the PA channel and provides substantial stabilization of the PA-LF_N complex.

To investigate the details of the interaction between the PA channel and LF_N, we engineered point mutations into residues localized in either LF_N binding subsite and estimated their relative energetic contribution to channel binding (Figure 2.7a,b). Several mutations localized in the carboxy-terminal binding subsite, PA R200S, I207S, and H211A, disrupt LF_N binding by 1-1.5 kcal mol^{-1}. These residues form two binding "hotspots", i.e., locations where point mutations disrupt binding most severely (Clackson & Wells 1995). By contrast, the mutations, F202S and P205S, located between these two carboxy-terminal-site hotspots have minimal effects on LF_N binding, reflecting that LF_N's carboxy terminus does not make substantial contact with these residues (Figure 2.6b). The LF_N Y236A mutant most appreciably perturbs PA-channel binding and represents the LF_N hotspot in the carboxy-terminal subsite interaction. Other adjacent LF_N residues in the carboxy-terminal subsite interaction have minimal effects on PA channel binding.

We then investigated the relative energetic contribution of residues localized in the α-clamp binding subsite (Figure 2.7a,b). We find that PA Arg178 comprises the major hotspot site in PA's α clamp, where the R178A mutation destabilizes the complex by 2.9 kcal mol^{-1}. While the aromatic PA mutant, F464S, destabilizes LF_N binding at the α-clamp site by 0.7 kcal mol^{-1}, the PA F236S mutant does not. Additionally, we find that none of 23 point mutations introduced into LF_N α1 and β1 destabilizes the LF_N-PA channel complex. Interestingly, the mutation, LF_N M40A, stabilizes the complex 1.3 kcal mol^{-1} (Figure 2.7b). These results indicate contrasting binding energetic behaviors for the two different LF_N-binding subsites. At the carboxy-terminal subsite, a classical interface is observed, where specific LF_N and PA side chains comprise the respective hotspots on either interface. At the α-clamp subsite, while we identify PA Arg178 as a major hotspot residue, no clear hotspot can be identified on LF_N α1/β1. These observations suggest that the stabilizing interactions in the α-clamp subsite do not involve specific LF_N side chains, but rather the ~2.5 kcal mol^{-1} of binding stabilization is due to the formation of nonspecific contacts and the more general exclusion of SASA.

The PA α clamp possesses nonspecific binding activity. The robustness of the binding interaction is intriguing given the paucity of specific α-clamp interactions. To test the specificity of the α-clamp interaction, we either replaced the entire LF_N α1/β1 sequence with other non-

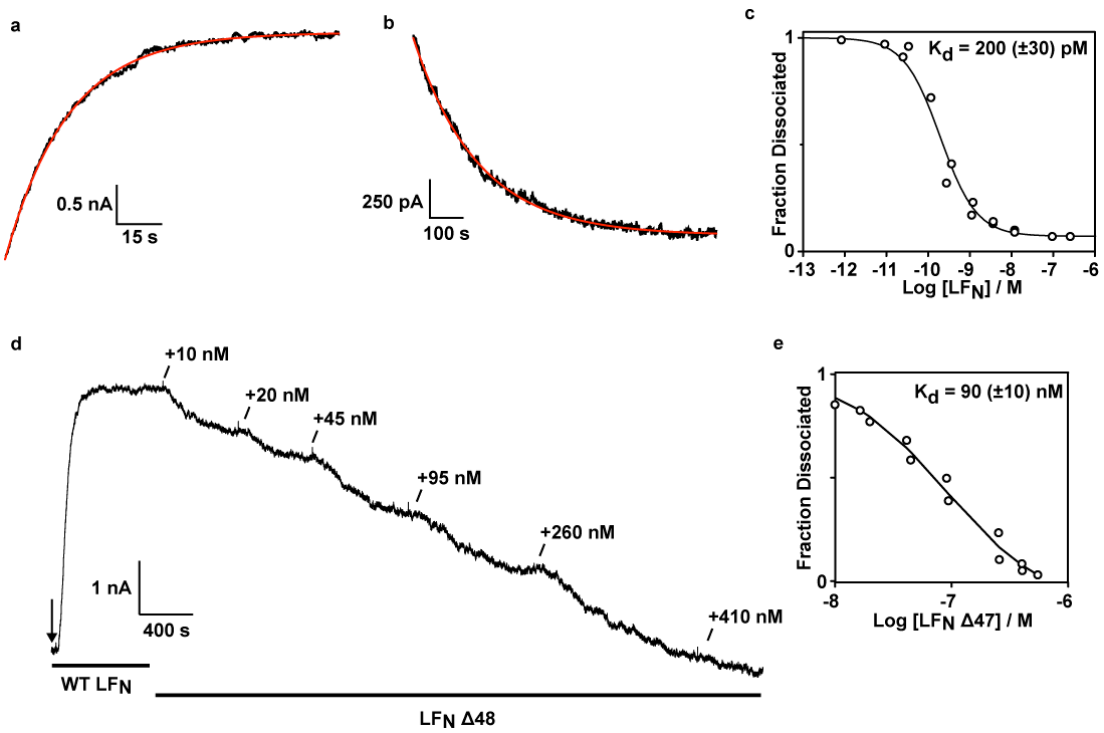

Figure 2.11. Equilibrium and kinetic measurements of LF_N binding to the PA channel. Channel binding was measured by recording equilibrium and kinetic changes in current, I, in the absence of an applied potential. Representative **(a)** association and **(b)** dissociation kinetics for WT LF_N binding to WT PA channels at pH 7.4. Association kinetics were measured at 3 nM WT LF_N. Dissociation was initiated by perfusing the cis chamber and/or adding 1 μM Δ47 LF_N competitor. Each current versus time, t, record (black line) was fit (red line) to a single exponential, $I = I_o\exp(-kt) + c$, to obtain the rate constant, k. I_o is the amplitude, and c is an offset constant. In the depicted example in (a), the association rate constant, k_a, is determined by $k_a = k/[L]$, and the k_a is 1.63×10^7 ($\pm3\times10^5$) M^{-1} s^{-1}. From (b), the dissociation rate constant, k_d, is 0.00518 ($\pm1\times10^{-5}$) s^{-1}. **(c)** Equilibrium binding currents as a function of WT LF_N ligand concentration, $[L]$, fit to a single-site binding model, $I = I_o/(1 + K_d/[L])$. The equilibrium dissociation constant values, K_d, determined in kinetic (318 pM) and equilibrium experiments (200 pM) are in good agreement. **(d)** Depicted is an example of equilibrium-binding-competition experiment used to determine the apparent K_d of ligands that bind PA but cannot block its conductance. After PA-channel insertion stabilized, the pH was adjusted to 7.4. Then 0.8 nM WT LF_N was added (indicated by the arrow). Once the WT-LF_N-conductance block stabilized, small increments of Δ47 LF_N were added. As Δ47 LF_N competes for WT LF_N sites, the channels reopen. Increments of added Δ47 LF_N are indicated. **(e)** The degree of reopening as a function of total Δ47 LF_N concentration was used to measure Δ47 LF_N's K_d, which is 90 (±10) nM.

homologous sequences from LF and EF or introduced multiple mutations into α1/β1 (Figure 2.12). Interestingly, we find that these LF_N α1/β1 replacements bind with similar affinities as WT LF_N (differing by 0.2 to 1.0 kcal mol^{-1}, Figure 2.7d). Furthermore, multisite LF_N mutants in which the buried hydrophobic face of α1/β1 is replaced with either four Ser residues (LF_N HS1) or four Phe residues (LF_N Ar1) bind PA with similar affinity as WT LF_N (Figure 2.7d), indicating that the α clamp also binds non-amphipathic helices. Finally, we find that these LF_N α1/β1-replacement constructs bind 1.3-2.4 kcal mol^{-1} less tightly to PA R178A relative to WT PA (Figure 2.7d), thereby confirming that this nonspecific-binding activity is localized to the α-clamp subsite. Thus the α clamp binds a broad array of sequences, providing 1.5-4 kcal mol^{-1} of stabilization (depending upon the identity of the α1/β1 sequence).

LF_N must unfold to bind the α-clamp subsite. Our crystal structure and thermodynamic binding data indicate that the α-clamp subsite binds nonspecifically to unfolded protein substrates. This model is well supported by several additional lines of evidence. First, the thermodynamic comparison of WT LF_N and the truncated Δ*n* LF_N mutants is appropriate because these mutants have similar folded secondary structure content as WT LF_N (Figure 2.9b). Moreover, the Δ47 LF_N construct binds similarly to PA R178A as WT PA (Figure 2.7d), confirming that the Δ47 LF_N truncation does not bind at the α-clamp site, as implied by the structure (Figure 2.2a). Second, fusions of LF's amino terminus and DTA (LF_{1-60}-DTA and LF_{1-20}-DTA) are sufficient to bind to the α-clamp site, since their affinity for the PA channel is disrupted by the PA R178A mutation (Figure 2.7d and Figure 2.13). This result indicates that the α clamp is an independent binding site capable of binding to unstructured sequences at the amino-terminus of a substrate. Third, knowing that LF_N α1/β1 unfolds upon binding PA (Figure 2.14a), we engineered the double mutant, LF_N I39C E72C ($LF_N^{C39\text{-}C72}$, which forms a disulfide bond that prevents α1/β1 unfolding). Interestingly, $LF_N^{C39\text{-}C72}$ has 10^4-fold reduced affinity for PA channels under non-reducing conditions (Figure 2.14b); however, under reducing conditions (in the presence of dithiothreitol, DTT), $LF_N^{C39\text{-}C72}$ binds with the same affinity as WT LF_N (Figure 2.14b). We also kinetically observe a DTT-dependent $LF_N^{C39\text{-}C72}$ blockade of PA channels (Figure 2.15a). Therefore, LF_N must unfold α1 and β1 to properly bind the α clamp and interact stably with PA oligomers.

Binding to PA induces strain and disorder into LF_N. We then asked how the unfolding of LF_N α1/β1 on the surface of PA affects the remaining folded structure of LF_N. First, we measured the stability of the Δ*n* LF_N mutants using chemical denaturant titrations probed by CD at 222 nm (CD_{222}). The Δ*n* mutants' stabilities are estimated by fitting the CD_{222}-probed titration data to a four-state equilibrium unfolding model ($N \leftrightarrow I \leftrightarrow J \leftrightarrow U$) (Krantz *et al.* 2004) (Figure 2.15b and Table 2.2). We find the truncation mutants possess native (*N*), intermediate (*I* and *J*), and unfolded (*U*) states. The truncations, however, destabilize the *N* state by ~1.2 kcal mol^{-1}, where the deletion of the α1 helix is more destabilizing than the deletion of the β1 strand (Figure 2.14c). Second, we compared the crystallographic atomic displacement parameters (*B* factors) of bound LF_N with free LF_N (1J7N (Pannifer *et al.* 2001)). In this analysis, we calculate the relative change in normalized *B* factor (ΔB_{norm}) for each LF_N residue upon binding PA (Figure 2.14d). The β2-β4 sheet and surrounding helices increase in B_{norm} upon binding PA, whereas α1/β1 decrease in B_{norm} (Figure 2.14d). To corroborate these ΔB_{norm} values, we measure changes in backbone and side chain mobility using fluorescence anisotropy (FA). LF_N mutants with unique Cys substitutions were labeled with thiol-reactive fluorescent probes. Upon binding WT PA_7 oligomers, the fluorescent probes attached to LF_N's α1/β1 structures show gains in normalized relative FA (FA_{norm}), and conversely, probes in the β2-β4 sheet show losses in FA_{norm}

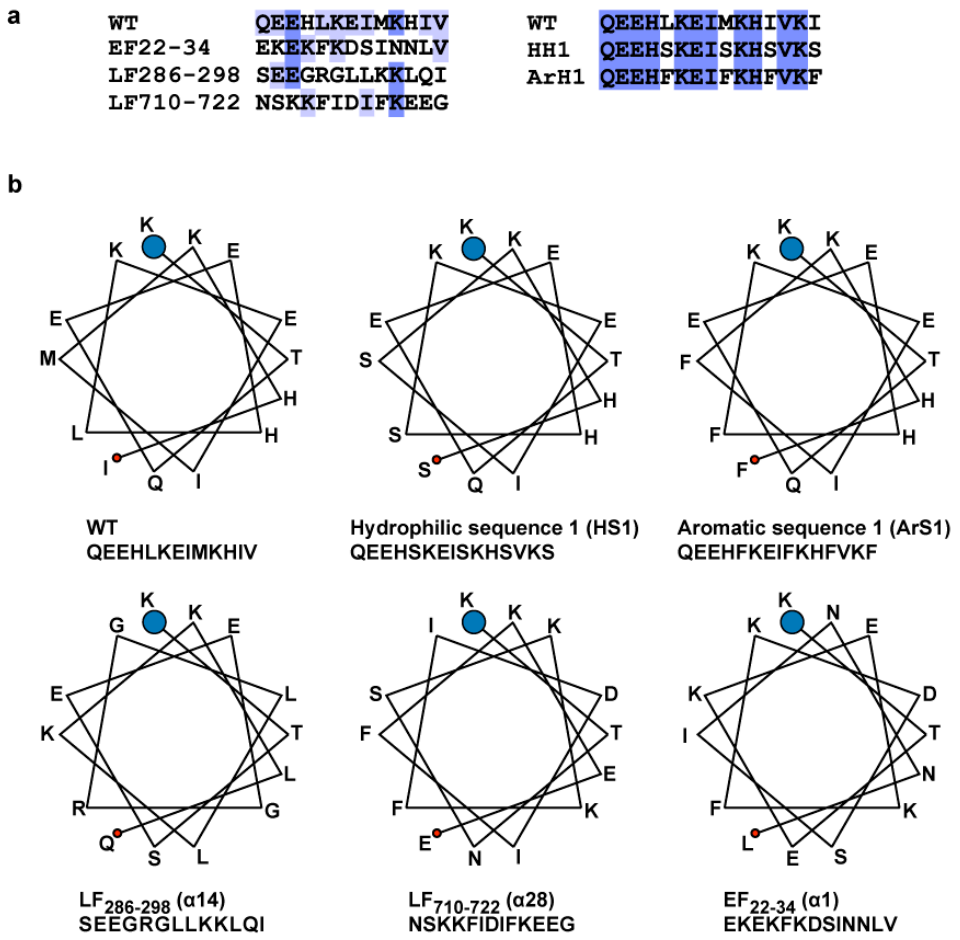

Figure 2.12. Sequence-alignment and helical-wheel analysis of α1/β1-replacement sequences. (a) A multiple sequence alignment of the 13- or 15-residue sequences engineered into the α1 position of LF_N (WT). Residues are colored according to the BELVU convention (Sonnhammer & Hollich 2005): highly-conserved (dark blue), conserved (light blue), and non-conserved (white). The CLUSTALW (Thompson *et al.* 1994) pairwise-percent-identity scores (relative to WT) are: HS1 (73%), ArS1 (73%), $LF_{286-298}$ α14 (23%), $LF_{710-722}$ α28 (23%), and EF_{22-34} α1 (31%). **(b)** Helical wheels are depicted, where the amino-terminal end (large blue circle at front of page) and carboxy-terminal end (small red circle at the back of the page) are indicated.

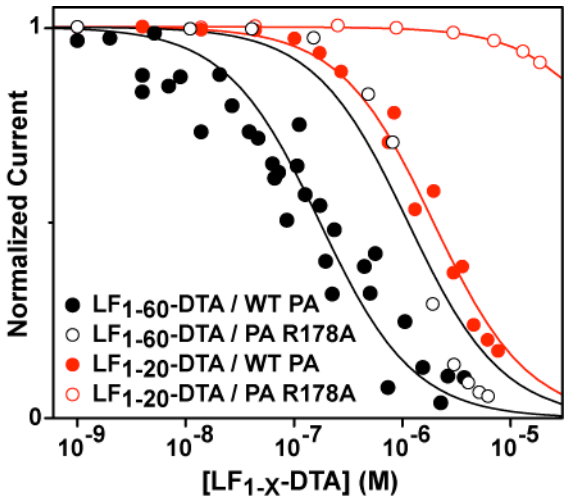

Figure 2.13. The first 20 or 60 amino-terminal residues of LF_N are sufficient to bind the PA channel at the α-clamp subsite. Binding curves for $LF_{1\text{-}60}$-DTA (black) and $LF_{1\text{-}20}$-DTA (red) using either WT PA channels (●) or PA R178A channels (○). In these experiments, PA channels are inserted into a planar lipid bilayer bathed in asymmetric KCl solutions buffered in 10 mM potassium phosphate, pH 7.4. An additional 0.1 M equivalent of potassium chloride was added only to the cis solution. Aliquots of $LF_{1\text{-}60}$-DTA or $LF_{1\text{-}20}$-DTA were added and the final equilibrium currents were recorded at a ΔΨ of 0 mV. Equilibrium curves were fit to a one-binding-site model to calculate the K_d: $LF_{1\text{-}60}$-DTA with WT PA, K_d = 0.16 (±0.02) μM; $LF_{1\text{-}20}$-DTA with WT PA, K_d = 1.9 (±0.1) μM; $LF_{1\text{-}60}$-DTA with PA R178A, K_d = 1.1 (±0.2) μM; and $LF_{1\text{-}20}$-DTA with PA R178A, K_d = 190 (±4) μM.

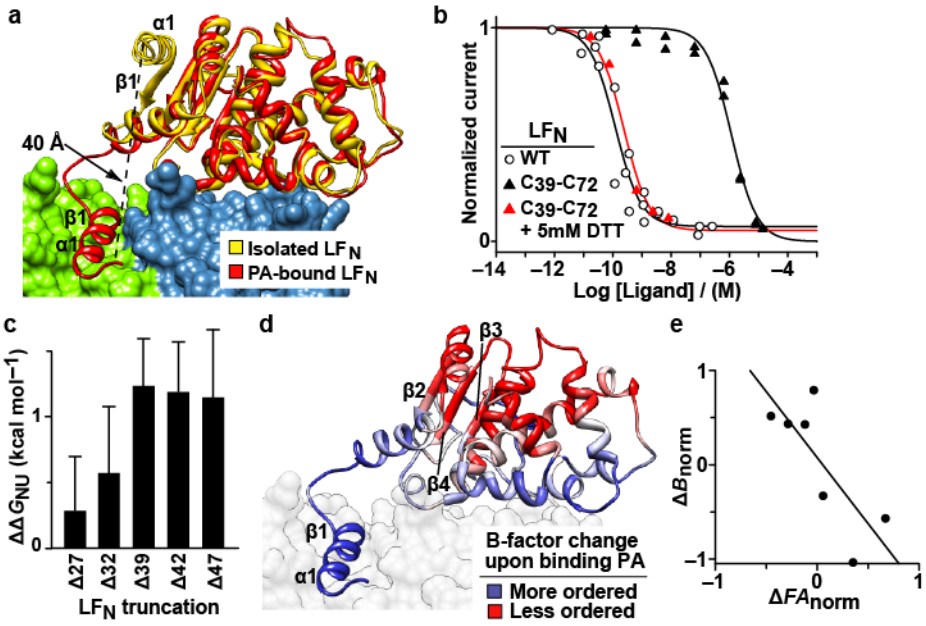

Figure 2.14. Dynamics and thermodynamics of the pre-translocation unfolding of LF_N. (a) Rendering of LF_N's unfolding transition on the surface of the PA_NPA_C dimer (green and blue, respectively). Free LF_N (gold) (PDB 1J7N (Pannifer *et al.* 2001)) is Cα-aligned to the LF_N in the $PA_8(LF_N)_4$ complex (red). **(b)** $LF_N^{C39\text{-}C72}$ binding to WT PA channels (pH 7.4, 0 mV) in the presence of 5 mM DTT (red ▲) and in the absence of DTT (black ▲). A WT LF_N binding curve (○) is also shown. Normalized equilibrium currents were fit to single-site binding model to obtain K_d values: WT LF_N, K_d = 120 (±30) pM; $LF_N^{C39\text{-}C72}$, K_d = 1.2 (±0.1) μM; and $LF_N^{C39\text{-}C72}$ + 5 mM DTT, K_d = 240 (±60) pM. **(c)** Equilibrium stability measurements (pH 7.5, 20 °C) of amino-terminal deletions of LF_N (Δ*n* LF_N). Equilibrium free energy differences ($\Delta\Delta G_{NU}$) were obtained from denaturant titration data fit to a four-state equilibrium unfolding model (Krantz *et al.* 2004) (Figure 3.13b), where $\Delta\Delta G_{NU} = \Delta G_{NU}(\Delta n) - \Delta G_{NU}(WT)$. Error bars are the mean ±s.d. (*n* = 3-4). Fit parameters are listed in Table 2.2. **(d)** Residues in LF_N are colored by their differences in normalized *B* factor (ΔB_{norm}), which is obtained by comparing the model of free LF_N (1J7N, structure 1) and LF_N in complex with PA (structure 2) using $\Delta B_{norm} = B_{1,i} / \langle B_1 \rangle - B_{2,i} / \langle B_2 \rangle$. The $\langle B \rangle$ is the average *B* factor for the entire chain. **(e)** ΔB_{norm} is plotted against the normalized fluorescence anisotropy (FA) change (ΔFA_{norm}) for 7 different site-specifically-labeled residues (37, 48, 72, 126, 164, 199, and 242) in LF_N. $\Delta FA_{norm} = FA_{1,i} / \langle FA_1 \rangle - FA_{2,i} / \langle FA_2 \rangle$, where free LF_N and the LF_N-PA oligomer complex are state 1 and state 2, respectively. The linear fit is significant (p = 0.04). Raw anisotropy changes upon binding the PA oligomer for these labeled LF_N are shown in Figure 2.16.

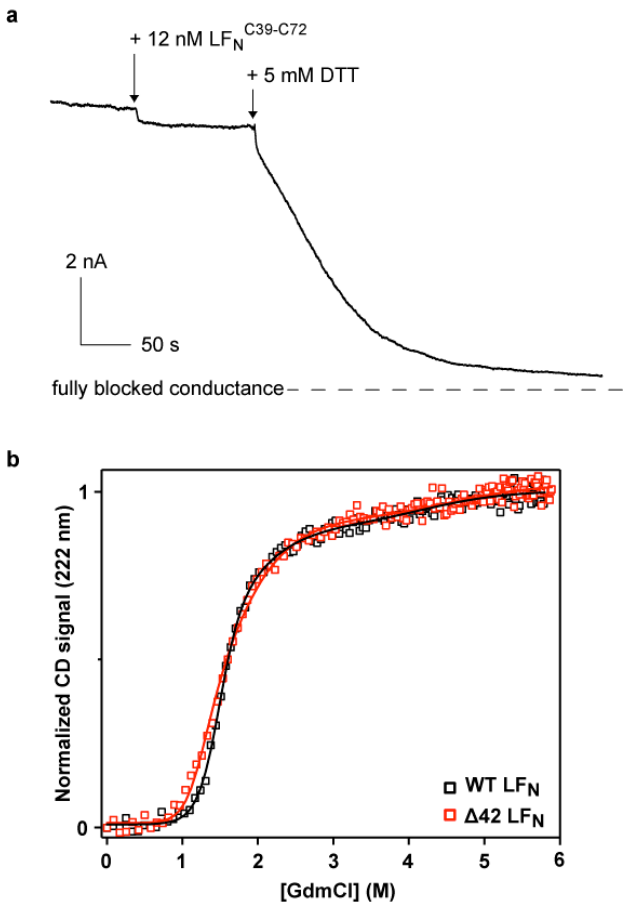

Figure 2.15. LF_N unfolding increases its affinity for PA but reduces its stability. (a) A planar bilayer recording at a ΔΨ of 0 mV of PA channels bathed in 10 mM potassium phosphate at pH 7.4. To generate a current, the salt concentration was asymmetric: 100 mM potassium chloride was added to the cis side, and 0 mM potassium chloride was added to the trans side of the bilayer. The recording begins after the PA current stabilized. Then 12 nM $LF_N^{C39-C72}$ was added to the cis side of the membrane (the side to which PA channels were inserted), a small decrease in PA current was established. Once this minor $LF_N^{C39-C72}$ block stabilized, 5 mM DTT was added to the cis side of the membrane. A rapid and nearly complete block of the PA current was subsequently observed. **(b)** Equilibrium denaturant titration profiles of WT (black squares) and Δ42 LF_N (red squares) probed by CD at 222 nm. The buffer is 20 mM sodium phosphate, 0.75 M trimethylamine *N*-oxide, pH 7.5. The normalized titration data are fit (solid line) to a four-state model ($N \leftrightarrow I \leftrightarrow J \leftrightarrow U$), which has been described elsewhere (Krantz *et al.* 2004). See Table 3.2 for a listing of all curve fit parameters.

Table 2.2. Thermodynamic stability free-energy parameters for Δ*n* LF_N truncations

LF_N	ΔG_{NU}	ΔG_{IU}	ΔG_{JU}	$\Delta\Delta G_{NU}$
WT	-12.6 (±0.4)	-7.2 (±0.4)	-4.7 (±0.3)	0
Δ27	-12.3 (±0.2)	-6.7 (±0.3)	-4.4 (±0.3)	0.3 (±0.4)
Δ32	-12.0 (±0.4)	-6.7 (±0.2)	-4.9 (±0.5)	0.6 (±0.5)
Δ39	-11.33 (±0.08)	-6.6 (±0.3)	-4.6 (±0.2)	1.2 (±0.4)
Δ42	-11.4 (±0.2)	-6.81 (±0.04)	-4.80 (±0.02)	1.2 (±0.4)
Δ47	-11.4 (±0.4)	-6.7 (±0.3)	-4.6 (±0.1)	1.2 (±0.5)

The four-state equilibrium free energy parameters, ΔG_{NU}, ΔG_{IU}, ΔG_{JU}, in kcal mol^{-1} are obtained from fitting equilibrium denaturant titration profiles probed by CD. The buffer and conditions are 20 mM sodium phosphate, 0.75 M trimethylamine *N*-oxide, pH 7.5, and 20 °C. The guanidinium chloride denaturant sensitivities (*m* values) defining each thermodynamic transition between states are fixed to constant values (where m_{NI}, m_{IJ}, m_{JU} are 3.59, 1.37, and 1.05 kcal mol^{-1} M^{-1}, respectively) consistent with previous values (Krantz *et al.* 2004). The four-state fit model ($N \leftrightarrow I \leftrightarrow J \leftrightarrow U$) has been described elsewhere (Krantz *et al.* 2004). Example curve fits are shown in Figure 2.15b. The equilibrium free energy differences ($\Delta\Delta G_{NU}$) are computed as $\Delta\Delta G_{NU} = \Delta G_{NU}(MUT) - \Delta G_{NU}(WT)$.

(Figure 2.16a). Overall, these ΔFA_{norm} values inversely correlate with ΔB_{norm} values (p value of 0.04, Figure 2.14e), confirming that the more dynamic regions in the crystal are also dynamic in solution. Therefore, we conclude that the ~2.5 kcal mol^{-1} of stabilization gained when α1/β1 binds to the α-clamp site not only offsets the ~1.2 kcal mol^{-1} of thermodynamic destabilization imparted by the unfolding of α1/β1 but also accounts for the observed entropic increases in strain and disorder throughout LF_N's remaining folded structure.

The role of the α clamp in protein translocation. To determine the role of the α clamp during protein translocation, we use planar lipid bilayer electrophysiology, which records changes in PA conductance as substrate-blocked channels translocate their substrates and reopen (Krantz *et al.* 2005; Krantz *et al.* 2006; Thoren *et al.* 2009). We examined 37 point mutations in PA and LF_N. Of the 13 PA mutants tested, we find that the α-clamp mutant, PA F202S, slows LF_N translocation 20-fold, or 1.7 kcal mol^{-1} (Figure 2.17a). A subset of the LF_N point mutations (H35A, M40A, and H42A), which point toward either face of the α-clamp cleft (Figure 2.6a), inhibit translocation 0.8-1.7 kcal mol^{-1} (Figure 2.17a). These translocation defects are observed for both PA_7 and PA_8 channels (Figure 2.18). Conversely, other buried α1 sites (LF_N Leu36, Ile39, and Ile43) are tolerant to substitution and do not affect protein translocation (Figure 2.17a). Interestingly, we find that the observed positional translocation defects are alleviated with the inclusion of a bulky group at position 40 (Thoren *et al.* 2009) and a positively-charged residue at positions 35 and 42 (Figure 2.17a). All of the LF_N α1/β1 replacements translocate similarly to WT LF_N (Figure 2.17a). We conclude, therefore, that efficient LF_N unfolding and translocation are catalyzed by the aromatic α-clamp residue (PA Phe202); however, the LF_N α1/β1 sequence itself has rather minimal charge and steric requirements.

The broad substrate specificity of the α clamp led us to ask which PA residues facilitate translocation of full-length LF, a more complex, multidomain substrate. LF has a different rate-limiting step than LF_N and requires a greater driving force (Krantz *et al.* 2006); therefore, we measure its translocation kinetics under a ΔpH and ΔΨ. We find the PA α-clamp mutants, F202S and P205S, reduce LF translocation efficiency, ε, by ~60% (where $\varepsilon = A_{obs}/A_{exp}$, A_{exp} and A_{obs} are the expected and observed amplitudes, respectively, Figure 2.20). The PA mutants F236S and F202S inhibit the rate of LF translocation (Figure 2.17b). Interestingly, these PA mutants do not appreciably affect LF_N binding (Figure 2.7a), and only PA F202S inhibits LF_N translocation (Figure 2.17a). Finally, we find PA R178A is defective in LF_N binding but not defective in translocation. We conclude that hydrophobic and aromatic residues surrounding the α clamp (Figure 2.21) catalyze the translocation of LF.

2.4 Discussion

Some models (Krantz *et al.* 2005; Zimmer *et al.* 2008) propose that nonspecific clamping sites are critical features of unfolding machines. In general, unfoldases are thought to denature proteins by applying mechanical forces (Thoren *et al.* 2009) and transiently trapping partially unfolded conformations in nonspecific binding sites (Krantz *et al.* 2005). Unfolded protein, however, is inherently more complex than folded protein, especially in terms of its configurational flexibility and combinatorial complexity. Therefore, a translocase channel would have to accommodate an ever-changing array of possible chemistries and configurations as the unfolded chain is translocated. An elegant solution to this problem may be that unfolded sequences adopt a more rigid and uniform α-helical or β-strand conformation upon binding to an unfoldase, as we observe in the PA-LF_N complex (Figure 2.6a). Indeed we find that PA's α clamp can bind to a broad array of amino acid sequences (Figure 2.7d). This nonspecific

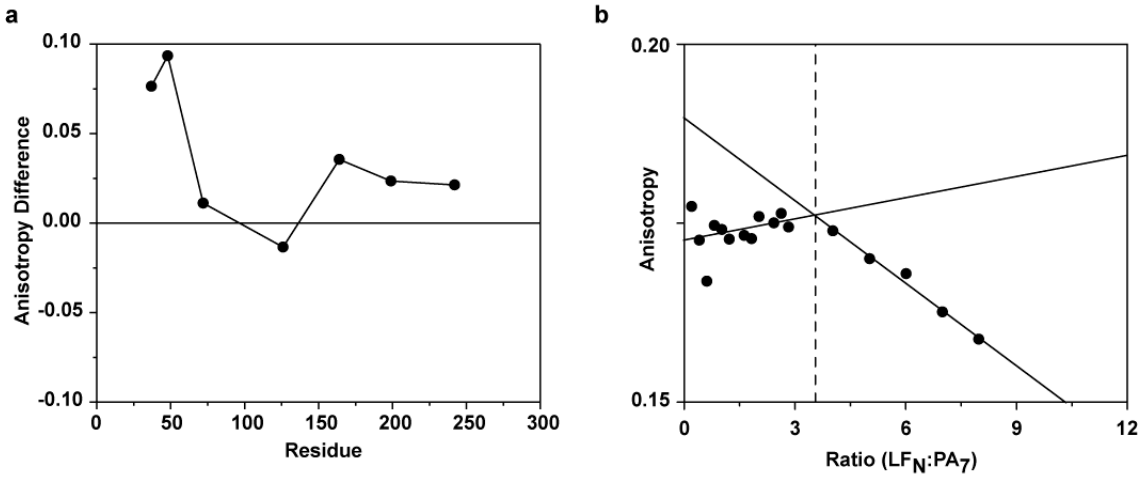

Figure 2.16. Fluorescence anisotropy changes upon PA-LF$_N$ complex formation. (a) Seven site-directed Cys mutants in LF$_N$ were labeled with EDANS (designated LF$_N$*.) The FA of each free LF$_N$* (30 nM) and each LF$_N$* in complex with PA$_7$ oligomer (30 nM each) at pH 7.5 were recorded. The plotted difference in anisotropy (ΔFA) was calculated according to $\Delta FA = FA$(LF$_N$*:PA$_7$ complex) – FA(LF$_N$*). **(b)** The FA signal for PA-LF$_N$* complex formation is saturable at the appropriate stoichiometry. PA$_7$ oligomers (30 nM) were adjusted to pH 5.5 in 0.1% (w/v) DBM, and the binding partner, LF$_N$ V48C*EDANS, was titrated. Upon reaching equilibrium, each sample's FA signal was recorded. The saturation in the FA signal occurs at a LF$_N$:PA$_7$ ratio of 3.4 (±0.2), which is consistent with the 3:1 stoichiometry of the complex reported using other methods (Mogridge *et al.* 2002; Kintzer *et al.* 2009).

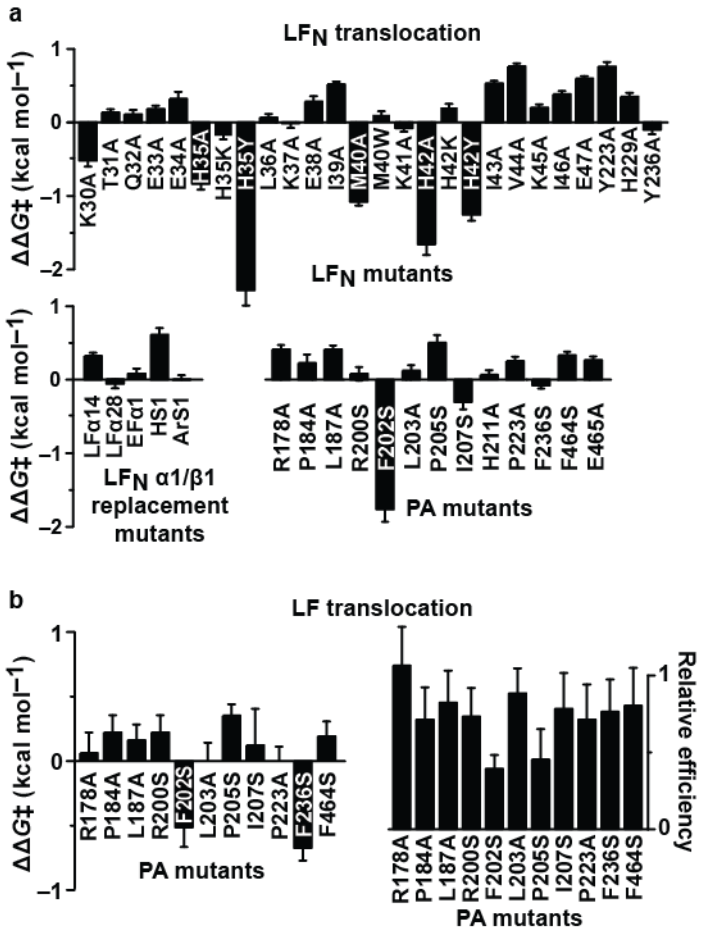

Figure 2.17. The role of the α clamp in LF_N and LF translocation. Planar lipid bilayer translocation results for various mutant channels and substrates. **(a)** Differences in translocation activation energy ($\Delta\Delta G^{\ddagger}$) for (top) LF_N mutants, (bottom left) LF_N α1/β1-replacement mutants, and (bottom right) PA mutants are shown. The reference state is WT LF_N:WT PA. $\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}(\text{WT}) - \Delta G^{\ddagger}(\text{MUT})$, and $\Delta G^{\ddagger} = RT \ln t_{1/2} / c$. The $t_{1/2}$ value is the time for half of the protein to translocate, and c is a 1-sec reference constant. All LF_N translocation rates were measured at symmetrical pH 5.6, $\Delta\Psi$ = 40 mV. A negative value indicates the rate of translocation slowed upon mutation. The relative translocation efficiencies for these LF_N translocations are given in Figure 3.16b. **(b)** Full-length LF translocation at pH_{cis} = 6.1, pH_{trans} = 7.4, ΔpH = 1.3, $\Delta\Psi$ = 20 mV. (left) $\Delta\Delta G^{\ddagger}$ values and (right) relative translocation efficiencies ($\varepsilon_{MUT}/\varepsilon_{WT}$) for mutant PA channels. Individual LF translocation records are shown in Figure 2.21. Error bars in (a-b) are the mean ±s.d. (n = 2-12).

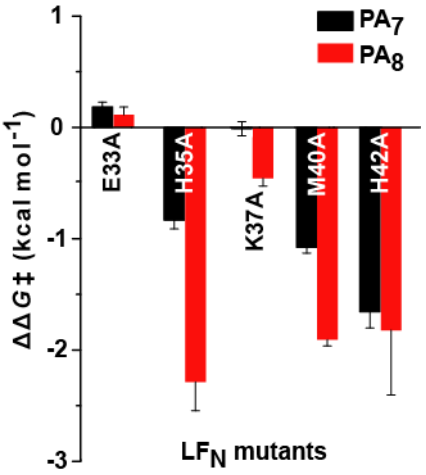

Figure 2.18. LF_N translocates similarly via PA_7 and PA_8 channels. WT LF_N and the indicated LF_N mutants (MUT) were translocated at symmetrical pH 5.6, $\Delta\Psi$ = 40 mV using either heptameric (PA_7, black) or octameric (PA_8, red) channels. The translocation activation energy difference ($\Delta\Delta G^{\ddagger}$) was calculated according to $\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}(\text{WT}) - \Delta G^{\ddagger}(\text{MUT})$, where $\Delta G^{\ddagger} = RT \ln t_{1/2} / c$. The $t_{1/2}$ is the time (in seconds) for half of the protein to translocate; c is a 1-sec reference constant; R is the gas constant; and T is the temperature. A negative $\Delta\Delta G^{\ddagger}$ indicates the rate of translocation is slower for the LF_N mutant. Error bars are the mean ±s.d. (n = 2-6).

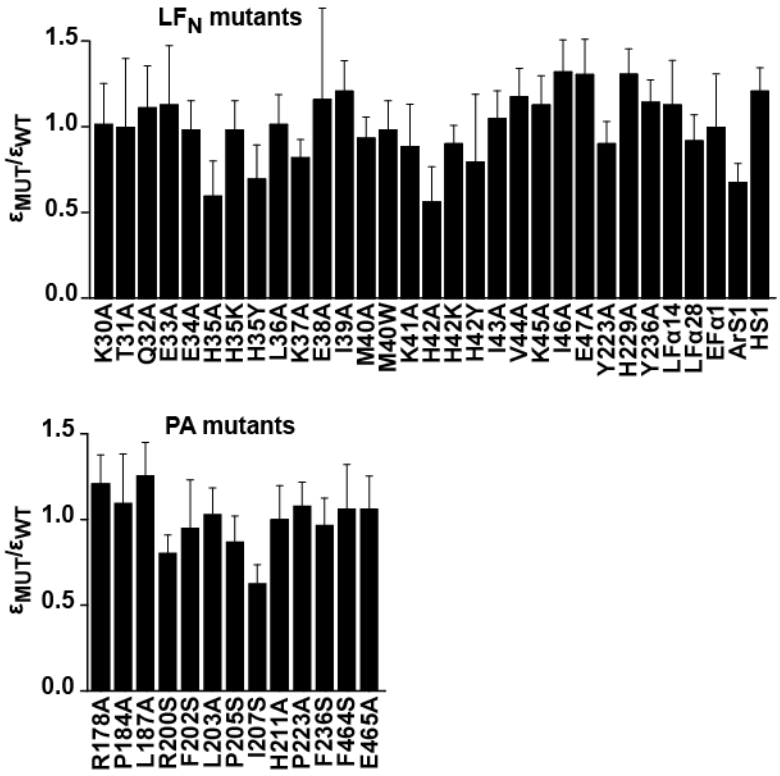

Figure 2.19. Translocation efficiencies for LF_N and mutants on PA channels. LF_N translocation was measured at symmetrical pH 5.6 and a $\Delta\Psi$ of 40 mV. The translocation efficiency (ε) is given as the fraction of channels that successfully translocate, where WT LF_N translocation efficiency is ~60% under these conditions. The relative translocation efficiency ($\varepsilon_{MUT}/\varepsilon_{WT}$) is the ratio of the mutant and WT translocation efficiencies. (Left) $\varepsilon_{MUT}/\varepsilon_{WT}$ for LF_N point mutants and $\alpha1/\beta1$-replacement mutants translocated via WT PA; and (Right) $\varepsilon_{MUT}/\varepsilon_{WT}$ for WT LF_N translocated via mutant PAs. Errors are the mean ±s.d. (n = 2-6).

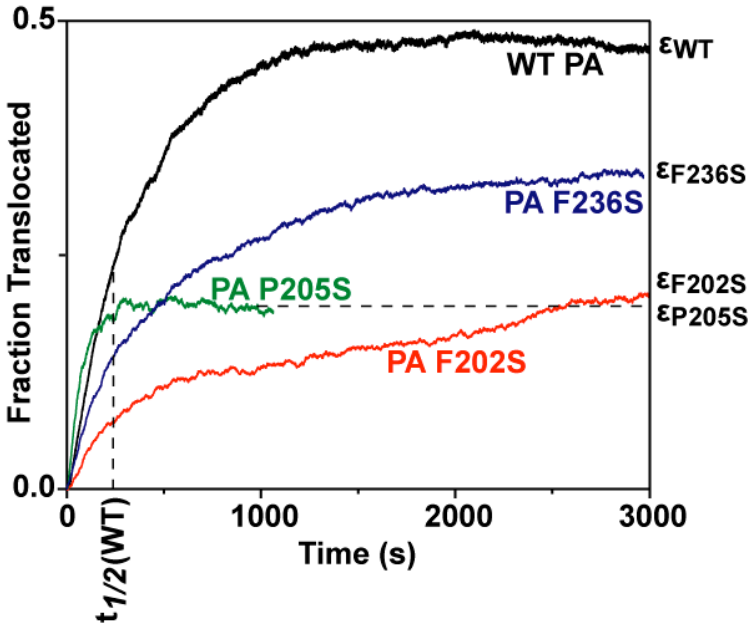

Figure 2.20. LF translocation records. LF translocation was measured using planar lipid bilayer electrophysiology under a 1.3-unit ΔpH (pH_{cis} = 6.1, pH_{trans} = 7.4) and a 20-mV ΔΨ. WT LF was translocated via WT PA (black) and the PA mutants, PA F202S (red), PA P205S (green), and PA F236S (blue). The time for half of the translocated LF to translocate ($t_{1/2}$) is indicated for WT PA. The efficiency, ε, (shown for each PA) is the ratio of the final amplitude and expected amplitude. For WT LF translocation via WT PA, ε is ~50%.

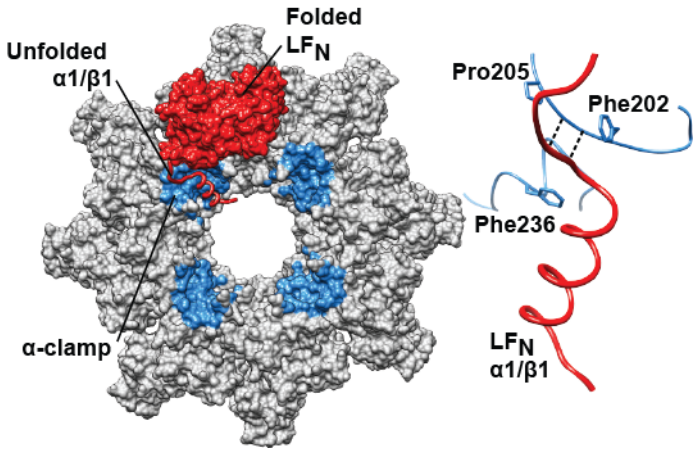

Figure 2.21. The α clamp catalyzes translocation by interacting with unfolded structure. (left) LF_N α1/β1 (red ribbon) unfolds from the structured carboxy-terminal subdomain (red surface) by binding into the α-clamp site (cyan surface) on the PA oligomer (gray surface). The interaction is comprised of nonspecific interactions. The α-clamp sites orient the unfolded structure toward the central pore, where the protein is translocated. (right) Residues in PA's α-clamp site (cyan) that affect LF_N and/or LF translocation are rendered as sticks. LF_N α1/β1 (red ribbon) and parallel β-sheet hydrogen bonds (black dotted lines) between LF_N β1 and PA β13 are shown.

binding activity likely reflects the general helical shape complementarity of the α-clamp site, which excludes ~1000 $Å^2$ on PA without making specific side-chain-side-chain interactions. Additionally, backbone hydrogen bonds, which are ubiquitous features of polypeptides, can provide nonspecific contact points between the translocase and substrate, as we observe between LF_N β1 and PA_N β13 (Figures 2.6a and 2.17b).

Broad peptide-binding specificity has been observed in other systems, including calmodulin (Meador *et al.* 1992; Meador *et al.* 1993); the ClpXP adapter, SspB (Levchenko *et al.* 2003; Levchenko *et al.* 2005); the chaperone, GroEL/ES (Landry & Gierasch 1991; Wang, Z. *et al.* 1999; Li *et al.* 2009); and the unfoldase, ClpA/Hsp100 (Hinnerwisch *et al.* 2005). For calmodulin, which is analogous structurally to the PA oligomer's α-clamp cleft, multiple peptide helices are recognized by the cleft formed by its twin Ca^{2+}-ion binding sites. The ClpXP adapter, SspB, binds multiple unstructured carboxy-terminal degradation signal tags in various conformations in a cleft. The chaperone complex, GroEL/ES, can bind to various amphipathic helices and strands. A substrate-binding site, identified in the unfolding machine ClpA/Hsp100, is located above the ϕ-clamp-type site and may be analogous to the α-clamp site on the PA oligomer.

Our structure provides new insight on how a nonspecific polypeptide clamp can unfold its substrate. By binding to LF_N in multiple locations using nonspecific interactions [i.e., in the α clamp (Figure 2.6a) and ϕ clamp (Krantz *et al.* 2005)], LF_N can be partially unfolded (Figure 2.14a) and maintained in a more strained (Figure 2.14d,e) and less stable conformation (Figure 2.14c-e). The region of LF_N that is most destabilized upon binding PA (Figure 2.14d,e) coincides with LF_N's β2-β4 sheet, which was previously reported as the mechanical breakpoint, or structure that is rate-limiting to the unfolding step of translocation (Thoren *et al.* 2009). Therefore, we infer the α-clamp site stabilizes unfolding intermediates, introduces strain into the mechanical breakpoint, and feeds unfolded structure into the central ϕ-clamp site.

We estimate that the costs associated with binding to the α-clamp site (Figure 2.7a-c) may be offset by orienting the substrate toward the central lumen (Figure 2.21), reducing the stability of the substrate (Figure 2.14c), and minimizing the diffusional mobility of unstructured regions before (Figure 2.14d,e) or during translocation (Krantz *et al.* 2005). We expect that nonspecific-clamping sites should lessen the counterproductive diffusive motions expected for large sections of unfolded polypeptide chain by maintaining contact with the unfolded chain and further reducing backbone conformational entropy, thus allowing the ΔΨ/ΔpH driving force to efficiently unfold (Thoren *et al.* 2009) and translocate proteins (Krantz *et al.* 2006) (Figure 2.17a,b). Although the α clamp forms a stable complex with unfolded structure, this intermediate does not represent a thermodynamic trap. Rather populating partially unfolded translocation intermediates would lower a much greater overall rate-limiting barrier expected in the absence of such intermediates, thereby allowing translocation to proceed on a biologically reasonable timescale.

2.5 Materials and Methods

Plasmids and proteins. Site-directed mutagenesis was performed using the commercial Quikchange procedure (Agilent Technologies) (Zheng *et al.* 2004). WT PA and PA mutants, including the construct used in the crystallization experiments, $PA^{\Delta MIL}$ [in which the membrane insertion loop (residues 303-324) was deleted and replaced with a type II turn sequence (Kintzer *et al.* 2009)], were expressed and purified as described (Krantz *et al.* 2005). Heptameric and octameric PA oligomers were produced as described (Kintzer *et al.* 2009).

LF, LF_N, His_6-DTA and mutants thereof were purified from overexpressing bacteria using standard Ni^{2+}-nitrilotriacetic-acid-(NTA)-affinity chromatography and Q-sepharose, anion-exchange chromatography (GE Healthcare, USA) (Krantz *et al.* 2006). Their six-histidine (His_6) tags were removed with bovine α-thrombin treatment (0.5 units/mg of protein) for 30 minutes at room temperature in 20 mM Tris (pH 8), 150 mM NaCl, 2 mM $CaCl_2$ and 1 M dextrose.

Amino-terminal truncation mutants of LF_N were made by PCR amplifying the truncated sequence and cloning the sequence into pET15b-LF_N (Lacy *et al.* 2002) via the 5' Nde I and 3' BamH I sites. These constructs are named Δ*n* LF_N, where *n* designates the amino acids that were deleted from the amino terminus. Note that due to the design of the thrombin cleavage site in the pET15b vector, thrombin cleavage leaves an additional GSHM sequence at the amino terminus of all pET15b LF_N constructs.

Fusions of LF and DTA were produced by introducing an in-frame Sac I restriction site into the pET15b-DTA vector (Blanke *et al.* 1996) prior to the DTA reading frame. The first 20 or 60 residues of LF [including the His_6 tag encoded in the pET15b-LF_N vector (Lacy *et al.* 2002)] were subcloned into the Sac I engineered DTA vector at the vector-encoded 5' Nco I site and the silent 3' Sac I site. These His_6-tagged LF_{1-20}-DTA and LF_{1-60}-DTA fusions and His_6-tagged DTA were purified from overexpressing bacteria using Ni^{2+}-NTA-affinity chromatography, Blue-sepharose chromatography (GE Healthcare), and Q-sepharose, anion-exchange chromatography. The His_6 tag was not removed from DTA, but the His_6 tag was removed from the His_6-tagged LF_{1-20}-DTA and LF_{1-60}-DTA fusions as described above. Note the unstructured His_6 tag on the amino terminus of His_6-tagged DTA is MGSSHHHHHHSSGLVPRG.

All LF and LF_N α1/β1-replacement constructs were made using a three-step, gene-synthesis procedure, according to the following scheme:

HM ^{1}AGGHGDVGMHVKEKEKNKDENKRKDEERNKT ^{32}QEEHLKEIMKHIV
^{45}KIEVKGEEAVKKEAAEKLLEKVPSDVLEMYKAIGGKIYI ^{84}**VD**

The bold face pairs of amino acids on either end are encoded by the restriction sites, Nde I and a silent Sal I site (V84 and D85), respectively, which are used for cloning. Superscripted numbers indicate the numbering convention of LF residues in 1J7N (Pannifer *et al.* 2001). The underlined sequence (residues 32-44) is the guest site, which is replaced with the following peptides:

$LF_{286-298}$ (α14) `SEEGRGLLKKLQI`
$LF_{710-722}$ (α28) `NSKKFIDIFKEEG`
EF_{22-34} (α1) `EKEKFKDSINNLV`

In the case of Aromatic Sequence 1 and Hydrophilic Sequence 1, the guest site is residues 32-46 ^{32}QEEHLKEIMKHIVK^{46}I, which is replaced by:

Aromatic Sequence (ArS1) `QEEHFKEIFKHFVKF`
Hydrophilic Sequence 1 (HS1) `QEEHSKEISKHSVKS`

Overlapping oligonucleotides encoding the desired sequences with the α1/β1 replacement were synthesized (Elim Biopharmaceuticals, Inc., Hayward, CA) and amplified by two rounds of polymerase chain reaction (PCR). In Round I, 20 nM of nested oligonucleotides with consistent annealing temperatures of ~55 °C were amplified in a standard PCR reaction. In Round II, 1 μL

of the PCR product made in Round I was amplified with the two outermost PCR primers (1 μM each) to make the synthetic double-stranded DNA fragment. These LF_N α1/β1-replacement synthetic DNA fragments were ligated via a 5' Nde I site and 3' Sal I site into either the pET15b-LF(Sal I) or the pET15b-LF_N(Sal I) construct, which contain an in-frame, silent Sal I restriction site in LF or LF_N at V84 and D85. The synthetic LF and LF_N constructs were purified and their His_6 tags were subsequently removed as described above.

Synthesis and purification of the $PA_8(LF_N)_4$ complex used in structural studies. $PA^{\Delta MIL}$ monomer was treated with trypsin at a ratio of 1:1000 (wt/wt) for 15 minutes at room temperature and then inhibited with soybean trypsin inhibitor at 1:100 (wt/wt) and phenylmethylsulphonyl fluoride (PMSF) at 0.1 mM. Crude mixtures of oligomeric $PA^{\Delta MIL}$ were produced by anion-exchange chromatography (Kintzer *et al.* 2009). Homogeneous $PA^{\Delta MIL}$ octamer was made by incubating this oligomeric mixture in 74 mM sodium acetate, 7 mM Tris, 0.62 M NaCl, 37 mM tetrabutylammonium bromide, 7% ethanol, 0.07% *n*-dodecyl-β-D-maltopyranoside, pH 5.7 and then microcentrifuging for 20 minutes (14k RPM) (Kintzer *et al.* 2009). Soluble $PA^{\Delta MIL}$ octamer (judged pure by electron microscopy) was complexed with LF_N at a one-to-one molar ratio with respect to PA monomer, purified over S200 gel filtration in 20 mM Tris, 150 mM NaCl, pH 8.0, and tested for homogeneity by nanoESI-MS (Figure 2.1b).

Crystallization and X-ray diffraction. The $PA_8(LF_N)_4$ complex was crystallized by the hanging-drop, vapor-diffusion method (McPherson 1976). Prior to crystallization, the protein complex was incubated with 20 mM ATP on ice for 10 minutes and then mixed one-to-one with well solution (13-17% (w/v) polyethylene glycol with average molecular weight 3000 Da, 100 mM cacodylic acid, 200 mM $MgCl_2$, pH 6.7-7.3.) Rectangular prisms grew overnight at 19 °C, maturing to dimensions of 100-300 μm. Crystals were harvested in a one-to-one mixture of well solution and cryoprotectant (50% (v/v) glycerol, 20 mM Tris-Cl, 150 mM NaCl, pH 8) and plunged into liquid N_2. X-ray diffraction data were collected at a wavelength of 1.1159 Å at 100 K on a Quantum 315r CCD detector at beamline 8.3.1 at the Lawrence Berkeley National Laboratory Advanced Light Source (MacDowell *et al.* 2004). A single crystal, belonging to the $P42_12$ space group, diffracted X-rays to 3.1 Å and had the unit cell dimensions, 178.4, 178.4, and 240.4 Å for *a*, *b*, and *c*, respectively (Table 2.1). The diffraction data (99.8% complete) were indexed and scaled in HKL2000 (Otwinowski & Minor 1997).

Model refinement. The $PA_8(LF_N)_4$ complex structure was solved by molecular replacement (MR) using PHASER (Storoni *et al.* 2004). The MR search model was a loop-stripped PA dimer from 3HVD (Kintzer *et al.* 2009). Two PA dimers were found in the asymmetric unit. Rigid-body and TLS refinement using PHENIX (Adams *et al.* 2004) produced F_o–F_c electron density consistent with a helical bundle that aligned to LF_N α2, α4, α9, and α10. Rounds of polyalanine-model building in COOT (Emsley & Cowtan 2004) and refinement in PHENIX revealed that the identified polyalanine secondary structure elements aligned well with a model of LF_N [LF residues 51-250 (PDB 1J7N (Pannifer *et al.* 2001))]. All of LF_N's secondary-structure elements, except the amino-terminus (LF_{1-28}) and the carboxy-terminal helix (α12), were identified and independently refined as rigid bodies to produce the initial model of the PA_2LF_N ternary complex. LF_N α1, β1, α2, α3, α4, and α5 and the loop regions, LF_{84-117}, $LF_{162-169}$ and $LF_{197-204}$, required either partial or extensive modeling to properly align them with the observed electron density. LF_{29-50} (α1/β1) was manually built extending from α2 (residue 51). Rounds of model building in COOT were followed by coordinate and *B*-factor refinement with non-crystallographic symmetry restraints in PHENIX. Backbone torsion angles were refined using the Torsion Optimization Procedure (TOP) provided by H. Gong, E. Haddadian, T.

Sosnick, and K. Freed at the University of Chicago. Molprobity analysis (Davis *et al.* 2007) of the structure shows that 91% of residues are in the favored Ramachandran regions, yielding an overall Molprobity score of 2.88 (87th percentile for a 3.10 (±0.25) Å resolution structure). Surface burial calculations and molecular graphics were computed in CHIMERA (Pettersen *et al.* 2004). The final model and refinement statistics are shown in Table 2.1.

Protein Data Bank accession code. The structure factors and coordinates for the $PA_8(LF_N)_4$ complex have been deposited in the PDB with the accession code 3KWV.

Planar lipid bilayer electrophysiology. Planar lipid bilayer currents were recorded using an Axopatch 200B amplifier (Molecular Devices Corp., Sunnyvale, CA) (Kintzer *et al.* 2009; Thoren *et al.* 2009). Membranes were painted on a 100-μm aperture of a 1-mL, white-Delrin cup with 3% (w/v) 1,2-diphytanoyl-*sn*-glycerol-3-phosphocholine (Avanti Polar Lipids, Alabaster, AL) in *n*-decane. *Cis* (side to which the PA oligomer is added) and *trans* chambers were bathed in various buffers as required. By convention, $\Delta\Psi \equiv \Psi_{cis} - \Psi_{trans}$ ($\Psi_{trans} \equiv 0$ V), and $\Delta pH \equiv pH_{trans} - pH_{cis}$.

To monitor LF_N, $LF_{1\text{-}20}$-DTA or $LF_{1\text{-}60}$-DTA binding to the PA channel, we first inserted PA channels into a planar lipid bilayer bathed in asymmetric KCl solutions buffered in 10 mM potassium phosphate ($[\text{added KCl salt}]_{cis}$ = 100 mM, $[\text{added KCl salt}]_{trans}$ = 0 mM, pH_{cis} = 6.5, pH_{trans} = 7.40). Once PA channel insertion was complete the cis buffer was perfused and exchanged to pH 7.40, 100 mM KCl. (The pH of the cis and trans buffers were matched to 0.01 units.) LF_N, $LF_{1\text{-}20}$-DTA or $LF_{1\text{-}60}$-DTA was then added to the cis side of the membrane at small increments, allowing for binding equilibrium to be maintained. Final current (I) levels were recorded, and the equilibrium current-block versus ligand concentration, [L], curves were fit to a simple single-binding site model, $I = I_o/(1 + K_d/[L]) + c$, to obtain the equilibrium dissociation constant, K_d, where I_o is the current amplitude and c is an offset.

In kinetic binding experiments, the rate of ligand binding and dissociation were recorded and fit to a single-exponential function, $I = I_o\exp(-kt) + c$, to obtain the observed rate constant, k, where I_o is the amplitude and c is an offset constant. The kinetic association rate constant, k_a, was computed using $k_a = k/[L]$. Dissociation of the ligand from the current-blocked complexes was initiated by perfusing the cis compartment with 5-10 volumes of buffer or by adding a 1-μM excess of Δ47 LF_N to compete with ligand binding. (The truncated form does not block the current when it binds to the PA channel.) The K_d could then be calculated from the kinetic rate constants using $K_d = k_d/k_a$. Kinetically and thermodynamically determined K_ds were self-consistent. Refer to Figure 2.11 for specific examples of these analyses.

The K_d for Δ27, Δ32, Δ39, Δ42, and Δ47 LF_N were deduced in equilibrium competition experiments with WT LF_N-PA channel complexes. PA channels were first inserted and then 0.1-0.8 nM of WT LF_N was added to the cis compartment. Once equilibrium was established the Δn LF_N competitor was added in increments. The degree of channel reopening established upon equilibration as a function of the competitor concentration was used to assess each competitor's K_d.

All LF_N translocation experiments were carried out as described previously using a universal pH bilayer buffer system (UBB: 10 mM oxalic acid, 10 mM phosphoric acid, 10 mM MES, 1 mM EDTA, and 100 mM KCl) at a symmetrical pH 5.6 (Thoren *et al.* 2009). Two to six replicate experiments were conducted for each mutant to establish the time (in seconds) for half of the substrate to translocate ($t_{1/2}$). The individual kinetic effects of LF_N mutations (MUT) were assessed by comparing the activation energy of translocation ($\Delta G‡$) at 40 mV for the mutant and WT LF_N, where $\Delta G‡ = RT \ln t_{1/2} / c$. R is the gas constant, T is the temperature, and c is 1

second. The change in $\Delta G‡$ ($\Delta\Delta G‡$) is reported at $\Delta\Delta G‡ = \Delta G‡(WT) - \Delta G‡(MUT)$. Efficiency, ε, was also obtained from each translocation record by the relation, $\varepsilon = A_{obs}/A_{exp}$, where A_{obs} is the observed amplitude of channels that reopened (or translocated), and A_{exp} is the expected amplitude if all of the channels reopened (or translocated).

LF translocation experiments were carried out similarly except that a 1.3-unit ΔpH was also applied during translocation. The pH of the UBB in the cis and trans chambers was adjusted to apply the proton gradient, where pH_{cis} = 6.1 and pH_{trans} = 7.4. The ΔΨ was 20 mV. LF translocation was assessed by $t_{1/2}$ and ε as described for LF_N. Relative translocation efficiency for each mutant was calculated as $\varepsilon_{MUT}/\varepsilon_{WT}$.

A special protocol was devised to analyze the PA R178A mutant due to LF's rapid dissociation from the channel. When LF was added to the channel, a 1.3-unit ΔpH was applied at a ΔΨ of 0 mV during the perfusion step. Translocation was then initiated by stepping the ΔΨ to 20 mV. This method, however, can only clamp one LF in the channel, presumably by engaging the substrate with the ΔpH-dependent mechanism involving the ϕ clamp (Krantz *et al.* 2005; Krantz *et al.* 2006). Thus the other substrates can dissociate during perfusion.

Equilibrium unfolding titrations. Guanidinium chloride titrations were performed on the Δ*n* LF_N truncations in 10 mM sodium phosphate, 0.75 M trimethylamine *N*-oxide, pH 7.5, 20 °C as described (Krantz *et al.* 2004; Thoren *et al.* 2009). Each titration point was probed by circular dichroism (CD) spectroscopy at 222 (±2) nm using a Jasco J-810 spectropolarimeter (Easton, MD). The CD-probed curves fit to a four-state thermodynamic model ($N \leftrightarrow I \leftrightarrow J \leftrightarrow U$) (Krantz *et al.* 2004).

Fluorophore labeling of LF_N and PA. Prior to all dye-modification reactions, excess DTT was removed from the Cys-substituted PA or LF_N proteins by buffer exchange on a G25 desalting column, equilibrated in 20 mM Tris-Cl pH 7.5, 150 mM NaCl. DTT-free PA_{83} K563C or a Cys-substituted LF_N was labeled with 10 molar equivalents of either 5-[2-[(2-Iodo-1-oxoethyl)amino]ethylamino]-1-naphthalenesulfonic acid (IEDANS), Alexa fluor 555 C_5 maleimide (AF_{555}), or Alexa fluor 647 C_5 maleimide (AF_{647}) (Invitrogen, USA) in the presence of 100 μM tris(2-carboxyethyl)phosphine (TCEP) (Sigma-Aldrich) at room temperature for 3 hours. The reaction was quenched with 5 mM DTT, and the labeled proteins were purified on a G25 desalting column to remove free-dye molecules. The labeling efficiency for each dye was determined by comparing the free-dye absorbance maximum and protein absorbance at 280 nm. The labeling efficiency was judged to be >90% in each case. For the IEDANS labeling of various Cys-substituted LF_Ns, we found the modification was >95% complete also by MALDI mass spectrometry.

Fluorescence anisotropy (FA). Fluorescence anisotropy (FA) was used to report on the changes in fluorophore mobility for LF_N labeled with IAEDANS. LF_N with single Cys residues introduced at specific sites (residues 37, 48, 72, 126, 164, 199, and 242) were modified with IEDANS dye and purified as described above. FA was measured using a FluoroMax-3 spectrofluorometer equipped with moveable linear polarizers. The excitation wavelength was 360 (±10) nm, and the emission wavelength range was 460-560 nm. The emission intensity value was the average intensity over this range. Each FA value, *a*, is calculated from the emission intensities from the parallel ($F_{\|}$) and perpendicular ($F\perp$) arrangement of the excitation and emission polarizers by $a = F_{\|} - F\perp / (F_{\|} + 2F\perp)$; however, a *G*-factor correction was applied to account for differences in sensitivities for the two different optical paths in the instrument. The FA signal change upon binding PA was not due to non-specific protein-protein associations, since the LF_N V48C*EDANS signal change is saturable at a 3:1 stoichiometry (LF_N:PA heptamer) (Figure

2.16b), which is consistent with the number of LF_N molecules that bind to PA_7 (Mogridge *et al.* 2002; Kintzer *et al.* 2009).

Normalized *B*-factor change and normalized FA-signal change calculations. Normalized differences in FA change (ΔFA_{norm}) or *B*-factor change (ΔB_{norm}) in LF_N upon binding PA were calculated by determining the mean change in the FA signal or *B* factor for all probed sites and then calculating the ratio of each individual signal change to this mean value. For the ΔB_{norm} calculation, $\Delta B_{norm} = B_{1,i} / \langle B_1 \rangle - B_{2,i} / \langle B_2 \rangle$. The $B_{1,i}$ and $B_{2,i}$ values are the average *B* factor for all atoms in the i^{th} residue in the free LF_N (structure 1) and the LF_N in complex with PA (structure 2), respectively. $\langle B_1 \rangle$ and $\langle B_2 \rangle$ are the average *B* factors for all residues in the entire LF_N chain taken from each respective structure. Likewise, for the ΔFA_{norm} calculation, $\Delta FA_{norm} = FA_{1,i} / \langle FA_1 \rangle - FA_{2,i} / \langle FA_2 \rangle$. This calculation is treated in an analogous manner, except only the 7 Cys-substituted, fluorescently-probed LF_N residues were considered.

FRET-probed, PA-assembly assay. A 1:1 mixture of dye-labeled $_n$PA ($_n$PA K563C*AF_{555} and $_n$PA K563C*AF_{647}) was diluted to 10 nM in buffer (10 mM sodium cacodylate, 100 mM potassium chloride, pH 7.4 at room temperature) either in the presence or absence of 10 nM LF_N or Δ47 LF_N, following a prior method (Wigelsworth *et al.* 2004). Assembly was reported by the increase in the 668-nm and 566-nm emission intensity ratio (F_{668}/F_{566}) upon excitation at 555 nm, which reached a steady state in about one hour. The excitation and emission bandwidths were 2 nm. Emission intensity ratios were collected throughout the record at 5-minute intervals on a FluoroMax-3 spectrofluorometer (Horiba Jobin Yvon, Edison, NJ) using a 1×1-cm cuvette. Curves were fit with a second-order kinetic model described previously (Wigelsworth *et al.* 2004).

NanoESI-MS. Mass spectra of the protein complexes were acquired as described previously (Kintzer *et al.* 2009) using a quadrupole time-of-flight (Q-Tof) mass spectrometer with a Z-spray ion source (Q-Tof Premier, Waters, Milford, MA). Ions were formed using nanoelectrospray (nano-ESI) emitters prepared by pulling borosilicate capillaries (1.0 mm O.D./0.78 mm I.D., Sutter Instruments, Novato CA) to a tip I.D. of ~1 μm with a Flaming/Brown micropipette puller (Model P-87, Sutter). The instrument was calibrated with CsI clusters. The protein solution for the stoichiometry determinations was concentrated to 10 μM followed by dialysis into 10 mM ammonium bicarbonate, pH 7.8. Immediately prior to mass analysis, the solution was diluted 1:1 with 150-300 mM ammonium acetate, pH 7.8. A 0.127-mm-diameter platinum wire was inserted through the capillary into the solution, and electrospray was maintained by applying a 1-1.3 kV potential relative to instrument ground. Raw data were smoothed three times using the Waters MassLynx software mean smoothing algorithm with a window of 50 *m*/*z* (mass-charge ratio).

Electron microscopy. Each PA oligomer was diluted into an EM buffer (20 mM Tris, 150 mM NaCl, pH 8), making a final concentration of 20-30 nM with respect to PA monomer. 400 mesh copper grids were successively covered by a holey carbon film and a continuous carbon film. 4 μl of the diluted PA oligomer sample were applied to a freshly glow-discharged support grid for 30 s and then stained in 5 successive drops (75μl) of 2% uranyl acetate (Sigma-Aldrich). Negative-stain EM images were recorded with a Tecnai 12 (FEI Company, Hillsboro, OR) operated at 120 kV at 49,000× magnification. Images were taken using a CCD camera (2.13 Å/pixel specimen scale). Particle images were selected for each data set using automatic or manual particle picking using boxer in EMAN (Ludtke *et al.* 1999). Reference-free processing was done using the software package, SPIDER (Frank *et al.* 1996). Images were subjected to three successive cycles of multi-reference alignment, multivariate statistical analysis, and

classification (Stark *et al.* 1995; van Heel *et al.* 1996). The last classification was done using only the lowest order eigenvectors, as described elsewhere (White *et al.* 2004), to separate the data by size and by the heptameric and octameric oligomerization states. A second method of image processing was used whereby crystal-structure-reference images were made from two-dimensional projections of low resolution density maps generated from the crystal structures of the PA heptamer (Lacy *et al.* 2004) and octamer (Kintzer *et al.* 2009) using SPIDER (Frank *et al.* 1996). Crystal-structure-referenced images were aligned and classified using the lowest order eigenvectors as stated above. All final class-average images were manually inspected for their oligomer number. The number of particles per classification was used to determine the percentages of heptamers and octamers in each sample. About 2000 to 10000 particles were analyzed per sample. Each method of classification, reference-free or crystal-structure-referenced, produced similar results (±2%) (Kintzer *et al.* 2009). The reported percentages of heptameric and octameric PA are given as the mean of the referenced and reference-free analysis.

Macrophage cytotoxicity. LT cytotoxicity was monitored by an enzyme-coupled lactose dehydrogenase (LDH) release assay (Decker & Lohmann-Matthes 1988). Immortalized bone marrow macrophages from 129 mice (a gift from the Vance Lab at UC Berkeley) were grown to confluence in RPMI 1640 medium (Invitrogen) supplemented with 10% fetal bovine serum (Invitrogen), 100 units/mL penicillin (Sigma-Aldrich), and 100 μg/mL streptomycin (Sigma-Aldrich) in a humid, 5%-CO_2 atmosphere at 37 °C. One day prior to conducting assays, cells were trypsinized and re-plated at 10^5 cells/well. Cells were treated in triplicate with varying concentrations of LF and constant concentrations of PA. Dilutions of LF were prepared in ice-cold phosphate-buffered saline (PBS). Toxin-treated cells were incubated for 4 hours at 37 °C. Plates were then centrifuged at 1400 RPM, and 50 μL of supernatant were removed and added to a new 96-well plate. The supernatant was incubated with 20 μL of lactate solution (36 mg/mL in PBS) and 20 μL of *p*-iodonitrotetrazolium chloride (2 mg/mL in PBS with 10% dimethyl sulfoxide). The enzymatic reaction was started with the addition of 20 μL of nicotinamide adenine dinucleotide (NAD^+)/diaphorase solution (13.5 units/ml diaphorase and 3 mg/ml NAD^+). After 15 minutes, the products were analyzed on a spectrophotometric microplate reader (Bio-Rad Laboratories, Richmond, CA) at 490 nm. The change in the absorbance signal is proportional to the number of lysed cells, where the amount of LDH released was normalized to the value obtained in wells treated with 1% Triton X-100 detergent. Effective-concentration-for-50%-lysis values (EC_{50}) were determined by fitting normalized cell lysis versus PA concentration data in ORIGIN (OriginLab Corp., Northampton, MA).

PA-DTA-fusion cytotoxicity assays were carried out using J774 mouse macrophage cells grown in Dulbecco's Modified Eagle's Medium with 5% fetal bovine serum (FBS), 200 μM glutamine, and 10 μg/ml gentamycin. A 96-well plate was seeded with $5x10^4$ cells per well (in 100μl media) 16-20 hours before toxin application and incubated at 37 °C and 5% CO_2. LF-DTA fusions at varying concentrations and PA at a final concentration of 1 μg/ml were added to the cell culture (diluted in PBS supplemented with 1% bovine serum albumin). Plates were incubated at 37 °C and 5% CO_2 for 20-24 hours. Cell lysis was assayed using the LDH release assay described above.

Chapter 3

Domain flexibility modulates the heterogenous assembly mechanism of anthrax protective antigen

3.1 Abstract

Individually, the three protein components of anthrax toxin are nontoxic, but when they assemble, they form active holotoxin complexes. The role of the protective antigen (PA) component of the toxin is to deliver the two other enzyme components, lethal factor (LF) and edema factor (EF), across the plasma membrane and into the cytoplasm of target cells. PA is produced as a proprotein, which must be proteolytically activated, and generally cell-surface activation is mediated by a furin-family protease. Activated PA can then assemble into one of two non-interconverting oligomers, a homoheptamer and homooctamer, which have unique properties. Herein we describe molecular determinants that influence the stoichiometry of PA in toxin complexes. By tethering PA Domain 4 to Domain 2 with two different length cross-links, we can control the relative proportions of PA heptamers and octamers. The longer cross-link favors octamer formation, whereas the shorter one favors formation of the heptamer. X-ray crystal structures of PA (up to 1.45 Å resolution), including these cross-linked PA constructs, reveal that a hinge-like movement of Domain 4 correlates with the relative preference for each oligomeric architecture. Furthermore, we report the conformation of the flexible loop containing the furin-cleavage site and show that for efficient processing, the furin site cannot be moved ~5 or 6 residues within the loop. We propose that there are different orientations of Domain 4 relative to the main body of PA that favor the formation of either the heptamer or the octamer.

3.2 Introduction

Anthrax toxin (Atx) (Young & Collier 2007; Thoren & Krantz 2011) is a virulence factor secreted by pathogenic strains of *Bacillus anthracis*. Atx consists of three nontoxic protein components (Smith & Keppie 1954; Stanley & Smith 1961; Beall *et al.* 1962), namely protective antigen (PA, 83-kDa), lethal factor (LF, 90-kDa), and edema factor (EF, 89-kDa). The PA component assembles into a ring-shaped oligomer capable of forming a membrane-spanning translocase channel, which delivers the two enzyme components, LF and EF, into the cytosol of a host cell. LF is a Zn^{2+}-dependent protease (Duesbery *et al.* 1998; Duesbery & Vande Woude 1999; Pannifer *et al.* 2001), which cleaves host-cell mitogen-activated protein kinase kinases (Duesbery *et al.* 1998; Duesbery & Vande Woude 1999). PA, LF, and EF are individually nontoxic; however, LF plus PA creates lethal toxin (LT), which can alter normal cell function and may even cause death (Pezard *et al.* 1991). EF is Ca^{2+}- and calmodulin-dependent adenylyl cyclase (Leppla 1982; Leppla 1984; Drum *et al.* 2002). Similarly, PA plus EF produces edema toxin (ET), which induces tissue swelling and may also cause death.

The PA component contains four different domains (Domain 1-4), which are critical to various stages of toxin assembly and translocation. To form cytotoxic complexes, PA must first assemble with LF and/or EF. Two different types of assembly pathways have been described: a cell-surface pathway (for review, see ((Young & Collier 2007)) and a plasma-based/extracellular pathway (Kintzer *et al.* 2009; Kintzer *et al.* 2010). On the surface of host cells, PA binds to one of two known Atx receptors: ANTXR1 (Bradley *et al.* 2001) and ANTXR2 (Scobie *et al.* 2003). The PA-ANTXR2 interaction (Lacy *et al.* 2004; Santelli *et al.* 2004) is rather stable and dissociates with a half-life measured in days (Wigelsworth *et al.* 2004). The interaction involves

Domain 2 and Domain 4 in PA, such that the latter domain coordinates the receptor's Ca^{2+} or Mg^{2+} dependent metal ion adhesion site (Lacy *et al.* 2004; Lacy *et al.* 2004; Santelli *et al.* 2004; Wigelsworth *et al.* 2004; Scobie *et al.* 2006). ANTXR2-bound PA is subsequently cleaved by a furin-family protease yielding the proteolytically-activated or nicked form, called $_{n}$PA. After a 20-kDa portion of $_{n}$PA (PA_{20}) dissociates, the remaining 63-kDa, receptor-bound portion (PA_{63}) self-assembles into an ~2:1 mixture of heptameric (PA_7) (Milne *et al.* 1994; Petosa *et al.* 1997; Lacy *et al.* 2004) and octameric (PA_8) (Kintzer *et al.* 2009; Feld *et al.* 2010; Kintzer *et al.* 2010) ring-shaped oligomers. The complexes are endocytosed (Abrami *et al.* 2003) and transferred to an acidic compartment (Friedlander 1986), where the lower pH conditions induce the PA oligomers to form transmembrane translocase channels (Blaustein *et al.* 1989; Miller *et al.* 1999; Katayama *et al.* 2008). LF and EF translocate through the PA channel to enter the cytosol of the host cell.

Recent studies (Kintzer *et al.* 2009; Feld *et al.* 2010; Kintzer *et al.* 2010) investigating the molecular mechanism of PA oligomerization have revealed that the assembly of the octameric PA oligomer is largely dependent on the degree to which even-numbered PA intermediates (PA_2 and PA_4) are populated. Co-assembly factors such as LF, EF, or dimeric ANTXR2 (dANTXR2) presumably favor the formation of these PA_2 intermediates and, correspondingly, PA_8 formation (Kintzer *et al.* 2009) (Figure 3.1a). Interestingly, when PA's membrane insertion loop (MIL) is deleted ($PA^{\Delta MIL}$), these mutant subunits can form 5-10 fold more octamer than heptamer—even in the absence of co-assembly factors. The structural explanation for the propensity of $PA^{\Delta MIL}$ to form the octamer is not well understood. Alignment of the PA subunits from the heptameric (Lacy *et al.* 2004) and octameric (Kintzer *et al.* 2009) oligomer structures reveals that the conformation of Domain 4 (D4, residues 596-735) is quite flexible (Figure 3.1b). Relative to the main body of the protein, PA D4 may be treated as a rigid body whose "pitch" may be described as a rotation around the "hinge" joining Domain 3 (D3, residues 458-595) and D4. While oligomerization may occur at a large contiguous surface containing residues from Domain 1 (D1, residues 1-258) (Lacy *et al.* 2004; Kintzer *et al.* 2009; Feld *et al.* 2010), Domain 2 (D2, residues 259-457) (Lacy *et al.* 2004; Kintzer *et al.* 2009) and D3 (Mogridge *et al.* 2001; Lacy *et al.* 2004; Kintzer *et al.* 2009), it is unknown how the orientation of D4 may modulate the oligomerization pathway and determine the stoichiometry of PA in the complex.

Mammalian furin-type proteases are members of the prohormone convertase (PC) family (Nakayama 1997; Steiner 1998; Thomas 2002; Henrich *et al.* 2003). A close homolog of furin in yeast, kexin, functions analogously by cleaving prohormone proteins (Fuller *et al.* 1989; Fuller *et al.* 1989; Holyoak *et al.* 2003). Furin, the most well-characterized PC (EC 3.4.21.75) (Henrich *et al.* 2003), cleaves its substrates at a cationic, R-X-(R/K/X)-R↓, consensus site (where the scissile bond is indicated with an ↓) (Remacle *et al.* 2008). Furin is involved in diverse cellular processes, such as homeostasis and embryogenesis, and it is expressed in all tissues and cells as a secreted membrane *N*-linked-type glycoprotein that matures in the trans-Golgi network. Furin and its yeast homolog, kexin, each anchor to the cell membrane via a homologous single-pass transmembrane helix located on its carboxy terminus (Fuller *et al.* 1989; Fuller *et al.* 1989). As a membrane protein, furin can cycle between the cell surface and the endosomal compartment (Molloy *et al.* 1994; Thomas 2002). Furin functions to proteolytically process many types of membrane-bound and soluble factors, including: growth factors, extracellular matrix proteases, viral proteins and bacterial toxins.

The role of furin and related PCs in viral and bacterial pathogenesis is widespread. Many bacterial and viral proproteins require proteolytic activation prior to maturation into either active

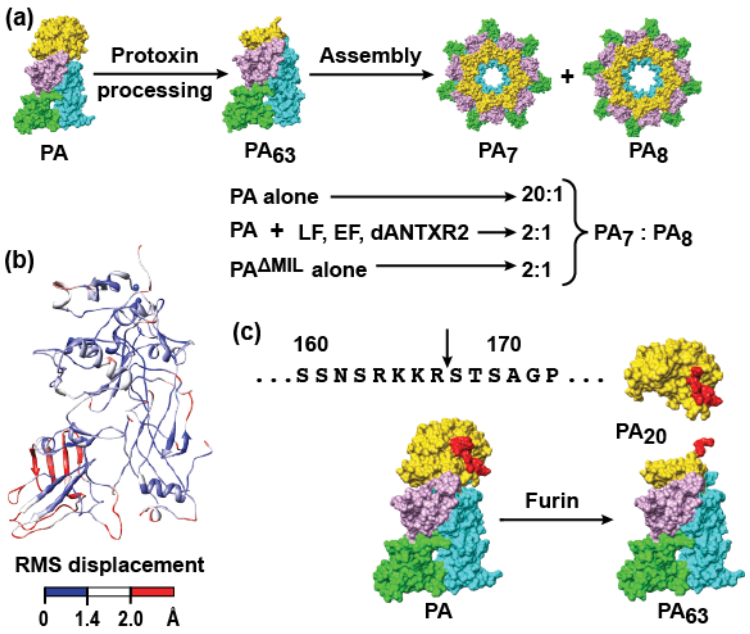

Figure 3.1. Anthrax toxin assembly is heterogeneous and requires proteolytic processing. (a) Cartoon depiction of anthrax toxin assembly. PA domains are colored as follows: D1 (gold), D2 (cyan), D3 (violet), and D4 (green). Upon oligomerization, PA forms both PA_7 and PA_8 complexes. The oligomeric ratio of PA complexes (PA_7:PA_8) decreases from about 20:1 for unliganded PA assembly to 2:1 in the presence of ligand (LF, EF, or dANTXR2). The $PA^{\Delta MIL}$ construct can form complexes at a 2:1 oligomeric ratio; however, it does so in the absence of ligand. **(b)** Global Cα alignment of a PA subunit (chain G) from the PA_7 crystal structure (PDB 1TZO) (Lacy *et al.* 2004) onto a PA subunit (chain G) from the PA_8 crystal structure (PDB 3HVD) (Kintzer *et al.* 2009). The ribbons are colored based on the RMS displacement of the Cα positions. Small, intermediate and large RMS displacements are colored blue, white and red, respectively. **(c)** (Above) The PA sequence containing the furin-cleavage site, where the scissile bond is indicated with an arrow. (Below) The furin-dependent cleavage reaction of the PA proprotein releases the 20-kDa fragment (PA_{20}) and yields the 63-kD portion, called PA_{63}. PA_{63} ultimately assembles into PA_7 and PA_8 oligomers.

enzymes or assembled virulence factors. Generally, but not always, furin-activated proteolysis permits these pathogenic proteins to subsequently fuse with or penetrate the host cell membrane. Included amongst these virulence factors are the bacterial toxins (Gordon & Leppla 1994), anthrax (Molloy *et al.* 1992), shiga (Garred *et al.* 1995), diphtheria (Tsuneoka *et al.* 1993), tetanus, botulinum, *Clostridium septicum* α-toxin (Gordon *et al.* 1997), and aerolysin (Abrami *et al.* 1998); the influenza A virus fusion protein, hemagglutinin (Stieneke-Grober *et al.* 1992); human immunodeficiency virus-1 envelope glycoprotein, gp160 (Hallenberger *et al.* 1992); flaviviruses, such as tick-borne encephalitis and West Nile (Stadler *et al.* 1997); and filoviruses, such as Marburg and Ebola (Volchkov *et al.* 1998).

In order for Atx to assemble, PA must first be cleaved by a furin-type protease. Cleavage occurs after the sequence, ^{164}RKKR↓S, in a solvent-exposed loop within D1 (Molloy *et al.* 1992). Selective (Remacle *et al.* 2010) and nonselective (Sarac *et al.* 2004; Komiyama *et al.* 2005) inhibition of furin and furin-like proteases strongly impairs Atx function. Furthermore, substrate-mapping studies of many related furin-type proteases show that PA's RKKR furin-recognition sequence (Figure 3.1c) has maximized the motif for recognition by a broad number of these furin-type PC enzymes (Remacle *et al.* 2008). While NMR techniques have elucidated the structure of a furin cleavage site in a 19-residue peptide from gp160 (Oliva *et al.* 2002), atomic-resolution structural information for furin-cleavage sites in proviral and protoxin proteins have been challenging to obtain, because the sites are typically found on flexible loop structures. To gain an understanding on the structural basis of toxin assembly and furin activation we crystallized a variant of anthrax toxin PA and solved its structure to 1.45-Å resolution.

3.3 Results

Crystal structure of the PA$^{\Delta MIL}$ monomer. For our initial crystallographic studies, we chose a PA construct (PA$^{\Delta MIL}$) in which the MIL in D2 (residues 303-324) was replaced with the Pro-Gly dipeptide (which generally favors the formation of a Type II turn). The highly dynamic MIL is absent from previous structures of the PA monomer (Petosa *et al.* 1997; Santelli *et al.* 2004; Wimalasena *et al.* 2010); therefore, we assumed its absence would improve crystal quality. Furthermore, we had previously solved the structure of PA_8 to 3.2-Å resolution (Kintzer *et al.* 2009) and the $PA_8(LF_N)_4$ complex to 3.1-Å resolution (Feld *et al.* 2010) using the PA$^{\Delta MIL}$ construct. These PA$^{\Delta MIL}$ crystal forms diffracted to higher resolution than the PA_7 crystal forms using the wild-type (WT) PA construct, which previously diffracted to 3.6 Å (Lacy *et al.* 2004) and 4.5 Å (Petosa *et al.* 1997).

The PA$^{\Delta MIL}$ construct yielded large, well-diffracting orthorhombic crystals belonging to the $P2_12_12_1$ space group and diffracted X-rays to 1.45 Å. The structure was solved using molecular replacement; one PA protein was identified in the asymmetric unit. Interestingly, the unit cell dimensions and space group are identical to what Petosa *et al.* referred to as "Crystal form 1". From their study, Crystal form 1 did not diffract as well as "Crystal form 2," the basis for the PA structure with Protein Data Bank (PDB) accession code 1ACC (Petosa *et al.* 1997). Coordinates for Crystal form 1 were not deposited in the PDB, however. Contiguous electron density was identified for residues 15-735 of PA$^{\Delta MIL}$ (Figure 3.2a). Loop regions lacking in the 1ACC search model were built manually in the experimental electron density using standard methods.

The resolution of our structure allowed us to refine all of PA's atomic coordinates individually, and the atomic displacement parameters (ADP) were refined anisotropically. Numerous solvent molecules were resolved, and their ADP values were also refined

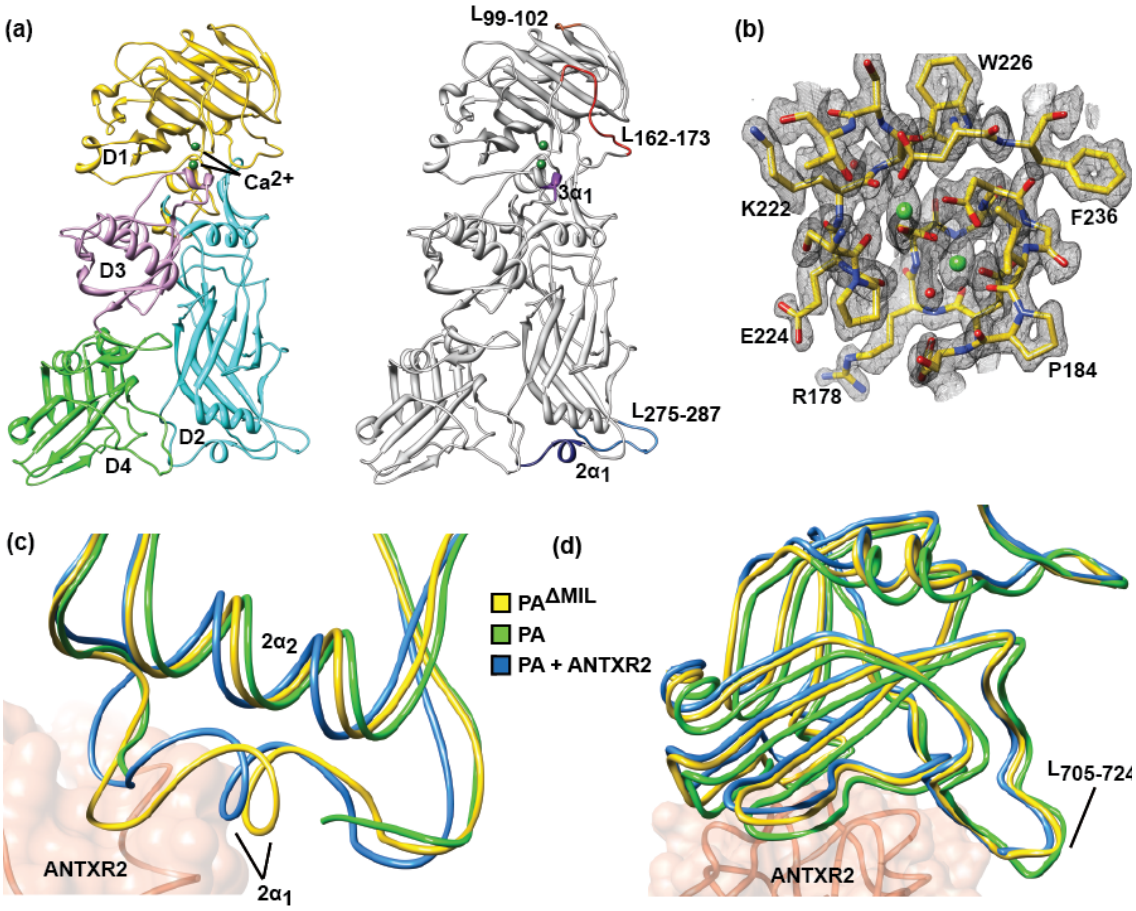

Figure 3.2. The structure of PA$^{\Delta MIL}$. (a) (left) Ribbons depiction of PA$^{\Delta MIL}$. D1 (gold), D2 (cyan), D3 (violet), D4 (green), and calcium ions (dark green). (right) Loop structures not present in 1ACC (Petosa *et al.* 1997) are indicated as follows: $L_{99\text{-}102}$ (orange), $L_{162\text{-}173}$ (red) containing the furin-cleavage site, $L_{275\text{-}287}$ (blue), the loop containing $2\alpha_1$ (navy blue), and $3\alpha_1$ (purple). **(b)** Sample $2F_o$-F_c electron density calculated at the end of refinement to 1.45 Å for the twin Ca^{2+}-coordination site contoured at σ=2.5. **(c)** LSQ Cα alignment of residues 16-340 for PA monomer (green, 1ACC) (Petosa *et al.* 1997), ANTXR2-bound PA (blue, 1T6B) (Santelli *et al.* 2004) and PA$^{\Delta MIL}$ (yellow). ANTXR2 (coral) from IT6B is included for reference. **(d)** D4 relation for PA monomer, ANTXR2-bound PA, and PA$^{\Delta MIL}$ (colored as in panel c) LSQ Cα superposition for residues 25-595.

anisotropically. Overall, the model is in good agreement with the electron density, with an R value of 19.4% and an R_{free} value of 21.6% (Table 3.1). The resulting structure closely resembles previous PA monomer structures. D1 contains a jelly-roll motif with peripheral small helices and loops, including the furin-cleavage site-containing $L_{162\text{-}174}$, as well as a twin-Ca^{2+} coordination site resembling the classical EF hand motif (Figure 3.2b). D2 contains a Greek-key motif implicated in ultimately forming the 14- or 16-stranded β-barrel channel, through which LF and EF translocate. Additionally, a loop containing $2\alpha_1$ was resolved, which has been implicated in binding to ANTXR2 (Lacy *et al.* 2004; Santelli *et al.* 2004; Scobie *et al.* 2006). D3 contains a mixed β-sheet and associated helices, including $3\alpha_1$, which was not present in 1ACC. D4 contains an immunoglobulin-like β-sandwich and is the most flexible of the PA domains.

Ordered loop regions. We resolved continuous electron density for $PA^{\Delta MIL}$ from residues 15 to 735 at the carboxy-terminus. While acknowledging that the membrane insertion loop (residues 305-324) is missing by design, this is the most complete PA model to date. (N.B. the numbering convention of 1ACC is retained.) While our structure and the three reported monomeric structures, PA (PDB 1ACC) (Petosa *et al.* 1997), the ANTXR2-bound PA (PDB 1T6B) (Santelli *et al.* 2004), and the 2-fluoro-histidine substituted PA (PDB 3MHZ) (Wimalasena *et al.* 2010), backbone align with close agreement with a root mean square (RMS) deviation < 1 Å, a number of surface-loop regions not modeled in 1ACC are present in the $PA^{\Delta MIL}$ structure. Residues 99-102 and 512-515 are not modeled in 1ACC; however, they are present in 1T6B, while the latter is present in 3MHZ. Residues 163 and 168-173, a region that includes the furin proteolytic-cleavage site, are not present in any deposited structure of PA, both monomeric and oligomeric. Interestingly, residues 159-161 are in a different conformation in 1ACC relative to our model, indicating that the loop containing the furin-cleavage site can adopt multiple conformations. Residues 162 and 164-167 are found in 3MHZ in approximately the same conformation as reported here. Also D2 residues 275-287 are present, modeled here as an extended loop whose atoms have ADP values comparable to the average protein atom. Residues 343-350, which are not present in 1ACC, are modeled in both 1T6B and 3MHZ.

Receptor-binding loop-helix $2\alpha_1$. In the published PA monomer structure (1ACC) (Petosa *et al.* 1997), residues 343-350 are not modeled. In the co-crystal structure of PA and ANTXR2 (1T6B) (Santelli *et al.* 2004) and the structure of the 2-fluoro-histidine substituted PA (3MHZ) (Wimalasena *et al.* 2010), these residues are well-ordered in a loop-helix motif (Figure 3.2c). Presumably, this ordering is due to the strong van der Waals interaction of PA residue L320 with ANTXR2 L154, which contributes to the strong energetics of PA-receptor binding. This loop-helix is also well-ordered for $PA^{\Delta MIL}$; the degree of order in this structure may reasonably be explained by the crystal contacts between $2\alpha_1$ and $3\alpha_1$. $PA^{\Delta MIL}$ and ANTXR2-bound PA from 1T6B align well and are in good agreement (RMS <0.3 Å). Alignment of the structures by Cα of residues 16-340 results in a slight shift of $2\alpha_1$ in the $PA^{\Delta MIL}$ structure relative to 1T6B. This movement frees up potential steric clashes of PA $2\alpha_1$ with the extracellular PA-binding domain of ANTXR2; therefore, the positioning of $2\alpha_1$ in the structure presented here likely represents the biological conformation in the absence of receptor and agrees well with the 3MHZ structure (Wimalasena *et al.* 2010).

Changes in orientation of PA D4. In the structure of PA in complex with ANTXR2 (1T6B), D4 undergoes a rigid-body conformational change, resulting in a slight tilting of the domain relative to the unbound structure of PA (1ACC) (Petosa *et al.* 1997; Santelli *et al.* 2004). This movement is actually required so that ANTXR2 may bind PA at two discontinuous contact points, i.e., at $2\alpha_1$ in D2 and around D683 in D4 (Lacy *et al.* 2004; Santelli *et al.* 2004).

Table 3.1 Data collection and refinement statistics

Data collection	**PA$^{\Delta MIL}$ pH 8.5**	**PA$^{\Delta MIL}$ C_{337}-C_{664}**	**PA C_{337}-C_{664}-DCA**	**PA$^{\Delta MIL}$ pH 6.5**
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, b, c (Å)	72.36, 93.39, 118.08	71.56, 94.60, 119.80	71.29, 93.69, 117.85	71.40, 93.93, 117.85
α, β, γ (º)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.1159	1.1159	1.1159	1.1159
Resolution (Å)	23.0-1.45(1.49-1.45) *	25.06-2.06(2.09-2.06)#	24.9-1.83 (1.87-1.83)$	9.99-1.70 (1.73-1.70)%
$R_{merge.}$ (%)	9.6(74.8)	11.4(86.8)	12.0(75.7)	8.5(85.5)
$I / \sigma I$	13.9(1.9)	15.7(2.5)	20.9(4.3)	22.1(2.1)
Completeness (%)	90.3(81.0)	99.9(99.2)	99.7(96.0)	99.8(98.0)
Redundancy	6.5(3.4)	7.6(7.7)	14.2(10.5)	8.1(7.4)
Wilson B	27.000	36.331	26.490	27.278
Refinement				
Resolution (Å)	23.0-1.45	23.86-2.06	24.9-1.83	9.99-1.70
No. reflections	126,263	50,702	70,160	87,171
R_{work}, R_{free} (%)	19.4/ 21.6	20.7 / 23.4	19.3 / 22.6	19.4 / 21.7
No. atoms				
Protein	5647	5589	5616	5590
Ligand/ion	17	2	13	2
Water	348	208	422	398
Average B-factors				
Protein	21.7	47.8	36.6	40.3
Ligand/ion	29.5	31.0	33.4	27.0
Water	29.9	42.9	39.3	43.0
R.m.s deviations				
Bond lengths (Å)	0.008	0.008	0.016	0.007
Bond angles (º)	1.151	1.041	1.087	1.053
Ramachandran (%)				
Favored	98.1	97.9	98.1	97.1
Outliers	0	0	0	0
MolProbity score	1.51	1.93	1.58	1.86

* Values in parenthesis refer to data in the highest resolution shell (1.49-1.45 Å)
# Values in parenthesis refer to data in the highest resolution shell (2.13-2.09 Å)
$ Values in parenthesis refer to data in the highest resolution shell (1.87-1.83 Å)
% Values in parenthesis refer to data in the highest resolution shell (1.73-1.70 Å)

Interestingly, a backbone alignment of the $PA^{\Delta MIL}$ structure with the ANTXR2-bound PA structure from D1-D3 reveals that the orientation of D4 relative to the main body (D1-D3) does not significantly deviate. Even though both the $PA^{\Delta MIL}$ and 1ACC structures are not bound to receptor, the orientation of D4 in $PA^{\Delta MIL}$ more closely resembles that of the receptor-bound structure, 1T6B (Figure 3.2d). Thus the orientation of D4 is the most significant structural change we can identify in the $PA^{\Delta MIL}$ structure.

Crystal structures of cross-linked PA monomers. In order to test the structural consequence of the orientation of D4 relative to the main body of PA, we engineered a pair of Cys mutations into the D2-D4 interface. We selected the residues, S337C and N664C, and incorporated the mutations into both the WT PA and $PA^{\Delta MIL}$ backgrounds. We then produced two different length cross-links. For the shorter-length cross-link, we allowed a disulfide bond to form between the two Cys residues, and we refer to this cross-link as C_{337}-C_{664}. For the longer-length cross-link, we used the small, thiol-specific cross-linker, 1,3-dichloroacetone (DCA) (Yin *et al.* 2000), and we refer to this DCA link as C_{337}-C_{664}-DCA (Figure 3.3a). We solved the crystal structures of both $PA^{\Delta MIL}$ C_{337}-C_{664} and PA C_{337}-C_{664}-DCA (Table 3.1). We found that C_{337}-C_{664} contains a disulfide bond (Figure 3.3b), while PA C_{337}-C_{664}-DCA contains a dithioether cross-link consistent with the DCA modification (Figure 3.3c). The two structures agree well with other PA monomer structures. We then performed a global, Cα alignment of $PA^{\Delta MIL}$ C_{337}-C_{664} to $PA^{\Delta MIL}$ to assess whether the orientation of D4 was impacted by the cross-links. We find that domains D2 and D4 showed the largest displacement, resulting in maximum RMS displacements greater than 0.8 Å (Figure 3.3d). This deviation increases with distance from the hinge between D3 and D4, indicating that the deviation is the result of a change in the hinge angle relating D4 with the main body of the protein. When we performed a similar RMS-displacement analysis, comparing PA C_{337}-C_{664}-DCA to $PA^{\Delta MIL}$. We found that the backbone positions did not move appreciably (Figure 3.3d). Finally, for the purpose of comparison, we aligned the Cα's of PA from 1ACC to $PA^{\Delta MIL}$ to assess the RMS deviation contribution of the MIL. While surface loops and other protein periphery contributed to large RMS-displacements in excess of 11 Å, the backbone atoms comprising the major secondary structures in D1, D2 and D3 show relatively strong agreement. We found that most of the major deviations of backbone positions (> 1.2 Å) were localized in D4 (Figure 3.3d).

Crystal structure of $PA^{\Delta MIL}$ at pH 6.5. During our analysis of structural changes in the relative orientation of D4, we realized that certain structures were obtained at slightly alkaline pH values of ~8.5 and other crystals were obtained at slightly acidic pH values of ~6.5. In order to rule out the possibility that the D4 movements we observe are simply due to differences in the pH of the crystallization condition, we crystallized $PA^{\Delta MIL}$ at pH 6.5. The structure was of similar quality to the others described thus far (Table 3.1). We then compared the structure of $PA^{\Delta MIL}$ at pH 6.5 to that obtained at pH 8.5 by computing the RMS displacement of the Cα backbone positions between the two different pH conditions. We found that the orientation of D4 is identical in the pH 8 and pH 6.5 crystal forms, and this change in pH does not dictate the relative orientation of D4 in PA (Figure 3.3d).

The D2-D4 interface modulates the propensity of PA to form octamers. In order to study the contribution of PA's D2-D4 interface to its oligomeric preference, we co-assembled native and cross-linked PA and $PA^{\Delta MIL}$ constructs with WT LF_N and analyzed the assembly products by negative-stain electron microscopy (Figure 3.4). We obtained about ~1500-5000 images of ring-shaped particles in each sample; these images were processed using single-particle alignment, classification, and class-averaging image analysis. From these

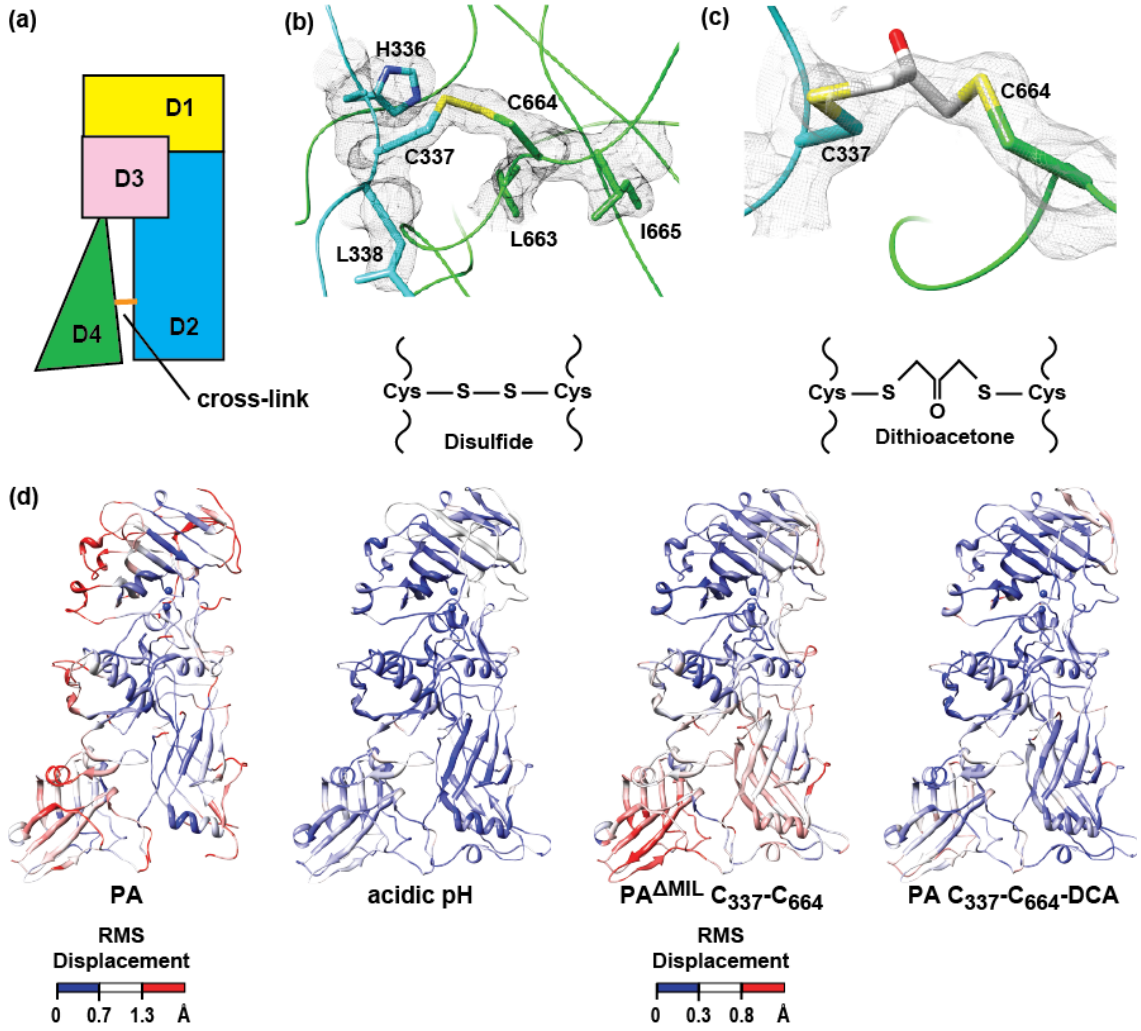

Figure 3.3. Cysteine cross-linking of PA D2 and D4. (a) Cartoon showing the location of the Cys-Cys cross-link introduced into PA, where domains are colored as in Fig. 1a. Two different cross-links were introduced via either a disulfide tether or the bifunctional cross-linking reagent, 1,3-dichloroactetone. Simulated-annealing $2F_o$-F_c electron density map calculated at the end of refinement for **(b)** $PA^{\Delta MIL}$ C_{337}-C_{664} and **(c)** PA C_{337}-C_{664}-DCA, contoured at σ=2.0 and σ=1.5, respectively, for which Cys337 and Cys664 and the bridging acetone cross-linker have been omitted. Sulfur and oxygen atoms are colored yellow and red, respectively. (d) Global Cα alignment of $PA^{\Delta MIL}$ to (left to right) PA from PDB accession code 1ACC, $PA^{\Delta MIL}$ acidic pH, $PA^{\Delta MIL}$ C_{337}-C_{664} and (right) PA C_{337}-C_{664}–DCA and the corresponding RMS deviations colored from low (blue) to high (red).

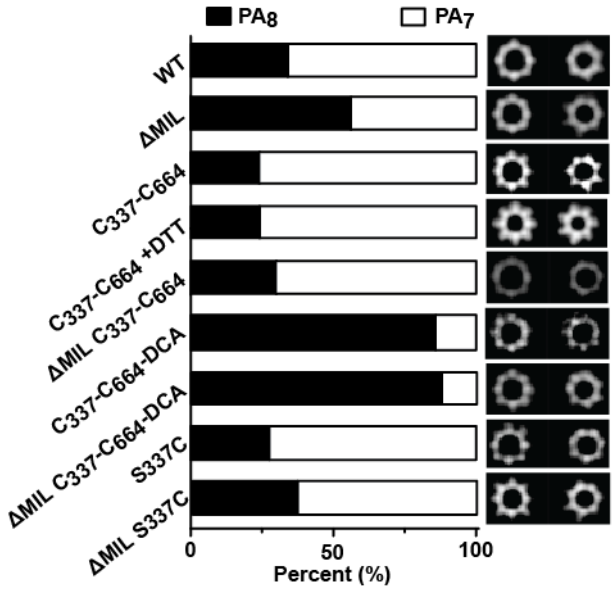

Figure 3.4. The interface of PA D2 and D4 controls oligomeric stoichiometry. The percentages of PA_7 (white) and PA_8 (black) complexes are given following the assembly of various PA constructs in the presence of LF_N. Representative class-average images for octamers (left) and heptamers (right) are shown at the right. The number (*N*) of particles analyzed are 2218, 3241, 2534, 2147, 2054, 1972, 1535, 1750 and 4845 for PA WT, $PA^{\Delta MIL}$, PA C_{337}-C_{664}, PA C_{337}-C_{664}+DTT, $PA^{\Delta MIL}$ C_{337}-C_{664}, PA C_{337}-C_{664}-DCA, $PA^{\Delta MIL}$ C_{337}-C_{664}-DCA, PA S337C and $PA^{\Delta MIL}$ S337C, respectively. Note: "+DTT" indicates that 10 mM of DTT was included in the oligomerization reaction.

analyses, we determined the proportion of heptameric and octameric oligomers in each sample (Kintzer *et al.* 2009; Feld *et al.* 2010; Kintzer *et al.* 2010; Kintzer *et al.* 2010b).

Our data show that two separate processes control oligomeric preference. First, we find that when we co-assemble $PA^{\Delta MIL}$ with LF_N, the resulting PA oligomer products are enriched with ~55% octamers (Figure 3.4). Under identical solution conditions, when we co-assemble WT PA with LF_N we find only ~35% of the resulting complexes are octameric (Kintzer *et al.* 2009). By further comparison, when WT and $PA^{\Delta MIL}$ assemble on their own, each produces ~2% and 24% octamer, respectively (Kintzer *et al.* 2009). Taken together these results imply two different pathways lead to octamer formation, and those mechanisms are additive. Second, the length of the cross-link between D4 and D2 changes the propensity of PA to form octameric complexes (Figure 3.4). The shorter C_{337}-C_{664} disulfide cross-link produces ~24% and ~30% octamer in the PA and $PA^{\Delta MIL}$ backgrounds, respectively, which is slightly less than non-cross-linked PA or $PA^{\Delta MIL}$ when co-assembled with LF_N (Figure 3.4). Interestingly, extending the length of the cross-link by 3 carbon atoms in the C_{337}-C_{664}-DCA construct resulted in complexes that were highly enriched with octamers (>85%) for both the WT PA and $PA^{\Delta MIL}$ backgrounds (Figure 3.4). Control constructs in which a single Cys residue was introduced showed no preference for the formation of the octameric complex. We surmise that cross-linking D4 to D2 alters the degree of conformational restriction in D4. Therefore, the interface between D2 and D4 plays an important role in determining the oligomeric preference of PA.

Correlation of D4 conformation and oligomeric heterogeneity. Since D4 appears to have the largest global RMS deviation when comparing the various PA monomer structures (Figure 3.3d), we hypothesize that its position may provide the structural basis for determining PA oligomeric stoichiometry. To correlate D4 conformations with oligomeric stoichiometry, we aligned the Cα atoms of D1-D3 residues for PA (derived from 1ACC), acidic $PA^{\Delta MIL}$, $PA^{\Delta MIL}$ C_{337}-C_{664}, and $PA^{\Delta MIL}$ C_{337}-C_{664}-DCA to the corresponding residues from $PA^{\Delta MIL}$. The center-of-mass (COM) for each construct's D2 and D4 was determined, accounting for differences in sequence. We then defined the hinge angle (θ) for D4 movements. θ is defined by the center of the oligomer's lumen (COL), the COM of D2, and the COM of D4. These centers were projected to a plane orthogonal to the central pore axis through the COL, and the COM of D2 is the angle vertex (Figure 3.5a). We find that θ increases relative to $PA^{\Delta MIL}$ for the WT PA and $PA^{\Delta MIL}$ C_{337}-C_{664}, which gave rise to a lower fraction of octamers, increasing by 2.5 degrees and 0.8 degrees, respectively. In contrast, θ for $PA^{\Delta MIL}$ at pH 6.5 and $PA^{\Delta MIL}$ C_{337}-C_{664}-DCA, the constructs that form an excess of octamer, increases by only 0.15 and 0.3 degrees, respectively (Figure 3.5b). Furthermore, the distance displaced (Δd) between the COM of D4, relative to $PA^{\Delta MIL}$, changes by 1.4 Å and 0.5 Å for WT PA and $PA^{\Delta MIL}$ C_{337}-C_{664}, respectively, but it changes by only 0.2 Å for both $PA^{\Delta MIL}$ at pH 6.5 and PA C_{337}-C_{664}-DCA (Figure 3.5b). Therefore, we conclude that there are two D4 conformations, termed Pro-PA_7 and Pro-PA_8, which differ in the linearity of D4 with respect to D2 and the oligomer lumen. These different conformations can give rise to lower and higher fractions of PA_8, respectively (Figure 3.5c).

The furin-cleavage site. The loop containing the furin-cleavage site is ordered in our $PA^{\Delta MIL}$ structure (Figure 3.6a). The structure of the loop begins as a β-hairpin with the preceding loop, as R164 hydrogen bonds back to S161. Notably, none of the residues make contact with residues in the PA_{63} domains, although the loop is in relatively close in proximity to the D2 loop containing F464. This result is expected given the fact that cleavage at the site functions to release PA_{20} from the rest of PA; and previous studies reported that the presence of the furin-site loop does not affect the dissociation kinetics of the PA_{20} fragment (Christensen *et al.* 2005).

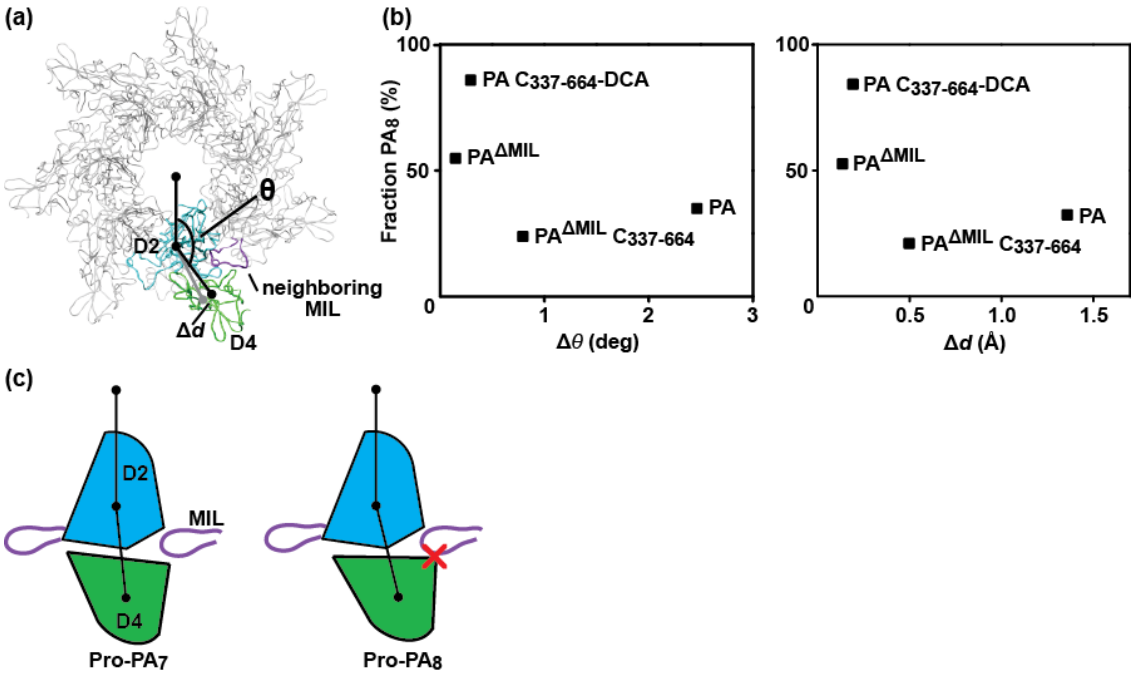

Figure 3.5. Molecular basis for PA oligomeric stoichiometry. (a) A model showing the angle, θ, which is used as a metric relating the observed movement of D4 relative to the amino-terminal domains of the PA_{63} moiety. θ is defined by the center of the oligomeric lumen (COL) and the centers of mass (COM) of D2 and D4. Also indicated in the model is the metric for the displacement of the D4 COM, called Δ*d*. The PA_7 coordinates, containing a model of the MIL, are from 1TZO (Lacy *et al.* 2004). D2, D4, and the MIL of the adjacent PA subunit are colored cyan, green, and purple, respectively. **(b)** Deviations in θ (Δθ) (left) and Δ*d* (right) are plotted against the percentage of PA_8 oligomer produced (as shown in Figure 3.4). The deviations are computed relative to the reference model from $PA^{\Delta MIL}$: $\Delta\theta = \theta(PA^{\Delta MIL}) - \theta(PA_i)$, and $\Delta d = d(PA^{\Delta MIL}) - d(PA_i)$, where PA_i is the PA monomer measured. **(c)** A molecular mechanism for PA oligomerization, where the hinge-like movement of D4 can lead to the population of two different population of PA that favor the formation of PA_7 and PA_8, called Pro-PA_7 (left) and Pro-PA_8 (right), respectively. In Pro-PA_8 D4 is rotated and translated relative to D4 for Pro-PA_7. The Pro-PA_8 conformation is sterically inhibited by the presence of the adjacent PA subunit's MIL during assembly. Presumably, the intramolecular dithiolacetone bridge linking D4 to D2 favors the Pro-PA_8 conformation in both the presence and absence of a neighboring MIL.

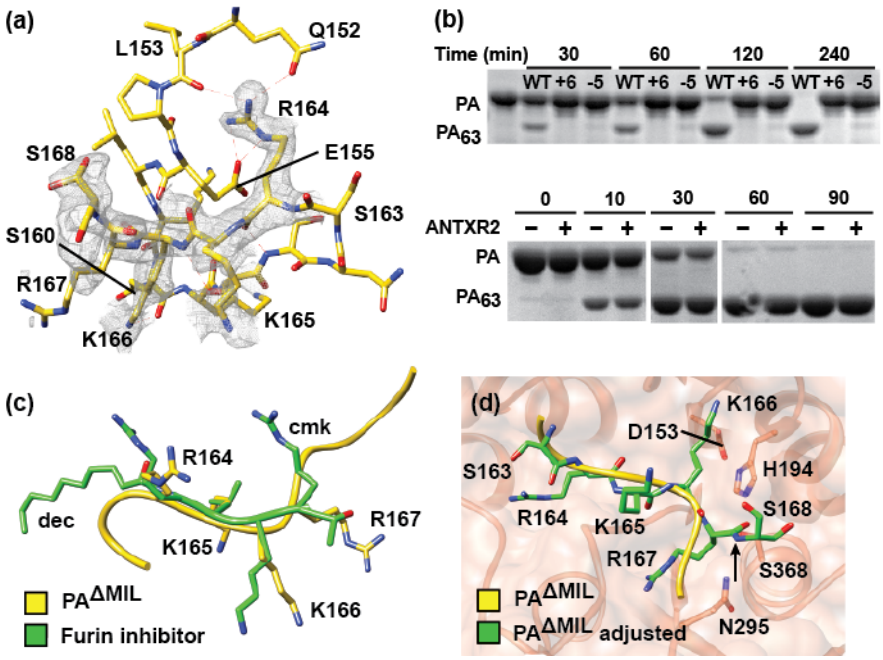

Figure 3.6. Structural basis for furin-dependent cleavage of PA. (a) Positioning of the furin-cleavage site loop relative to the rest of PA_{20}. Electrostatic interactions are represented as dashed red lines; carbon (yellow), oxygen (red), and nitrogen (blue) atoms are indicated. Sample composite simulated-annealing electron density (gray mesh) calculated at 1.45 Å at the end of refinement for PA residues 164-168 is contoured at σ=1. **(b)** SDS-PAGE assay for furin processing of (top) PA_{83}, PA^{+6}, and PA^{-5} and (bottom) PA and ANTXR2-bound PA. Proteins and incubation times as well as molecular weights corresponding to PA and PA_{63} are indicated above and to the left, respectively. **(c)** LSQ all-atom alignment of PA residues 164-167 with the furin-inhibitor peptide from 1P8J (Henrich *et al.* 2003). **(d)** Placement of the aligned (yellow) PA furin-cleavage site in the active site of furin (1P8J) (Henrich *et al.* 2003). Rotation around the peptide bond linking residues K166 and R167 (green) relieves steric clashes and positions the scissile bond in proximity to the catalytic S368 of furin.

Several residues in the furin-cleavage loop make contact with PA_{20}. Most prominent is R164, whose guanidino group forms a salt bridge with E155, as well as hydrogen bonds with the ε-amide of Q152 and the backbone carbonyl of L153. The γ-hydroxyl of S160 forms a hydrogen bond with the amide nitrogen of K166, and the ε-amino group of K166 hydrogen bonds to the backbone of Q158. The side chain for R167, which is the P1 site residue for furin, is not well ordered; however, its backbone carbonyl forms a hydrogen bond with Q115. Finally, K117 enters a hydrogen-bond network with the carbonyl oxygens of T169 and P173 and loosely approaches one of the conformers for S168. We conclude that the furin-site loop is freely accessible to solvent, makes numerous contacts with PA_{20} residues, and yet forms few interactions with PA_{63}.

We then asked whether protoxin processing by furin is dependent on the positioning of the site along the extended loop region. Using WT PA, we produced constructs in which the RKKR sequence was moved either six residues toward the N-terminus of PA (PA^{+6}) or five residues toward to the C-terminus (PA^{-5}). Samples were incubated with furin at various time points and analyzed by SDS-PAGE. While furin-dependent processing of WT PA is evident after 30 minutes and is essentially complete after 3 hours, virtually no processing was observed for either PA^{+6} or PA^{-5} (Figure 3.6b). Therefore, we conclude that the positioning of the furin-cleavage site is important to proper protoxin processing, and all sites within the loop are not equally accessible to furin.

Furin-dependent PA processing is not ANTXR2-dependent. We were intrigued by the degree to which the furin-cleavage site was ordered in these crystal structures. Is ordering of the ANTXR2 binding site linked to the ordering of the furin-cleavage site in PA? We suspected that crystal packing contacts resulted in the stabilization of 3α1, which is a structure implicated in ANTXR2-binding. Also it is known that anthrax toxin assembly is coordinated on cell surfaces by means of cholesterol-rich microdomains (Abrami *et al.* 2003). Could receptor binding alter the dynamics of the furin-cleavage site to either stimulate or inhibit cleavage to allow PA to assemble within microdomains? We tested this hypothesis by monitoring furin-dependent processing of both PA and PA in complex with ANTXR2 using SDS-PAGE (Figure 3.6b). We used both monomeric and dimeric ANTXR2 (dANTXR2) constructs (Kintzer *et al.* 2009; Kintzer *et al.* 2010b) to also test whether clustering of PA monomers affected the rate of furin processing (data from dANTXR2 not shown). The disappearance of the full-length PA band, and corresponding appearance of the PA_{63} band occurs at similar rates, that is, either when PA is free of receptor or it is complexed with dimeric or monomer versions of ANTXR2. Therefore, we conclude that the binding of PA to ANTXR2 has no effect on the rate of furin processing.

3.4 Discussion

Elucidating the thermodynamic mechanisms governing the formation of specific macromolecular architectures in the cell has remained a significant barrier to understanding the assembly and function of molecular machinery. Macromolecular assembly is a ubiquitous cellular process central to normal physiology as well as the mechanisms of microbial pathogenesis. Viruses and bacterial virulence factors, including anthrax toxin, must properly assemble for function. Because each PA subunit contains four folded domains, a number of functions including host-receptor recognition, oligomerization, channel formation, enzyme subunit binding, and enzyme activity may be attributed to each individual domain and/or inter-domain cross-talk. The fact that anthrax toxin assembles heterogeneously is a testament to the inherent complexity and flexibility involving the interactions of these multidomain subunits

during assembly. Heterogenous assembly mechanisms can lead to new functions for the macromolecular machine, such as altered complex stability, dynamics and function. Furthermore, the complexity of the intra-domain and inter-domain interactions allow for allosteric regulators to bind and alter either the assembly or the function of the complex.

Due to the primacy of virus and virulence-factor assembly to pathogenesis, small-molecule inhibitors that disrupt or alter assembly may seem at first glance to be reasonable drug-development strategies; however, these types of therapeutics have lagged considerably relative to those that target metabolic pathways and traditional enzyme activities. The lack of assembly-targeted drugs is, in part, due to the difficulty of assembly assays compared to enzyme assays or binary-component binding reactions. But more likely, the interfacial surface areas that define the thermodynamic interactions between subunits in these molecular machines are too extensive to impair or inhibit with a small-molecule drug by means of direct interference. Thus, the most-likely path to developing small molecules that inhibit virulence factor assembly would likely involve the triggering of a cryptic and/or native site that allosterically controls assembly. The allosteric-modulation strategy enables the development of a more practically-sized molecule to modulate the assembly pathway. It is in this spirit that we investigate the molecular mechanism of anthrax toxin assembly.

Protoxin maturation by furin. Since many viruses and toxins are secreted as proproteins, they must be proteolytically processed or activated prior to assembly. This mechanism ensures that these virulence factors can localize properly and infect targeted cells. In our structure of $PA^{\Delta MIL}$, we modeled residues 162-173, which contain the site for furin-dependent cleavage and subsequent activation of PA (Figure 3.6a). Furin has been implicated as the protease responsible for the proprotein processing of a myriad of disease pathogens (Hallenberger *et al.* 1992; Molloy *et al.* 1992; Stieneke-Grober *et al.* 1992; Tsuneoka *et al.* 1993; Gordon & Leppla 1994; Garred *et al.* 1995; Gordon *et al.* 1997; Stadler *et al.* 1997; Abrami *et al.* 1998; Volchkov *et al.* 1998). The crystal structure of furin, in complex with an inhibitor peptide, has been solved; thus the structural basis for furin specificity has been reported (Henrich *et al.* 2003). Aligning the PA sequence containing the furin-cleavage motif (S_{163}-R-K-K-R) to the dec-R-K-V-R-cmk inhibitor (Figure 3.6c), where the amino and carboxy termini of the inhibitor peptide were modified with the decyl (dec) and carboxymethyl ketone (cmk) groups, respectively, shows the backbone atoms are in close agreement, with the exception of the P1 site, R167. In order to align well with the inhibitor, the peptide bond between K166 and R167 must be rotated. We find that residues remain well within the β-strand region of the Ramachandran plot upon rotation (Ramachandran & Sasisekharan 1968). Given that R167 is freely accessible to solvent and its movement is not restricted in our structure of $PA^{\Delta MIL}$, we expect that this movement is not unreasonable. Successful rotation of R167 positions the scissile bond site in proximity to the catalytic Ser and His residues of furin and allows the residues carboxy to 167 to exit the furin catalytic site (Figure 3.6d). Furthermore, we show that moving the furin-cleavage site either upstream or downstream results in a loss of protoxin processing in vitro (Figure 3.6b). We propose that steric clashes with PA_{20} residues accounts for the specific requirement of the furin-cleavage site positioning on $L_{158\text{-}173}$. The atomic model of PA presented here provides a starting point for understanding the structural basis for furin-dependent cleavage of PA. Furthermore, being a protease-sensitive site, the ordering of the furin loop in the $PA^{\Delta MIL}$ structure suggested that there was some level of cross talk between the MIL, D2, and D4. We tested this hypothesis and found that receptor binding across D2 and D4 does not alter the rate of furin processing. Furin processing and PA assembly on cell surfaces, which appears to be

coordinated with cholesterol-rich microdomains (Abrami *et al.* 2003), may be controlled through other means.

Furin is a critical housekeeping enzyme involved in proprotein and prohormone maturation. Protoxins have taken advantage of the ubiquitous convertase, using it to activate virulence factors and infect a plethora of cell types. While furin-specific inhibitors have emerged as attractive therapeutics for preventing protoxin activation, such treatments may also obstruct normal cell processes that rely on furin processing. A recent study found that the efficacy of furin targets transforming growth factor beta (TGFβ1) and matrix metalloprotein (MT1-MMP) were not affected by furin inhibition, and it was proposed that the inherent redundancy of cell-surface convertases might compensate for the loss in furin activity (Remacle *et al.* 2010). However, other furin targets have not been studied, and a broader picture of the side effects associated with inhibiting furin activity remains unclear. Here, we report for the first time the atomic structure of a furin-cleavage site, which may lead to the development of drugs that are specifically tailored to inhibiting the furin-protoxin interaction, rather than furin itself. Indeed, one such inhibitor that binds to the viral envelope protein gp160 in a region proximal to the furin-cleavage site has been implicated as a possible treatment for HIV-1 (Murray *et al.* 2010). In this manner, the inhibitor interacts exclusively with the target antigen, preventing furin processing; thus, potential side effects arising from furin inhibition activity would be minimized. This avenue may be worth pursuing with anthrax toxin, as furin-loop accessibility can affect the rate of processing (Figure 3.6b) and drug binding may lead to altered conformational dynamics.

Molecular determinants of PA's heterogeneous oligomerization mechanism. Another goal of this study is to obtain a molecular-level understanding of anthrax toxin assembly. How does the PA subunit assemble into a heterogeneous mixture of two different oligomeric architectures? From this effort we expect to draw broader conclusions about macromolecular assembly paradigms and to inspire novel therapies for anthrax disease. Previous studies indicate that PA subunits form mixtures of PA_7 and PA_8 complexes, yielding of a PA_8 content of ~5% to 30% (Kintzer *et al.* 2009; Kintzer *et al.* 2010; Kintzer *et al.* 2010b). The PA_8 complex in particular is interesting because it is more pH-stable and thermostable than the more abundant PA_7 complex (Kintzer *et al.* 2010). PA_8 complexes can survive in the plasma fraction of bovine blood with an ~30-minute half life, while the PA_7 complexes precipitate and inactivate in a few minutes (Kintzer *et al.* 2010). Our present model is that allosteric control of the level of PA_8 complexes during pathogenesis may alter the level of toxicity achieved by the toxin (Kintzer *et al.* 2009; Kintzer *et al.* 2010). An alternative, non-mutually-exclusive model is that the heterogeneity allows for more efficient assembly on cell surfaces, preventing situations where the chance occurrence of odd- or even-numbered subcomplex stoichiometries would impede proper assembly if a strict oligomeric architecture were prerequisite for assembly (Kintzer *et al.* 2009). In either case, it is likely that some level of control is present in the system, and this control determines the ultimate stoichiometry of the oligomeric mixture produced.

In previous work, we demonstrated that two separate phenomena produce relatively heterogeneous PA oligomer populations: (i) Addition of protein assembly co-factors that facilitate the formation of dimeric PA intermediates can drive PA_8 formation up to as high as ~30%; these proteins include the dANTXR2, LF, and EF. (ii) The $PA^{\Delta MIL}$ construct tends to form ~30% PA_8 complexes in the absence of dimerization-enhancing protein co-factors (Kintzer *et al.* 2009). Mechanism (i) appears to have a straightforward intermolecular explanation, where dimerization favors the formation of the less prevalent even-numbered PA_8. However, mechanism (ii), albeit an artificial one, is not clear. Does the presence of the MIL trigger an

allosteric site that prevents octamer formation? Or does the MIL simply block octamerization by a steric mechanism, where it would interfere with D4 from an adjacent PA in the oligomer? In this report, we explored the mechanism (ii), which appears allosteric in nature. These results describe for the first time that a complex network of interactions exist within the PA monomer, which can coordinate oligomeric architecture. Interestingly, when $PA^{\Delta MIL}$ is co-assembled with LF_N, PA_8 complexes are enriched >50% (Figure 3.4), or 24% more than when $PA^{\Delta MIL}$ is assembled in isolation (Kintzer *et al.* 2009). This result indicates that the mechanisms (i) and (ii) are in fact additive. Furthermore, we produced a PA construct whose D4 is loosely tethered to the main body of the protein via a dithioacetone cross-link. Independent of the presence of the MIL, this construct yields a highly enriched PA_8 stoichiometry—in excess of 85% (Figure 3.4). Therefore, we find that removing the MIL and loosely tethering D4 to D2 have similar effects on PA_8 enrichment, albeit the latter is a noticeably stronger effect. On the other hand, tightly cross-linking D4 and D2 with a disulfide bond appears to inhibit the formation of the PA_8 complex relative to WT (Figure 3.4). Thus the conformation of individual PA subunits influences the outcome of an assembly reaction.

Based upon our structural observations that D4 can occupy a Pro-PA_7 and Pro-PA_8 conformation, we propose that these two conformations predispose the PA subunit to form either respective oligomeric architecture. The structure of $PA^{\Delta MIL}$ and various cross-linked variants reveal that the major conformational change in these monomers surrounds the hinge-like movement of D4 (Figure 3.5b-c). D4 can adopt two different conformations, Pro-PA_7 and Pro-PA_8, which affect PA oligomer stoichiometry. Interestingly, the two conformations can be populated with different length cross-links tethering D4 to D2. The Pro-PA_7 conformation is favored with the shorter disulfide bond cross-link, while the Pro-PA_8 conformation is favored by the 3-atom extension inherent to the dithioacetone cross-link. We suspect that restricting the conformational space of the D2-D4 pitch angle with the dithioacetone cross-link essentially lowers the kinetic barrier for PA to adopt the Pro-PA_8 conformation. Furthermore, the propensity of a particular PA monomer to adopt either D4 conformation is not limited to the presence of the MIL, as the cross-link phenomenon is consistent with both WT PA and $PA^{\Delta MIL}$ backgrounds (Figure 3.4). Thus steric interference by the MIL may not be the only factor influencing the oligomeric architecture, since its presence can be overridden by the D2-D4 dithioacetone cross-link (Figure 3.4). The Pro-PA_8 conformation populated by the dithioacetone cross-link, therefore, also controls the PA oligomerization interface in a way that favors the PA_8 species over the PA_7 species. In conclusion, we propose that the flexibility and relative orientation of D4 holds an important key in deciding the outcome of PA's heterogeneous oligomerization pathway. We suspect that other heterogeneous oligomeric systems may also have similar mechanisms that control product stoichiometry. Future work should address how small-molecule, drug-like molecules can influence virulence factor assembly, stoichiometry, and function.

3.5 Materials and Methods

Protein expression and purification. An expression plasmid encoding the PA deletion mutant, $PA^{\Delta MIL}$, was produced by removing residues 305-324 while simultaneously introducing the mutations V303P and H304G (Kintzer *et al.* 2009). The resulting PA construct replaces the membrane insertion loop (residues 303-324), leaving in its place a type-II turn; residue numbering is consistent with 1ACC (Petosa *et al.* 1997). Site-directed mutations of PA and $PA^{\Delta MIL}$ were produced using the QuikChange procedure (Agilent Technologies, Santa Clara, CA). PA and subsequent PA mutants were over-expressed in the periplasm of *Escherichia coli*

BL21(DE3) and purified as PA monomers over Q-sepharose anion exchange (GE Biosciences), as described (Kintzer *et al.* 2009). PA constructs used for crystallography were further purified over S200 gel filtration chromatography (GE Biosciences) in buffer A (20 mM Tris-Cl, 150 mM NaCl pH 8), and concentrated to 25 mg/mL.

For PA S337C N664C and $PA^{\Delta MIL}$ S337C N664C mutants, monomers were split into two fractions prior to storage at -80°C. The first fraction was allowed to form a disulfide bond under oxidizing conditions while 10 mM dithiothriotol (DTT, Gold Biotechnology, St. Louis, MO) was added to the second fraction. DCA-modified constructs were produced following a prior method (Yin *et al.* 2000). Cys-containing PA proteins were immobilized on Q-sepharose anion exchange using N_2-purged buffers. ~5 mg/mL fractions of DTT-free S337C N664C protein were incubated with 1 mM 1,3-dichloroacetone (Sigma-Aldrich, St. Louis, MO) on ice for 30 min, and the subsequent labeled protein was purified by S200 gel filtration in buffer A.

LF_N (LF residues 1-263) was over-expressed from a pET15b construct (Lacy *et al.* 2002) in the cytoplasm of *E. coli* BL21(DE3) and purified as described (Kintzer *et al.* 2009). The hexahistidine tags were removed by incubation with 0.5 units bovine α-thrombin (Enzyme Research, South Bend, IN) per mg of LF_N for 30 minutes at room temperature in buffer A supplemented with 2.5 mM $CaCl_2$ and 1 M dextrose.

Dimeric ANTXR2, corresponding to the soluble domain residues 40-217, was over-expressed from a pGEX vector (GE Healthcare) as a glutathione-S-transferase fusion protein in the cytoplasm of *E. coli* and purified on a glutathione-sepharose affinity column (GE Healthcare), as described (Wigelsworth *et al.* 2004). Monomeric ANTXR2 was produced by transferring residues 40-217 from pGEX into pET15b using the NdeI and BamHI restriction sites, yielding a N-terminal His-tag construct. Site-directed mutagenesis yielded the mutant C175A. The protein was over-expressed in *E. coli* and purified by Ni^{2+}-affinity chromatography, as described (Santelli *et al.* 2004).

The PA^{+6} and PA^{-5} mutants were made using a three-step, gene-synthesis procedure, in a manner similar to a previous procedure (Feld *et al.* 2010). Overlapping oligonucleotides encoding the desired sequences were synthesized (Elim Biopharmaceuticals, Inc., Hayward, CA) and amplified by two rounds of polymerase chain reaction (PCR). In Round I, 20 nM of nested oligonucleotides with consistent annealing temperatures of ~55 °C were amplified in a standard PCR reaction. In Round II, 1 μL of the PCR product made in Round I was amplified with the two outermost PCR primers (1 μM each) to make the synthetic double-stranded DNA fragment. These synthetic DNA fragments were ligated via a 5' Hind III site and 3' Kpn I site into the PA reading frame from a pET22b vector containing an in-frame, silent Kpn I restriction site in PA at V175.

Protein crystallization. Initial crystallization trials were carried out by screening ~400 sparse-matrix (Jancarik & Kim 1991) conditions using a Mosquito nanoliter robot (TTP Labtech, Cambridge, MA) to set 200 nL drops by the hanging-drop vapor-diffusion method (McPherson 1976) in 96-well format. Crystal conditions were further investigated using manual 24-well format trays using a 1:1 ratio of 1 or 2 μL drops of protein and reservoir solution. For $PA^{\Delta MIL}$ pH 8.5 and PA C_{337}-C_{664}-DCA the reservoir solution was composed of 15-25% (w/v) polyethylene glycol monomethyl ether (PEG ME) with an average molecular weight of 2000 Da, 100 mM Tris-Cl, 200 mM trimethylamine *N*-oxide, pH 8.3-8.7. These proteins formed trapezoidal prism-shaped crystals overnight, maturing to dimensions 300-800 μm. Crystals were harvested in an antifreeze solution containing a 1:1 mixture of 50% (v/v) PEG ME with an average molecular weight of 550 Da, and reservoir solution supplemented with 150 mM NaCl, $PA^{\Delta MIL}$ pH 6.5 and

$PA^{\Delta MIL}$ C_{337}-C_{664} reservoir solution contained 29-32% (v/v) pentaerythritol ethoxylate (15/4 EO/OH), 50 mM bis-tris Cl, 50 mM ammonium sulfate, pH 6.5-6.7. These proteins formed coffin-shaped crystals within 24 hours, maturing to dimensions 200-400 μm, and the crystals were harvested directly out of the drop. All crystal forms were flash-frozen by rapidly plunging the crystal into liquid N_2.

X-ray diffraction data collection, solution and refinement. X-ray diffraction data were collected at a wavelength of 1.1159 Å (11111 eV) and a temperature of 100 K at the Lawrence Berkeley National Lab's Advanced Light Source, Beamline 8.3.1 using a Quantum 315r CCD area detector (ADSC, Poway, CA) (MacDowell *et al.* 2004). Crystals containing $PA^{\Delta MIL}$ pH 8.5 diffracted X-rays to 1.45 Å in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 72.36 Å, b = 93.39, c = 118.08, and angles, α, β, and γ, of 90°. Crystals containing PA C_{337}-C_{664}-DCA diffracted X-rays to 1.83 Å in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 71.29 Å, b = 93.69, c = 117.85, and angles, α, β, and γ, of 90°. For both samples, a second data set was collected from the same crystal with a shorter exposure time to obtain the overloaded, low angle reflections. Crystals containing $PA^{\Delta MIL}$ C_{337}-C_{664} diffracted X-rays to 2.09 Å in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 71.56 Å, b = 94.60, c = 119.80, and angles, α, β, and γ, of 90°. Crystals containing $PA^{\Delta MIL}$ pH 6.5 diffracted X-rays to 1.70 Å in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 71.40 Å, b = 93.93, c = 117.85, and angles, α, β, and γ, of 90°. The diffraction data were indexed, scaled, and merged in HKL2000 (Otwinowski & Minor 1997). Molecular replacement for $PA^{\Delta MIL}$ pH 8.5 was carried out in PHASER (Storoni *et al.* 2004), using a loop-stripped model of PA (1ACC) (Petosa *et al.* 1997) as the search model. Missing loops with significant ($\sigma > 2.5$) F_o-F_c difference density were built manually in COOT (Emsley & Cowtan 2004), and the electron density was corroborated with composite simulated-annealing omit maps generated in PHENIX (Adams *et al.*). The coordinates and ADP values for all atoms were refined individually and anisotropically, respectively, in PHENIX, followed by further rounds of model building and solvent picking in COOT. Some of the alternate side-chain conformations were identified using RINGER (Lang *et al.* 2010), and occupancy refinement was carried out in PHENIX. Structure solution and refinement for PA C_{337}-C_{664}-DCA and $PA^{\Delta MIL}$ pH 6.5 was carried out similarly, accept that $PA^{\Delta MIL}$ was used as the search model, and ADP values were refined isotropically. A similar strategy was also employed for $PA^{\Delta MIL}$ C_{337}-C_{664} except that atomic coordinates were refined both individually and with rigid body groups, and ADP values were refined isotropically. Two molecules of 2-methoxyethanol were identified in the density for $PA^{\Delta MIL}$ pH 8.5 and PA C_{337}-C_{664}-DCA, as well as an acetone molecule linking Cys337 and Cys664 for PA C_{337}-C_{664}-DCA. The oxygen atom on the ketone was not resolved; therefore, its occupancy was set to 0. The final model geometry for all structures was validated using MOLPROBITY (Davis *et al.* 2007) and PROCHECK (Laskowski *et al.* 1993). Least-squares quotient (LSQ) Cα alignment of protein backbones was carried out in COOT (Emsley & Cowtan 2004). RMS deviations of protein backbone were computed with COMPAR in the CCP4 program suite (Collaborative Computational Project 1994). All molecular graphics were generated using CHIMERA (Pettersen *et al.* 2004).

Furin proteolysis of PA. WT PA and PA mutants (1 mg/mL) were incubated with 10 units of furin (New England Biolabs, Ipswitch, MA) per mg of PA in buffer A supplemented with 1 mM $CaCl_2$ at room temperature, as described (Christensen *et al.*). For the ANTXR2-PA proteolysis experiments, monomeric ANTXR2 C175A or dimeric GST-linked ANTXR2 was incubated with PA in a 4:1 molar ratio in buffer A supplemented with 1 mM $CaCl_2$ and 1 mM

$MgCl_2$ for 10 minutes at room temperature. The co-complex was then processed with 20 units of furin per mg of PA. Samples were separated on a 10% SDS-PAGE gel and stained with Coomassie Brilliant Blue (Bio-RAD Laboratories, Hercules, CA), and the Pageruler protein ladder (Fermentas Life Sciences, Glen Burnie, MD) was used to determine the molecular weights of the products.

LF_N-driven assembly of $_n$PA constructs. PA constructs (1 mg/mL) were nicked with trypsin (Sigma-Aldrich) at a ratio of 1:1000 (w/w) for 15 minutes at room temperature and then treated with soybean tryspin inhibitor (Worthington Biochemical, Lakewood, NJ) at 1:100 (w/w). The resulting $_n$PA mixture was co-assembled with LF_N in a 1:1 molar ratio for 1 hour at room temperature. Samples were purified over S200 gel filtration in buffer A.

Electron microscopy. Grid preparation, data collection, and image processing were performed as described (Kintzer *et al.* 2009). Briefly, each PA oligomer was diluted to 100 nM (with respect to PA monomer concentration) in buffer A supplemented with 0.001% (w/v) *n*-dodecyl-β-D-maltopyranoside (Affymetrix, Maumee, OH). PA complexes containing Cys mutations were diluted in buffer containing 5 mM DTT to improve image quality. 400 mesh copper grids were successively covered by continuous carbon film via a formvar support. 4 µl of PA sample was applied to the grid for 1 minute, washed in 3 successive drops of water, and then stained with 2% uranyl acetate (Sigma-Aldrich). Negative-stain EM images were recorded with a Tecnai 12 (FEI Company, Hillsboro, OR) operated at 120 kV at 49,000× magnification. Images were taken using a CCD camera at a 2.13-Å/pixel specimen scale. Particle images were selected for each data set using automatic or manual particle picking using boxer in EMAN (Ludtke *et al.* 1999). Reference-free processing was computed using SPIDER (Frank *et al.* 1996). Images were subjected to three successive cycles of multi-reference alignment, multivariate statistical analysis, and classification (Stark *et al.* 1995; van Heel *et al.* 1996), such that the last classification was done using only the lowest order eigenvectors (White *et al.* 2004). Following the last classification, we separated the images by size and by the heptameric and octameric oligomerization states. Also we used a second method of image processing, whereby crystal-structure-reference images were made from two-dimensional projections of low resolution density maps generated from the crystal structures of the PA heptamer (Lacy *et al.* 2004) and octamer (Kintzer *et al.* 2009) in SPIDER (Frank *et al.* 1996). Crystal-structure-referenced images were aligned and classified using the lowest order eigenvectors. Final class-average images were manually inspected as being heptameric, octameric, or unclassifiable "junk" (<5% of the total particles). The number of particles per identifiable classification was used to determine the percentages of heptamers and octamers. Approximately 1700-5000 total particles were analyzed per sample. Either classification procedure produces similar results for the analysis of the composition of PA oligomer samples, agreeing with other biophysical measures of oligomeric heterogeneity (Kintzer *et al.* 2009; Feld *et al.* 2010; Kintzer *et al.* 2010; Kintzer *et al.* 2010b).

Accession numbers. Coordinates and structure factors for $PA^{\Delta MIL}$ pH 8.5, $PA^{\Delta MIL}$ pH 6.5, $PA^{\Delta MIL}$ C_{337}-C_{664}, and PA_{83} C_{337}-C_{664}-DCA have been deposited in the PDB with accession codes 3TEW, 3TEX, 3TEY, and 3TEZ, respectively.

Chapter 4
Concluding remarks

Imagine a protein channel that is 50-150 Å long; during translocation, a frame of 15-50 residues is contained in the channel. The possible chemical complexity (considering the 20 natural amino acids) is enormous, ranging from 15^{20} to 50^{20} depending on channel length and polypeptide conformation. Therefore, it is reasonable to assume that protein translocases are not designed to *specifically* recognize each and every type of possible amino acid sequence. This type of design principle would be impossible to encode into the machine. Instead, the translocase may recognize broad structural or chemical properties of peptides. This type of sequence recognition is often called "nonspecific binding," and it refers to the ability of a binding site to recognize substrates using general features of the polymer. Even in the case of anthrax toxin—which only has two known natural substrates, LF and EF—the PA translocase is highly nonspecific and capable of translocating heterologous sequences (Milne *et al.* 1995; Blanke *et al.* 1996; Wesche *et al.* 1998; Feld *et al.* 2010). This nonspecificity comes as no surprise, given that the frame of sequence within the machine is continuously changing during translocation. Here we consider two classes of mechanisms that the anthrax toxin protein transporter system exploits in addressing the issue of nonspecificity:

Class I Hone in on general sequence properties, namely patches of sequence dense in hydrophobicity (i.e. sequence hydropathy) and/or a particular type of charge.

Class II Bind nonspecifically to sequences with similar steric shape, without regard to sequence chemistry.

For the Class I mechanism, we hypothesize that the machine can effectively *average* diverse residue chemistries into mean sequence properties for a given segment of peptide, thereby minimizing chemical complexity. The likely rational behind the ϕ-clamp site is its preference for hydrophobic and aromatic groups with little or no specificity for the precise geometrical arrangements of such groups (Krantz *et al.* 2005). Crude estimates of polypeptide binding attribute ~1.5 kcal mol^{-1} of binding free energy to the ϕ clamp (Feld *et al.* 2010). Additionally, the cation selectivity of PA implies that an anion-repulsion site resides within the channel, which repels negatively charged sequences and attracts positively charged sequences. Acidic residues in the translocating chain are responsible for ΔpH-driven translocation; however, the specific location and identity of these charges are much less critical (Brown *et al.* 2011).
By contrast, the α-clamp site is structurally designed to complement α-helical shape, rather than particular residue chemistries. Only one α-clamp residue, R178, makes any significant van der Waals contacts with LF_N α1 residues; the rest of the interaction is dominated by backbone contacts. Additional backbone interactions independent of side chain chemistry include two parallel-β-sheet hydrogen bonds with LF_N β1. Thus the α clamp employs a Class II mechanism of highly nonspecific binding, providing ~2.5 kcal mol^{-1} of binding free energy (Feld *et al.* 2010). Furthermore, this binding energy was achieved independently of the sequence composition (Fig. 4.1). The α clamp is capable of binding a wide variety of sequences, including non-amphipathic ones. Therefore, the α clamp derives its binding affinity from shape complementarity and a pair of backbone hydrogen bonding interactions, rather than sequence-

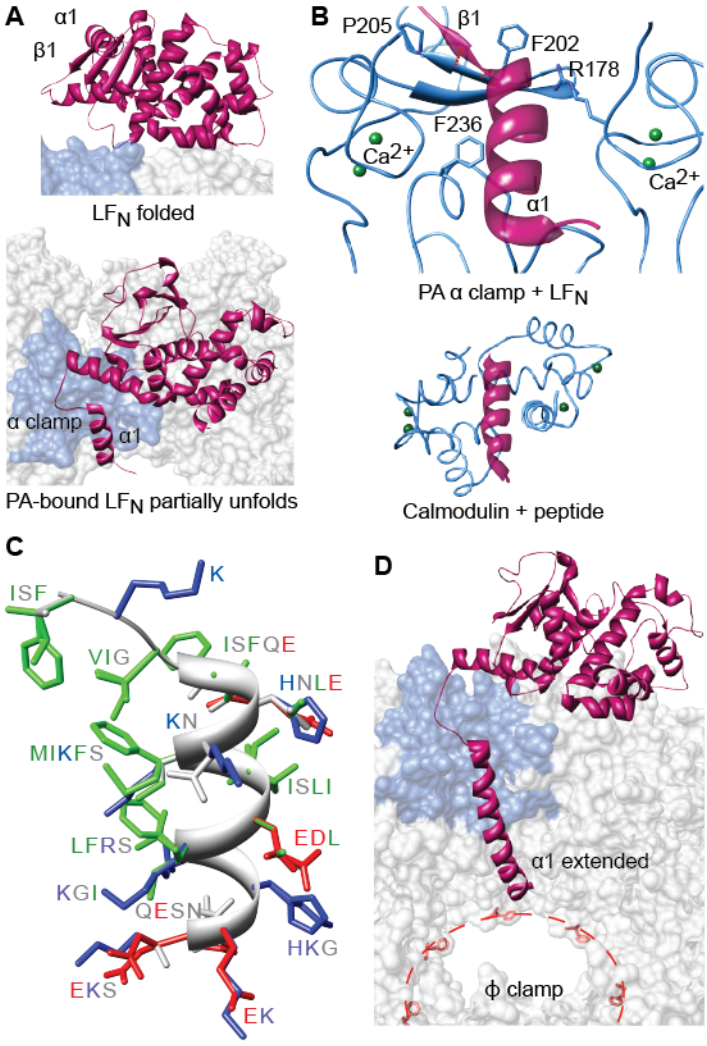

Figure 4.1. The α clamp. (A) LF_N (red-violet) changes conformations from (top) a folded (1J7N (Pannifer *et al.* 2001)) to (bottom) a partially unfolded (3KWV (Feld *et al.* 2010)) state, such that LF α1/β1 unfurl and bind into PA's α clamp (denim surface). **(B)** (top) A detailed view of the α clamp (denim ribbon) in complex with LF's α1 (red-violet ribbon) indicating the structural calcium ions (green) scaffolding the site. (bottom) Calmodulin in complex with a peptide helix (1CDM (Meador *et al.* 1993)), where the latter is colored analogously to the α-clamp structure. **(C)** The α clamp can bind nonspecifically to and translocate a variety of sequences illustrated in a helical structural alignment, which is colored by residue chemistry: basic (blue), acidic (red), polar (gray) and hydrophobic (green) (Feld *et al.* 2010). **(D)** Considering the Zimm-Bragg formalism (Zimm & Bragg 1959), the α clamp may act as an α-helix nucleation site. Elongated helical structure can then be fed into the ϕ clamp (red dotted line and F427 phenyl residues).

specific polar, charged, or hydrophobic contacts.

4.1 Configurational entropy, diffusion, and substrate orientation

Assume a 10-residue peptide is recognized during a cycle of translocation, and each residue can be in the α, β, or left-hand turn conformation. Such a peptide could effectively occupy ~10^5 possible conformations. Selecting specific subsets of possible conformations to bind and forcibly move during protein translocation would require work. For example, to select 10 conformations out of 10^5 would require ~4 kcal mol^{-1} of energy, independent of operating under a load. Adding chemical complexity to the system further increases the entropic penalty. Assuming the translocase requires a binary pattern of hydrophobic and hydrophilic sites within the polypeptide sequence, and as there are 10^3 possible binary patterns in a decamer, selecting 10 of these configurations would increase the required amount of energy to 7-8 kcal mol^{-1}.

A translocase thus has two mechanisms available to offset the entropic costs of binding the decamer segment: (i) making specific enthalpic interactions and/or (ii) through the dissociation of ordered solvent. The former is a mechanism of specificity and is unlikely. Specific interactions are weak, as they are offset by losses of peptide-solvent hydrogen bonds or ionic interactions. For the latter, many (Kauzmann 1959; Tanford 1968; Dill 1990) have considered that the hydrophobic effect is dominant in biomolecular interactions and is due to changes in solvation of the interacting groups. The effect is related to SASA displaced upon binding. For PA, the α clamp provides 2.5 to 4 kcal mol^{-1} of binding free energy through nonspecific binding (Feld *et al.* 2010). Given the SASA apparent in the structure is 1000 $Å^2$, we can estimate that the free energy gain per unit surface area is 2 to 4 cal mol^{-1} $Å^{-2}$. This value is 5 to 10 fold less than the 20 cal mol^{-1} $Å^{-2}$ calculated from a 1023-protein metaanalysis (Zhou & Zhou 2002). The disparity from the observed value for PA's α clamp may be due to changes in backbone configurational entropy, lower levels of ordered solvent released upon binding, or the introduction of strain and disorder elsewhere into the system. In any case, we may presume that nonspecific binding interactions stabilized through a general desolvation strategy can be used to do work, namely by ordering the translocating chain for further processing and/or by reducing the stability of the substrate protein. In order to achieve these aims, ~1000-2000 $Å^2$ of SASA was excluded upon binding. We contend that such interactions in a translocase would benefit from a uniform substructure, such as an α helix.

Even if we assume that the thermodynamic penalty is satisfied by some other means, the kinetic search process may limit translocation. Experimentally, protein substrates with ~250 to 750 residues can translocate in ~1 to 10 s (Huang *et al.* 1999; Burton *et al.* 2001; Kenniston *et al.* 2003; Kenniston *et al.* 2004; Krantz *et al.* 2006), and thus an interesting question is whether the translocase facilitates this search process. To speed up the search, the translocase may limit the conformational space of the translocating chain to a particular structure most suitable for its active site to bind. Assume the translocase engages more or less a helical structure to produce maximum force during translocation. A Class II-type structure may be able to facilitate the formation of helical structure during translocation. According to the Zimm-Bragg model for helix formation, α-helix stability, K_N, of an N-residue peptide is given as $K_N = \sigma s^N$ (Zimm & Bragg 1959). The initiation of α-helical structure, which is described by the factor, σ, is the limiting step. In turn, σ is dependent on the rotational degrees of freedom of the peptide chain. The equilibrium constant, K_N, increases with each additional residue added to the nascent helix. Thus helix formation is a nucleation process, where the equilibrium stability of a helical substrate would be greatly enhanced by the presence of a Class II structure that limits

conformational degrees of freedom, in the spirit of PA's α clamp (Fig 4.1). Helical structure has an additional advantage of possessing reduced conformational entropy relative to unstructured peptide, minimizing the potential for dissipative losses and allowing for a more efficient utilization of the available electrical/chemical free energy source. We conclude that helix-nucleation machinery in translocases (if oriented properly along the axis of the translocase) would provide a significant kinetic benefit to the translocation mechanism.

4.2 A proton-engine mechanism for translocation

Protons are dissipated by the PA translocase during translocation by means of the protonation and deprotonation of acidic residues in the substrate chain (Krantz *et al.* 2006; Brown *et al.* 2011). Similar to the Cl^-/H^+ antiporter, which exchanges two Cl^- ions for one H^+ ion (Ashcroft *et al.* 2009), protons are consumed by the PA translocase as a fuel to drive protein flux. How are directionality and force derived from the dissipation of H^+ ions down their gradient? In prior models, Brownian motion in the translocating protein is effectively biased by changes in protonation state of acidic residues in the protein substrate. We propose that sites like the ϕ clamp may *switch* conformation during the substrate protonation/deprotonation cycle and that this action plays a key role in governing the process of translocation. Given that a salt bridge likely positions the ϕ-clamp loop in an active conformation (Melnyk & Collier 2006), protonation could switch the interaction between on and off states. Furthermore, the loops containing the salt bridge residues (K397 and D426) as well as the active site ϕ clamp are confirmationally flexible, as evidenced in the published structures of the PA oligomer (Lacy *et al.* 2004; Kintzer *et al.* 2009; Feld *et al.* 2010) (Figure 4.2). While these structures are of the PA prechannel, it cannot be ruled out that the ϕ clamp in the channel state lacks this flexibility. Moreover, ϕ clamp flexibility is not a prerequisite for conformational switching, as the substrate may also exhibit similar switches in its protonation state. Regardless, we envision that the necessary force is generated in this mechanism through the biasing of Brownian motion via electrostatic repulsion, and the subsequent release of unfolded-state conformational entropy, ΔS_{config}, upon substrate extrusion from the channel.

In this model, the α-clamp site not only facilitates protein unfolding but also stabilizes and templates the formation of α-helical structure inside the PA channel (Fig. 4.1d). Furthermore, the internal diameter observed and predicted for 14- to 16-stranded β barrels is a suitable steric fit for peptide helices and likely favorable to helix initiation. For example, crystal structures of protein autotransporters are comprised of a β barrel, which contain α helices from the translocating chain within the lumen of the barrel (Oomen *et al.* 2004; Meng 2006). Why may this be important? We propose that ΔS_{config} will be much greater for a helix-to-coil transition. Thus during particular cycles of the transport pathway, the channel may maximize the amount of compact structures that form, including helical conformations. The activity of the α clamp supports this hypothesis (Feld *et al.* 2010), as well as recent work measuring the minimal length of polypeptide spanning the PA channel (Basilio *et al.* 2011). In the latter study, it was suggested using a streptavidin-biotin intermediate-capture approach that peptide chain spans the PA channel in an entirely non-helical, fully-extended conformation. Thus, the authors conclude that the substrate translocates in a fully-extended conformation devoid of helical structure. The streptavidin-capture data show that the rate of capture for the shortest, 33-residue probe is on the order of 1000 s, which is extremely slow, considering the 260-residue substrate, LF_N, translocates in ~10 s. The kinetics, however, suggest the opposite conclusion that the substrate is in a more compact structure >99% of the time, where the extended state is rare (i.e., < 1%). The

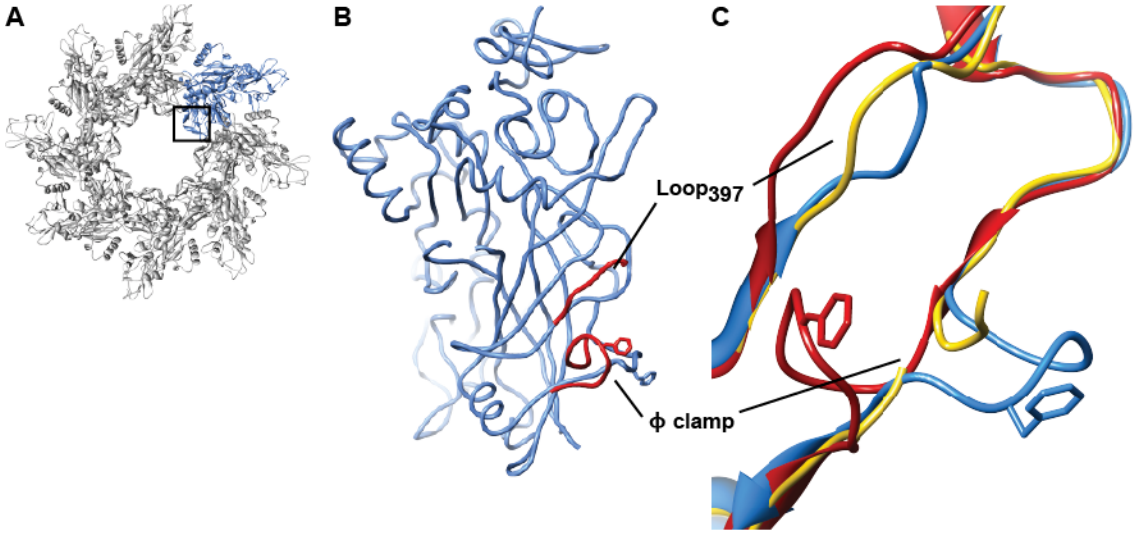

Figure 4.2. Loop flexibility in PA oligomers. (A) Ribbon depiction of the PA_8 oligomer colored grey with a subunit of interest colored blue. The black box indicates the location of the loops. **(B)** Backbone rendering of a PA subunit. Chain A from the LF_N-bound PA_8 3KWV (Feld *et al.* 2010) is colored blue and chain A from unbound PA_8 3HVD (Kintzer *et al.* 2009) is colored red. The loops containing K397 ($Loop_{397}$) and F427 (ϕ clamp) are indicated. **(C)** Zoom-in view of Loop397 and ϕ clamp colored as in **(B)** with the addition of chain A from PA_7 1TZO (Lacy *et al.* 2004). Residues 427-428 were not modeled in 1TZO.

authors further argue that longer-length substrates are less accessible to streptavidin capture because the chain will spend more time retracted in the channel. This explanation is in fact consistent with the chain experiencing a more compacted state. The compacted/coiled state is likely helical based on the Zimm-Bragg formalism (Zimm & Bragg 1959), the substrate desolvation activity of the upstream ϕ-clamp site (Krantz *et al.* 2005), the structure of the LF_N-liganded α clamp (Feld *et al.* 2010), and the known Ramachandran preferences (Ramachandran & Sasisekharan 1968) for polypeptides (i.e., anything less than an extended β conformation likely resides in the α-helical Ramachandran well, as the left-hand turn conformation is rare).

We suggest the translocating chain can fluctuate between more extended and more condensed states. The only requirement of our model is that during translocation the structure of polypeptide in the channel must be more compact than that immediately following extrusion from the channel; the absolute degree of helicity is not critical, as the peptide will be confined within the channel, regardless. When the polypeptide transitions from a more constrained helical state to a less constrained state, the realized increase in ΔS_{config} is significant. We propose a proton-engine mechanism, whereby the ΔpH, ϕ clamp, α clamp, and anionic-charge repulsion site in the PA channel coordinate protein translocation (Fig. 4.3):

Step 1. Anionic/deprotonated polypeptide loads into the upper vestibule of the PA channel by means of Brownian motion, binding into the α-clamp site as helix. The α clamp acts as a helix nucleating binding site, orienting and feeding the growing helical chain towards the channel lumen.

Step 2. The ϕ clamp switches to a *closed* position, tightly gripping the substrate polypeptide. Binding of unfolded chain at the ϕ clamp prevents backsliding (Krantz *et al.* 2005). The ϕ clamp impedes H^+ flow, reducing $[H^+]$ below the ϕ clamp. Acidic substrate residues above the clamp are subsequently protonated, and consequently less anionic in charge. The stability of the interaction at the ϕ clamp site may be further strengthened by the reduction in negative charge of the substrate.

Step 3. While the substrate polypeptide is tightly engaged at the ϕ clamp, the amino-terminal end of the peptide eventually begins to extrude from the end of the channel, where conformational space of the chain is less restricted. Translocation is thus thermodynamically favorable in the direction out of the channel due to the gain in $T\Delta S_{config}$. Importantly, the peptide can bypass the anionic charge repulsion site in the channel, since the peptide is now protonated at its Asp and Glu sites.

Step 4. The ΔpH at the ϕ clamp weakens as the substrate in the β barrel loses structure, solvent penetrates up the barrel, and H^+ dissipate out of the channel. The ϕ clamp switches to the *open* state, releasing bound peptide. Due to the charge selectivity of the channel, the peptide may only proceed through the channel until peptide and channel are no longer electrostatically compatible. At this point, translocation pauses until peptide recompresses into helix, the ϕ-clamp site can reset, and the deprotonated section of chain is protonated. The cycle will repeat in a fashion analogous to the thermodynamic cycles of a heat engine.

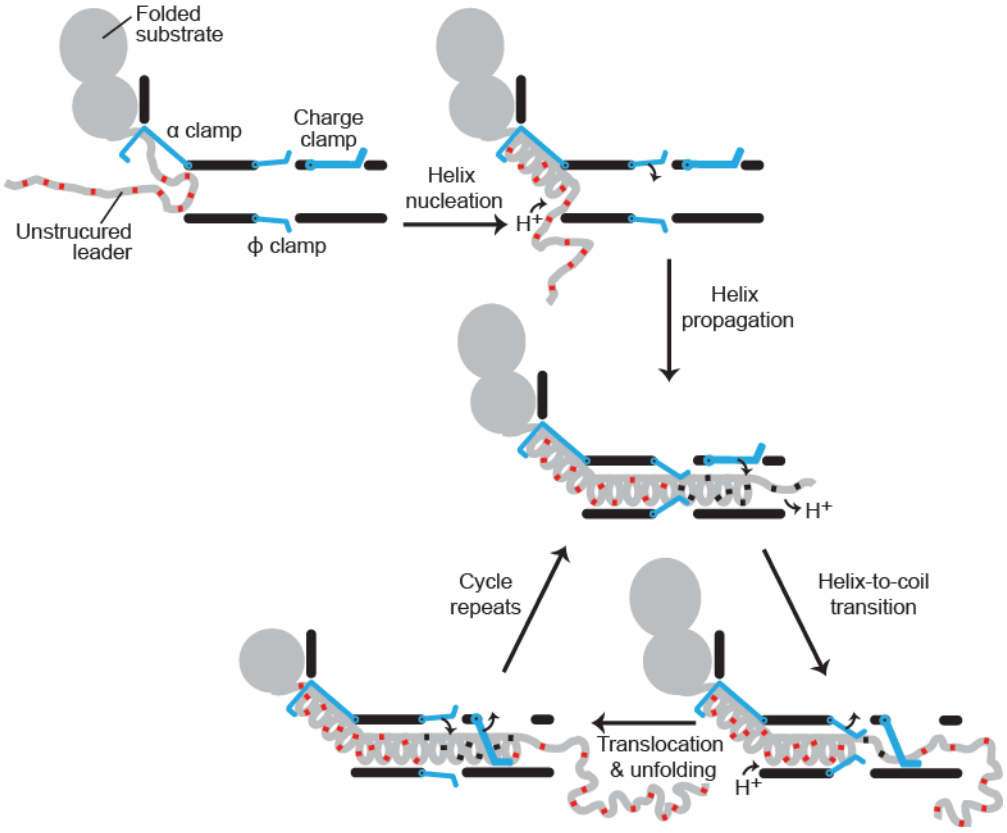

Figure 4.3. Translocation by a proton-driven engine. Folded substrate and unstructured leader sequence (gray) bind and dock to the PA channel. The α clamp nucleates and subsequently propagates helix formation. The ϕ clamp engages the compact translocating chain; deprotonated acidic residues (red) are then protonated (black). An increase in $T\Delta S_{config}$ outside the channel favors the transition of helix to unstructured random coil; acidic residues deprotonate in the higher pH of the cytosol. The charge clamp engages, permitting the passage of protonated acidic residues while preventing the retrotranslocation of deprotonated ones. Ungating of the ϕ clamp allows the chain to translocate, while the α clamp continues to stabilize and template unfolded polypeptide into helix. The cycle repeats until the substrate is fully unfolded and translocated.

4.3 An entropic force generation mechanism

How much force may be generated during this proton engine's power stroke? In this mechanism, the power stroke is defined by the distance over which potential energy, stored in the form of compressed and ordered substrate polypeptide conformations, is released into relaxed and disordered conformations. Assuming an α-helical compressed state, which has three-fold fewer degrees of freedom per residue than the fully random coil state, the mechanochemistry of this process is: $F = T\Delta S_{\text{config}} / d$, where d is the difference in length of peptide undergoing a α-helix-to-coil transition (2.2 Å per residue). F works out to ~20 pN—enough force to greatly accelerate protein unfolding reactions (Brockwell *et al.* 2003; Cecconi *et al.* 2005; Borgia *et al.* 2008; Crampton & Brockwell 2010). This calculation is an upper-limit. These forces are significant and higher than those estimated for BR models, where the substrate is a simple Brownian particle with only three translational degrees of freedom. The step size of the anthrax toxin translocase has yet to be reported; however, the ~10-nm-long β barrel likely places an order of magnitude limit on the power stroke length. Interestingly, the productive force generated over the range of the step as helix melts into random coil would not decrease over distance translocated. Conversely, the electrostatic force arising from the developing charge repulsion between the channel and peptide would decrease inversely as the square of the distance translocated. Here, we do not assume that the power stroke is largely dependent on the electrostatic repulsion between the substrate chain and the channel. Such an electrostatic force should not be ignored, although it would tend to occur over shorter distances.

The electrostatic repulsion may be considered the trigger that switches the translocase-substrate interaction between high-affinity and low-affinity modes, preventing unproductive motions. This electrostatic switch is akin to Maxwell's demon-operated trapdoor or any directionally biased ratchet. Conformational switching is generally what is observed for many ATPases and GTPases, and we expect switching to be a possible means both proton and ATP power sources are utilized. The available free energy dissipated by the pH gradient itself is more than adequate to supply the switch with the necessary power to drive the system and rectify the expansion direction of a more compacted α-helical peptide conformation to a more disordered state.

References

Abrami, L., Fivaz, M., Decroly, E., Seidah, N. G., Jean, F., Thomas, G., Leppla, S. H., Buckley, J. T. & van der Goot, F. G., (1998) The pore-forming toxin proaerolysin is activated by furin. *J. Biol. Chem.* **273**: 32656-61.

Abrami, L., Liu, S., Cosson, P., Leppla, S. H. & van der Goot, F. G., (2003) Anthrax toxin triggers endocytosis of its receptor via a lipid raft-mediated clathrin-dependent process. *J. Cell Biol.* **160**: 321-8.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H., (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**: 213-21.

Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. & Terwilliger, T. C., (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Rad.* **11**: 53-5.

Agrawal, A. & Pulendran, B., (2004) Anthrax lethal toxin: a weapon of multisystem destruction. *Cell Mol. Life Sci.* **61**: 2859-65.

Arora, N. & Leppla, S. H., (1993) Residues 1-254 of anthrax toxin lethal factor are sufficient to cause cellular uptake of fused polypeptides. *J. Biol. Chem.* **268**: 3334-41.

Arora, N. & Leppla, S. H., (1994) Fusions of anthrax toxin lethal factor with shiga toxin and diphtheria toxin enzymatic domains are toxic to mammalian cells. *Infect. Immun.* **62**: 4955-61.

Ashcroft, F., Gadsby, D. & Miller, C., (2009) Introduction. The blurred boundary between channels and transporters. *Philos Trans R Soc Lond B Biol Sci* **364**: 145-7.

Astumian, R. D., (1997) Thermodynamics and kinetics of a Brownian motor. *Science* **276**: 917-22.

Bann, J. G., (2012) Anthrax toxin protective antigen-Insights into molecular switching from prepore to pore. *Protein Sci* **21**: 1-12.

Barth, H., Aktories, K., Popoff, M. R. & Stiles, B. G., (2004) Binary bacterial toxins: biochemistry, biology, and applications of common Clostridium and Bacillus proteins. *Microbiol Mol Biol Rev* **68**: 373-402.

Basilio, D., Jennings-Antipov, L. D., Jakes, K. S. & Finkelstein, A., (2011) Trapping a translocating protein within the anthrax toxin channel: implications for the secondary structure of permeating proteins. *J Gen Physiol* **137**: 343-56.

Basilio, D., Juris, S. J., Collier, R. J. & Finkelstein, A., (2009) Evidence for a proton-protein symport mechanism in the anthrax toxin channel. *J Gen Physiol* **133**: 307-14.

Basilio, D., Kienker, P. K., Briggs, S. W. & Finkelstein, A., (2011) A kinetic analysis of protein transport through the anthrax toxin channel. *J Gen Physiol* **137**: 521-31.

Beall, F. A., Taylor, M. J. & Thorne, C. B., (1962) Rapid lethal effects of a third factor of anthrax toxin. *J. Bacteriol.* **83**: 1274-80.

Benson, E. L., Huynh, P. D., Finkelstein, A. & Collier, R. J., (1998) Identification of residues lining the anthrax protective antigen channel. *Biochemistry* **37**: 3941-8.

Bier, M., (2003) Processive motor protein as an overdamped brownian stepper. *Phys Rev Lett* **91**: 148104.

Blanke, S. R., Milne, J. C., Benson, E. L. & Collier, R. J., (1996) Fused polycationic peptide mediates delivery of diphtheria toxin A chain to the cytosol in the presence of anthrax protective antigen. *Proc. Natl Acad. Sci. U.S.A.* **93**: 8437-42.

Blaustein, R. O. & Finkelstein, A., (1990) Voltage-dependent block of anthrax toxin channels in planar phospholipid bilayer membranes by symmetric tetraalkylammonium ions. Effects on macroscopic conductance. *J Gen Physiol* **96**: 905-19.

Blaustein, R. O., Koehler, T. M., Collier, R. J. & Finkelstein, A., (1989) Anthrax toxin: channel-forming activity of protective antigen in planar phospholipid bilayers. *Proc. Natl Acad. Sci. U.S.A.* **86**: 2209-13.

Borgia, A., Williams, P. M. & Clarke, J., (2008) Single-molecule studies of protein folding. *Annu Rev Biochem* **77**: 101-25.

Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A., (2001) Identification of the cellular receptor for anthrax toxin. *Nature* **414**: 225-9.

Brockwell, D. J., Paci, E., Zinober, R. C., Beddard, G. S., Olmsted, P. D., Smith, D. A., Perham, R. N. & Radford, S. E., (2003) Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat Struct Biol* **10**: 731-7.

Brown, M. J., Thoren, K. L. & Krantz, B. A., (2011) Charge requirements for proton gradient-driven translocation of anthrax toxin. *J. Biol. Chem.* **286**: 23189-99.

Burton, R. E., Siddiqui, S. M., Kim, Y. I., Baker, T. A. & Sauer, R. T., (2001) Effects of protein stability and structure on substrate processing by the ClpXP unfolding and degradation machine. *Embo J* **20**: 3092-100.

Carnot, S. (1824). Reflexions sur la puissance motrice du feu et sur les machines propres a developper cette puissance. Paris, Bachelier.

Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S., (2005) Direct observation of the three-state folding of a single protein molecule. *Science* **309**: 2057-60.

Chauhan, V. & Bhatnagar, R., (2002) Identification of amino acid residues of anthrax protective antigen involved in binding with lethal factor. *Infect. Immun.* **70**: 4477-84.

Cheng, Y., (2009) Toward an atomic model of the 26S proteasome. *Curr. Opin. Struct. Biol.* **19**: 203-8.

Christensen, K. A., Krantz, B. A. & Collier, R. J., (2006) Assembly and disassembly kinetics of anthrax toxin complexes. *Biochemistry* **45**: 2380-6.

Christensen, K. A., Krantz, B. A., Melnyk, R. A. & Collier, R. J., (2005) Interaction of the 20 kDa and 63 kDa fragments of anthrax protective antigen: kinetics and thermodynamics. *Biochemistry* **44**: 1047-53.

Clackson, T. & Wells, J. A., (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* **267**: 383-6.

Collaborative Computational Project, N., (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**: 760-3.

Collier, R. J., (2009) Membrane translocation by anthrax toxin. *Mol Aspects Med* **30**: 413-22.

Crampton, N. & Brockwell, D. J., (2010) Unravelling the design principles for single protein mechanical strength. *Curr Opin Struct Biol* **20**: 508-17.

Cunningham, K., Lacy, D. B., Mogridge, J. & Collier, R. J., (2002) Mapping the lethal factor and edema factor binding sites on oligomeric anthrax protective antigen. *Proc. Natl Acad. Sci. U.S.A.* **99**: 7049-53.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C., (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**: W375-83.

Decker, T. & Lohmann-Matthes, M. L., (1988) A quick and simple method for the quantitation of lactate dehydrogenase release in measurements of cellular cytotoxicity and tumor necrosis factor (TNF) activity. *J. Immunol. Methods* **115**: 61-9.

Dill, K. A., (1990) Dominant forces in protein folding. *Biochemistry* **29**: 7133-7155.

Dill, K. A., Fiebig, K., M. & Chan, H. S., (1993) Cooperativity in protein-folding kinetics. *Proc. Natl. Acad. Sci. USA* **90**: 1942-1946.

Drazin, R., Kandel, J. & Collier, R. J., (1971) Structure and activity of diphtheria toxin. II. Attack by trypsin at a specific site within the intact toxin molecule. *J Biol Chem* **246**: 1504-10.

Drum, C. L., Yan, S. Z., Bard, J., Shen, Y. Q., Lu, D., Soelaiman, S., Grabarek, Z., Bohm, A. & Tang, W. J., (2002) Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature* **415**: 396-402.

Duesbery, N. S. & Vande Woude, G. F., (1999) Anthrax lethal factor causes proteolytic inactivation of mitogen-activated protein kinase kinase. *J. Appl. Microbiol.* **87**: 289-93.

Duesbery, N. S., Webb, C. P., Leppla, S. H., Gordon, V. M., Klimpel, K. R., Copeland, T. D., Ahn, N. G., Oskarsson, M. K., Fukasawa, K., Paull, K. D. & Vande Woude, G. F., (1998) Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280**: 734-7.

Einstein, A., (1905) Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* **17**: 549-560.

Emsley, P. & Cowtan, K., (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**: 2126-32.

Ezzell, J. W. & Abshire, T. G., (1992) Serum protease cleavage of *Bacillus anthracis* protective antigen. *J. Gen. Microbiol.* **138**: 543-9.

Falnes, P. O. & Sandvig, K., (2000) Penetration of protein toxins into cells. *Curr Opin Cell Biol* **12**: 407-13.

Feld, G. K., Kintzer, A. F., Tang, II, Thoren, K. L. & Krantz, B. A., (2012) Domain flexibility modulates the heterogeneous assembly mechanism of anthrax toxin protective antigen. *J Mol Biol* **415**: 159-74.

Feld, G. K., Thoren, K. L., Kintzer, A. F., Sterling, H. J., Tang, II, Greenberg, S. G., Williams, E. R. & Krantz, B. A., (2010) Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers. *Nature Struct. Mol. Biol.* **17**: 1383-90.

Feynman, R. P., Leighton, R. B. & Sands, M. (1963). The Feynman Lectures on Physics. Reading, MA, Addison-Wesley. **1:** 46.1-46.9.

Finkelstein, A., (2009) Proton-coupled protein transport through the anthrax toxin channel. *Philos Trans R Soc Lond B Biol Sci* **364**: 209-15.

Fischer, A., Holden, M. A., Pentelute, B. L. & Collier, R. J., (2011) Ultrasensitive detection of protein translocated through toxin pores in droplet-interface bilayers. *Proc Natl Acad Sci U S A* **108**: 16577-81.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A., (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**: 190-9.

Friedlander, A. M., (1986) Macrophages are sensitive to anthrax lethal toxin through an acid-dependent process. *J. Biol. Chem.* **261**: 7123-6.

Fuller, R. S., Brake, A. & Thorner, J., (1989) Yeast prohormone processing enzyme (KEX2 gene product) is a Ca2+-dependent serine protease. *Proc Natl Acad Sci U S A* **86**: 1434-8.

Fuller, R. S., Brake, A. J. & Thorner, J., (1989) Intracellular targeting and structural conservation of a prohormone-processing endoprotease. *Science* **246**: 482-6.

Garred, O., van Deurs, B. & Sandvig, K., (1995) Furin-induced cleavage and activation of Shiga toxin. *J. Biol. Chem.* **270**: 10817-21.

Gennerich, A. & Vale, R. D., (2009) Walking the walk: how kinesin and dynein coordinate their steps. *Curr Opin Cell Biol* **21**: 59-67.

Glick, B. S., (1995) Can Hsp70 proteins act as force-generating motors? *Cell* **80**: 11-4.

Gordon, V. M., Benz, R., Fujii, K., Leppla, S. H. & Tweten, R. K., (1997) Clostridium septicum alpha-toxin is proteolytically activated by furin. *Infect. Immun.* **65**: 4130-4.

Gordon, V. M. & Leppla, S. H., (1994) Proteolytic activation of bacterial toxins: role of bacterial and host cell proteases. *Infect. Immun.* **62**: 333-40.

Grigoriev, S. M., Muro, C., Dejean, L. M., Campo, M. L., Martinez-Caballero, S., Kinnally, K. W. & Kwang, W. J. (2004). Electrophysiological Approaches to the Study of Protein Translocation in Mitochondria. International Review of Cytology, Academic Press. **Volume 238:** 227-274.

Hallenberger, S., Bosch, V., Angliker, H., Shaw, E., Klenk, H. D. & Garten, W., (1992) Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* **360**: 358-61.

Harsman, A., Bartsch, P., Hemmis, B., Kruger, V. & Wagner, R., (2011) Exploring protein import pores of cellular organelles at the single molecule level using the planar lipid bilayer technique. *Eur J Cell Biol* **90**: 721-30.

Henrich, S., Cameron, A., Bourenkov, G. P., Kiefersauer, R., Huber, R., Lindberg, I., Bode, W. & Than, M. E., (2003) The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat. Struct. Biol.* **10**: 520-6.

Henry, J. P., Juin, P., Vallette, F. & Thieffry, M., (1996) Characterization and function of the mitochondrial outer membrane peptide-sensitive channel. *J Bioenerg Biomembr* **28**: 101-8.

Hinnerwisch, J., Fenton, W. A., Furtak, K. J., Farr, G. W. & Horwich, A. L., (2005) Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell* **121**: 1029-41.

Hoch, D. H., Romero-Mira, M., Ehrlich, B. E., Finkelstein, A., DasGupta, B. R. & Simpson, L. L., (1985) Channels formed by botulinum, tetanus, and diphtheria toxins in planar lipid bilayers: relevance to translocation of proteins across membranes. *Proc Natl Acad Sci U S A* **82**: 1692-6.

Holyoak, T., Wilson, M. A., Fenn, T. D., Kettner, C. A., Petsko, G. A., Fuller, R. S. & Ringe, D., (2003) 2.4 A resolution crystal structure of the prototypical hormone-processing protease Kex2 in complex with an Ala-Lys-Arg boronic acid inhibitor. *Biochemistry* **42**: 6709-18.

Horwich, A. L. & Fenton, W. A., (2009) Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q Rev Biophys* **42**: 83-116.

Huang, S., Ratliff, K. S. & Matouschek, A., (2002) Protein unfolding by the mitochondrial membrane potential. *Nat. Struct. Biol.* **9**: 301-7.

Huang, S., Ratliff, K. S., Schwartz, M. P., Spenner, J. M. & Matouschek, A., (1999) Mitochondria unfold precursor proteins by unraveling them from their N-termini. *Nat. Struct. Biol.* **6**: 1132-8.

Jancarik, J. & Kim, S. H., (1991) Sparse matrix sampling: A screening method for crystallization of proteins. *J. Appl. Cryst.* **24**: 409-411.

Janowiak, B. E., Fischer, A. & Collier, R. J., (2010) Effects of introducing a single charged residue into the phenylalanine clamp of multimeric anthrax protective antigen. *J Biol Chem* **285**: 8130-7.

Jennings-Antipov, L. D., Song, L. & Collier, R. J., (2011) Interactions of anthrax lethal factor with protective antigen defined by site-directed spin labeling. *Proc. Natl. Acad. Sci. U.S.A.* **108**: 1868-73.

Katayama, H., Janowiak, B. E., Brzozowski, M., Juryck, J., Falke, S., Gogol, E. P., Collier, R. J. & Fisher, M. T., (2008) GroEL as a molecular scaffold for structural analysis of the anthrax toxin pore. *Nature Struct. Mol. Biol.* **15**: 754-60.

Katayama, H., Wang, J., Tama, F., Chollet, L., Gogol, E. P., Collier, R. J. & Fisher, M. T., (2010) Three-dimensional structure of the anthrax toxin pore inserted into lipid nanodiscs and lipid vesicles. *Proc. Natl Acad. Sci. USA* **107**: 3453-7.

Kauzmann, W., (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**: 1-64.

Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T., (2003) Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**: 511-20.

Kenniston, J. A., Burton, R. E., Siddiqui, S. M., Baker, T. A. & Sauer, R. T., (2004) Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *J Struct Biol* **146**: 130-40.

Kintzer, A. F., Sterling, H. J., Tang, I. I., Abdul-Gader, A., Miles, A. J., Wallace, B. A., Williams, E. R. & Krantz, B. A., (2010) Role of the protective antigen octamer in the molecular mechanism of anthrax lethal toxin stabilization in plasma. *J. Mol. Biol.* **399**: 741-58.

Kintzer, A. F., Sterling, H. J., Tang, I. I., Williams, E. R. & Krantz, B. A., (2010b) Anthrax toxin receptor drives protective antigen oligomerization and stabilizes the heptameric and octameric oligomer by a similar mechanism. *PLoS ONE* **5**: e13888.

Kintzer, A. F., Thoren, K. L., Sterling, H. J., Dong, K. C., Feld, G. K., Tang, I. I., Zhang, T. T., Williams, E. R., Berger, J. M. & Krantz, B. A., (2009) The protective antigen component of anthrax toxin forms functional octameric complexes. *J. Mol. Biol.* **392**: 614-629.

Klimpel, K. R., Molloy, S. S., Thomas, G. & Leppla, S. H., (1992) Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin. *Proc Natl Acad Sci U S A* **89**: 10277-81.

Koehler, T. M. & Collier, R. J., (1991) Anthrax toxin protective antigen: low-pH-induced hydrophobicity and channel formation in liposomes. *Mol Microbiol* **5**: 1501-6.

Komiyama, T., Swanson, J. A. & Fuller, R. S., (2005) Protection from anthrax toxin-mediated killing of macrophages by the combined effects of furin inhibitors and chloroquine. *Antimicrob. Agents Ch.* **49**: 3875-82.

Krantz, B. A., Finkelstein, A. & Collier, R. J., (2006) Protein translocation through the anthrax toxin transmembrane pore is driven by a proton gradient. *J. Mol. Biol.* **355**: 968-79.

Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R., (2002) Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J Mol Biol* **324**: 359-71.

Krantz, B. A., Melnyk, R. A., Zhang, S., Juris, S. J., Lacy, D. B., Wu, Z., Finkelstein, A. & Collier, R. J., (2005) A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309**: 777-81.

Krantz, B. A., Trivedi, A. D., Cunningham, K., Christensen, K. A. & Collier, R. J., (2004) Acid-induced unfolding of the amino-terminal domains of the lethal and edema factors of anthrax toxin. *J. Mol. Biol.* **344**: 739-56.

Lacy, D. B., Lin, H. C., Melnyk, R. A., Schueler-Furman, O., Reither, L., Cunningham, K., Baker, D. & Collier, R. J., (2005) A model of anthrax toxin lethal factor bound to protective antigen. *Proc. Natl Acad. Sci. U.S.A.* **102**: 16409-14.

Lacy, D. B., Mourez, M., Fouassier, A. & Collier, R. J., (2002) Mapping the anthrax protective antigen binding site on the lethal and edema factors. *J. Biol. Chem.* **277**: 3006-10.

Lacy, D. B., Wigelsworth, D. J., Melnyk, R. A., Harrison, S. C. & Collier, R. J., (2004) Structure of heptameric protective antigen bound to an anthrax toxin receptor: a role for receptor in pH-dependent pore formation. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 13147-51.

Lacy, D. B., Wigelsworth, D. J., Scobie, H. M., Young, J. A. & Collier, R. J., (2004) Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: an anthrax toxin receptor. *Proc. Natl Acad. Sci. U.S.A.* **101**: 6367-72.

Landry, S. J. & Gierasch, L. M., (1991) The chaperonin GroEL binds a polypeptide in an alpha-helical conformation. *Biochemistry* **30**: 7359-62.

Lang, P. T., Ng, H. L., Fraser, J. S., Corn, J. E., Echols, N., Sales, M., Holton, J. M. & Alber, T., (2010) Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* **19**: 1420-31.

Langevin, P., (1908) Sur la théorie du mouvement brownien. *C. R. Acad. Sci. (Paris)* **146**: 530-533.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M., (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 283-291.

Lemons, D. S. & Gythiel, A., (1997) Paul Langevin's 1908 paper ''On the Theory of Brownian Motion'' ['' Sur la the´orie du mouvement brownien,'' C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]. *Am J Phys* **65**: 1079-1081.

Leppla, S. H., (1982) Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. *Proc. Natl Acad. Sci. U.S.A.* **79**: 3162-6.

Leppla, S. H., (1984) *Bacillus anthracis* calmodulin-dependent adenylate cyclase: chemical and enzymatic properties and interactions with eucaryotic cells. *Adv. Cyclic Nucl. Prot.* **17**: 189-98.

Levchenko, I., Grant, R. A., Flynn, J. M., Sauer, R. T. & Baker, T. A., (2005) Versatile modes of peptide recognition by the AAA+ adaptor protein SspB. *Nat. Struct. Mol. Biol.* **12**: 520-5.

Levchenko, I., Grant, R. A., Wah, D. A., Sauer, R. T. & Baker, T. A., (2003) Structure of a delivery protein for an AAA+ protease in complex with a peptide degradation tag. *Mol. Cell* **12**: 365-72.

Li, Y., Gao, X. & Chen, L., (2009) GroEL Recognizes an Amphipathic Helix and Binds to the Hydrophobic Side. *J. Biol. Chem.* **284**: 4324-31.

Liebermeister, W., Rapoport, T. A. & Heinrich, R., (2001) Ratcheting in post-translational protein translocation: a mathematical model. *J Mol Biol* **305**: 643-56.

Ludtke, S. J., Baldwin, P. R. & Chiu, W., (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**: 82-97.

Lum, R., Niggemann, M. & Glover, J. R., (2008) Peptide and protein binding in the axial channel of Hsp104. Insights into the mechanism of protein unfolding. *J. Biol. Chem.* **283**: 30139-50.

MacDowell, A. A., Celestre, R. S., Howells, M., McKinney, W., Krupnick, J., Cambie, D., Domning, E. E., Duarte, R. M., Kelez, N., Plate, D. W., Cork, C. W., Earnest, T. N., Dickert, J., Meigs, G., Ralston, C., Holton, J. M., Alber, T., Berger, J. M., Agard, D. A. & Padmore, H. A., (2004) Suite of three protein crystallography beamlines with single superconducting bend magnet as the source. *J. Synchrotron Rad.* **11**: 447-55.

Mahendran, K. R., Romero-Ruiz, M., Schlosinger, A., Winterhalter, M. & Nussberger, S., (2012) Protein Translocation through Tom40: Kinetics of Peptide Release. *Biophys J* **102**: 39-47.

Maillard, R. A., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C. M., Hodges, C., Martin, A. & Bustamante, C., (2011) ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**: 459-69.

Martin, A., Baker, T. A. & Sauer, R. T., (2008) Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat. Struct. Mol. Biol.* **15**: 1147-51.

Matlack, K. E. S., Misselwitz, B., Plath, K. & Rapoport, T. A., (1999) BiP Acts as a Molecular Ratchet during Posttranslational Transport of Prepro-alpha Factor across the ER Membrane. *Cell* **97**: 553-564.

Matouschek, A., (2003) Protein unfolding--an important process in vivo? *Curr. Opin. Struct. Biol.* **13**: 98-109.

Maxwell, J. C. (1871). Theory of Heat. Mineola, New York, Dover Publications, Inc.

McPherson, A., Jr., (1976) The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. *Methods Biochem. Anal.* **23**: 249-345.

Meador, W. E., Means, A. R. & Quiocho, F. A., (1992) Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. *Science* **257**: 1251-5.

Meador, W. E., Means, A. R. & Quiocho, F. A., (1993) Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science* **262**: 1718-21.

Melnyk, R. A. & Collier, R. J., (2006) A loop network within the anthrax toxin pore positions the phenylalanine clamp in an active conformation. *Proc. Natl Acad. Sci. U.S.A.* **103**: 9802-7.

Melnyk, R. A., Hewitt, K. M., Lacy, D. B., Lin, H. C., Gessner, C. R., Li, S., Woods, V. L., Jr. & Collier, R. J., (2006) Structural determinants for the binding of anthrax lethal factor to oligomeric protective antigen. *J. Biol. Chem.* **281**: 1630-5.

Meng, G., Surana, N.K., St. Geme, J.W. 3rd, Waksman, G., (2006) Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter. *EMBO J.* **25**: 2297-304.

Merdanovic, M., Clausen, T., Kaiser, M., Huber, R. & Ehrmann, M., (2011) Protein quality control in the bacterial periplasm. *Annu Rev Microbiol* **65**: 149-68.

Miller, C. J., Elliott, J. L. & Collier, R. J., (1999) Anthrax protective antigen: prepore-to-pore conversion. *Biochemistry* **38**: 10432-41.

Milne, J. C., Blanke, S. R., Hanna, P. C. & Collier, R. J., (1995) Protective antigen-binding domain of anthrax lethal factor mediates translocation of a heterologous protein fused to its amino- or carboxy-terminus. *Mol. Microbiol.* **15**: 661-6.

Milne, J. C., Furlong, D., Hanna, P. C., Wall, J. S. & Collier, R. J., (1994) Anthrax protective antigen forms oligomers during intoxication of mammalian cells. *J. Biol. Chem.* **269**: 20607-12.

Mogridge, J., Cunningham, K. & Collier, R. J., (2002) Stoichiometry of anthrax toxin complexes. *Biochemistry* **41**: 1079-82.

Mogridge, J., Mourez, M. & Collier, R. J., (2001) Involvement of domain 3 in oligomerization by the protective antigen moiety of anthrax toxin. *J. Bacteriol.* **183**: 2111-6.

Molloy, S. S., Bresnahan, P. A., Leppla, S. H., Klimpel, K. R. & Thomas, G., (1992) Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-

Arg and efficiently cleaves anthrax toxin protective antigen. *J. Biol. Chem.* **267**: 16396-402.

Molloy, S. S., Thomas, L., VanSlyke, J. K., Stenberg, P. E. & Thomas, G., (1994) Intracellular trafficking and activation of the furin proprotein convertase: localization to the TGN and recycling from the cell surface. *EMBO J.* **13**: 18-33.

Montal, M., (2009) Translocation of botulinum neurotoxin light chain protease by the heavy chain protein-conducting channel. *Toxicon* **54**: 565-9.

Mourez, M., Yan, M., Lacy, D. B., Dillon, L., Bentsen, L., Marpoe, A., Maurin, C., Hotze, E., Wigelsworth, D., Pimental, R. A., Ballard, J. D., Collier, R. J. & Tweten, R. K., (2003) Mapping dominant-negative mutations of anthrax protective antigen by scanning mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 13803-13808.

Movileanu, L., Schmittschmitt, J. P., Scholtz, J. M. & Bayley, H., (2005) Interactions of peptides with a protein pore. *Biophys J* **89**: 1030-45.

Murray, E. J., Leaman, D. P., Pawa, N., Perkins, H., Pickford, C., Perros, M., Zwick, M. B. & Butler, S. L., (2010) A low-molecular-weight entry inhibitor of both CCR5- and CXCR4-tropic strains of human immunodeficiency virus type 1 targets a novel site on gp41. *J. Virol.* **84**: 7288-99.

Nakayama, K., (1997) Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochem. J.* **327 (Pt 3)**: 625-35.

Nassi, S., Collier, R. J. & Finkelstein, A., (2002) PA_{63} channel of anthrax toxin: an extended β-barrel. *Biochemistry* **41**: 1445-50.

Navon, A. & Ciechanover, A., (2009) The 26 S proteasome: from basic mechanisms to drug targeting. *J. Biol. Chem.* **284**: 33713-8.

Nguyen, T., (2004) Three-dimensional model of the pore form of anthrax protective antigen. Structure and biological implications. *J Biomol Struct Dyn.* **22**: 253-65.

Oliva, R., Leone, M., Falcigno, L., D'Auria, G., Dettin, M., Scarinci, C., Di Bello, C. & Paolillo, L., (2002) Structural investigation of the HIV-1 envelope glycoprotein gp160 cleavage site. *Chemistry* **8**: 1467-73.

Oomen, C. J., Van Ulsen, P., Van Gelder, P., Feijen, M., Tommassen, J. & Gros, P., (2004) Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* **23**: 1257-66.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology. C. W. Carter Jr. and R. M. Sweet. New York, Academic Press, Inc. **276: Macromolecular Crystallography, part A:** 307-326.

Page, M. I. & Jencks, W. P., (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc Natl Acad Sci U S A* **68**: 1678-83.

Pannifer, A. D., Wong, T. Y., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska, J., Lacy, D. B., Collier, R. J., Park, S., Leppla, S. H., Hanna, P. & Liddington, R. C., (2001) Crystal structure of the anthrax lethal factor. *Nature* **414**: 229-33.

Pentelute, B. L., Barker, A. P., Janowiak, B. E., Kent, S. B. & Collier, R. J., (2010) A semisynthesis platform for investigating structure-function relationships in the N-terminal domain of the anthrax Lethal Factor. *ACS Chem Biol* **5**: 359-64.

Pentelute, B. L., Sharma, O., Collier, R.J., (2011) Chemical dissection of protein translocation through the anthrax toxin pore. *Angew Chem Int Ed Engl.* **50**: 2294-6.

Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C., (1997) Crystal structure of the anthrax toxin protective antigen. *Nature* **385**: 833-8.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E., (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**: 1605-12.

Pezard, C., Berche, P. & Mock, M., (1991) Contribution of individual toxin components to virulence of *Bacillus anthracis*. *Infect. Immun.* **59**: 3472-7.

Prakash, S. & Matouschek, A., (2004) Protein unfolding in the cell. *Trends Biochem Sci* **29**: 593-600.

Qa'dan, M., Christensen, K. A., Zhang, L., Roberts, T. M. & Collier, R. J., (2005) Membrane insertion by anthrax protective antigen in cultured cells. *Mol Cell Biol* **25**: 5492-8.

Ramachandran, G. N. & Sasisekharan, V., (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**: 283-438.

Rapoport, T. A., (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**: 663-9.

Reimann, P., Van den Broeck, C., Linke, H., Hanggi, P., Rubi, J. M. & Perez-Madrid, A., (2002) Diffusion in tilted periodic potentials: Enhancement, universality, and scaling. *Phys Rev E Stat Nonlin Soft Matter Phys* **65**: 031104.

Remacle, A. G., Gawlik, K., Golubkov, V. S., Cadwell, G. W., Liddington, R. C., Cieplak, P., Millis, S. Z., Desjardins, R., Routhier, S., Yuan, X. W., Neugebauer, W. A., Day, R. & Strongin, A. Y., (2010) Selective and potent furin inhibitors protect cells from anthrax without significant toxicity. *Int. J. Biochem. Cell. B.* **42**: 987-95.

Remacle, A. G., Shiryaev, S. A., Oh, E. S., Cieplak, P., Srinivasan, A., Wei, G., Liddington, R. C., Ratnikov, B. I., Parent, A., Desjardins, R., Day, R., Smith, J. W., Lebl, M. & Strongin, A. Y., (2008) Substrate cleavage analysis of furin and related proprotein convertases. A comparative study. *J. Biol. Chem.* **283**: 20897-906.

Santelli, E., Bankston, L. A., Leppla, S. H. & Liddington, R. C., (2004) Crystal structure of a complex between anthrax toxin and its host cell receptor. *Nature* **430**: 905-8.

Sarac, M. S., Peinado, J. R., Leppla, S. H. & Lindberg, I., (2004) Protection against anthrax toxemia by hexa-D-arginine in vitro and in vivo. *Infect. Immun.* **72**: 602-5.

Sauer, R. T. & Baker, T. A., (2011) AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* **80**: 587-612.

Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., Hersch, G. L., Joshi, S. A., Kenniston, J. A., Levchenko, I., Neher, S. B., Oakes, E. S., Siddiqui, S. M., Wah, D. A. & Baker, T. A., (2004) Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**: 9-18.

Schueler-Furman, O., Wang, C. & Baker, D., (2005) Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. *Proteins* **60**: 187-94.

Scobie, H. M., Rainey, G. J. A., Bradley, K. A. & Young, J. A., (2003) Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *Proc. Natl Acad. Sci. U.S.A.* **100**: 5170-4.

Scobie, H. M., Wigelsworth, D. J., Marlett, J. M., Thomas, D., Rainey, G. J., Lacy, D. B., Manchester, M., Collier, R. J. & Young, J. A., (2006) Anthrax toxin receptor 2-dependent lethal toxin killing in vivo. *PLoS Pathog.* **2**: e111.

Senzel, L., Huynh, P. D., Jakes, K. S., Collier, R. J. & Finkelstein, A., (1998) The diphtheria toxin channel-forming T domain translocates its own NH2-terminal region across planar bilayers. *Journal of General Physiology* **112**: 317-24.

Shen, Y., Zhukovskaya, N. L., Guo, Q., Florian, J. & Tang, W. J., (2005) Calcium-independent calmodulin binding and two-metal-ion catalytic mechanism of anthrax edema factor. *Embo J.* **24**: 929-41.

Simon, S. M., Peskin, C. S. & Oster, G. F., (1992) What drives the translocation of proteins? *Proc. Natl Acad. Sci. USA* **89**: 3770-4.

Smith, H. & Keppie, J., (1954) Observations on experimental anthrax: demonstration of a specific lethal factor produced in vivo by *Bacillus anthracis*. *Nature* **173**: 689.

Smoluchowski, M., (1906) Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Annalen der Physik* **21**: 756–780.

Smoluchowski, M., (1912) Experimentell nachweisbare, der Ublichen Thermodynamik widersprechende Molekularphenomene. *Phys. Zeitshur.* **13**: 1069-1079.

Sonnhammer, E. L. & Hollich, V., (2005) Scoredist: a simple and robust protein sequence distance estimator. *BMC Bioinformatics* **6**: 108.

Stadler, K., Allison, S. L., Schalich, J. & Heinz, F. X., (1997) Proteolytic activation of tick-borne encephalitis virus by furin. *J. Virol.* **71**: 8475-81.

Stanley, J. L. & Smith, H., (1961) Purification of factor I and recognition of a third factor of the anthrax toxin. *J. Gen. Microbiol.* **26**: 49-66.

Stark, H., Mueller, F., Orlova, E. V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R. & van Heel, M., (1995) The 70S *Escherichia coli* ribosome at 23 A resolution: fitting the ribosomal RNA. *Structure* **3**: 815-21.

Steiner, D. F., (1998) The proprotein convertases. *Curr. Opin. Chem. Biol.* **2**: 31-9.

Stieneke-Grober, A., Vey, M., Angliker, H., Shaw, E., Thomas, G., Roberts, C., Klenk, H. D. & Garten, W., (1992) Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J.* **11**: 2407-14.

Storoni, L. C., McCoy, A. J. & Read, R. J., (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60**: 432-8.

Sullivan, M. V. (1946). Activated Amusement Device. U. S. P. Office, Clay, W. C.

Tanford, C., (1968) Protein denaturation. *Adv. Protein Chem.* **23**: 121-282.

Thomas, G., (2002) Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat. Rev. Mol. Cell Biol.* **3**: 753-66.

Thompson, J. D., Higgins, D. G. & Gibson, T. J., (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-80.

Thomson, S. W., (1874) Kinetic Theory of the Dissipation of Energy. *Nature* **9**: 441-444.

Thoren, K. L. & Krantz, B. A., (2011) The unfolding story of anthrax toxin translocation. *Mol. Microbiol.* **80**: 588-95.

Thoren, K. L., Worden, E. J., Yassif, J. M. & Krantz, B. A., (2009) Lethal factor unfolding is the most force-dependent step of anthrax toxin translocation. *Proc. Natl Acad. Sci. U.S.A.* **106**: 21555-60.

Tian, L., Holmgren, R. A. & Matouschek, A., (2005) A conserved processing mechanism regulates the activity of transcription factors Cubitus interruptus and NF-kappaB. *Nat Struct Mol Biol* **12**: 1045-53.

Tomkiewicz, D., Nouwen, N. & Driessen, A. J., (2007) Pushing, pulling and trapping--modes of motor protein supported protein translocation. *FEBS Lett* **581**: 2820-8.

Tsuneoka, M., Nakayama, K., Hatsuzawa, K., Komada, M., Kitamura, N. & Mekada, E., (1993) Evidence for involvement of furin in cleavage and activation of diphtheria toxin. *J. Biol. Chem.* **268**: 26461-5.

Van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A., (2004) X-ray structure of a protein-conducting channel. *Nature* **427**: 36-44.

van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M., (1996) A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**: 17-24.

Volchkov, V. E., Feldmann, H., Volchkova, V. A. & Klenk, H. D., (1998) Processing of the Ebola virus glycoprotein by the proprotein convertase furin. *Proc. Natl Acad. Sci. U.S.A.* **95**: 5762-7.

Wang, J., Song, J. J., Franklin, M. C., Kamtekar, S., Im, Y. J., Rho, S. H., Seong, I. S., Lee, C. S., Chung, C. H. & Eom, S. H., (2001) Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure* **9**: 177-84.

Wang, Z., Feng, H., Landry, S. J., Maxwell, J. & Gierasch, L. M., (1999) Basis of substrate binding by the chaperonin GroEL. *Biochemistry* **38**: 12537-46.

Wesche, J., Elliott, J. L., Falnes, P. O., Olsnes, S. & Collier, R. J., (1998) Characterization of membrane translocation by anthrax protective antigen. *Biochemistry* **37**: 15737-46.

White, H. E., Saibil, H. R., Ignatiou, A. & Orlova, E. V., (2004) Recognition and separation of single particles with size variation by statistical analysis of their images. *J. Mol. Biol.* **336**: 453-60.

Wickner, W. & Schekman, R., (2005) Protein translocation across biological membranes. *Science* **310**: 1452-6.

Wigelsworth, D. J., Krantz, B. A., Christensen, K. A., Lacy, D. B., Juris, S. J. & Collier, R. J., (2004) Binding stoichiometry and kinetics of the interaction of a human anthrax toxin receptor, CMG2, with protective antigen. *J. Biol. Chem.* **279**: 23349-56.

Wimalasena, D. S., Janowiak, B. E., Lovell, S., Miyagi, M., Sun, J., Zhou, H., Hajduch, J., Pooput, C., Kirk, K. L., Battaile, K. P. & Bann, J. G., (2010) Evidence that histidine protonation of receptor-bound anthrax protective antigen is a trigger for pore formation. *Biochemistry* **49**: 6973-83.

Yin, L., Krantz, B., Russell, N. S., Deshpande, S. & Wilkinson, K. D., (2000) Nonhydrolyzable diubiquitin analogues are inhibitors of ubiquitin conjugation and deconjugation. *Biochemistry* **39**: 10001-10.

Young, J. A. & Collier, R. J., (2007) Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annu. Rev. Biochem.* **76**: 243-65.

Zhang, S., Finkelstein, A. & Collier, R. J., (2004) Evidence that translocation of anthrax toxin's lethal factor is initiated by entry of its N terminus into the protective antigen channel. *Proc. Natl Acad. Sci. USA* **101**: 16756-61.

Zhang, S., Udho, E., Wu, Z., Collier, R. J. & Finkelstein, A., (2004) Protein translocation through anthrax toxin channels formed in planar lipid bilayers. *Biophys. J.* **87**: 3842-9.

Zheng, L., Baumann, U. & Reymond, J. L., (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* **32**: e115.

Zhou, H. & Zhou, Y., (2002) Stability scale and atomic solvation parameters extracted from 1023 mutation experiments. *Proteins* **49**: 483-92.

Zimm, G. H. & Bragg, J. K., (1959) Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **31**: 526-535.

Zimmer, J., Nam, Y. & Rapoport, T. A., (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**: 936-43.