ELUCIDATING THE FOLDING PATHWAY OF OUTER MEMBRANE PROTEIN A

by
Emily Jane Danoff

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
July, 2014

© 2014 Emily Jane Danoff
All Rights Reserved
Abstract

Understanding how a protein folds to its functional structure is a central question of biophysics that is still not completely understood. Proteins that reside in membranes face the additional complication of inserting into the lipid bilayer while forming the correct conformation. Membrane proteins are essential for a number of biological processes and many diseases are linked to misfolding of these proteins, so gaining a better understanding of how membrane proteins attain their native structures is of great importance to biomedical research. Although most membrane proteins in the cell require the assistance of folding machinery to be properly folded and inserted, β-barrel outer membrane proteins are capable of spontaneously folding to the native state in the absence of folding factors, indicating an intrinsic folding mechanism dependent only on the amino acid sequence and the membrane environment. We have investigated several aspects of the intrinsic folding pathway for the model bacterial outer membrane protein OmpA, including the conformation and interactions of the aqueous unfolded state, the extent of secondary structure formation during the folding and insertion process, and the presence of off-pathway intermediates. In studying the conformations of the unfolded state, we found that the periplasmic domain of the protein, which has typically been ignored in folding studies, behaves as an independent folding unit and helps reduce the self-association of the unfolded transmembrane barrel domain. We also determined that the unfolded barrel domain has no regular structure and an expanded conformation, indicating that structure can only form upon interaction with a membrane. We next investigated the folding and membrane insertion of the OmpA β-barrel and developed a
comprehensive kinetic model to describe the pathways and folding intermediates of the protein. CD measurements revealed a partially inserted, penultimate state with a higher content of \( \beta \)-sheet structure than the native state. Based on our data and previous work we have proposed a detailed folding model for \( \beta \)-barrel proteins that is facilitated by the presence of defects in the lipid bilayer and secondary structure formation in the protein. We also identified several off-pathway intermediate states that give rise to additional exponential phases in the data.

**Thesis Advisor:** Dr. Karen Fleming

**Second Reader:** Dr. Doug Barrick

**Thesis Committee:** Dr. David Shortle (Chair)

Dr. Vince Hilser
Dr. Herschel Wade
To my family
I would like to thank everyone who has helped me during this challenging journey. I am deeply grateful to my advisor, Dr. Karen Fleming, for her continuous support and encouragement. She serendipitously caused me to come to Johns Hopkins to study biophysics by sending my undergraduate advisor information about the program, and her mentorship has guided me through the research and writing process these past years. I have enjoyed discussing science with her, learning from her, arguing with her, and singing karaoke with her. I would also like to thank the members of my defense committee: Dr. Doug Barrick, Dr. David Shortle, Dr. Vince Hilser, and Dr. Herschel Wade for useful discussions and suggestions. Other faculty members have also influenced me during my time at Hopkins, whether teaching classes I have attended, serving on my thesis review committees throughout the years, or providing feedback on my research at poster sessions, and I am grateful to them all. I also want to thank Dr. Pat Fleming for his insights and advice, and frequent help in rearranging the lab conference room.

Major thanks go to the members of the Fleming lab, past and present, with whom I have had the pleasure of working and being friends. Dr. Nancy Burgess, Dr. Alex Ebie Tan, and Dr. Preston Moon welcomed me to the lab and helped set my feet upon the detail-oriented research path. I have also enjoyed working with Dustin Himmerich, Mike Genuaidi, Will Chung, Dr. Nathan Zaccai, Dr. Dennis Gessmann, Sarah Kempka McDonald, Cliff Sandlin, Ashlee Plummer, and numerous undergraduate and rotation
students during my time in the Fleming lab. Special thanks to Dennis, Sarah, and Ashlee, for cheering me on and providing chocolate when times were tough.

Thank you to the dedicated administrators (past and present) for the Program in Molecular Biophysics and the Biophysics Department. Ranice Crosby, Kathie Kolish, Jerry Levin, Jessica Bailey, Lexie Ebert, Ken Rutledge, Chris Altizer, and Nancy Foltz kept everything running smoothly and have helped me immensely throughout the years.

Lastly I want to thank my family, without whom I would not be where I am today. Thank you to my parents, my brothers, and sisters-in-law for loving and supporting me, and for giving me a niece and nephews to dote on. Thank you also to my wonderful cat, Carmen, for welcoming me home every night with a purr.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>analytical ultracentrifugation</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>BAM</td>
<td>β-barrel assembly machinery</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>diC&lt;sub&gt;9&lt;/sub&gt;PC</td>
<td>1,2-dinonanoyl-&lt;i&gt;sn&lt;/i&gt;-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>diC&lt;sub&gt;10&lt;/sub&gt;PC</td>
<td>1,2-didecanoyl-&lt;i&gt;sn&lt;/i&gt;-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>diC&lt;sub&gt;11&lt;/sub&gt;PC</td>
<td>1,2-diundecanoyl-&lt;i&gt;sn&lt;/i&gt;-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>diC&lt;sub&gt;12&lt;/sub&gt;PC</td>
<td>1,2-dilauroyl-&lt;i&gt;sn&lt;/i&gt;-glycero-3-phosphocholine (also abbreviated DLPC)</td>
</tr>
<tr>
<td>diC&lt;sub&gt;13&lt;/sub&gt;PC</td>
<td>1,2-ditridecanoyl-&lt;i&gt;sn&lt;/i&gt;-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>diC&lt;sub&gt;14&lt;/sub&gt;PC</td>
<td>1,2-dimyristoyl-&lt;i&gt;sn&lt;/i&gt;-glycero-3-phosphocholine (also abbreviated DMPC)</td>
</tr>
<tr>
<td>diC&lt;sub&gt;18:1&lt;/sub&gt;PC</td>
<td>1,2-dioleoyl-&lt;i&gt;sn&lt;/i&gt;-glycero-3-phosphocholine (also abbreviated DOPC)</td>
</tr>
<tr>
<td>DSC</td>
<td>differential scanning calorimetry</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>GdnHCl</td>
<td>guanidine hydrochloride</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>H-bond</td>
<td>hydrogen bond</td>
</tr>
<tr>
<td>IM</td>
<td>bacterial inner membrane</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
</tr>
<tr>
<td>LUV</td>
<td>large unilamellar vesicle</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OM</td>
<td>bacterial outer membrane</td>
</tr>
<tr>
<td>OMP</td>
<td>outer membrane protein</td>
</tr>
<tr>
<td>PC</td>
<td>phosphatidylcholine</td>
</tr>
<tr>
<td>PE</td>
<td>phosphatidylethanolamine</td>
</tr>
<tr>
<td>PG</td>
<td>phosphatidylycerol</td>
</tr>
<tr>
<td>POTRA</td>
<td>polypeptide transport associated domain</td>
</tr>
<tr>
<td>PPOE</td>
<td>predetermined pathway with optional errors</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SUV</td>
<td>small unilamellar vesicle</td>
</tr>
<tr>
<td>SV</td>
<td>sedimentation velocity</td>
</tr>
<tr>
<td>s*</td>
<td>apparent sedimentation coefficient</td>
</tr>
<tr>
<td>s*$_{20,w}$</td>
<td>sedimentation coefficient at standard conditions (at 20 ºC, in water)</td>
</tr>
<tr>
<td>TB</td>
<td>terrific broth</td>
</tr>
<tr>
<td>TCEP</td>
<td>tris(2-carboxyethyl) phosphine hydrochloride</td>
</tr>
<tr>
<td>ThT</td>
<td>thioflavin T</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane</td>
</tr>
<tr>
<td>T$_{m}$</td>
<td>melting phase transition temperature</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>Protein</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>BamA</td>
<td>β-barrel assembly machinery subunit A in the outer membrane</td>
</tr>
<tr>
<td>BR</td>
<td>bacteriorhodopsin</td>
</tr>
<tr>
<td>CFTR</td>
<td>cystic fibrosis transmembrane conductance regulator</td>
</tr>
<tr>
<td>DAGK</td>
<td>diacylglycerol kinase</td>
</tr>
<tr>
<td>DegP</td>
<td>periplasmic serine protease with independent chaperone function</td>
</tr>
<tr>
<td>FadL</td>
<td>long-chain fatty acid transporter in the outer membrane</td>
</tr>
<tr>
<td>FepA</td>
<td>TonB-dependent receptor responsible for ferric enterobactin transport across the outer membrane</td>
</tr>
<tr>
<td>FhuA</td>
<td>TonB-dependent receptor responsible for ferrichrome transport across the outer membrane</td>
</tr>
<tr>
<td>FomA</td>
<td>major outer membrane protein of <em>Fusobacterium nucleatum</em></td>
</tr>
<tr>
<td>GPCR</td>
<td>G-protein-coupled receptor</td>
</tr>
<tr>
<td>OmpA</td>
<td>outer membrane protein A, responsible for anchoring the outer membrane to the peptidoglycan in the periplasm (the full-length protein is sometimes referred to as OmpA325)</td>
</tr>
<tr>
<td>OmpA171</td>
<td>the N-terminal β-barrel domain of OmpA, residues 1-171</td>
</tr>
<tr>
<td>OmpA_{per}</td>
<td>the C-terminal periplasmic domain of OmpA, residues 172-325</td>
</tr>
<tr>
<td>OmpC</td>
<td>outer membrane porin C, trimeric non-specific porin</td>
</tr>
<tr>
<td>OmpF</td>
<td>outer membrane porin F, trimeric non-specific porin</td>
</tr>
<tr>
<td>OmpG</td>
<td>outer membrane porin G, pH-dependent monomeric non-specific porin</td>
</tr>
<tr>
<td><strong>OmpLA</strong></td>
<td>outer membrane phospholipase A, exhibits lipolytic activity upon calcium binding and dimerization</td>
</tr>
<tr>
<td><strong>OmpT</strong></td>
<td>outer membrane protein T, protease activity against antimicrobial peptides</td>
</tr>
<tr>
<td><strong>OmpW</strong></td>
<td>outer membrane protein W, channel for small, hydrophobic molecules</td>
</tr>
<tr>
<td><strong>OmpX</strong></td>
<td>outer membrane protein X, virulence protein associated with adhesion and neutralizing host defenses</td>
</tr>
<tr>
<td><strong>PagP</strong></td>
<td>PhoP/PhoQ-activated gene product, palmitoyl transferase for lipid A</td>
</tr>
<tr>
<td><strong>PhoE</strong></td>
<td>outer membrane phosphoporin E, trimeric non-specific porin</td>
</tr>
<tr>
<td><strong>Sec61</strong></td>
<td>hetero-oligomeric channel component of eukaryotic translocon</td>
</tr>
<tr>
<td><strong>SecA</strong></td>
<td>motor domain of bacterial translocon</td>
</tr>
<tr>
<td><strong>SecB</strong></td>
<td>bacterial chaperone that directs polypeptides to the translocon post-translationally</td>
</tr>
<tr>
<td><strong>SecYEG</strong></td>
<td>bacterial homolog of Sec61, channel component of bacterial translocon</td>
</tr>
<tr>
<td><strong>Skp</strong></td>
<td>seventeen kilodalton protein, periplasmic chaperone</td>
</tr>
<tr>
<td><strong>SRP</strong></td>
<td>signal recognition particle, directs ribosome-nascent polypeptide complex to the translocon</td>
</tr>
<tr>
<td><strong>SurA</strong></td>
<td>survival factor A protein, periplasmic chaperone and peptidyl-prolyl isomerase</td>
</tr>
</tbody>
</table>
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td></td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>Abbreviations</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td>List of Proteins</td>
<td></td>
<td>ix</td>
</tr>
<tr>
<td>Table of Contents</td>
<td></td>
<td>xi</td>
</tr>
<tr>
<td>List of Tables</td>
<td></td>
<td>xiii</td>
</tr>
<tr>
<td>List of Figures</td>
<td></td>
<td>xiv</td>
</tr>
</tbody>
</table>

**Chapter 1**  Introduction

1.1 Overview and Perspectives  
1.2 Membrane Protein Structure and Function  
1.3 Membrane Protein Biogenesis  
1.4 The Intrinsic Folding Pathway of Outer Membrane Proteins  
1.5 Overview of Thesis  
1.6 Figures  

**Chapter 2**  The soluble, periplasmic domain of OmpA folds as an independent unit and reduces the self-association propensity of the unfolded OmpA transmembrane β-barrel

2.1 Abstract  
2.2 Introduction  
2.3 Materials and Methods  
2.4 Results  

70  
71  
74  
85
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Discussion</td>
<td>100</td>
</tr>
<tr>
<td>2.6 Acknowledgements</td>
<td>104</td>
</tr>
<tr>
<td>2.7 Tables and Figures</td>
<td>105</td>
</tr>
</tbody>
</table>

**Chapter 3**  
The OmpA transmembrane β-barrel folds via a multi-step mechanism with a partially inserted intermediate state containing extensive β-sheet structure

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Abstract</td>
<td>121</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>122</td>
</tr>
<tr>
<td>3.3 Materials and Methods</td>
<td>127</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>140</td>
</tr>
<tr>
<td>3.5 Discussion</td>
<td>195</td>
</tr>
<tr>
<td>3.6 Acknowledgements</td>
<td>208</td>
</tr>
<tr>
<td>3.7 Tables and Figures</td>
<td>209</td>
</tr>
</tbody>
</table>

**Chapter 4**  
Concluding Remarks

References  
Appendix: Fitting Equations

Copyright Permission Letters

Curriculum Vita
List of Tables

Chapter 2  The soluble, periplasmic domain of OmpA folds as an independent unit and reduces the self-association propensity of the unfolded OmpA transmembrane β-barrel

2.1 Primers used for cloning 105

2.2 Molecular weights, partial specific volumes, and extinction coefficients calculated in Sednterp 106

Chapter 3  The OmpA transmembrane β-barrel folds via a multi-step mechanism with a partially inserted intermediate state containing extensive β-sheet structure

3.1 CD signal values for unfolded OmpA 171 209

3.2 Kinetic rate constants for diC12PC model development 210

3.3 Kinetic rate constants and CD signals for final diC12PC model 211

3.4 Kinetic rate constants and CD signals for diC13PC model 212

3.5 Kinetic rate constants and CD signals for diC11PC model 213

3.6 Kinetic rate constants and CD signals for diC9PC model 214

3.7 Kinetic rate constants and CD signals for diC10PC model 215

3.8 Kinetic rate constants and CD signals for diC14PC at 25 ºC model 216

3.9 Kinetic rate constants and CD signals for diC14PC at 24 ºC model 217

3.10 Kinetic rate constants and CD signals for diC14PC at 26 ºC model 218

3.11 Attributes of OmpA main pathway folding species 219
List of Figures

**Chapter 1  Introduction**

1.1  $\alpha$-helical and $\beta$-barrel transmembrane protein structures  60
1.2  Membrane protein biogenesis  62
1.3  Thermodynamic model for $\alpha$-helical membrane protein folding  64
1.4  Secondary and tertiary structures of OmpA and PagP  65
1.5  Folding pathways proposed for OmpA  67
1.6  Folding pathways proposed for PagP  69

**Chapter 2  The soluble, periplasmic domain of OmpA folds as an independent unit and reduces the self-association propensity of the unfolded OmpA transmembrane $\beta$-barrel**

2.1  Structural models of OmpA constructs  107
2.2  CD wavelength spectra of OmpA constructs  108
2.3  Unfolding titration of OmpA$_{Per}$ measured by CD  110
2.4  Typical primary sedimentation velocity scans  111
2.5  Typical sedimentation velocity data analyzed by g(s*) and c(s*)  112
2.6  OmpA self-association as a function of concentration and urea  113
2.7  The periplasmic and barrel domains do not interact *in trans*  114
2.8  Time-dependence of OmpA self-association  115
2.9  Analysis of OmpA$_{171}$ CD kinetics  116
2.10  Thioflavin T binding indicates cross-$\beta$ structure  117
2.11  TEM images of oligomeric OmpA$_{171}$  118
2.12  Delayed folding of OmpA$_{325}$ and OmpA$_{171}$  119
2.13 Sedimentation coefficients extrapolated to the absence of urea

Chapter 3 The OmpA transmembrane β-barrel folds via a multi-step mechanism with a partially inserted intermediate state containing extensive β-sheet structure

3.1 Densitometry of folding gels demonstrates the “gone” state

3.2 Dividing by the boiled band intensity is a more accurate method of calculating fractional values

3.3 Electron microscopy images of LUVs for all lipids

3.4 The kinetics of the folded, unfolded, and gone species are highly dependent on bilayer thickness

3.5 DSC of diC₁₄PC LUVs reveals a phase transition at 24 ºC

3.6 Folding kinetics in diC₁₄PC are fastest at the transition temperature

3.7 Unfolded OmpA₁₇₁ has no regular structure in the absence of a bilayer

3.8 LUVs exhibit a CD signal

3.9 CD spectra of folded OmpA₁₇₁ indicate β-sheet structure and an aromatic exciton interaction

3.10 Processing of raw CD kinetics data

3.11 CD kinetics at 216 nm and 230 nm are complex and dependent on bilayer thickness

3.12 CD kinetics in diC₁₄PC are fastest at the transition temperature

3.13 The initial wavelength spectrum of OmpA₁₇₁ in diC₉PC reveals an intermediate with high β-sheet content

3.14 OmpA₁₇₁ adopts the same β-structured conformation in diC₁₃PC at 4 ºC as was observed in diC₉PC

3.15 The β-structured intermediate forms quickly in diC₁₃PC at 4 ºC and converts rapidly to the native state at 25 ºC
3.16 A two-step sequential mechanism is inadequate to fit the SDS-PAGE data for OmpA\textsubscript{171} folding into diC\textsubscript{12}PC

3.17 Models of increasing complexity better describe the diC\textsubscript{12}PC gel data

3.18 Final model and fitted kinetic data for folding into diC\textsubscript{12}PC

3.19 Variability of rate constants determined for diC\textsubscript{12}PC model

3.20 Model and fitted kinetic data for folding into diC\textsubscript{13}PC

3.21 Model and fitted kinetic data for folding into diC\textsubscript{11}PC

3.22 Model and fitted kinetic data for folding into diC\textsubscript{9}PC

3.23 Model and fitted kinetic data for folding into diC\textsubscript{10}PC

3.24 Model and fitted kinetic data for folding in diC\textsubscript{14}PC at 25 °C

3.25 Model and fitted kinetic data for folding in diC\textsubscript{14}PC at 24 °C

3.26 Model and fitted kinetic data for folding in diC\textsubscript{14}PC at 26 °C

3.27 CD signals of mechanism species are consistent across all lipid conditions

3.28 Structural mechanism proposed for OmpA β-barrel folding
Chapter 1

Introduction

1.1 Overview and Perspectives

Lipid membranes are an essential component of cellular life as they provide a selectively-permeable barrier that separates the contents of a cell (or organelle) from the external environment. Communication with the outside world is carried out through proteins associated with the cell membrane called membrane proteins, and includes such functions as transport of metabolites, interactions with intracellular and extracellular species, and transduction of signals into the cell. The significance of these proteins is underscored by the finding that 20-30% of all genes in many organisms’ genomes encode membrane proteins (Krogh et al., 2001), and more than 50% of the drugs currently on the market target membrane proteins (Overington et al., 2006). Furthermore, dysfunction of membrane proteins has been implicated in many human diseases, most often through disrupted targeting or folding of these proteins into the correct membrane (Sanders & Myers, 2004). Therefore elucidating the process by which membrane proteins fold is of significant interest.

Despite their clear importance, progress in understanding membrane protein folding has lagged far behind that of soluble proteins, mostly due to the experimental challenges associated with membrane protein studies. Because they reside in the non-polar lipid bilayer, membrane proteins contain a predominance of hydrophobic residues, which can lead to problems with solubility and denaturation in \textit{in vitro} experiments. β-
barrel membrane proteins, which are found in the outer membranes of Gram-negative bacteria, mitochondria, and chloroplasts, are composed of sequences of alternating hydrophobic and hydrophilic residues and thus have a decreased propensity for aggregation and are more easily denatured. In addition, these proteins are capable of spontaneously folding and inserting into synthetic lipid bilayers, making them ideal candidates for investigations of the intrinsic folding mechanism of membrane proteins. β-barrel outer membrane proteins are worth studying in their own right as well, as they are often involved in bacterial pathogenesis and mitochondrial function. From a basic science perspective, elucidating the driving forces behind β-barrel folding will also provide valuable information about protein sequence-structure-function relationships, protein-bilayer interactions, and the role of the hydrophobic effect in protein folding.

1.2 Membrane Protein Structure and Function

The two classes of transmembrane protein structures: α-helical and β-barrel

Membrane proteins can be grouped into categories based on their mode of association with the membrane and their structure. Proteins that do not extend into the hydrophobic interior of the lipid bilayer but are associated with the membrane surface (or other membrane proteins) through non-covalent interactions are known as peripheral membrane proteins. These proteins can be separated from the membrane by changes in ionic strength or pH (Singer & Nicolson, 1972). In contrast, proteins that insert partially or fully into the membrane are called integral membrane proteins; harsher methods such as adding detergents or denaturants are necessary to remove these proteins from the
membrane. Integral membrane proteins that traverse the entire lipid bilayer one or more times are called transmembrane (TM) proteins. Other integral membrane proteins extend only partway into the membrane with a segment of their peptide chain, or are covalently attached to a lipid molecule in one leaflet of the bilayer (so-called lipoproteins)(Cross, 1990).

Biological membranes are predominantly composed of a double layer of phospholipid molecules, with the nonpolar fatty acid tails of the lipids directed toward the interior of the bilayer and the polar head groups on the outside (Gorter & Grendel, 1925). The hydrophobic nature of the bilayer core excludes water and other polar solutes, thus creating a barrier around the cell. In TM proteins that traverse the bilayer, the amino acids that reside in the membrane interior generally possess hydrophobic side chains in order to have an energetically favorable interaction with the nonpolar lipid acyl chains (Ulmschneider et al., 2005). In addition, the amino and carbonyl groups of the peptide backbone must form hydrogen bonds (H-bonds) within the protein, due to water molecules being unavailable to interact with them. This requirement for intramolecular H-bonding results in TM protein structures being restricted to two main classes: α-helical and β-barrel* (Kennedy, 1978).

The α-helix and β-sheet are the two canonical types of protein secondary structure (Pauling & Corey, 1951; Pauling et al., 1951). These repeating patterns of H-bonding are observed in nearly all proteins, soluble and membrane-associated, and provide a stable framework for a protein’s global fold. The majority of TM proteins adopt the α-helix conformation, in some cases with only a single membrane-spanning helix,

* A current database of all available membrane protein structures is maintained by Stephen White’s group at http://blanco.biomol.uci.edu/mpstruc.
and in others with multiple traversing segments (called polytopic) forming a bundle of helices (for example, see bacteriorhodopsin (BR) in Figure 1.1A). Each α-helix contains a series of H-bonds between the carbonyl of residue $i$ and the amino group of residue $i+4$, resulting in the twisting of the peptide chain to form a right-handed helix shape (Figure 1.1B)(Richardson & Richardson, 1989). This structure efficiently satisfies all H-bonding partners in the peptide backbone and is thus ideal for the water-excluding bilayer environment. The geometry of the α-helix results in the amino acid residue side chains extending out perpendicularly to the helix axis (Figure 1.1B). The side chains are therefore either in contact with the nonpolar lipid tails or with other side chains of adjacent helices, and thus are predominantly of the hydrophobic class of amino acids. With a rise per residue of 1.5 Å for the α-helix structure, a chain of ~20 residues is required to traverse a typical membrane, which has a width of ~30 Å (Haltia & Freire, 1995; White & Wimley, 1999). This continuous stretch of hydrophobic residues in the primary sequence has been widely used with much success to develop algorithms for TM protein prediction (Jayasinghe et al., 2001), and is also largely responsible for the process by which TM segments are inserted into biological membranes (see below). In addition, the continuous hydrophobic segments of α-helical membrane proteins result in a high propensity for aggregation in the absence of the bilayer environment, making studies of these proteins challenging.

β-sheet secondary structure is formed by H-bonding between adjacent segments of polypeptide that form β-strands. The amino and carbonyl moieties that H-bond come from separate strands and thus are further apart in sequence than the H-bonding partners in an α-helix (Figure 1.1D). The strands can be aligned parallel or anti-parallel to each
other, with slightly different patterns of H-bonding resulting (Richardson & Richardson, 1989). The arrangement of β-strands within a β-sheet allows all of the H-bonding partners to be satisfied with the exception of the strands on either end of the sheet. In TM proteins that adopt a β-sheet conformation, the edge strands are accommodated by the sheet twisting into a barrel shape that allows the first and last β-strands to H-bond to each other, thus forming a β-barrel (see OmpF in Figure 1.1C) (Wimley, 2003). β-barrel TM proteins have a meander topology of anti-parallel β-strands, with the number of strands ranging from 8 to 22 (Schulz, 2000). The β-strands are connected on the intracellular side by short turns and on the extracellular side by long loops. The majority of β-barrels have an even number of strands, but a TM β-barrel with 19 strands has been discovered (Bayrhuber et al., 2008).

The architecture of the β-strand structure leads to the residue side chains extending perpendicularly on alternate sides of the β-sheet plane (Figure 1.1D) (Richardson & Richardson, 1989). In a β-barrel, this results in each transmembrane strand having every other residue face the hydrophobic lipid environment of the bilayer while the intervening residues face the barrel interior. Similar to α-helical TM proteins, the lipid-facing residues are predominantly hydrophobic (Wimley, 2002). But in contrast to α-helical proteins, β-barrel sequences are more difficult to recognize and predict due to the lack of continuous sections of hydrophobic residues. Nevertheless, the alternating inside-outside dyad repeat pattern has been utilized to develop algorithms for β-barrel prediction with some success (Wimley, 2002; Zhai & Saier, 2002). This amphipathic amino acid composition also makes β-barrel TM proteins more amenable to in vitro
studies due to a decreased propensity for aggregation and the ability to be solubilized by denaturing agents (discussed further below)(Stanley & Fleming, 2008).

Interestingly, membrane proteins that form the β-barrel structure are only found in a few specific environments: the outer membranes of Gram-negative bacteria † (Buchanan, 1999; Koebnik et al., 2000), mitochondria (Colombini, 2012; Paschen et al., 2005; Walther et al., 2009), and chloroplasts (Schleiff, Eichacker, et al., 2003; Schleiff, Soll, et al., 2003). The occurrence of β-barrel membrane proteins in the outer membranes of mitochondria and chloroplasts is thought to be a reflection of the evolutionary origin for these organelles as prokaryotic endosymbionts (Gray et al., 1999). β-barrel membrane proteins have a thinner belt of hydrophobic residues on their exterior than α-helical membrane proteins (24-27 Å as opposed to 30 Å), which is consistent with bacterial outer membranes being thinner than inner membranes and eukaryotic membranes (Lomize et al., 2006; Tamm et al., 2004). Due to the smaller bilayer width and the more extended conformation of β-strands compared to α-helices (3.3 Å per residue), only 9-11 residues are required in each strand to traverse the membrane (depending on the tilt of the strand) (White & Wimley, 1999; Wimley, 2002). The shorter TM sequences in β-barrel membrane proteins are another reason they are harder to predict than α-helical TM proteins.

† The cell envelope of Gram-negative bacteria is composed of two membranes: the inner membrane (IM) and the outer membrane (OM). These are separated by an aqueous compartment called the periplasm. The cell wall is located in the periplasm and consists of a thin layer of peptidoglycan. The inner membrane is composed entirely of phospholipids while the outer membrane has an asymmetrical structure with the inner leaflet being composed of phospholipids and the outer leaflet being composed of lipopolysaccharide (LPS) (Bos et al., 2007; Ruiz et al., 2005).
Membrane protein functions and roles in disease

Transmembrane proteins have a host of different functions, and are often grouped into three main categories: transport proteins, enzymes, and receptors (although these functions are not mutually exclusive) (Almén et al., 2009). Both α-helical and β-barrel membrane proteins have been identified with these functions.

As the name implies, transport membrane proteins are responsible for allowing the transport of solutes across the membrane. Some transport proteins, such as ion channels, achieve this through passive diffusion; channel proteins allow a solute to flow down its electrochemical gradient, often regulating the flow through some gating mechanism. Other transporters use energy to perform active transport and pump a solute against its electrochemical gradient (Stein, 1990). A large number of transport proteins are α-helical membrane proteins. A few examples include the voltage-gated potassium channel, KcsA (Doyle et al., 1998), the ligand-gated ion channel, nicotinic acetylcholine receptor (Unwin, 2005), and the Na⁺/K⁺ ATPase pump (Skou, 1957). α-helical transport proteins form the membrane-traversing channel for their substrate with a bundle of amphipathic helices that enclose a hydrophilic pore through the membrane. β-barrel membrane proteins also frequently serve as transport proteins, with the hydrophilic channel extending through the center of the barrel. For example, OmpF, OmpC, and PhoE serve as general porins in the bacterial outer membrane, and are responsible for the “leaky” nature of this membrane compared to the inner membrane or eukaryotic plasma membranes (Koebnik et al., 2000; Tamm et al., 2004). The β-barrels FhuA and FepA are active transporters from the outer membrane, and are responsible for the uptake of iron-siderophore complexes (Buchanan, 1999). To accommodate the passage of
macromolecules, β-barrel transporters are on the larger side of the scale, with 16-22 strands. The smaller β-barrels (such as OmpA, with only 8 strands) have densely packed cores from inward-pointing side chains and serve functions other than as channels (see below).

The second major function of TM proteins is as enzymes. A well-characterized example is the electron transport chain in cellular respiration. The proteins making up this chain reside in the inner mitochondrial membrane and participate in a series of redox reactions to pass electrons between them, eventually reducing oxygen and leading to the generation of ATP, the cellular energy currency (Saraste, 1999). The formation of ATP is driven by protons flowing down their electrochemical gradient through the membrane protein ATP synthase (Boyer, 1997), and this gradient is produced by the electron transport chain proteins pumping the protons across the membrane in conjunction with the redox reactions. Thus these membrane protein enzymes couple their activity to transport functions, providing an example of overlap between the categories delineated above.

Several bacterial β-barrel proteins have been identified with enzymatic functions. PagP is a transacylase that is responsible for transferring a palmitate chain from a phospholipid to lipid A as part of the biosynthetic pathway for lipopolysaccharide (LPS) (Bishop et al., 2000); OmpLA is a phospholipase that cleaves phospholipids that have erroneously entered the outer leaflet of the outer membrane (Dekker et al., 1997); OmpT is a protease that cleaves antimicrobial peptides (Vandeputte-Rutten et al., 2001).

The third functional category for TM proteins is as receptors. These proteins transduce external signals into the cell by binding a ligand (such as a hormone or
neurotransmitter) and undergoing some change that initiates an intracellular response (Krauss, 1999). There are three main mechanisms by which receptors transduce a signal. The first two again demonstrate overlap with the other two types of TM protein functions already discussed: ion channel-linked and enzyme-linked behavior. Ligand-gated ion channels, such as the nicotinic acetylcholine receptor (Unwin, 2005), transduce a signal upon ligand binding by undergoing a conformational change that allows a particular ion type to flow into the cell and set off a signal cascade. In contrast, enzyme-linked receptors respond to a ligand binding by catalyzing some chemical reaction on the intracellular side of the membrane. An example of this type of receptor would be epidermal growth factor receptor (Carpenter & Cohen, 1979), a tyrosine kinase that upon binding epidermal growth factor dimerizes and autophosphorylates several of its tyrosine residues, leading to further protein interactions inside the cell and downstream effects.

The third main type of TM receptor is the G-protein-coupled receptor, or GPCR. GPCRs constitute the largest family of membrane proteins and mediate a vast array of eukaryotic cellular responses, including the physiological processes of vision, olfaction, and taste (Rosenbaum et al., 2009). The basic mechanism involves the GPCR undergoing a conformational change upon binding of ligand, which leads to the activation of an associated trimeric G-protein on the intracellular side of the membrane and causes it to bind GTP, leading to further downstream signaling events (Gilman, 1995).

The majority of receptor TM proteins are α-helical but there are a few examples of β-barrel proteins with receptor function. Mentioned above as examples of active transporters, the outer membrane β-barrels FhuA and FepA also utilize receptor function as an essential component of their transport ability. These proteins are examples of TonB-
dependent receptors because their coupling to the integral inner membrane protein TonB is what provides the energetic driving force for iron-siderophore uptake against its concentration gradient. Essentially, binding of ligand leads to a conformational change in the β-barrel protein that is communicated to the TonB-ExbB-ExbD complex, which utilizes the chemiosmotic potential of the inner membrane to drive further conformational changes in the β-barrel receptor and induce substrate passage (Ferguson & Deisenhofer, 2002).

Besides the three main functions described above, some TM β-barrels fall into an additional category of function related to virulence. For example, the bacterial outer membrane protein OmpX plays a role in adhesion and invasion of eukaryotic cells as well as neutralizing host defense mechanisms, presumably through presenting a protruding edge of β-sheet structure as a binding partner to host surface proteins (Fernández et al., 2004; Vogt & Schulz, 1999). The protein OmpA predominantly serves a structural role in Gram-negative bacteria by linking the outer membrane to the peptidoglycan layer through a soluble periplasmic domain (Koebnik, 1995; Park et al., 2012), but OmpA has also been implicated in pathogenesis of a number of bacteria, with functions of adhesion, invasion, and evasion of host defenses (Confer & Ayalew, 2013). A number of secreted bacterial toxins have also been identified with β-barrel structure, such as α-hemolysin from *Staphylococcus aureus* (Song et al., 1996) and perfringolysin from *Clostridium perfringens* (Ramachandran et al., 2002). These toxins kill cells by inserting as β-barrel pores into the membrane and causing unregulated flow of ions, small molecules and water into and out of the cell. In contrast to the native β-barrels of the bacterial outer membrane that are typically formed by a single polypeptide sequence, these toxins form
pores upon binding the host membrane by assembling into oligomeric complexes, with each monomer contributing one β-hairpin (two strands) to the barrel structure (Wimley, 2003). Evidently TM proteins serve a variety of important functions for cellular life. Therefore it is not surprising that when TM proteins are not correctly folded and assembled to their native state and location, disease states can result.

In the past 30 years, medical genetics has identified over 140,000 different genetic lesions associated with human inherited diseases, many of which correspond to missense mutations in protein open reading frames that lead to a single amino acid change in the resultant protein (Stenson et al., 2014). Interestingly, the most common mechanism by which point mutations cause disease is by disruption of normal protein folding or trafficking, thus resulting in loss of that protein’s function, and less often by direct perturbation of a protein’s function (Sanders & Myers, 2004). Many disease-linked mutations occur in TM proteins, and correspondingly result in improper folding and trafficking of these proteins to the correct membrane location. Probably the best studied example is the cystic fibrosis-causing chloride channel, CFTR (cystic fibrosis transmembrane conductance regulator). It has been found that CFTR proteins carrying the most common mutation responsible for cystic fibrosis, ∆F508, are not processed correctly and consequently are not delivered to the plasma membrane of epithelial cells (Denning et al., 1992), although the chloride conductance of these channels appears to be unaffected by the mutation when studied in isolation (Li et al., 1993). Many other diseases are also associated with misassembly of plasma membrane proteins, such as retinitis pigmentosa (a type of hereditary blindness caused by misfolding of rhodopsin
proteins in the retina)(Garriga et al., 1996), Charcot-Marie-Tooth disease (a hereditary motor and sensory neuropathy caused by misfolding of the membrane protein PMP22 in myelinating Schwann cells)(Naef & Suter, 1999), and familial hypercholesterolemia (a condition that contributes to coronary atherosclerosis and is caused by misfolding of LDL receptors)(Jeon et al., 2001).

With membrane proteins playing such crucial roles in biological processes and human health, it is therefore of great importance that researchers continue to study and dissect these complex macromolecules and their biogenesis. Armed with better knowledge of how TM proteins correctly attain their native state and location, we will be able to further develop treatments for diseases caused by incorrect membrane protein processing and assembly. One such strategy could be designing drugs that bind and stabilize the native state of a membrane protein, thus decreasing the likelihood of misassembly leading to disease (Sanders & Myers, 2004).

Besides being directly responsible for many genetic human diseases through a loss of some important innate function, membrane proteins are also largely responsible for diseases resulting from bacterial infections. Many of the β-barrel TM proteins from bacteria are involved in virulence and have been implicated in a variety of bacterial infections, including middle ear infections caused by Haemophilus influenzae (Reddy et al., 1996), gastritis caused by Helicobacter pylori (Dorrell et al., 1999; Peck et al., 1999), and whooping cough caused by Bordetella pertussis (Smith et al., 2001). The process by which β-barrel TM proteins are trafficked to the outer membrane and folded to their native state is therefore of profound interest to biomedical science because preventing this process could be a useful strategy for combating bacterial infections.
In addition to the direct medical benefits of understanding membrane protein folding, this topic has major relevance for basic science in general. The underlying principles of how a polypeptide sequence encodes structure and function for a protein and how the protein attains its native state are still yet to be fully elucidated, and the extension of these principles to membrane protein folding is even less well understood. Knowing the thermodynamic and kinetic driving forces that govern membrane protein folding will be of invaluable use, certainly for treating and preventing human diseases but also for future biotechnology, bioengineering, and protein design purposes (Stanley & Fleming, 2008).

1.3 Membrane Protein Biogenesis

*α*-helical membrane proteins are inserted by the Sec translocon complex

In eukaryotic cells, many *α*-helical TM proteins are first inserted into the membrane of the endoplasmic reticulum (ER) and subsequently modified and sorted to the correct location through secretory vesicles (Rapoport et al., 2004). As depicted in Figure 1.2A, the insertion process usually takes place co-translationally and is mediated by a protein-conducting channel termed the translocon (which is itself an *α*-helical membrane protein). TM proteins are directed toward the translocon by a highly hydrophobic N-terminal signal sequence that emerges first from the ribosome and is recognized by a signal recognition particle (SRP). Signal sequences are typically cleaved off by signal peptidase once proteins are translocated across the membrane, but for integral membrane proteins the hydrophobic signal sequence often ends up being a
transmembrane segment of the protein (Alder & Johnson, 2004). After the SRP directs
the ribosome-nascent-polypeptide complex to the ER via the SRP receptor, the ribosome
docks with the translocon and translation proceeds, directing the nascent chain through
the complex.

The core component of the translocon is the hetero-oligomeric complex Sec61,
which forms a water-filled channel across the membrane (Deshaies & Schekman, 1987;
Görlich et al., 1992). The crystal structure for the homologous archaeal translocon,
SecYEβ, suggests that TM helices exit the channel through a lateral gate into the lipid
bilayer (Van den Berg et al., 2004). It is not completely clear how the translocon
differentiates between TM segments that need to be transferred to the surrounding bilayer
and secretory proteins that are passed completely through the channel into the lumen of
the ER, but hydrophobicity certainly plays a role. It has been proposed that the nascent
polypeptide directly interacts with the lipids surrounding the translocon during
translocation due to the lateral gate “breathing” open and closed. When the gate is
opened, the polypeptide is exposed to the nonpolar bilayer environment, and if it is a
hydrophobic stretch of amino acids destined to become a TM segment, the chain
partitions into the bilayer and folds into an α-helix (Hessa et al., 2005; Rapoport et al.,
2004). In polytopic α-helical membrane proteins, each TM segment transfers into the
bilayer as it is translocated and then forms tertiary contacts with other inserted TM
segments.

The translocon is highly conserved across all species (Rapoport et al., 1996). The
bacterial homolog, SecYEG (Ito, 1984), integrates α-helical TM proteins into the inner
membrane of Gram-negative bacteria and translocates secretory, periplasmic, and outer
membrane proteins into the periplasmic space where they are further sorted to their destinations. As in eukaryotes, inner membrane proteins are generally targeted to the translocon by SRP binding to their signal sequence and integrated into the membrane co-translationally (Driessen & Nouwen, 2008).

Outer membrane β-barrel proteins are assisted by a folding machinery including periplasmic chaperones and the BAM complex

β-barrel membrane proteins that reside in the outer membranes of Gram-negative bacteria (hereafter referred to as OMPs for outer membrane proteins) have a slightly more complicated biogenesis, as they must cross the inner membrane and the periplasmic space before inserting into the outer membrane (Figure 1.2B). OMPs are synthesized in the cytoplasm and directed to the translocon via a signal sequence, just as for α-helical membrane proteins. However, OMPs (and other bacterial secretory proteins) are mostly targeted to the translocon post-translationally, via the chaperone SecB (Baars et al., 2006). Proteins following this pathway tend to possess less hydrophobic signal sequences that are rich in aromatic and basic amino acids. These signal sequences are bound preferentially by trigger factor, and thereby prevented from interacting with SRP and being co-translationally inserted into the membrane, as for α-helical TM proteins (Ullers et al., 2006). The nascent peptides subsequently associate with SecB, which maintains them in an unfolded state after their translation is completed. SecB directs the proteins to the translocon through binding to SecA, a peripheral motor domain associated with SecYEG. The proteins are then translocated across the inner membrane through the translocon channel by SecA via ATPase activity (Driessen & Nouwen, 2008; du Plessis
et al., 2011; Holland, 2004). OMPs avoid being inserted into the inner membrane due to the fact that their TM sequences alternate between hydrophilic and hydrophobic residues, in contrast to the hydrophobic stretches that occur in α-helical TM segments.

After translocating the inner membrane, the signal sequence is cleaved off by signal peptidase (Zwizinski & Wickner, 1980) and the proteins are released into the aqueous periplasmic space. OMPs must then traverse the periplasm, cross the peptidoglycan layer, and fold and insert into the outer membrane. The molecular details of this pathway are not fully understood but several key proteins involved in the process have been identified: the periplasmic chaperones Skp, SurA, and DegP, and the BAM (β-barrel assembly machinery) complex in the outer membrane (Figure 1.2B).

The trimeric chaperone protein Skp has been shown to bind unfolded OMPs soon after their emergence from the translocon and it is thought to prevent these proteins from aggregating in the periplasm (R. Chen & Henning, 1996; Harms et al., 2001; Schäfer et al., 1999). This is supported by the fact that Skp maintains OMPs in an unfolded conformation and inhibits aggregation in vitro (McMorran et al., 2013; Walton et al., 2009; Walton & Sousa, 2004). The peptidylprolyl isomerase SurA has also been implicated in OMP biogenesis, as strains with mutations in the surA gene exhibit OMP assembly defects (Lazar & Kolter, 1996; Missiakas et al., 1996; Rouviere & Gross, 1996), and SurA interacts with OMPs and has chaperone activity (independent of its PPlase activity) in vitro (Behrens et al., 2001; S. Wu et al., 2011). DegP is a periplasmic protease that has been demonstrated to also have chaperone activity (Spiess et al., 1999) and it has been implicated in OMP assembly based on genetic assays (Rizzitello et al., 2001). Interestingly, individual null mutations of the skp or surA genes produce strains
that are still viable (albeit with decreased OMPs or OM defects), but the combination of both mutations is synthetic lethal (Rizzitello et al., 2001). This result has led to the proposal that Skp and SurA serve redundant roles in OMP biogenesis and therefore represent parallel pathways by which OMPs cross the periplasm. DegP has been suggested to be part of the Skp-mediated pathway, because the double mutation of degP and skp is not lethal, but the combination of degP and surA mutations is lethal (Rizzitello et al., 2001). The parallel pathways model has been further refined to propose that certain OMPs follow the SurA-mediated pathway preferentially, including the abundant OMPs OmpA, OmpF, and LamB, but these proteins can be efficiently assembled by the Skp/DegP pathway in the absence of SurA (Denoncin et al., 2012; Vertommen et al., 2009).

However, a different model has also been proposed, wherein Skp and SurA act sequentially in the same pathway for OMP biogenesis, with Skp interacting with the unfolded OMP early in the pathway and SurA facilitating OMP folding in conjunction with the BAM complex late in the pathway (see below)(Bos et al., 2007; Walther et al., 2009). The viability of the individual skp and surA mutants is explained in this model by the idea that in the absence of one chaperone, the other chaperone is still sufficient to facilitate OMP folding; the double mutation is lethal because the cell is unable to tolerate the loss of both chaperones. The role of DegP in OMP biogenesis has also been contested, with a possible explanation for the lethal degP surA double mutation being that the protease activity of DegP is essential to remove toxic unfolded OMP aggregates from the periplasm that build up in the absence of SurA (Walther et al., 2009). To be
representative of both possible chaperone pathway models, the process is depicted in a
general manner in Figure 1.2B.

It should be noted that the periplasm is also the location of the peptidoglycan cell
wall. This mesh-like structure (also called the murein sacculus) encircles the bacterial cell
and plays a critical role in maintaining cell shape and preventing osmotic and mechanical
lysis (Scheurwater & Burrows, 2011). Peptidoglycan is composed of glycan strands
cross-linked by peptides, arranged in a single disorganized layer that lies parallel to the
cell surface (Gan et al., 2008). The presence of the cell wall between the inner and outer
membranes of Gram-negative bacteria might appear to be a barrier to OMP passage, but
in fact peptidoglycan has been shown to be a porous structure that allows the passage of
proteins less than 100 kDa (Demchick & Koch, 1996; Vazquez-Laslop et al., 2001).
Therefore it has been assumed that OMP-chaperone complexes cross the peptidoglycan
layer with no hindrance. Accordingly, this structure is omitted from Figure 1.2B for
clarity.

The final stage of OMP biogenesis is to fold and insert into the outer membrane.
In recent years a protein complex has been identified that is essential for this process in vivo, termed the BAM complex. The key component of the BAM complex is the highly
conserved OM-embedded β-barrel BamA (Gentle et al., 2004; Walther et al., 2009). This
protein was identified in Neisseria meningitidis (originally called Omp85), but it is found
in all Gram-negative bacteria and was shown to be essential for cell viability and OMP
assembly (Voulhoux et al., 2003; Werner & Misra, 2005; T. Wu et al., 2005). Homologs
have also been identified in eukaryotes: Tob55/Sam50 in mitochondria (Paschen et al.,
2005) and Toc75 in chloroplasts (Soll & Schleiff, 2004). All BamA proteins are
composed of a C-terminal TM β-barrel domain (Ni et al., 2014; Noinaj et al., 2013) and an N-terminal periplasmic domain consisting of varying numbers of polypeptide transport associated (POTRA) sub-domains (Gatzeva-Topalova et al., 2008; S. Kim et al., 2007). Interestingly, bacterial BamA proteins contain five POTRA domains whereas the chloroplast homolog contains three and the mitochondrial homolog contains only one (McMorran et al., 2014; Walther et al., 2009).

The BamA from *Escherichia coli* (formerly called YaeT) was found to associate with four lipoproteins, BamBCDE (formerly YfgL, NlpB, YfiO, and SmpA, respectively), which are anchored to the inner leaflet of the OM and make up the rest of the BAM complex (T. Wu et al., 2005). The complex can be separated into two sub-complexes of BamAB and BamCDE (Ricci et al., 2012), which can be reassembled *in vitro* to produce a functional BAM complex (Hagan et al., 2010). BamB is thought to interact with BamA through POTRA domains 2 and 3 while BamCDE interacts with BamA via BamD binding to POTRA domain 5 (see Figure 1.2B)(Selkris et al., 2013).

The exact mechanism by which the BAM complex facilitates the folding and insertion of OMPs into the outer membrane is not clear but recent work has provided some clues. It is thought that OMP/chaperone complexes are targeted to the BAM complex by a highly conserved C-terminal sequence in the OMP, of which the final phenylalanine residue has been shown to be particularly important for OMP assembly (de Cock et al., 1997; Struyve et al., 1991). Recognition of the targeting sequence has been attributed to BamA, but the structural details of the interaction and its role in OMP folding are unknown (Gessmann et al., 2014; Robert et al., 2006). In addition to recognizing the OMP substrate, BamA has been shown to catalyze the OMP folding
reaction (Gessmann et al., 2014), and a few models for how this takes place have been proposed. In one hypothesized model, the first and last \( \beta \)-strands (1 and 16) of BamA separate and the exposed H-bonding partners serve as a template for the substrate OMP to form \( \beta \)-sheet structure. This model is supported by molecular dynamics simulations showing an opening between \( \beta \)-strands 1 and 16 of BamA (Noijaj et al., 2013). It is debated whether the substrate passes through the interior of the BamA barrel or enters the membrane from the periplasm at the BamA-lipid interface (K. H. Kim et al., 2012). The opening between \( \beta \)-strands 1 and 16 could also serve as a gate for substrate passage between the barrel interior and the membrane (Ni et al., 2014), but there is no conclusive evidence that the OMP passes through the BamA barrel. An alternative model is that the function of BamA is simply to destabilize the membrane adjacent to it, which OMPs take advantage of to insert and fold via an intrinsic mechanism (see below)(Gessmann et al., 2014). This possibility is supported by the observation of substantial membrane thinning next to \( \beta \)-strands 1 and 16 of BamA in molecular dynamics simulations (Noijaj et al., 2013).

The function of the POTRA domains in OMP folding catalysis is unclear, although some studies have suggested a role in stabilizing substrate OMPs by \( \beta \)-augmentation (S. Kim et al., 2007). The POTRA domains also serve as docking points for the other proteins of the BAM complex, but the roles of these lipoproteins are also unclear. Only BamD is essential for cell viability, but deletion of the other Bam proteins results in OMP assembly defects, indicating that the five proteins constitute the most efficient complex (Malinverni et al., 2006; Ruiz et al., 2005; Sklar et al., 2007). BamB has been proposed to serve as a scaffold protein that helps orient the POTRA domains for
interaction with substrate OMPs and other Bam proteins, as well as possibly interacting with OMPs by β-augmentation (Heuck et al., 2011; Noinaj et al., 2011). The BamCDE sub-complex has been suggested to facilitate a conformational switch in the BamA protein that is important for its catalytic function (Ricci et al., 2012; Rigel et al., 2013).

An interesting conundrum of OMP assembly is the fact that there is no known energy source in the periplasm, so the catalysis of OMP folding must take place without the input of energy. This contrasts with the process of translocation across the IM driven either by the translating ribosome or the SecA motor protein, which utilize GTP or ATP as an energy source. However, OMPs have been found to possess extremely high thermodynamic stabilities (Moon & Fleming, 2011) and it has been proposed that this creates an energy “sink” that drives OMP sorting, dissociation from chaperones, and insertion into the OM (Moon et al., 2013). Furthermore, many OMPs have been shown to be able to fold to the native state in synthetic membranes in the absence of chaperones or the BAM complex (Buchanan, 1999; Burgess et al., 2008; Huysmans et al., 2007; Surrey & Jähnig, 1992), raising the question of why the folding machinery is even required in vivo. The answer seems to lie in the fact that the native lipid head groups in the bacterial OM (phosphoethanolamine and phosphoglycerol) impose a high kinetic barrier to OMP folding, resulting in folding rates too slow to be compatible with the bacterial growth rate (Gessmann et al., 2014). Therefore the role of BamA is to lower this kinetic barrier and accelerate OMP folding, possibly through the mechanisms discussed above. It should be noted that the BAM complex and the periplasmic chaperones are probably not providing essential conformational instructions to the OMPs, because these proteins’ ability to independently fold to the functional form in vitro indicates the existence of an intrinsic
folding mechanism. Most likely, the chaperones help prevent off-pathway reactions, such as aggregation, that compete with folding, and BamA helps facilitate the intrinsic mechanism, perhaps by providing a β-sheet template to help get folding started, or perhaps just by destabilizing the membrane enough so the OMP can take advantage and fold into it. The intrinsic folding pathway utilized by OMPs is the focus of this dissertation, and the following section will discuss the previous work in this area.

1.4 The Intrinsic Folding Pathway of Outer Membrane Proteins

Membrane protein folding

In the 1950s and 60s, Anfinsen demonstrated the in vitro refolding of the protein ribonuclease from a completely denatured state (Anfinsen et al., 1961; Sela et al., 1957). These experiments provided evidence that the amino acid sequence of a protein is all that is needed to define its native three-dimensional structure, and that the native conformation resides at a free energy minimum, or thermodynamic equilibrium (Anfinsen, 1973). Known as the “thermodynamic hypothesis”, these concepts have been firmly established to be true for the vast majority of soluble proteins (Englander et al., 2007). However, it was initially unclear whether the same principles would apply to proteins that reside in membranes. The lipid bilayer is a unique hydrophobic and anisotropic environment that is very different from the aqueous solution that soluble proteins exist in. Furthermore, membrane proteins utilize complex macromolecular machines to be inserted into membranes in vivo. Therefore it would not have been surprising if membrane proteins were subject to different driving forces directing their
assembly, and their native states were kinetically-trapped instead of being at a free energy minimum (Popot & Engelman, 1990; Stanley & Fleming, 2008). Nevertheless, it was shown in the years subsequent to Anfinsen’s discovery that membrane proteins are also equilibrium structures. In 1981, the Khorana group demonstrated the renaturation of acid-unfolded bacteriorhodopsin (BR) to a fully functional state in phospholipid bilayers in the absence of any assembly machinery (Huang et al., 1981). They also achieved reassembly of active protein from denatured proteolytic fragments (Huang et al., 1981), and since then a number of other α-helical TM proteins have been shown to be capable of assembling from fragments (Popot & Engelman, 2000). The thermodynamic hypothesis was extended to β-barrel TM proteins with the demonstration that OmpF and OmpA could be folded into lipid bilayers from a denatured state (Dornmair et al., 1990; Eisele & Rosenbusch, 1990; Surrey & Jähnig, 1992) or from polypeptide fragments in vitro (Debnath et al., 2010). In recent years, many other OMPs have also been successfully folded into synthetic membranes (Buchanan, 1999; Burgess et al., 2008). Therefore it is generally accepted that membrane proteins, like soluble proteins, are at a free energy minimum. The driving forces behind membrane protein folding, though, are still likely to be different from those for soluble proteins because they reside in the fundamentally distinct bilayer milieu.

With the establishment that protein structure is defined by amino acid sequence (and environment in the case of membrane proteins), the next logical question is how a protein reaches its native fold. In 1968, Levinthal postulated the existence of defined folding pathways based on the observation that a protein sampling every possible conformation available to the polypeptide backbone would take longer than the age of the
universe to find its native structure, when in reality most proteins fold on a sub-second timescale (Levinthal, 1968). Therefore, he suggested that local interactions form rapidly within a folding protein, which then serve as nucleation points for further folding. In the years since, enormous progress has been made in understanding the process of protein folding, although many questions remain (Sosnick & Barrick, 2011). In the last thirty years the concept of the funnel-shaped folding energy landscape has emerged, which proposes that the energetically downhill nature of the conformational search drives protein folding. The funnel interpretation conveys that a variety of pathways lead to the ultimate lowest-energy native state, and the protein can pass through different transition states and intermediates on these parallel pathways (Onuchic & Wolynes, 2004). The kinetic heterogeneity often observed in protein folding kinetics (i.e., multiple exponential phases with distinct rate constants are required for fitting) has been interpreted as evidence of parallel pathways, as such behavior can be attributed to different fractions of the protein population folding via different intermediates at different rates. However, there is also substantial support for the idea that proteins follow a dominant folding pathway that proceeds through sequential incorporation of native structural elements termed “foldons” (Hu et al., 2013). This behavior is driven by the cooperativity of foldon units and the principle of sequential stabilization, which orders the pathway based on existing structure guiding the formation of the next structural unit in each step. Multi-exponential kinetics can be explained within the framework of this model by the existence of off-pathway misfolded states along the productive folding pathway. Termed the “predetermined pathway with optional errors,” or PPOE model, slower kinetic phases arise from the probabilistic population of these misfolded states, which then must be
corrected and resume the folding process via the same folding pathway (Bédard et al., 2008; Krishna & Englander, 2007).

Although great progress has been made in understanding protein folding in the last 50 years, the overwhelming majority of the data and conclusions are based on the study of soluble proteins. Membrane proteins reside in a hydrophobic lipid bilayer, and therefore could be subject to different motivating factors when attaining their native structures. Studies of how integral membrane proteins fold and insert into the bilayer have lagged far behind those for soluble proteins due to two main challenges: the difficulty of denaturing membrane proteins and the need for a membrane-mimetic environment to obtain the native state (Stanley & Fleming, 2008).

To study a protein’s transition between the unfolded and folded states, it is necessary to populate the unfolded state, which is often achieved by denaturing the native protein. However, it has been found that the conventional denaturation methods used in the soluble protein field, such as chemical denaturants and temperature, often do not completely denature membrane proteins (Haltia & Freire, 1995). For example, neither 8 M urea nor 6 M GdnHCl (guanidine hydrochloride) perturb the tertiary structure of folded BR (G. Q. Chen & Gouaux, 1999), and the addition of the detergent SDS (sodium dodecyl sulfate) or heating BR result in a loss of tertiary structure but substantial retention of helical content (Brouillette et al., 1987; London & Khorana, 1982). A further complication is that α-helical TM proteins are frequently insoluble in chemical denaturants in the absence of detergents or lipids, due to the continuous stretches of hydrophobic amino acids that compose TM helices (Stanley & Fleming, 2008). β-barrel TM proteins offer a unique advantage over α-helical TM proteins in this respect as they
are generally both soluble and fully denatured by chemical denaturants due to their alternating hydrophobic/hydrophilic pattern of amino acids. For this reason, much recent progress has been made in exploring membrane protein folding by utilizing OMPs as folding subjects (discussed further below).

Another major challenge to membrane protein folding studies is the requirement for a bilayer-like environment into which the protein can fold. The most commonly used membrane-mimetics in the field are detergent micelles and lipid vesicles. Although micelles are useful for efficiently reconstituting membrane proteins in a native form, they are geometrically and chemically different from biological membranes and therefore are not ideal structures in which to study the folding process. As bilayer structures, lipid vesicles are better representatives of the membrane; however, these entities introduce other experimental challenges, such as light scattering that interferes with spectroscopic measurements, and tumbling that is too slow for conventional NMR measurements. Still, advancing experimental techniques are allowing such challenges to be overcome and the field to progress.

Despite the technical difficulties, progress has been made in elucidating the folding pathways of integral membrane proteins. In 1990, Popot and Engelman proposed the two-stage model to describe the assembly of $\alpha$-helical membrane proteins (Popot & Engelman, 1990). This model postulates that $\alpha$-helical TM protein folding can be separated into two energetically distinct stages: the insertion of individually stable $\alpha$-helices into the bilayer, and the lateral association of these helices to form the tertiary structure of helical bundles. Independent $\alpha$-helices are stabilized by hydrophobic interactions between the lipid acyl chains and nonpolar amino acid side chains, and intra-
helix H-bonds strengthened by the low dielectric environment of the membrane. Helix-helix interactions are driven by side chain packing, electrostatics, loops between helices, and ligand binding (Popot & Engelman, 2000). This model has since been extended to include further stages of membrane protein maturation, such as prosthetic group incorporation and higher order oligomerization (Engelman et al., 2003). With the establishment that helix association is a separate process from helix formation, much work has gone into examining the second stage of helix-helix interactions, and studies of oligomeric single-pass TM proteins such as glycophorin A have identified amino acid sequences and lipid determinants that drive lateral interactions of α-helices (Mackenzie, 2006). It should be noted that the denaturants mentioned above as not fully unfolding α-helical TM proteins are useful in accessing the helix association stage of membrane protein assembly, and these agents have been utilized, for example, in studies of glycophorin A and diacylglycerol kinase (DAGK)(Lemmon et al., 1992; Nagy et al., 2001).

Studying the initial helix folding/insertion step of the two-stage model is more challenging, for the reasons discussed above. Furthermore, from a biological perspective, the mechanism by which a polypeptide independently transfers from an aqueous solution to across the bilayer doesn’t seem relevant as it is now known that this process is facilitated in the cell by the translocon apparatus. However, there are cases of α-helical TM proteins spontaneously inserting across membranes without the use of the translocon, such as colicins (Cramer et al., 1995) and diphtheria toxin (Zhan et al., 1995). In addition, framing the folding process in reference to an unfolded aqueous state aids in thermodynamic analysis and determination of the energetic driving forces behind
membrane protein folding. For this reason, White and Wimley formulated a four-step thermodynamic cycle to describe \(\alpha\)-helical membrane protein folding, starting with the aqueous unfolded state partitioning into the interfacial region of the bilayer, then folding into an \(\alpha\)-helix, inserting across the membrane, and ultimately associating with other helices (Figure 1.3)(White & Wimley, 1999). The cyclic aspect of the model arises from the theoretical “water path” the protein could also follow, where folding and association take place in the aqueous environment instead of the membrane. Such states are important for thermodynamic calculations, but are highly disfavored in reality and are therefore omitted from Figure 1.3 for clarity. Numerous groups have examined the processes of partitioning, folding, and helix insertion across the membrane, often utilizing short model peptides for these studies, and much information has been gained regarding sequence and bilayer effects and the energetics associated with these steps (Mackenzie, 2006; White & Wimley, 1999).

Although extremely useful for delineating thermodynamic states and facilitating calculations of energetic parameters involved in membrane protein folding, White and Wimley’s scheme is not a kinetic model for \(\alpha\)-helical TM protein assembly. There could be additional intermediate states populated along the folding pathway that are not included in the thermodynamic pathway. In addition, consideration must be given to energetic barriers and transition states that exist along the pathway as key components of understanding the folding process, which can only be accessed through kinetics studies. Kinetics measurements of \(\alpha\)-helical membrane protein folding have been carried out for a few proteins, such as BR and DAGK; these proteins have been found to follow multi-step pathways and have folding rates strongly influenced by bilayer properties (Allen et al.,
2004; Nagy et al., 2001). Although these studies have provided important first steps into elucidating the folding pathways of α-helical TM proteins, many questions still remain, and interpreting results is continually complicated by misfolding and aggregation of the denatured state.

β-barrel TM proteins are more amenable to in vitro folding studies due to their decreased aggregation propensity and ability to be completely denatured by high concentrations of chemical denaturants. For this reason, a number of studies have been conducted with these proteins, most commonly utilizing bacterial OMPs as folding subjects, and the general features of the intrinsic OMP folding pathway have been described. However, there are still caveats associated with such experiments, including a tendency of OMPs to aggregate in low denaturant concentrations, and unanswered questions about the folding process remain. The following sections review the past progress in elucidating OMP folding pathways and specify the unknown features that this work aims to address.

Studies of OmpA have revealed β-barrel folding intermediates and a concerted mechanism of membrane insertion

The most extensively studied OMP has been OmpA, due to its high natural abundance in E. coli. OmpA was originally identified as one of the major transmembrane proteins of the bacterial outer membrane by several groups and correspondingly has been known by several other names: II* (Garten et al., 1975), O-10 (Nakamura & Mizushima, 1976), 3a (Schnaitman, 1974), B (Reithmeier & Bragg, 1974), d (Van Alphen et al., 1977), and tolG protein (Chai & Foulds, 1974). It was observed early on that OmpA acts
as a receptor for several bacteriophages and colicins (Chai & Foulds, 1974; Datta et al., 1977; Morona et al., 1984) and that it serves as a mediator in F-dependent conjugation (Van Alphen et al., 1977), thus demonstrating its surface-exposure and location in the OM. OmpA has also been implicated in bacterial pathogenesis, with functions relating to adhesion, invasion, and evasion of host defenses (Confer & Ayalew, 2013). The primary physiological role of OmpA is thought to be helping to maintain the structural integrity of the bacterial cell envelope by anchoring the peptidoglycan layer to the OM through interactions with its globular periplasmic domain. This function is supported by the observation that cells lacking OmpA and Lpp (a major lipoprotein that also interacts with peptidoglycan) have substantial defects in integrity and morphology, including “blebbing” of the OM and detachment of peptidoglycan from the OM (Sonntag et al., 1978). Sequence analysis has shown that the periplasmic domain of OmpA contains a peptidoglycan-associating motif (De Mot & Vanderleyden, 1994; Koebnik, 1995), and a recent study used crystallography and NMR to verify this interaction and localize the binding site to two conserved residues in the OmpA periplasmic domain and diaminopimelate in peptidoglycan (Park et al., 2012).

OmpA is initially synthesized with a 21-residue N-terminal signal sequence that is cleaved off upon passage through the IM (Beck & Bremer, 1980; Movva et al., 1980). Targeting of OmpA to the translocon has been shown to occur post-translationally via the chaperone SecB (Baars et al., 2006; Lecker et al., 1990). The mature OmpA protein has 325 residues (R. Chen et al., 1980), with the first 171 residues comprising the N-terminal TM β-barrel domain and the next 154 residues making up the C-terminal periplasmic domain. The two-domain structure was first proposed based on two key observations:
digestion of native OmpA with trypsin or pronase results in a portion of the protein remaining protected by the membrane and active as a phage receptor (Schweizer et al., 1978), and gene truncation experiments revealed that only the first ~180 residues are required for incorporation into the OM and phage receptor activity (Bremer et al., 1982). The membrane-embedded domain was predicted to span the bilayer eight times in an anti-parallel β-sheet conformation based on the locations of mutations in isolated phage-resistant protein variants (Morona et al., 1984) and circular dichroism (CD) and Raman spectroscopy indicating a high content of β-sheet structure in the N-terminal domain of the native protein (Nakamura & Mizushima, 1976; Vogel & Jähnig, 1986). The predicted topology and secondary structure were proven to be quite accurate with the determination of the N-terminal domain structure by crystallography (Pautsch & Schulz, 1998, 2000) and NMR (Arora et al., 2001; Cierpicki et al., 2006). Shown in Figure 1.4A, E. coli OmpA_{171} (residues 1-171 of the mature protein) was found to form an eight-stranded β-barrel with long extracellular loops and short periplasmic turns (three of the four loops had unresolved portions in the high-resolution crystal structure, indicating they have high mobility; these residues are colored tan in Figure 1.4A). The crystal structure of the periplasmic domain for the E. coli OmpA has not been determined, but structures of homologous proteins from the “OmpA-like” family have been solved (Grizot & Buchanan, 2004; Park et al., 2012) and contain mixed α/β secondary structure, in agreement with CD measurements (Sugawara et al., 1996).

There is some evidence that OmpA can function as a non-specific pore, as it has been shown that a minor subset of reconstituted OmpA allows diffusion of small solutes into proteoliposomes (Sugawara & Nikaido, 1992, 1994) and OmpA incorporated into
planar lipid bilayers exhibits ion conductance (Saint et al., 1993). The crystal structure of the OmpA β-barrel does not contain an internal continuous passage for water or solutes (Pautsch & Schulz, 1998), but molecular dynamics and mutagenesis studies have suggested that the switching of a central internal salt bridge serves as a gating mechanism and can create an aqueous passage through the protein (Bond et al., 2002; Hong et al., 2006). It has also been suggested that OmpA can fold into a larger pore conformation based on the observation that reconstituted OmpA can form two temperature-sensitive conductance states in planar bilayers (Arora et al., 2000; Zakharian & Reusch, 2003). The lower conductance state has been attributed to the N-terminal barrel domain of the two-domain structure because it can occur in the absence of the C-terminal periplasmic domain. However, the higher conductance state is only formed when the periplasmic domain is present, so it has been suggested that this state corresponds to a larger pore formed by both the N-terminal and C-terminal parts of the protein in a 16-stranded β-barrel, the topology of which has been proposed previously (Reusch, 2012; Stathopoulos, 1996). There is also evidence that the major porin of Pseudomonas aeruginosa, OprF, a homolog of OmpA, can form a one-domain channel conformation in addition to a two-domain conformer akin to the two-domain OmpA structure (Sugawara et al., 2012). Still, the OmpA one-domain state has not been isolated and characterized, and the physiological significance of this conformation is unknown, so it will be discussed no further here.

OmpA, along with several other OMPs, was observed early on to be “heat-modifiable” when undergoing polyacrylamide gel electrophoresis in the presence of SDS (SDS-PAGE), and this property has been exploited as a powerful method of monitoring
OmpA conformation. Specifically, it was found that OmpA isolated from the OM migrated to a larger apparent molecular weight during SDS-PAGE (~33-35 kDa depending on the exact electrophoresis conditions) when it was heated to 100 °C in SDS before electrophoresis compared to the apparent molecular weight when it was not heated (~28-30 kDa) (Garten et al., 1975; Reithmeier & Bragg, 1974). Further studies proposed that the heat-modifiability was due to a conformational change brought about by boiling in SDS, and the smaller (non-heated) and larger (heated) apparent molecular weight species corresponded to the folded and denatured forms of the protein, respectively (Heller, 1978; Nakamura & Mizushima, 1976). The identities of the differently migrating bands of OmpA as well as many other heat-modifiable OMPs have since been verified by numerous experimental observations: the non-heated, faster-migrating form retains the same amount of β-sheet structure as in the absence of SDS, while the heated, slower-migrating form is converted to α-helical structure, similar to soluble proteins denatured in SDS (Nakamura & Mizushima, 1976); the slower form migrates to a position consistent with the actual molecular weight of the protein (35.3 kDa in the case of OmpA), indicating behavior when undergoing SDS-PAGE similar to soluble proteins (on which the molecular weight migration positions are based), as would be expected more of the denatured state than the folded state; the faster-migrating form exhibits a protected fragment after trypsin digestion, corresponding to the membrane-embedded barrel domain, while the slower-migrating form is completely digested (Schweizer et al., 1978); and the faster-migrating form maintains phage receptor activity, while the slower-migrating form does not (Schweizer et al., 1978).
The reason for OmpA’s resistance to denaturation in SDS has been attributed to a large kinetic barrier to unfolding in SDS (with a half-life of ~3 years) due to the high stability of the β-barrel structure, which is overcome by heating (Ohnishi & Kameyama, 2001). It is not fully understood why the folded form migrates faster than the unfolded form in SDS-PAGE, but it is probably due to the combination of a difference in the amount of SDS detergent bound by the two states, and a difference in shape, with the folded conformation being more compact than the unfolded state (Ohnishi et al., 1998; Reithmeier & Bragg, 1977). Some α-helical membrane proteins also exhibit aberrant migration by SDS-PAGE and this behavior has been attributed to unusually high or low amounts of detergent binding (Rath et al., 2009). Interestingly, some OMPs exhibit an opposite gel-shift behavior, with the folded state migrating slower than the unfolded state. This occurs for the smaller, eight-stranded β-barrel OMPs including OmpX, PagP, and OmpA171 (the barrel-only form of OmpA)(Burgess et al., 2008; Huysmans et al., 2007; Ried et al., 1994).

Early studies of the folding and assembly of OmpA investigated the protein’s in vivo path to the OM using SDS-PAGE, protease protection, and phage sensitivity to monitor formation of the native state. Initial work was aimed at determining the features of the amino acid sequence necessary for proper targeting and insertion. It was found that the eighth β-strand of the N-terminal barrel (putative at that point but later verified by the crystal structure) is important for proper assembly into the OM, with deletion or mutation of the strand preventing this from occurring (Freudl et al., 1985; Klose, MacIntyre, et al., 1988; Klose, Schwarz, et al., 1988). However, membrane incorporation could be restored when the eighth β-strand was substituted with at least 9 residues of amphipathic or
hydrophobic sequence, indicating that the physicochemical properties of the strand are more important than the exact sequence for OmpA assembly (Klose et al., 1989). Mutational studies of the other β-strands also determined that four of the five lipid-facing residues need to be hydrophobic and charged residues and prolines are generally not tolerated in the transmembrane portions of strands (Koebnik, 1999a). The extramembranous loops and turns were found to be non-essential for outer membrane incorporation, although mutations of periplasmic turns resulted in reduced efficiency and kinetics of OmpA assembly (Koebnik & Krämer, 1995). Shortening of all four extracellular loops had no effect on OM incorporation (Koebnik, 1999b).

The lipid constituent of the outer leaflet of the OM, lipopolysaccharide, was initially thought to be an essential component for OmpA assembly into the membrane because the addition of LPS to heat-denatured OmpA or an unfolded intermediate form resulted in formation of the faster-migrating, protease-protected form (Freudl et al., 1986; Schweizer et al., 1978). However, it was subsequently shown that addition of the detergent octylglucoside could bring about the same conformational change, indicating that LPS is not essential for OmpA folding. Additionally, the protein could then be reconstituted into synthetic membranes and maintain all of the native characteristics, showing that LPS is also not necessary for proper membrane insertion (Dornmair et al., 1990). It has further been demonstrated that OmpA can fold to the native state in a variety of detergent types, as long as the concentration of detergent is above the critical micelle concentration (Kleinschmidt, Wiener, et al., 1999). These results indicate that OmpA requires a supramolecular assembly of amphipathic molecules to provide a hydrophobic environment for folding, but that the specific OM environment is not
essential. This conclusion has allowed a multitude of experiments to be carried out on OmpA in vitro using membrane-mimetic environments, including structural work and folding studies (discussed below).

These experiments also demonstrated that OmpA does not achieve its conformational change until it interacts with the membrane. In other words, OmpA does not become faster-migrating and protease-protected while still in the periplasm, a notion supported by the observation of an intermediate form in vivo, termed imp-OmpA for “immature processed.” This intermediate was found to accumulate in the periplasm of E. coli cells during overexpression of OmpA. It was processed, meaning the signal sequence had been cleaved off upon translocation through the inner membrane, but was immature because it did not exhibit a gel-shift nor was it protected from protease digestion. Pulse-chase labeling demonstrated that the intermediate was subsequently converted to mature OmpA, being incorporated into the OM and demonstrating faster migration by SDS-PAGE, protease protection, and phage receptor activity (Freudl et al., 1986).

All of these experiments provided clues to the biological pathway and assembly of OmpA, but in vivo experiments were not amenable to more detailed structural studies of the conformational change associated with membrane incorporation. For this reason, a series of in vitro studies were conducted to probe the secondary and tertiary structure formation of OmpA and the existence of intermediate states.

Surrey and Jähnig’s in vitro folding studies of OmpA were the first to investigate the folding and insertion of a β-barrel membrane protein directly into a membrane from an aqueous denaturant-unfolded state (Surrey & Jähnig, 1992). It had previously been demonstrated that OmpA and the porin OmpF could be refolded into detergent micelles
and then transferred to bilayers (Dornmair et al., 1990; Eisele & Rosenbusch, 1990), but they sought to study the process of folding directly into a membrane as it more closely mimics the situation in vivo, and is therefore more physiologically relevant. Surrey and Jähnig demonstrated that upon dilution of denaturant, OmpA spontaneously folds into small unilamellar vesicles (SUVs) composed of the phospholipid diC14PC (DMPC) with a conformation indistinguishable from the natively folded protein (based on demonstration of the SDS-PAGE gel-shift, trypsin protection, phage sensitivity, and CD and fluorescence spectroscopy). In addition, they discovered that OmpA folds into membranes in vitro unidirectionally, with the same final orientation as the native state (i.e., the extracellular side translocating across the bilayer to the inside of the vesicle and the periplasmic side remaining on the outside of the vesicle). This conclusion was based on the observation that all of the folded protein was accessible to cleavage of the periplasmic domain by trypsin (leaving the membrane-protected barrel domain); if some of the protein had folded with the opposite orientation, then the periplasmic domain would have been located on the inside of the vesicles and would not have been removed by trypsin. This is exactly what was observed when OmpA was first refolded in detergent micelles before being reconstituted into bilayers: the periplasmic domain was removed for only 50% of the protein population, indicating that the protein was inserted in the membrane in both orientations. Unidirectional folding was also observed when only the N-terminal domain was folded into SUVs (a small piece of the C-terminus was removed by Glu-C endoproteinase for all proteins, again demonstrating a single orientation), indicating that the unique orientation was not a result of the periplasmic domain’s
presence. Therefore, it was concluded that inserting in the native orientation must be an inherent part of the intrinsic folding pathway of OmpA (Surrey & Jähnig, 1992).

Surrey and Jähnig also observed a membrane-adsorbed intermediate when folding OmpA into SUVs of diC\textsubscript{14}PC below the main lipid phase transition temperature. Under this condition the lipid bilayer is in the highly ordered gel phase, and it appeared that the protein was unable to insert into the membrane based on the lack of a gel-shift and complete digestion by trypsin. However, the protein exhibited similar CD and fluorescence spectra to the folded/inserted state, indicating the same degrees of β-sheet content and hydrophobic environment for the tryptophan (Trp) residues (Surrey & Jähnig, 1992). This partially folded, membrane-adsorbed state was further investigated by Rodionova et al. using Fourier transform infrared spectroscopy, and it was verified that the membrane-adsorbed state had a similar amount of β-sheet structure as the folded state. Interestingly, fluorescence quenching experiments utilizing phospholipids containing bromine at different membrane depths indicated that the Trps in the adsorbed state were more buried in the bilayer than those in the folded conformation, which appeared to be near the surface of the membrane (Rodionova et al., 1995). It should be noted that OmpA contains five Trp residues, all of which are located in the N-terminal β-barrel domain (indicated in red in Figure 1.4A), so the behavior of the periplasmic domain is invisible to experiments utilizing tryptophan fluorescence. In addition, the measured fluorescence is an average of the fluorescence due to each Trp residue. It can be seen from the crystal structure of OmpA\textsubscript{171} in Figure 1.4A that all five Trps are located in the interfacial regions of the bilayer, so the quenching experiments that showed the Trps were located near the membrane surface in the folded state were consistent with the
later determined structure. The finding that the Trps in the “membrane-adsorbed” state were, on average, more buried than the native state Trps indicates that this intermediate state is actually partially inserted into the membrane, although not to the degree where it would be protected from proteolysis.

In a subsequent study, Surrey and Jähnig investigated the kinetics of OmpA folding into diC_{14}PC SUVs, which was the first study of the kinetics of refolding of a transmembrane protein directly into a bilayer (Surrey & Jähnig, 1995). Based on CD spectroscopy, the authors suggested that upon dilution of OmpA from high denaturant into water, the protein immediately (<1 s) collapses to a partially folded state, which aggregates on a slower timescale. They demonstrated that aggregation competes with folding and that it can be suppressed by raising the pH to 10 or by increasing the lipid to protein ratio, findings that proved to be useful for designing future folding studies. Folding into SUVs appeared to follow similar biphasic kinetics by CD and fluorescence, indicating the two techniques were monitoring the same conformational change. Based on these data, Surrey and Jähnig proposed a simple kinetic scheme for OmpA folding consisting of an initial collapse to a partially folded state in water, U_{AQ}, which can aggregate, followed by slower conversion to a partially folded and inserted state upon interaction with membranes, I_{M} (on a timescale of minutes), and even slower conversion to the folded, fully inserted state, F (on a timescale of hours)(Figure 1.5A)(Surrey & Jähnig, 1995). This model was consistent with the previous observation of a “membrane-adsorbed” intermediate state when the membranes were in the gel phase, because this intermediate could correspond to the partially folded and inserted state that precedes the native state in the scheme.
Although this scheme seems reasonable based on the kinetics data, one problem with the model is the assumption that OmpA collapses to a partially folded state in water. A factor not addressed is that half of the protein’s mass is contained in the soluble periplasmic domain, which presumably folds outside of the membrane. It is reasonable to assume that the periplasmic domain would be unfolded by high denaturant and then refold upon dilution into water, and this structure would certainly contribute to the CD spectrum obtained under these conditions. Therefore it cannot be concluded that the membrane domain partially folds in water (although the possibility cannot be excluded either) because the periplasmic domain folding likely contributes to the observation of secondary structure formation in the CD spectrum.

Kleinschmidt and Tamm built upon Surrey and Jähnig’s model by examining the folding kinetics of OmpA into SUVs of diC_{18:1}PC (DOPC) by SDS-PAGE and fluorescence spectroscopy at different temperatures (Kleinschmidt & Tamm, 1996). They concluded that there are actually two membrane-associated intermediate states that form before insertion and folding to the native barrel, based on the observation of three kinetic phases in the fluorescence kinetics at low temperatures. No formation of the faster-migrating, trypsin-protected species was observed by SDS-PAGE at the lowest temperature over the course of 8 h, but during this time the fluorescence spectrum became more intense and blue-shifted, indicating some degree of folding and membrane association. The authors further verified that the intermediate states were on-pathway by conducting temperature-jump folding experiments and demonstrating that the partially folded conformations populated at low temperatures readily converted to the inserted, fully folded conformation when the temperature was raised. Kleinschmidt and Tamm
therefore extended the folding scheme of Surrey and Jähnig to include a second membrane-associated intermediate state, $I_{M2}$, that forms on a timescale of minutes to hours and converts to the native state over the course of hours (Figure 1.5B) (Kleinschmidt & Tamm, 1996).

Kleinschmidt and colleagues next investigated the process of barrel folding and insertion across the membrane by utilizing a technique called “time-resolved distance determination by fluorescence quenching” to monitor the behavior of the tryptophan residues as they crossed the bilayer (Kleinschmidt, den Blaauwen, et al., 1999; Kleinschmidt & Tamm, 1999). To facilitate this, the authors constructed five variants of OmpA, each with all but one of the Trp residues replaced by a phenylalanine, and thereby created single-Trp mutants. By measuring the fluorescence kinetics of each variant folding into diC$_{18:1}$PC SUVs containing the quencher bromine at different bilayer depths, they determined the time-resolved location of each Trp in the bilayer as the protein folded. Their analysis led to two major conclusions: OmpA folds via three membrane-associated intermediate states with different degrees of membrane insertion (thus extending their previous model; see Figure 1.5C), and the insertion of each β-hairpin across the bilayer to form the β-barrel occurs simultaneously, thus indicating a highly concerted folding mechanism (Kleinschmidt, den Blaauwen, et al., 1999).

The membrane-associated intermediates were distinguished based on the locations of the Trp residues in the bilayer; in $I_{M1}$ all five Trps appeared to be the farthest from the bilayer center (14-16 Å), in $I_{M2}$ the Trps had all penetrated slightly into the bilayer (~10 Å from the center), in $I_{M3}$ all but one of the Trps had penetrated approximately to the center of the bilayer, and in the native state the Trps were again located ~10 Å from the
bilayer center (presumably all but one on the opposite bilayer edge from where they inserted). The authors’ illustration of the intermediate states indicating the degrees of Trp penetration into the membrane is reproduced in Figure 1.5C. With its Trps the farthest from the bilayer center, intermediate IM1 was proposed to be a relatively disordered state adsorbed to the membrane surface. Intermediate IM2 was predicted to have formed some β-strands oriented parallel to the membrane surface and be partially inserted to the interfacial region of the bilayer, so the authors suggested this state be thought of as a “molten disk” (Kleinschmidt, den Blaauwen, et al., 1999). The intermediate IM3 was found to be more inserted into the bilayer (with the Trps penetrating to the center of the membrane), presumably with more β-structure, and hence the authors termed it a “molten globule,” in parallel to the partially folded intermediate state proposed for soluble proteins (Ohgushi & Wada, 1983). This state was also consistent with the partially folded and inserted intermediate characterized by Rodionova et al. previously (Rodionova et al., 1995), because the Trps in this conformation were more deeply buried in the membrane than the Trps in the native state. The tryptophan that did not penetrate to the bilayer center was Trp7, which is the only Trp located on the periplasmic side of the membrane in the crystal structure (Figure 1.4A). Therefore, the observation that Trp7 never crossed the bilayer is highly consistent with the native structure of OmpA. The locations of all five Trps at ~10 Å from the bilayer center in the final folded state is also in good agreement with the location of the Trps in the crystal structure (Pautsch & Schulz, 2000).

The conclusion that the β-barrel of OmpA is formed by a concerted mechanism was based on the fact that all four extracellularly located Trps crossed the bilayer with essentially the same kinetics. As each of these Trps is located on a separate β-hairpin, it
was concluded that each hairpin crosses the bilayer simultaneously. This concerted folding mechanism was intriguing because it contrasted with the previously proposed two-stage model for α-helical membrane proteins (Popot & Engelman, 1990). The two-stage model suggests that α-helical bundle membrane proteins are inserted into the membrane first as individual α-helices, and then associate laterally to form the native structure. However, a concerted mechanism seems more plausible for β-barrel proteins because β-hairpins are unlikely to be stable individually in a bilayer as hydrophobic α-helices are; individual β-hairpins would have unsatisfied H-bonds on both edges and hydrophilic side chains extending into the bilayer on one side. By inserting simultaneously, and forming the barrel as the protein crosses the membrane, all of the H-bonding partners in the peptide backbone would be satisfied within the hydrophobic bilayer environment, and the polar side chains would be sequestered within the barrel lumen, away from the nonpolar lipids. Interestingly, the proposed OmpA folding model does have similarities to the thermodynamic model of White and Wimley, which suggests α-helices form in the interfacial region of the bilayer, parallel to the membrane surface, before inserting across the membrane (White & Wimley, 1999). OmpA likewise appears to form at least some β-strand secondary structure while bound to the membrane surface or partially inserted, which then fully crosses the bilayer to form the native barrel.

Kleinschmidt and colleagues subsequently further investigated OmpA β-barrel formation by using site-directed fluorescence quenching to monitor specific β-strand associations within the barrel (Kleinschmidt et al., 2011). The authors employed some of the single-Trp mutants used in the previous study to create double mutants with residues in the strands adjacent to the Trps mutated to cysteines and labeled with nitroxyl spin
labels. The spin label quenches tryptophan fluorescence when the two are brought into close proximity, so by monitoring the fluorescence quenching over the course of OmpA folding for different Trp-Cys pairs, the authors determined the rates of association for the “trans” and “cis” parts of β-strands 1, 2, 3 and 8. Trans refers to the end of the barrel that is on the extracellular side of the membrane in the final structure (and therefore translocates the membrane) and cis refers to the end of the barrel that remains on the periplasmic side. No quenching was observed when the protein was unfolded in urea, and only a minimal amount of quenching upon dilution to aqueous buffer, indicating no strand association under these conditions. The presence of membranes (diC18:1PC SUVs) was required for more substantial quenching, with maximal quenching occurring at higher temperatures. At lower temperatures, where the early membrane-associated intermediate states were presumably trapped, low to moderate levels of quenching indicated some degree of adjacent β-strands coming into proximity of each other. Interestingly, the trans pairs appeared to be closer together than the cis pairs at this stage of folding. In addition, the residue pairs between strands 1 and 8 (cis and trans) exhibited more quenching than under aqueous conditions, showing that the first and last β-strands of the barrel are in closer proximity in the membrane-associated stage than in the aqueous unfolded state. At higher temperatures, the maximal levels of quenching were observed, indicating the closest associations of each pair of strands, and consistent with the formation of the native inserted β-barrel. It was found that the trans pairs of residues associated more closely before the cis pairs, and that the trans pairs between strands 1, 2 and 3 came together slightly before the trans pair for strands 1 and 8. The cis pair between strands 1 and 2 was the slowest to associate (of the pairs examined, not
necessarily of all the strands). Taken together, these results indicate that the trans ends of the \( \beta \)-strands associate and insert across the membrane first, with barrel closure occurring in concert with insertion, and the cis ends of the strands associate last (Kleinschmidt et al., 2011). This model is consistent with the previous concerted mechanism of barrel insertion and allowed the authors to refine their interpretation of the structures of folding intermediates \( I_{M2} \) and \( I_{M3} \), as shown in Figure 1.5D.

The Kim group recently conducted complementary studies on OmpA barrel formation using Förster resonance energy transfer (FRET) instead of quenching (Kang et al., 2012). Specific residues in single-Trp variants were mutated to cysteines and labeled with the fluorophore dns, which acts as an acceptor of tryptophan fluorescence. By monitoring the FRET efficiency during OmpA folding (into diC\(_{14}\)PC SUVs) for different Trp-Cys(dns) pairs, the authors monitored the evolution of distances across the barrel pore and across the bilayer. Their results showed that the barrel pore forms during the initial stage of insertion, followed by insertion across the bilayer, which is consistent with the concerted mechanism proposed by Kleinschmidt and colleagues. In addition, they observed a much slower change in FRET signal, which was attributed to protein relaxation in the bilayer, such as pore expansion or changes in local solvation (Kang et al., 2012).

The studies of OmpA discussed thus far have all utilized small unilamellar vesicles\(^\dagger\) for folding because it was observed that OmpA did not fold into large unilamellar vesicles (LUVs) of diC\(_{14}\)PC or diC\(_{18:1}\)PC under the conditions used. For this

\(^\dagger\) Small unilamellar vesicles (SUVs) are generally defined as 30-50 nm in diameter and are formed by ultrasonication of lipids. In contrast, large unilamellar vesicles (LUVs) are formed by extrusion through a polycarbonate membrane, the pore size of which defines the vesicle diameter. The experiments discussed here all utilize 100 nm diameter LUVs.
reason it was proposed that the high curvature of SUVs helps facilitate OmpA insertion and folding, due to the strained nature of the membranes inducing a higher prevalence of membrane defects compared to less curved bilayers (Surrey & Jähnig, 1992; Tamm et al., 2004). However, it was demonstrated by Kleinschmidt and Tamm that OmpA can fold into LUVs composed of the phospholipids diC_{10}PC, diC_{11}PC, or diC_{12}PC (DLPC), and that the kinetics of folding are dependent on the lipid acyl chain length; the fastest folding was observed in the lipids with the shortest chain length and therefore composing the thinnest bilayers (Kleinschmidt & Tamm, 2002a). It has been shown experimentally that thinner bilayers are more permeable to solutes (Blokh et al., 1975; Paula et al., 1996) and simulations have demonstrated that thinner bilayers have a much higher incidence of spontaneous pore formation than thicker bilayers (Bennett et al., 2014). Therefore it seems probable that OmpA utilizes these bilayer pores/defects to initiate insertion into the membrane, thus causing the fastest folding to occur in the thinnest, most defect-prone bilayers. This conclusion is also consistent with the accelerated folding observed in SUVs of longer chain lipids because the high curvature induces the defects that facilitate protein insertion.

It should be noted that acyl chain lengths of 10, 11, and 12 generally do not occur in biological membranes in nature, but the measured hydrophobic thicknesses of these bilayers are well matched to the thicknesses of the hydrophobic surfaces of OMPs (24-27 Å)(Fleming et al., 2012; Lomize et al., 2006). Moreover, the outer membranes of Gram-negative bacteria are thought to be thinner than typical eukaryotic membranes and inner bacterial membranes, due to the outer leaflet being composed of LPS, which contains shorter acyl chains. Recent molecular dynamics simulations further support this
prediction as it was shown that an asymmetric bilayer composed of LPS and *E. coli*-native phospholipids has a hydrophobic thickness of ~25 Å (E. L. Wu *et al.*, 2014). Therefore bilayers composed of lipids with 10-12 carbons in their acyl chains are reasonable representatives of the native membrane environment for OMPs. In addition, the lower curvature of LUV bilayers compared to SUVs more closely mimics the bacterial membrane (Kleinschmidt & Tamm, 2002a), and thus makes LUVs preferable for studying the folding reaction in membranes. For these reasons, most subsequent folding studies of OMPs have utilized LUVs of diC\textsubscript{12}PC or shorter chain lipids.

**Folding studies of other OMPs have revealed further complexities**

The intrinsic folding pathway has been investigated for a handful of other OMPs besides OmpA, and both similarities and dissimilarities have been found. One such OMP studied in some detail has been PagP from *E. coli*. The function of PagP is to transfer a palmitate chain from a phospholipid to lipid A of LPS in the outer leaflet of the OM. This behavior has been implicated in virulence of pathogenic bacteria because it reinforces the hydrocarbon core of the membrane and helps protect the organism from host immune responses (Bishop, 2005). The structure of PagP has been determined by crystallography (Ahn *et al.*, 2004) and NMR (Hwang *et al.*, 2002), and it was found to be an eight-stranded β-barrel with an N-terminal amphipathic α-helix that packs against the barrel, as shown in Figure 1.4B. Interestingly, PagP is predicted to be tilted by ~25 ° relative to the membrane normal, based on the hydrophobic surface of the barrel and the locations of the outward-facing tryptophan residues, which are known to reside in the interfacial regions of the bilayer in all other OMPs (shown in red in Figure 1.4B)(Ahn *et al.*, 2004). This
tilting results in some of the TM β-strands being longer than others, as illustrated in the topology diagram of Figure 1.4B (note that the lower bilayer edge is curved in this diagram to denote that the strands are membrane-embedded). The residues that were unresolved in the crystal structure are colored tan.

Huysmans and colleagues demonstrated that PagP could be refolded in vitro to the native conformation in detergent micelles and lipid bilayers (diC12PC SUVs and LUVs), based on the SDS-PAGE gel-shift, CD and fluorescence spectroscopy, and an activity assay (Huysmans et al., 2007). CD measurements were utilized to monitor the formation of β-sheet secondary structure, as well as tertiary structure due to the presence of an exciton interaction in the folded barrel that gives rise to a peak in the CD spectrum at 232 nm. CD bands in this region have previously been attributed to interacting aromatic residues (Kuwajima et al., 1991; Manning & Woody, 1989), but such signals are usually masked in the presence of the much stronger signal from α-helical secondary structure (Grishina & Woody, 1994; Woody, 1994). However, in β-sheet rich proteins containing neighboring aromatic residues, such as PagP, a so-called exciton couplet can be observed in the CD spectrum. A similar peak at 232 nm has also been seen for the OMP OmpLA when folded into detergent micelles, indicating a similar aromatic interaction in the native barrel (Dekker et al., 1995). In PagP, the signal has been attributed to two specific residues, Tyr26 and Trp66, which are brought into proximity across the barrel pore (Khan et al., 2007). For this reason the exciton interaction only occurs when the barrel is folded, and its signal in the CD spectrum is thus a sensitive indicator of tertiary structure. Huysmans et al. used the CD signal at 232 nm as well as SDS-PAGE to investigate the role of the N-terminal α-helix in the folding of PagP, and they found that the helix is not
essential for folding of the β-barrel, but it does help stabilize the barrel, as evidenced by a reduced resistance to thermal denaturation in mutants lacking the helix or helix-barrel interactions (Huysmans et al., 2007).

In a ground-breaking study, Huysmans and colleagues next interrogated the transition state for membrane insertion of PagP by measuring the folding kinetics and thermodynamic stability of 19 mutants of the protein and performing φ-value analysis (Huysmans et al., 2010). The authors found that the transition state was highly polarized, with the C-terminal half of the β-barrel being more structured than the N-terminal half. Based on these results, a tilted insertion mechanism was proposed for PagP, the authors’ illustration of which is reproduced in Figure 1.6A. It is unclear whether this mechanism is unique to PagP, with its distinctive tilted native conformation, or if other β-barrels also experience a polarized transition state during insertion. There is no indication that the concerted mechanism proposed for OmpA involves a tilted transition state, but work on OmpA has focused on intermediate states and not examined any transition states so it is still an open question. Interestingly, the state preceding the transition state in this study was found to be membrane-adsorbed but lacking regular structure (designated as IM in Figure 1.6A), making it similar to the initial membrane-associated states IM1 and IM2 identified for OmpA (Figure 1.5C and D). No partially folded and inserted intermediate IM3 was observed for PagP, as the kinetics and thermodynamics were well-described by a simple two-state folding reaction. It should be noted though that the kinetic parameters were determined from unfolding measurements, in high denaturant (8.8-10 M urea) because the folding kinetics were multiphasic below 7.8 M urea (Huysmans et al., 2010). Under these conditions, it is possible that IM3 is not populated and therefore was not
observed. Thus it cannot be ruled out that a partially folded intermediate state is not an obligatory intermediate during the folding process of PagP, as it is for OmpA.

Huysmans and colleagues subsequently investigated the multiphasic folding kinetics of PagP into diC_{12}PC LUVs as a function of lipid to protein ratio (LPR) and bilayer composition (Huysmans et al., 2012). They found that at low LPRs, the fluorescence kinetics exhibited a burst phase (attributed to fast membrane adsorption) followed by two exponential phases. Based on interrupted folding experiments (Kiefhaber, 1995) the authors suggested that the multi-exponential behavior was due to parallel folding pathways (Figure 1.6B). Parallel pathways have also been proposed for another OMP, FomA from *Fusobacterium nucleatum*, which exhibited bi-exponential kinetics for the formation of the native state by SDS-PAGE (Pocanschi et al., 2006). In both cases, the parallel pathways were hypothesized to arise from multiple unfolded starting states for the protein. Specifically, at low LPRs (and low denaturant) a population of protein could fail to adsorb to the bilayer surface and instead become trapped in a hydrophobically collapsed state in solution. This state would take longer to undergo the conformational change required for proper membrane adsorption and insertion than the state that immediately associated with the membrane, and thereby follow a slower folding pathway to the native state (Huysmans et al., 2012; Pocanschi et al., 2006). However, rather than being indicative of independent parallel pathways, such a situation is better described by the PPOE (predetermined pathway with optional errors) model proposed for soluble proteins (Krishna & Englander, 2007). In the absence of any evidence that the protein progresses through different intermediate states along different folding pathways, it is likely that the slower folding species follows the same membrane insertion process,
and the slow rate is due to the rate-limiting step of correction of the misfolded state. In addition, multiphasic behavior does not necessarily stem from misfolding of the initial state, but could be due to errors occurring at other points along the pathway and originating from membrane-associated intermediate states.

Huysmans et al. also probed the effect of bilayer composition on the folding of PagP and found that the inclusion of diC\textsubscript{12}PE in the LUVs caused a reduction in the folding kinetics. Fitting of the timecourses revealed that the burst phase was decreased and the amplitude of the slow phase was increased at the expense of the fast phase (Huysmans et al., 2012). This was interpreted as protein being shifted to the slower folding pathway, but in the context of the PPOE model it would indicate an increased population of the off-pathway misfolded state. Slower folding kinetics have also been observed for several other OMPs in the presence of PE lipid head groups (Gessmann et al., 2014; Patel et al., 2009). Due to their chemical structure, PE head groups can engage in H-bonding with neighboring lipids, resulting in denser lipid packing and less fluid membranes than for PC lipids (Boggs, 1987; Suits et al., 2005). The decreased fluidity of the bilayer is most likely the reason that OMP folding is greatly slowed in the presence of PE, thus further emphasizing the important role that the physical properties of the membrane play in OMP folding. Because the inner leaflet of the OM of Gram-negative bacteria is composed of predominantly PE lipids (~88%)(Harwood & Russell, 1984), the reduced folding into these lipids also underscores the important role of chaperones and the BAM complex \textit{in vivo} to facilitate OMP folding.

The folding pathways of OmpA and PagP have been studied in the most detail, but \textit{in vitro} folding has been investigated for a few other OMPs as well. A comparative
study of OMP folding was conducted by Burgess et al. for nine different OMPs from *E. coli*: OmpX, OmpW, OmpA, PagP, OmpT, OmpLA, FadL, BamA (referred to as Omp85), and OmpF (Burgess et al., 2008). Interestingly, this work demonstrated a wide variety of behaviors for the different OMPs with regard to folding kinetics, resistance to thermal denaturation, and the effects of bilayer properties such as acyl chain length, saturation, and curvature on folding. In general, higher folding efficiencies were observed for OMPs in thinner and more highly curved bilayers (i.e., SUVs), in agreement with previous work on OmpA (Kleinschmidt & Tamm, 2002a; Surrey & Jähnig, 1995). However, the specific folding behaviors under different conditions were unique for each OMP. For example, while the folding kinetics were slowed down in LUVs of longer chain lipids for all OMPs, the folding kinetics for a particular bilayer thickness were drastically different between OMPs (Burgess et al., 2008). This observation demonstrates that although they have all evolved to reside in the same biological membrane, the OMPs from *E. coli* exhibit unique folding behavior that must be a result of their different amino acid sequences. How the different folding abilities of OMPs are accommodated *in vivo* is probably related to chaperone function and needs to be explored in more detail. Still, these results do not rule out the possibility that OMPs fold by the same intrinsic pathway, with differing kinetics being the result of different rate constants and different propensities to form misfolded states. It has been shown that OMPs aggregate at low denaturant concentrations in the absence of membranes to differing degrees, so this self-association propensity could affect individual OMPs’ folding behavior differently (Ebie Tan et al., 2010).
Another interesting result of the Burgess work was the appearance of a lag phase in OMP folding kinetics monitored by SDS-PAGE in thicker bilayers. In the presence of diC₁₂PC LUVs, OmpX, OmpA, PagP, OmpLA, and BamA all exhibited a lag in the appearance of the folded state by gel (Burgess et al., 2008). This is strongly indicative of a multi-step mechanism, which is consistent with the multiple intermediate states previously identified for OmpA (Kleinschmidt, den Blaauwen, et al., 1999; Kleinschmidt & Tamm, 1996). Evidently, one or more kinetic steps is slowed down in thicker bilayers, causing a delay in the formation of the native state. It should be noted that the lack of a lag in thinner bilayers does not necessarily mean these steps do not occur in thin membranes, but could indicate certain steps are kinetically silent because they occur too fast to allow substantial population of intermediate states.

The Burgess study clearly demonstrates that OMPs follow the same folding trends when folding into synthetic bilayers in vitro, although the specific kinetic behavior is unique between OMPs. Therefore it is reasonable to propose that all OMPs fold by the same intrinsic mechanism as OmpA, but more detailed investigations are needed to verify the folding pathway for other OMPs.

In contrast to the folding studies described so far, which were all conducted on bulk populations and relied on chemical denaturants to populate the unfolded state, the Müller group has examined the unfolding and folding of single molecules of OMPs using mechanical force. Single-molecule force spectroscopy was utilized to demonstrate the step-wise unfolding behavior of OmpG (Sapra et al., 2009), KpOmpA (the OmpA homolog from Klebsiella pneumonia)(Bosshart et al., 2012), and FhuA (Thoma et al., 2012) as these molecules were pulled from one terminus out of a lipid bilayer. Fitting of
the series of force peaks in force-distance curves indicated that the OMPs typically unfolded in segments corresponding to individual β-hairpins, with multiple β-hairpins sometimes unfolding together as well. This was markedly different from the unfolding behavior of the soluble β-barrel protein GFP (green fluorescent protein), which occurred in one major step. However, molecular dynamics simulations revealed that this difference was due to the soluble protein continuously reorienting itself so that each β-hairpin experienced the least load as it unfolded (Hensen & Müller, 2013). Conversely, the restrictive environment of the lipid bilayer caused the membrane-embedded β-barrels to remain oriented perpendicularly to the bilayer plane and thus require relatively large forces (150-250 pN) to unfold each β-hairpin.

Interestingly, the authors also observed refolding of OmpG and KpOmpA upon relaxation of the tensile force after mechanical unfolding. For OmpG, the protein appeared to re-insert into the membrane β-hairpin by β-hairpin (Damaghi et al., 2011), demonstrating a fundamentally different folding process from the concerted mechanism proposed for OmpA. However, the details of the refolding mechanism for KpOmpA were not reported (Bosshart et al., 2012), so it is difficult to conclude whether step-wise refolding is a general mechanism for OMPs under the single-molecule force spectroscopy conditions used. In addition, FhuA was unable to refold after relaxation, instead populating misfolded states with unfolding force-distance profiles distinctly different from that for natively folded FhuA (Thoma et al., 2012).

The step-wise unfolding in response to mechanical force and subsequent refolding of KpOmpA has interesting biological implications, because OmpA is responsible for anchoring the OM to the peptidoglycan via its periplasmic domain. Cell-cell adhesion
(mediated through the extracellular loops of OmpA and other OMPs) as well as osmotic pressure and membrane deformation undoubtedly subject the peptidoglycan-bound domain of OmpA to substantial tensile forces. It is possible that this stress is relieved by segments of the OmpA TM β-barrel unfolding from the OM, thus extending the polypeptide while still retaining a strong anchor in the membrane (Bosshart et al., 2012). The spontaneous refolding of the protein after removal of the mechanical stress would then allow its continued function.

Questions about the OMP folding pathway remaining to be answered

Great progress has been made in elucidating the folding pathway of TM β-barrel proteins but several key issues still remain. One uncertainty is the conformation of the folding-competent unfolded state formed upon dilution of denaturant (UAQ). Initial work proposed that under aqueous conditions, OmpA immediately undergoes hydrophobic collapse to a partially folded intermediate state (Surrey & Jähnig, 1995). However, there is no experimental evidence for this conformation being compact, such as hydrodynamic shape measurements, nor is there conclusive evidence that this state has secondary structure. The CD measurements were conducted on the full-length OmpA protein, which contains the soluble periplasmic domain, so it is unclear to what extent this globular domain was contributing to the appearance of secondary structure in the CD spectrum. To determine the actual structure content of the unfolded barrel domain, studies need to be conducted on this portion of the protein in the absence of the soluble domain. An additional complicating factor is the tendency of UAQ to self-associate and form non-native aggregates, as was observed in early studies of OmpA (Surrey & Jähnig, 1995) and
more recently for other OMPs (Ebie Tan et al., 2010). It is possible that the self-associated state contains non-native secondary structure, so investigations of the conformation of $U_{AQ}$ also need to include verification of the oligomeric state of the protein.

Furthermore, it has been shown that aggregation of the unfolded state competes with folding, but this reaction has not been extensively investigated nor has it been included in analysis of the apparent folding kinetics. In many folding studies of OMPs, final folding efficiencies less than 100% have been observed, but it is unclear if this is a result of competing irreversible aggregation of $U_{AQ}$ or population of some other misfolded off-pathway state. Even when close to 100% folding is achieved and no aggregated protein is observed at the end of the folding reaction, self-association could still be occurring in a competitive manner and thereby affecting the observed folding rates. For these reasons, the aggregation of $U_{AQ}$ needs to be investigated in more detail: either conditions should be found where self-association is shown to be completely eliminated, or the reaction should be modeled more quantitatively into the OMP folding mechanism.

Another issue not fully resolved is the extent of secondary structure formation during the folding and insertion process. Early work on OmpA indicated that a membrane-adsorbed intermediate (formed in the presence of gel phase bilayers) contained substantial $\beta$-sheet structure (Rodionova et al., 1995; Surrey & Jähnig, 1992). With the extension of the folding mechanism to include three membrane-associated intermediate states, it was assumed that this $\beta$-structured conformation corresponded to the third intermediate state that forms before total insertion and $\beta$-barrel formation ($I_{M3}$)
(Tamm et al., 2004). However, it has not been verified that I_{M3} contains native-like β-
structure or that I_{M1} and I_{M2} do not. Kleinschmidt and Tamm reported that the formation
of secondary structure (measured by CD) and tertiary structure (measured by SDS-
PAGE) were synchronized based on the kinetics having similar apparent rate constants
(Kleinschmidt & Tamm, 2002a). In the context of the concerted insertion mechanism, it
seems reasonable that formation of the native amount of β-sheet content occurs in
conjunction with formation of the inserted β-barrel. But this result does not exclude the
formation of a different amount of β-structure in the intermediate states, and detailed CD
kinetics of the entire folding mechanism have not so far been investigated. In addition,
the experiments on OmpA have all included the periplasmic domain, so it is unclear how
the secondary structure of this domain influences the observed CD kinetics. Therefore,
more detailed investigations of the kinetics of β-strand formation are needed to fully
describe the structural features of the OMP folding pathway.

Lastly, there is some ambiguity over the presence of parallel pathways in the
folding mechanism for OMPs. Although multi-exponential kinetics can indicate
independent reaction pathways, the presence of optional off-pathway misfolded states can
also result in multiple exponential phases in the measured data. More rigorous kinetic
modeling would be useful in differentiating between these possibilities. Monitoring
folding kinetics under the exact same conditions by orthogonal methods, such as SDS-
PAGE and CD, would allow global fitting of the data to specific kinetic schemes and
further investigation of the conformational features of unfolded and intermediate states.
1.5 Overview of Thesis

Chapter 2: The soluble, periplasmic domain of OmpA folds as an independent unit and reduces the self-association propensity of the unfolded OmpA transmembrane β-barrel

To probe the role of the periplasmic domain in OmpA folding and the conformations and interactions of the unfolded TM β-barrel domain, we have studied the two domains in isolation. Using CD, we determined that the periplasmic domain is an independent folding unit with mixed α/β secondary structure and it has a thermodynamic stability of -6.2 (±0.1) kcal mol\(^{-1}\) at 25 ºC. We utilized AUC to investigate oligomerization of the unfolded barrel domain in low denaturant and determined that the periplasmic domain reduces the self-association propensity of the barrel domain when covalently attached, thereby identifying a novel chaperone function for the periplasmic domain that may be important for folding in vivo. In addition, we found that the oligomeric state of the OmpA barrel domain contains β-sheet structure and binds the dye Thioflavin T, indicating an amyloid-like structure. Finally, we investigated the conformation of the monomeric, folding-competent unfolded form of the barrel domain and determined that it contains no regular structure and has an expanded shape. This result refutes the previous assumption that the barrel domain collapses to a compact, partially folded state prior to folding into membranes.
Chapter 3: The OmpA transmembrane β-barrel folds via a multi-step mechanism with a partially inserted intermediate state containing extensive β-sheet structure

To gain a better understanding of the intrinsic β-barrel folding pathway, we investigated the folding of the OmpA TM domain into lipid bilayers of increasing thickness. We monitored the formation of secondary and tertiary structure formation using CD and SDS-PAGE, and globally fit the data to complex kinetic schemes. This revealed the presence of multiple intermediate states, including a penultimate state that is partially inserted and contains a higher β-sheet content than the native state. We also identified several off-pathway intermediate states that give rise to multi-exponential kinetics, including an anomalously migrating species by SDS-PAGE that we have termed the “gone” state. We conclude this chapter by proposing a detailed kinetic mechanism for OmpA β-barrel folding and discussing the structural features of each intermediate, based on our data and previous studies of OMP folding.
1.6 Figures

Figure 1.1 α-helical and β-barrel transmembrane protein structures

(A) An example of an α-helical transmembrane protein, bacteriorhodopsin (BR) from *Halobacterium salinarum* (PDB id 1C3W)(Luecke et al., 1999). Colors indicate separate subunits of the native homotrimer. The membrane is shown in grey. (B) A single α-helix (helix A of BR) shown on the left in cartoon representation with the side chains as sticks, and on the right with just the backbone as sticks and the hydrogen bonds indicated by dashed lines. (C) An example of a β-barrel transmembrane protein, OmpF (porin) from *E.*
coli (PDB id 2OMF) (Cowan et al., 1992). Only one subunit of the native trimer is shown for clarity. The membrane is shown in grey. (D) A section of β-sheet (sections of strands 11-13 of OmpF) shown on the left in cartoon representation with the side chains as sticks, and on the right with just the backbone as sticks and the hydrogen bonds indicated by dashed lines. Structure images were made using MacPyMOL.
Figure 1.2 Membrane protein biogenesis

The biological pathways by which α-helical and β-barrel transmembrane proteins are inserted into cellular membranes. (A) α-helical membrane proteins are inserted cotranslationally via the Sec translocon. The ribosome-nascent-polypeptide complex are
directed to the translocon by SRP (shown here after dissociating from the complex). The polypeptide chain (green line) is translocated through the translocon channel and hydrophobic TM segments (green rectangle) pass laterally into the bilayer to form an α-helix bundle. The depicted final membrane protein structure is the potassium channel KcsA (PDB id 1BL8)(Doyle et al., 1998). (B) β-barrel OMP proteins from Gram-negative bacteria follow a complex pathway to reach the outer membrane. The unfolded polypeptide chain (labeled uOMP, orange) is directed to the SecYEG translocon post-translationally via the chaperone SecB and translocated through the IM channel by the motor domain SecA. Upon reaching the periplasm, the uOMP is thought to interact with the chaperone proteins Skp, SurA, and DegP, which deliver the uOMP to the BAM complex in the outer membrane (peptidoglycan layer not shown for clarity). The POTRA domains of BamA are denoted P1-5 and the associated lipoproteins are labeled with the corresponding letter designation. OMP folding and insertion into the OM is catalyzed by the BAM complex through an unknown mechanism. The depicted final OMP structure is the porin OmpF (PDB id 2OMF)(Cowan et al., 1992).
Figure 1.3 Thermodynamic model for α-helical membrane protein folding

A four-step thermodynamic model developed by White and Wimley to describe α-helical membrane protein folding. The membrane is shown in grey with the area between the black lines corresponding to the hydrophobic core region and the areas adjacent to the black lines corresponding to the slightly more polar interfacial regions of the bilayer. The first step is partitioning of the unfolded polypeptide to the interfacial region, followed by folding into a stable α-helix, insertion across the bilayer, and helix-helix association. Figure adapted from (White & Wimley, 1999).
Figure 1.4  Secondary and tertiary structures of OmpA and PagP

Topology diagrams (left) and crystal structures (right) of (A) OmpA$_{171}$ (PDB id 1QJP) (Pautsch & Schulz, 2000) and (B) PagP (PDB id 1THQ)(Ahn et al., 2004). In the topology diagrams, squares indicate β-sheet structure, with grey squares and bold font
signifying that the residue side chains point out into the lipid environment of the bilayer and white squares signifying side chains point into the pore of the barrel. Residues in tan circles were unresolved in the crystal structures. Tryptophan residues are colored red and shown in stick representation in the crystal structures. The dashed line extending from the C-terminus of the OmpA_{171} diagram in (A) indicates the continuation of the sequence to the soluble periplasmic domain. The membrane is indicated by grey horizontal lines. The bottom bilayer edge in (B) is curved to indicate the portions of the sequence embedded in the membrane. This is necessary due to the differing lengths of β-strands stemming from PagP’s tilted orientation in the membrane. Crystal structure images were made using MacPyMOL.
Figure 1.5  Folding pathways proposed for OmpA

(A) Surrey and Jähnig’s initial multi-step model for OmpA folding into lipid bilayers. $U_D$ is the unfolded state in high denaturant, $U_{AQ}$ is the unfolded state in aqueous buffer after the dilution of denaturant, Agg indicates an aggregated state(s) formed by $U_{AQ}$, $I_M$ is a
membrane-associated intermediate state, and F is the native, fully inserted folded state. Adapted from (Surrey & Jähnig, 1995). (B) Kleinschmidt and Tamm’s extended model incorporating a second membrane-associated intermediate state, $I_{M2}$. Adapted from (Kleinschmidt & Tamm, 1996). (C) Kleinschmidt and colleagues’ further extended model incorporating a third membrane-associated intermediate state, $I_{M3}$. The off-pathway aggregation reaction of $U_{AQ}$ is omitted for simplicity, but is assumed to still be possible. The illustration shows the concerted mechanism of β-barrel insertion and the positions of tryptophan residues (circles) within the bilayer. Trp7 is specifically labeled and never crosses the bilayer core. Mechanism adapted from (Kleinschmidt & Tamm, 1999). Image reproduced from (Tamm et al., 2004) with permission. (D) Kleinschmidt and colleagues’ refined model for β-barrel insertion. Colored circles indicate positions of Trp and Cys residues used in site-directed fluorescence quenching. Image reproduced from (Kleinschmidt et al., 2011) with permission.
Figure 1.6  Folding pathways proposed for PagP

(A) Huysmans and colleagues’ illustration of the PagP tilted insertion process via a polarized transition state (TS). The starting state is an unstructured membrane-associated conformation (IM). Reproduced from (Huysmans et al., 2010) in accordance with PNAS copyright policy. (B) General folding pathway for PagP proposed by Huysmans and colleagues. The first step is fast membrane adsorption of the unfolded aqueous state (UAQ) followed by a second step of folding and membrane insertion. The authors propose that this step can follow one of two parallel pathways, resulting in fast and slow phases in the folding kinetics. Reproduced from (Huysmans et al., 2012) with permission.

§ http://www.pnas.org/site/aboutpnas/rightperm.xhtml
Chapter 2

The soluble, periplasmic domain of OmpA folds as an independent unit and reduces the self-association propensity of the unfolded OmpA transmembrane β-barrel**

2.1 Abstract

OmpA is one of only a few transmembrane proteins whose folding and stability have been investigated in detail. However, only half of the OmpA mass encodes its transmembrane β-barrel; the remaining sequence is a soluble domain that is localized to the periplasmic side of the outer membrane. To understand how the OmpA periplasmic domain contributes to the stability and folding of the full-length OmpA protein, we cloned, expressed, purified and studied the OmpA periplasmic domain independently of the OmpA transmembrane β-barrel region. Our experiments showed that the OmpA periplasmic domain exists as an independent folding unit with a free energy of folding equal to -6.2 (±0.1) kcal mol⁻¹ at 25 °C. Using circular dichroism, we determined that the OmpA periplasmic domain adopts a mixed α/β secondary structure, a conformation that

has previously been used to describe the partially folded non-native state of the full-length OmpA. We further discovered that the OmpA periplasmic domain reduces the self-association propensity of the unfolded barrel domain, but only when covalently attached (in cis). *In vitro* folding experiments showed that self-association competes with β-barrel folding when allowed to occur before the addition of membranes, and the periplasmic domain enhances the folding efficiency of the full-length protein by reducing its self-association. These results identify a novel chaperone function for the periplasmic domain of OmpA that may be relevant for folding *in vivo*. We have also extensively investigated the self-association reaction of unfolded OmpA; we found that the oligomeric form of the transmembrane domain contains β-sheet structure and binds Thioflavin T, indicating an amyloid-like structure. Kinetic experiments revealed that the transmembrane domain must form a critical nucleus comprised of three molecules before undergoing further oligomerization to form large molecular weight species. Finally, we studied the conformation of the unfolded OmpA monomer and found that the folding-competent form of the transmembrane region adopts an expanded conformation, which is in contrast to previous studies that have suggested a collapsed unfolded state.

### 2.2 Introduction

Over the past two decades, the transmembrane protein OmpA has been extensively investigated as a model for membrane protein folding. OmpA folds into a large number of hydrophobic environments, including many different detergents and lipid compositions (Dornmair *et al.*, 1990; Freudl *et al.*, 1986; Kleinschmidt & Tamm, 1996,
A folding pathway describing the conformational changes that OmpA must undergo to attain its native conformation in membranes was proposed in the 1990s (Kleinschmidt, den Blaauwen, et al., 1999; Surrey & Jähnig, 1995), and OmpA is one of only four transmembrane proteins whose thermodynamic stability has been measured in phospholipid vesicles (Hong & Tamm, 2004; Huysmans et al., 2010; Moon & Fleming, 2011; Moon et al., 2013).

However, examination of the sequence shows that full-length OmpA is actually a two-domain protein in which only the N-terminal half (OmpA_{171}, residues 1-171) is a membrane-embedded β-barrel (Pautsch & Schulz, 2000). In contrast, the remaining sequence of OmpA (172-325) comprises a soluble, periplasmic domain (OmpA_{Per}) that is thought to interact with peptidoglycan (De Mot & Vanderleyden, 1994; Koebnik, 1995). The contributions of OmpA_{Per} to the kinetic pathways and thermodynamic stability of the full-length OmpA protein (OmpA_{325}) have not been explicitly investigated and have, in fact, largely been ignored. One reason for this is that changes in the conformation of the periplasmic domain are mostly invisible to the methods that have been employed in studying OmpA (Hong et al., 2007; Hong et al., 2006; Hong & Tamm, 2004; Kleinschmidt & Tamm, 1996, 2002a; Kleinschmidt, Wiener, et al., 1999; Rodionova et al., 1995; Surrey & Jähnig, 1992, 1995; Tamm et al., 2004). Many folding investigations of OmpA use tryptophan fluorescence spectroscopy as a reporter of changes to the protein’s conformation and/or environment; however, all five of the tryptophan residues in OmpA are located in the transmembrane β-barrel region, while none are located in the periplasmic sequence (see Figure 2.1). Therefore, fluorescence studies do not detect
changes in the conformation of the periplasmic domain. The other principal method used to measure OmpA folding takes advantage of the fact that microbial β-barrel proteins exhibit a different migration on SDS-PAGE gels depending on whether they are folded or unfolded (Burgess et al., 2008; Nakamura & Mizushima, 1976). Because the periplasmic domain is soluble, it does not show this behavior, so changes to its conformation are not detected by this assay either.

This chapter was originally published as a contribution to a special issue commemorating the 25th annual meeting of the Gibbs Society of Biothermodynamics (Danoff & Fleming, 2011). In keeping with the Gibbs tradition of dissecting a system into its components and examining the individual contributions of each part to overall biological function, we examined the structural features and thermodynamic stability of the OmpA periplasmic domain in isolation, as well as its interactions and role in folding with regard to the β-barrel domain of OmpA. We used circular dichroism spectroscopy, sedimentation velocity, and in vitro folding to study the conformation and interactions of the individual domains of OmpA in the context of the behavior of the full-length protein. Our studies significantly revise previous ideas about the unfolded state conformations and interactions of OmpA and contribute to a more complete and accurate scheme for its kinetic folding pathway. In addition, we discovered a novel chaperone function of the periplasmic domain that may be relevant for OmpA folding in vivo.
2.3 Materials and Methods

Construction of the OmpA<sub>Per</sub> homology model

We submitted the amino acid sequence corresponding to the OmpA periplasmic domain to the Swiss-Model web server (http://swissmodel.expasy.org). This server used the Neisseria meningitidis RmpM OmpA-like domain structure (PDB id 1R1M) (Grizot & Buchanan, 2004) as a best match to template a homology model of the E. coli OmpA periplasmic domain. These two sequences share 37.1% sequence identity, which is higher than the 30% threshold sequence identity needed for accurate model building (Baker & Sali, 2001). The resultant model of the OmpA periplasmic domain is essentially superimposable upon the RmpM structure with the exception of one loop region in which RmpM has an extra turn of helix both preceding and following one turn (Figure 2.1A).

Cloning and expression of proteins

The mature form of full-length OmpA (OmpA<sub>325</sub>) and the N-terminal barrel domain (OmpA<sub>171</sub>) were PCR amplified using primers designed to include NdeI (5’) and BamHI (3’) sites (primers are listed in Table 2.1). The OmpA constructs were amplified using ExTaq polymerase (Takara) from an overnight growth of E. coli K12 MG1655. The PCR products were ligated into the pCR2.1-TOPO vector using a TOPO TA cloning kit (Invitrogen). QuikChange site-directed mutagenesis (Stratagene) was used to remove an internal BamHI site in OmpA<sub>325</sub>, which did not change the amino acid sequence. The plasmids were cut with NdeI and BamHI, and the insert was ligated into a pET11a vector. These plasmids were transformed into DH5α cells, and the sequences were confirmed by double-stranded DNA sequencing.
The periplasmic domain of OmpA (OmpA_{Per}) was PCR amplified using primers incorporating NdeI (5') and XhoI (3') sites. The gene was amplified from the OmpA\textsubscript{325} gene in pET11a and ligated into the TOPO vector. The plasmid was cut with restriction enzymes and ligated into a pET28b vector, which expresses the protein with an N-terminal His-tag followed by a TEV cleavage site. The plasmid was transformed into DH5\textalpha and the sequence was confirmed by DNA sequencing.

Plasmids for the three OmpA constructs were separately transformed into HMS174(DE3) cells (Novagen) and grown in 500 ml of TB medium at 37 °C with shaking to an optical density of 0.8 at 600 nm. Protein expression was induced by the addition of 1 mM IPTG and cells were incubated for 4-6 h at 37 °C with shaking before harvesting by centrifugation (5000 rpm, 15 min, 4 °C). Cell pellets were frozen for periods of time ranging from overnight to several days before undergoing further processing.

**Preparation of urea solutions**

Ultra-pure urea was purchased from Amresco. Urea solutions were prepared at a concentration of 10 M in water and deionized by adding AG 501-X8 resin (BioRad) at a ratio of 1 g resin for every 20 ml of solution. Solutions were stirred with the resin for 1 h at room temperature. The resin was removed by filtration and the urea used to prepare Urea Buffer (8 M urea, 20 mM Tris, pH 8). The final urea concentration was determined by refractometry (Warren & Gordon, 1966). For purifications of OmpA\textsubscript{325}, Urea Buffer contained 2 mM TCEP (Pierce). Urea Buffer was stored at -20 °C.
Purification of OmpA$_{325}$ and OmpA$_{171}$

A pellet from a 500 ml growth was resuspended in 25 ml of Lysis Buffer (50 mM Tris, pH 8, 40 mM EDTA) and lysed by French press. Brij-35 (Sigma) was added to a final concentration of 0.1%, and inclusion bodies were isolated by centrifugation at 5500 rpm for 30 min. These were washed twice by resuspension in 25 ml of Wash Buffer (10 mM Tris, pH 8, 1 mM EDTA) followed by pelleting by centrifugation under the same conditions. Purified inclusion bodies were resuspended a third time in Wash Buffer and split into four fractions before a final centrifugation. The supernatant was discarded and inclusion body pellets were stored at -20 ºC.

An inclusion body pellet was dissolved in 7 ml of Urea Buffer and clarified by centrifugation. The supernatant was filtered through a 0.45 µm pore (Millipore). Protein samples were further purified using a BioRad BioLogic DuoFlow Chromatography System. Samples were loaded onto an UNO Q6 continuous bed anion exchange column (BioRad) and eluted with an NaCl gradient in Urea Buffer. Protein-containing fractions were pooled and concentrated using centrifugal filtration (Millipore). Samples were de-salted and further purified using a Superdex 200 10/300 GL gel filtration column (GE Healthcare), run in Urea Buffer. Protein-containing fractions were pooled and concentrated as before. Protein concentration was determined by measuring the absorbance at 280 nm and using extinction coefficients calculated in Sednterp (Laue et al., 1992). These values are reported in Table 2.2. Purified unfolded OmpA$_{325}$ and OmpA$_{171}$ were aliquoted into microcentrifuge tubes and stored at -80 ºC until use.
**Purification of OmpA<sub>Per</sub>**

A pellet from a 500 ml growth was resuspended in 25 ml of Buffer A (20 mM Sodium Phosphate, pH 8, 500 mM NaCl, 20 mM imidazole) and a tablet of Complete EDTA-free protease inhibitor cocktail (Roche) added. Cells were lysed by French press and the lysate clarified by centrifugation at 5500 rpm for 30 min. The supernatant was retained and DNase (Roche) added to a final concentration of 2 µg ml<sup>-1</sup>. The sample was filtered through a 0.45 µm pore (Millipore). His-tagged OmpA<sub>Per</sub> was purified by loading the sample on a column packed with Ni Sepharose High Performance (GE Healthcare) and eluting with Buffer B (20 mM Sodium Phosphate, pH 8, 500 mM NaCl, 500 mM imidazole). The His-tag was removed by incubation overnight with TEV protease (Kapust <i>et al.</i>, 2001) (1:20 molar ratio TEV:OmpA<sub>Per</sub>) and the sample dialyzed into Buffer A. Cleaved protein was isolated by passing the sample through the Ni column again. Protein-containing fractions were pooled and dialyzed into 20 mM Tris, pH 8. The concentration was determined using the extinction coefficient listed in Table 2.2. OmpA<sub>Per</sub> was stored at 4 °C until use.

**Vesicle preparation**

1,2-Didecanoyl-<i>sn</i>-glycero-3-phosphocholine (diC<sub>10</sub>PC) lipids dissolved in chloroform (Avanti Polar Lipids) were dried to a thin film in glass vials under a gentle stream of nitrogen gas. The lipid films were evacuated overnight to remove residual solvent and stored at -20 °C until use. For vesicle preparation, lipid films were reconstituted in 20 mM Tris, pH 8 at a concentration of 10 mg ml<sup>-1</sup> and large unilamellar
vesicles (LUVs) were made by extruding reconstituted lipids 21 times though a 0.1 \( \mu m \) filter using a mini-extruder (Avanti)(Hope et al., 1985).

**Circular dichroism**

CD measurements were performed using an Aviv Circular Dichroism Spectrometer, Model 410 (Aviv Biomedical), with a custom inset detector to reduce the effects of light scattering. Wavelength spectra were recorded at 25 \(^\circ\)C between 200 and 280 nm in 1 nm increments, with an averaging time of 5 s. For each sample, three scans were recorded and averaged. A path length of 1 cm or 1 mm was used (Hellma cuvettes) and spectra of cuvettes containing only buffer were subtracted from sample spectra to correct for background signal. For samples containing LUVs, spectra of LUV-only mixtures were used for background subtraction. Data were converted to molar ellipticity, \([\theta]\), using the following equation:

\[
[\theta] = \frac{\theta}{10cl}
\]

where \( \theta \) is the ellipticity in mdeg, \( c \) is the concentration in M, and \( l \) is the path length in cm.

Thermodynamic measurements on OmpA\textsubscript{Per} were performed using an automated Hamilton Dispenser, MICROLAB 540B. In unfolding titrations, an appropriate amount of Urea Buffer was titrated into a folded sample followed by incubation at 25 \(^\circ\)C while stirring for 1 min before taking a measurement. The CD signal was monitored at 220 nm with an averaging time of 5 s. Measurements were corrected for protein dilution. Refolding titrations were performed in a similar manner, with the titrator dispensing buffer instead of denaturant, and the initial sample prepared in 7.5 M urea. Titration data
were fit in IGOR Pro (WaveMetrics) using the linear extrapolation method of Santoro & Bolen (Santoro & Bolen, 1988, 1992).

**CD kinetics of self-association**

CD signal kinetics were measured during OmpA\textsubscript{325} and OmpA\textsubscript{171} self-association at various protein concentrations and in 600 mM urea, 20 mM Tris, pH 8. The CD signal was monitored at 218 nm for approximately 16 hours, with the sample incubated at 25 °C without stirring. The interval between data points was 60 s and the time constant was 10 s. The CD signal was converted to fraction monomer using the values measured for completely monomeric protein and completely self-associated protein.

The method of initial rates was used to estimate the order of the self-association reaction (Kleinschmidt & Tamm, 2002a). For the general reaction,

\[ nA \rightarrow B \]

the rate of reaction is given by the following expression:

\[ \frac{-d[A]}{dt} = k[A]^n \]  

Equation 2.2

where \( k \) is the rate constant and \( n \) is the order of the reaction. At early time points, the concentration of A can be approximated as the starting concentration, \([A]_o\), so the initial rate of reaction can be expressed as:

\[ \left(\frac{d[A]}{dt}\right)_{t=0} = k[A]_o^n \]  

Equation 2.3

A double logarithmic plot of the initial rate as a function of \([A]_o\) is linear and has a slope of \( n \), the order of the reaction:
Equation 2.4

\[
\ln \left( - \frac{d[A]}{dt}_{t=0} \right) = n \ln ([A]_o) + \ln k
\]

To determine the order of the OmpA self-association reaction, the CD kinetics were plotted as monomer concentration versus time and a line fit to the earliest time points. The slopes were plotted on a double logarithmic plot as a function of total OmpA concentration, and a line was fitted to the data.

The kinetics data were further analyzed using a nucleated growth polymerization model to determine the critical nucleus size, \( n^* \). According to this model, at early time points the concentration of monomer units incorporated into polymers (\( \Delta \)) varies linearly with time squared (S. Chen et al., 2002):

\[
\Delta(t) = s(c)t^2
\]

where \( s(c) \) is the slope of this line and is a function of total monomer concentration, \( c \), and the critical nucleus, \( n^* \):

\[
s(c) \propto c^{n^*+2}
\]

A double logarithmic plot of \( s \) as a function of total monomer concentration, \( c \), is linear and has a slope of \( n^*+2 \).

To perform this analysis on the CD kinetics data, the data were transformed to plot associated OmpA versus time squared and a line fit to the earliest time points. The slopes were plotted on a double logarithmic plot as a function of total OmpA concentration, and a line was fitted to the data.

**Sedimentation velocity analytical ultracentrifugation**

Sedimentation velocity experiments were carried out in a Beckman XL-A analytical ultracentrifuge, using two-sector cells and an An60Ti rotor. All experiments
were conducted at a speed of 50,000 rpm and 25 °C. Sedimentation profiles were detected using absorbance optics operated in continuous mode. Protein molecular weights, partial specific volumes, extinction coefficients and buffer densities were calculated using Sednterp (Table 2.2) (Laue et al., 1992).

*Sedimentation velocity: self-association of unfolded OmpA\textsubscript{325} and OmpA\textsubscript{171}*

To study the concentration dependence of OmpA self-association, samples were prepared at varying protein concentrations in 20 mM Tris, pH 8 and various urea concentrations. OmpA\textsubscript{325} samples also contained 2 mM TCEP to eliminate disulfide bond mediated dimer formation. This was unnecessary for OmpA\textsubscript{171} because the two cysteine residues responsible for disulfide bonding are located in the periplasmic domain. After dilution of the protein from high urea to the final condition, the samples were loaded into the sedimentation velocity cells, placed in the rotor, and temperature equilibrated prior to rotor acceleration. To ensure consistency, the total time for these steps was controlled so that the rotor would start 30 min after protein dilution. Sedimentation velocity profiles were detected using the absorbance optics at a single wavelength adjusted between 227 nm and 235 nm to obtain an absorbance signal between 0.1 and 1.3.

Sedimentation velocity data were analyzed using the time derivative method of Stafford as implemented in DCDT+ (Philo, 2006). When analyzing the sedimentation data, the scans were divided into two analysis windows based on the apparent populated species: either monomer (late scans) or oligomer (early scans). However, care was taken to examine all scans for the presence of intermediate sedimentation coefficients. Sedimentation coefficient distribution peaks (g(s*)) obtained in DCDT+ at low s-values
were fitted to Gaussian equations to confirm they contained monomeric protein, and the concentration of monomer was determined by integrating the area under the monomeric $g(s^*)$ peak. Very broad distributions were observed at large $s$-values, and the entire oligomeric peak was integrated to determine the concentration of these species. The fraction monomer was calculated as the concentration of monomer divided by the sum of monomer and oligomers. Data were also analyzed by the $c(s^*)$ method as implemented in SedFit (Schuck, 2000) and the fraction monomer found by integration of the $c(s^*)$ curve agreed well with that found by $g(s^*)$ (Figure 2.5).

For a given urea concentration, the fraction monomer was plotted as a function of protein concentration and the curve was fit to the Hill equation (using IGOR Pro), which has the form:

$$y = base + \frac{max - base}{1 + \left(\frac{x_{1/2}}{x}\right)^n}$$

Equation 2.7

The midpoint concentration, $x_{1/2}$, for each curve was plotted as a function of urea concentration and a line fit to the data to extrapolate to the absence of urea.

*Sedimentation velocity: time-dependence of self-association*

To determine the kinetics of OmpA self-association, samples were prepared as above and incubated for various times before initiating centrifugation. Data were analyzed as described above to determine the fraction monomer at each time point.
Sedimentation velocity: hydrodynamic shape estimates

The urea dependences of the hydrodynamic shapes of unfolded OmpA_{325} and OmpA_{171} were analyzed by sedimentation velocity of samples at a concentration of 2 µM in various urea concentrations. The monomer peaks in the g(s*) curves were fit for molecular weight and sedimentation coefficient. After converting to s*_{20,w}, the s-values were plotted as a function of urea, and a linear extrapolation was used to determine the sedimentation coefficient in the absence of urea. Perrin’s equations as implemented in Sednterp were used to calculate the axial ratio of a prolate ellipsoid of revolution for the species.

Fluorescence measurements of Thioflavin T binding

Thioflavin T (ThT) was purchased from Sigma and dissolved in water to a concentration of 500 µM. Samples were prepared of ThT alone or with OmpA_{171} or OmpA_{325} at concentrations of 3 µM ThT and either 1 µM protein and 600 mM urea or 3 µM protein and 300 mM urea. All samples were prepared in 20 mM Tris, pH 8. Samples without protein or with 1 µM protein were measured immediately and samples with 3 µM protein were incubated for 16 h at 25 ºC without stirring before making measurements.

Fluorescence measurements were performed at 25 ºC with a 1 cm path length using an ISS PC1 spectrofluorometer. Excitation slits were 2.4 mm and emission slits were 2.0 mm. Excitation spectra were recorded from 350 to 480 nm with the emission monochromator set to 482 nm. Emission spectra were recorded from 460 to 600 nm with excitation at 450 nm. For each spectrum, three scans were recorded and averaged.
Transmission electron microscopy

Unfolded OmpA171 was incubated under self-associating conditions (3 μm protein, 300 mM urea, 20 mM Tris, pH 8) for 7 h. A negatively stained sample was prepared for TEM using a freshly ionized formvar/carbon-coated copper grid. The protein solution was adsorbed to the grid for 5 min, then washed six times with water. The grid was stained with 2% uranyl acetate for 1 min before removing excess stain. The sample was observed in an FEI Tecnai 12 TWIN electron microscope operating at 100 kV. Images were captured using an SIS Megaview III camera (Olympus). Width measurements of observed structures were made with the freely available ImageJ software.

Delayed folding experiments

OmpA325 or OmpA171 were diluted into a folding condition that lacked vesicles (2 μM or 5 μM protein, 600 mM urea) and incubated at 25 ºC for 30 min before the addition of LUVs at a final lipid to protein ratio of 800:1, to initiate folding. Samples underwent gentle stirring during folding. Certain folding mixtures also contained equimolar concentrations of OmpAPer (2 μM or 5 μM). Aliquots were removed from the folding reactions after 3 h and folding quenched by the addition of 4X SDS gel-loading buffer to a final concentration of 1X (Sambrook et al., 1989). The fraction folded was determined by SDS-PAGE (Laemmli, 1970) using pre-cast acrylamide gels from BioRad, staining with Coomassie Blue R-350 (GE Healthcare), and digital transmission scanning (Epson 4490). Densitometry was performed using ImageJ, and the fraction folded was calculated.
by dividing the intensity of the folded band by the sum of the intensities of the folded and unfolded bands.††

2.4 Results

The OmpA periplasmic domain adopts a mixed α/β secondary structure that can fold independently of the transmembrane region

Nearly twenty years ago, Surrey & Jähnig showed that OmpA adopts a mixed α/β secondary structure immediately upon dilution to a folding condition (<100 mM urea) from high (>6 M) urea concentrations (Surrey & Jähnig, 1995). This form has been referred to in the literature as a “partially folded” state from which folding begins in aqueous solutions with vesicles. However, it has always been unclear whether this secondary structure arises from conformations of the periplasmic domain, the transmembrane β-barrel domain, or both. This ambiguity is essential to resolve because elucidation of physically based folding schemes requires knowing the set of conformations that polypeptide sequences can sample along a folding trajectory.

Because the mixed α/β secondary structure arose immediately upon dilution of OmpA to a folding condition, we hypothesized that this secondary structure might represent the folding of the OmpA periplasmic domain rather than a non-native conformation adopted by the transmembrane β-barrel region. There is no crystal structure

†† In Chapter 3, fraction folded measured by SDS-PAGE is calculated as the intensity of the folded band divided by the intensity of a boiled sample, due to the protein populating another state that does not migrate as folded or unfolded. This methodology had not been developed at the time of the experiments in this chapter. However, this difference in analysis is not expected to affect the overall result or conclusions drawn from the delayed folding experiments.
of the periplasmic domain, so we began to address this question by constructing a homology model for OmpA<sub>Per</sub> using the structure of the OmpA-like protein, RmpM (Grizot & Buchanan, 2004). Shown overlaid with RmpM in Figure 2.1A, this model is indeed well described as mixed α/β secondary structure.

OmpA<sub>171</sub> had previously been shown to independently adopt the transmembrane β-barrel fold in the presence of detergents or lipids; its structure is shown in the top part of Figure 2.1B (Pautsch & Schulz, 2000). The modeled OmpA<sub>Per</sub> is shown adjacent to OmpA<sub>171</sub> at the same scale in a representation of the full-length OmpA (OmpA<sub>325</sub>).

To determine whether the observed mixed α/β structure of full-length OmpA was due to the periplasmic domain, the β-barrel domain, or both, we cloned, expressed and purified OmpA<sub>Per</sub>, OmpA<sub>171</sub>, and OmpA<sub>325</sub> independently of each other and measured the CD spectra of the three constructs. Figure 2.2A shows CD spectra of OmpA<sub>Per</sub> alone, folded in aqueous Tris buffer (solid blue) and unfolded in 6 M urea (dotted blue). It is apparent from these spectra that folded OmpA<sub>Per</sub> has mixed α/β structure, as indicated by the double trough at 208 and 222 nm (Nordén et al., 2010). In 6 M urea the protein has no regular structure, consistent with an unfolded conformation.

Panels B-D of Figure 2.2 show CD data for OmpA<sub>171</sub> (red), OmpA<sub>Per</sub> (blue), and OmpA<sub>325</sub> (green) under various buffer conditions (described below) used to investigate the conformations of the two domains. The mathematical sum of the spectra for OmpA<sub>171</sub> and OmpA<sub>Per</sub> is shown as a dashed purple line in each panel. The data in all cases demonstrate that the mixed α/β secondary structure characteristic of aqueous OmpA<sub>325</sub> arises entirely from OmpA<sub>Per</sub>, though the conformation of the OmpA<sub>171</sub> barrel region is dependent on the conditions employed.
Figure 2.2B shows spectra collected at protein concentrations of 1 µM in 20 mM Tris, pH 8, 600 mM urea and diC\textsubscript{10}PC LUVs at a lipid to protein ratio of 800:1. To obtain a measurable signal from such a low protein concentration, a 1 cm path length was used and reliable data could only be collected above 210 nm. However, a distinctive β-trough at 216 nm can be observed for OmpA\textsubscript{171} (red), indicating that the barrel folds in the presence of LUVs. In addition, the 222 nm trough indicative of α-helix is present in the OmpA\textsubscript{Per} spectrum (blue), indicating that it is folded in the presence of 600 mM urea (this result is also consistent with the unfolding titration described below and shown in Figure 2.3). The CD spectrum for full-length OmpA\textsubscript{325} (green) overlays well with the mathematical sum (dashed purple) of the spectra for OmpA\textsubscript{171} and OmpA\textsubscript{Per}, indicating that these are independently folding domains. Interestingly, the spectrum of OmpA\textsubscript{171} reveals a novel 230 nm positive peak that we only observe under conditions where the β-barrel is folded. A similar peak has previously been observed in the PagP transmembrane β-barrel and has been attributed to an exciton interaction between aromatic groups that only forms when the PagP β-barrel is folded (Khan \textit{et al.}, 2007). This positive peak is masked in OmpA\textsubscript{325} by the negative α-helix signal arising from OmpA\textsubscript{Per}.

Figure 2.2C shows spectra for these three constructs in the same buffer conditions as panel B, except lacking LUVs (1 µM protein, 20 mM Tris, pH 8, 600 mM urea). As expected, the periplasmic domain (blue) does not require vesicles for folding and is folded under these conditions as confirmed by the α-helix trough deflection at 222 nm. However, OmpA\textsubscript{171} (red) displays a CD spectrum containing no regular structure, and it lacks the 230 nm positive peak, showing it is unfolded in the absence of vesicles. The mathematical sum (dashed purple) of these two spectra overlays upon that of full-length
OmpA325 (green). Altogether, these data show that the “folding competent” unfolded state of the OmpA transmembrane β-barrel region is not mixed α/β, but rather contains no regular structure.

Figure 2.2D shows the same buffer conditions as panel C (20 mM Tris, pH 8, 600 mM urea, no LUVs), except the protein concentration is higher (10 µM) and the samples have been allowed to incubate overnight. To accommodate this higher protein concentration, a 1 mm path length was used and data were collected down to 202 nm. We show below that 10 µM protein in 600 mM urea is a condition that results in self-association of the OmpA325 and OmpA171 unfolded states to form high molecular weight oligomers. However, even in this oligomeric form, the CD spectra of the two independent domains (OmpA171 in red and OmpAPer in blue) sums to that of the full-length spectrum (green), indicating that self-association of OmpA unfolded states does not involve unfolding of its periplasmic domain. Interestingly, the CD spectra for self-associated OmpA171 and OmpA325 show broad negative peaks in the β-region of the CD spectrum, but these spectral shapes are distinctly different from the spectra of natively folded OmpA barrel in panel B. To determine if the non-native β-sheet structure of the self-associated state was due to formation of an amyloid-like structure, we conducted dye-binding experiments with Thioflavin T, which is known to specifically bind amyloid fibrils (see below).

As implied by the CD spectra, we expected that OmpAPer represented an independent folding domain. To demonstrate this, we determined its thermodynamic stability using chemical denaturation experiments. Folding of OmpAPer has been invisible in previous denaturation studies of OmpA because those studies used fluorescence
spectroscopy or SDS-PAGE to measure folded and unfolded populations, and changes to the conformation of the periplasmic domain are not detected by these methods (Hong & Tamm, 2004). Figure 2.3 shows a typical urea denaturation curve for OmpA_{Per} measured using CD spectroscopy. The data are well described by two-state linear extrapolation equations (Santoro & Bolen, 1988, 1992) and reveal a free energy of folding in the absence of denaturant equal to -6.2 (±0.1) kcal mol\(^{-1}\) with an m-value of -1.44 (±0.04) kcal mol\(^{-1}\) M urea\(^{-1}\). The same values were obtained from measurements on protein from a completely separate purification. A refolding titration was also performed to verify the path independence of this measurement (data not shown).

*The OmpA periplasmic domain acts in cis to reduce self-association of the unfolded β-barrel domain*

In 1992, Surrey & Jähnig showed that the aqueous, unfolded state (U_{AQ}) of OmpA can slowly self-associate to form large oligomers (Surrey & Jähnig, 1992). Because OmpA presumably folds as a monomeric entity, the interactions of OmpA unfolded states would be detrimental to folding and should be tightly controlled by the cell with chaperone proteins. Moreover, the presence of oligomeric unfolded states represents an additional kinetic step that must be considered in the development of *in vitro* folding pathways. Surrey & Jähnig formed the U_{AQ} state by diluting OmpA from high to low concentrations of urea in solutions that lacked vesicles (i.e., buffer conditions that would have strongly supported β-barrel folding if vesicles had been present) and observed by fluorescence spectroscopy that U_{AQ} slowly self-associated on the time scale of hours.
The Fleming lab previously investigated the $U_{AQ}$ interactions of OmpA and seven additional outer membrane proteins (OMPs) in buffers containing 1 M urea and found that OmpA was one of only a few OMPs to be entirely monomeric (Ebie Tan et al., 2010). This result was contrary to the early findings of Surrey & Jähnig, but the urea concentration was higher in our experiments, and we reasoned that 1 M urea could completely destabilize OmpA oligomers. Here we revisit this question of OmpA $U_{AQ}$ self-association by measuring the protein concentration dependence of its sedimentation coefficient distribution at lower concentrations of urea. Shown below, we did indeed observe self-association of OmpA unfolded states at urea concentrations more comparable to those used in the Surrey & Jähnig studies. Initially we controlled the set-up time for our sedimentation velocity experiments to be 30 minutes, to capture the physical parameters describing the “instantaneous” self-association that might compete with the productive folding of OmpA (when vesicles are present). However, we also showed that self-association continues over the course of 12-16 hours, and we examined the concentration dependence of the rate of self-association (see below).

Primary sedimentation velocity data for OmpA$_{325}$ at 3 µM and in 450 mM urea are shown in Figure 2.4, where the separation between two boundaries is clearly evident. Figure 2.5A shows the sedimentation coefficient distribution function obtained for the same dataset, using the time derivative method implemented in DCDT+ (Philo, 2006). Under these conditions it can be observed that OmpA$_{325}$ is a mixture of a ~2 S species, best described as monomer, and a polydisperse ensemble of large oligomers with sedimentation coefficients in the range of 20-60 S. As described in Materials and Methods, two separate analysis regions were used to obtain these distinct distributions.
Figure 2.5B shows similar results were obtained using the continuous c(s*) method in Sedfit (Schuck, 2000). In both cases, the monomer was well separated from the oligomeric species and we could integrate the area under the two regions separately to determine the fraction monomer. Moreover, in all buffer conditions the ~2 S species is well described by a single Gaussian fit (implemented in DCDT+) that returns a molecular weight corresponding to the OmpA monomer. Integrating small and large s-value regions as well as confirming that the ~2 S species was monomeric OmpA was straightforward for all protein and urea concentrations.

Using sedimentation velocity, the fraction monomer was measured as a function of protein and urea concentrations for OmpA_{325} and OmpA_{171} at the 30 minute time point. These data are plotted in Figure 2.6 and it can be observed that both proteins exhibit a sigmoidal dependence of self-association in three different urea concentrations: 300 mM (red triangles), 450 mM (orange circles), and 600 mM (gold diamonds) urea. For both OmpA_{325} and OmpA_{171}, low (<2) micromolar concentrations are required before the protein remains entirely monomeric for up to 30 minutes at these urea concentrations. Additionally, in all conditions OmpA_{171} self-associates at lower total protein concentrations than OmpA_{325}. The difference between these two proteins is the soluble periplasmic domain within OmpA_{325}, and these data therefore suggest that OmpA_{Per} must act to reduce the propensity of OmpA_{325} UAQ to oligomerize.

The data are well described by the Hill equation, which gives the midpoint protein concentrations for each urea condition (see Materials and Methods for equation). In Figure 2.6C we show that the midpoint concentrations vary linearly with the urea concentration. By extrapolating to 0 M urea, we estimate the midpoint of OmpA_{325} self-
interaction to be \( \sim 200 \) nM. This means that in the absence of urea, only 200 nM of OmpA\(_{325}\) is required to have half of the protein self-associated after 30 minutes. Consistent with a more stable interaction, the OmpA\(_{171}\) midpoint data extrapolates to negative protein concentrations below 0.12 M urea. This means that below 0.12 M urea there is no OmpA\(_{171}\) concentration low enough to allow half of the protein to remain monomeric (at the 30 min time point). However, it should be noted that this linear extrapolation might flatten out at extremely low protein concentrations that we are unable to experimentally access with the absorbance optics of the XL-A system. Nevertheless, oligomers of the OmpA barrel alone are clearly more stable than the full-length protein, indicating that the periplasmic domain acts to disfavor self-association when covalently attached to the barrel domain (\textit{in cis}).

To test whether the periplasmic domain is also able to influence OmpA self-association when not directly attached to the barrel domain (\textit{in trans}), we conducted sedimentation experiments with OmpA\(_{Per}\) added separately to solutions of OmpA\(_{325}\) or OmpA\(_{171}\). We would expect to see two indications in the sedimentation data if the periplasmic domain and barrel were to interact: a decrease in the fraction of oligomeric protein as measured by integration of the g(s\(^*\)) curve, and the presence of a new peak in the g(s\(^*\)) curve at an s-value consistent with the molecular weight of a complex. However, no significant change was observed in the fraction monomer at 30 minutes of either OmpA\(_{325}\) or OmpA\(_{171}\), with up to a four-fold molar excess of OmpA\(_{Per}\) (data not shown). Additionally, we did not observe a complex peak in the small s-value region for the same mixtures. Figure 2.7 shows representative data for mixtures with a two-fold excess of OmpA\(_{Per}\). We found that the g(s\(^*\)) curve is well described by a calculated curve.
for a non-interacting mixture (the sum of the curves for the two proteins alone, scaled according to the concentrations in the mixture). Altogether these data show that the periplasmic domain does not form a stable complex with the unfolded barrel domain at these micromolar concentrations, and it must be covalently attached to reduce self-association of the barrel domain.

The unfolded OmpA β-barrel associates with a rate-limiting, critical nucleus size of three molecules

Figure 2.6 shows a snapshot of OmpA self-association, as all of the data correspond to the fraction monomer present after only 30 minutes of incubation at low urea. We were also interested in determining the time-dependence of the self-association reaction and we utilized both sedimentation velocity and circular dichroism to monitor the extent of oligomerization over time. SV experiments were carried out for various total protein concentrations as described above, but with increasing incubation times, in order to sample different time points during the course of the reaction. The measured fraction monomer values of OmpA_{325} and OmpA_{171} as a function of time are shown as open diamonds in Figure 2.8.

Overlaid upon the sedimentation velocity data points are the CD traces representing the signal decay of monomeric OmpA_{325} and OmpA_{171}. The CD signal reflects the change in the secondary structure at 218 nm shown in the wavelength scans of Figure 2.2C and D. We measured the kinetics of this CD signal change over the course of ~16 hours for various OmpA concentrations. The curves were converted to fraction monomer and are plotted in Figure 2.8 as solid lines for comparison with the SV data. For
OmpA$_{171}$ (panel B), the time-dependent decrease in fraction monomer measured by CD agrees very well with the SV data, and it is evident that higher initial concentrations of protein lead to a more rapid decrease in monomer concentrations and a lower equilibrium value of fraction monomer. Both of these features are consistent with a self-association reaction. In contrast, the OmpA$_{325}$ CD curves do not agree with the SV data (panel A); the CD signal appears not to decay as fast as the fraction monomer time points measured by SV and therefore it would appear that the two techniques are not measuring the same physical process. One explanation for this observation is that the structural change that leads to a more negative CD signal is a separate process from the association into particles that sediment with high s-values. In the case of OmpA$_{171}$, which is composed of the barrel domain only, these processes apparently occur at the same rate. But for OmpA$_{325}$, which includes the soluble periplasmic domain, the change in secondary structure occurs more slowly than the association into faster-sedimenting particles.

To further investigate the nature of the OmpA$_{171}$ barrel self-association reaction, we used the method of initial rates to determine the order of the reaction (see Materials and Methods)(Kleinschmidt & Tamm, 2002a). Figure 2.9A shows a double logarithmic plot of the initial rate as a function of total protein concentration, determined from the OmpA$_{171}$ CD kinetics data. A linear fit to these data points gives a slope of 3.2, which corresponds to the order of the reaction. Therefore the rate-limiting step of self-association appears to involve the interaction of three OmpA$_{171}$ molecules. We attempted to fit the kinetic curves to the integrated rate equation for a third order reaction but this returned poor fits and varying values of rate constant (data not shown), suggesting the
kinetics are more complex and may involve multiple kinetic phases. However, it is still evident that the initial rate-limiting step involves three OmpA\textsubscript{171} molecules.

Interestingly, we never observed intermediate-sized species by sedimentation velocity so we reasoned that the protein could be undergoing a nucleated growth polymerization reaction where the nucleus is not appreciably populated (S. Chen et al., 2002). This type of reaction involves initial formation of an energetically unfavorable nucleus, followed by rapid elongation into polymer by the addition of monomer units. Analysis of the time-dependence of polymer formation allows determination of the critical nucleus size ($n^{*}$) necessary to facilitate downhill polymerization (see Materials and Methods). In brief, the concentration of monomers incorporated into polymer should have a linear dependence on time squared at very early time points. A double logarithmic plot of these slopes as a function of total monomer concentration will also be linear, with a slope equal to $n^{*}+2$. The resulting values determined from the OmpA\textsubscript{171} CD kinetics data are shown in Figure 2.9B and a linear fit to the data gives a slope of 4.6, or a critical nucleus of 2.6. Therefore ~three OmpA molecules must condense to form a nucleus that then undergoes further oligomerization to form higher molecular weight species, which we observe in the large s-value region of our SV data. This alternative analysis is highly consistent with the order of reaction we determined to be three, so it is clear that OmpA self-association involves the initial formation of a nucleus of ~three molecules before further association occurs.
Thioflavin T binding and electron microscopy indicate the oligomeric form has an amyloid fibril-like structure

Figure 2.2D shows that the self-associated form of the unfolded OmpA barrel has a CD spectrum indicative of β-sheet structure. Based on this observation and the kinetic data consistent with a nucleated growth polymerization mechanism, we hypothesized that the protein could be forming an amyloid-like configuration. Amyloid fibrils are highly ordered protein aggregates composed of a cross-β structure that are often associated with neurodegenerative conditions such as Alzheimer’s disease and Parkinson’s disease (Chiti & Dobson, 2006). Such structures can be identified by specific binding to the dye Thioflavin T (ThT); when bound to amyloid fibrils, ThT exhibits a novel excitation band ~450 nm with an intense fluorescence at 482 nm and it has been used extensively to indicate amyloid fibrils in vitro (LeVine, 1993; Naiki et al., 1989; Nilsson, 2004).

To determine if the OmpA oligomers bound ThT, we prepared mixtures of the protein and dye under self-associating and non-self-associating conditions and measured fluorescence excitation and emission spectra. Figure 2.10A shows excitation spectra for OmpA_{325} (top, green) and OmpA_{171} (bottom, red) in the presence of 3 μM ThT (emission collected at 482 nm). The solid lines correspond to a self-associating condition for both constructs as determined by SV (3 μM protein, 300 mM urea, 16 h incubation) while the dashed lines correspond to a condition where both proteins remain entirely monomeric (1 μM protein, 600 mM urea, measured immediately). It is clear that the monomeric condition gives excitation spectra similar to the spectrum for ThT alone (turquoise) while the self-associating condition induces an excitation peak at ~440 nm for both protein constructs. Figure 2.10B shows the emission spectra for the same conditions (excited at
450 nm) and it is evident that the self-associating condition exhibits a large emission peak at 482 nm that is not seen under the monomeric condition or for dye alone. We therefore conclude that the oligomeric form of the unfolded OmpA barrel domain binds ThT and has a cross-β, amyloid-like structure. Interestingly, both the excitation and emission peaks for oligomeric OmpA\textsubscript{171} (bottom) are larger than for oligomeric OmpA\textsubscript{325} (top), indicating a higher quantity of amyloid structure. This is consistent with the previous observation that OmpA\textsubscript{171} is more prone to self-association as seen by SV (Figure 2.6) and that the β-sheet structure forms more slowly in OmpA\textsubscript{325} (Figure 2.8).

An amyloid fibril-like structure for self-associated OmpA is further supported by the morphology observed by transmission electron microscopy. Representative images of unfolded OmpA\textsubscript{171} under associating conditions are shown in Figure 2.11 and short worm-like fibers are seen. The average width of these structures was found to be 12.9 (±2.2) nm, which is consistent with the typical amyloid fibril width of 7-13 nm (Chapman \textit{et al.}, 2002; Chiti & Dobson, 2006; Serpell \textit{et al.}, 2000). Interestingly, smaller spherical objects were also observed, which could correspond to the oligomeric precursors to amyloid fibrils, called protofibrils. Protofibril species with spherical or annular structures have been observed previously for many amyloid fibril forming proteins (Conway \textit{et al.}, 2000; Kayed \textit{et al.}, 2004; Lashuel \textit{et al.}, 2003; Shin \textit{et al.}, 2008; Walsh \textit{et al.}, 1999).

\textit{OmpA self-association reduces folding efficiency when there is a delay in the addition of membranes}

To determine the effect of oligomer formation on β-barrel folding, we conducted a delayed folding assay in which OmpA\textsubscript{325} and OmpA\textsubscript{171} were allowed to incubate for 30
minutes prior to the addition of diC_{10}PC LUVs. If self-association of unfolded states competes with folding, we reasoned the folding of OmpA_{171} should be reduced compared to OmpA_{325} because it has a greater propensity for self-association. Figure 2.12 shows the fraction folded of OmpA_{325} (panel A) and OmpA_{171} (panel B) at a total concentration of 2 μM or 5 μM in 600 mM urea, 20 mM Tris, pH 8. The first set of bars shows that both proteins at both concentrations fold with efficiencies close to one when there is no delay time (i.e., membranes are present when the protein is diluted into the folding mixture). When there is a 30 minute delay before the addition of vesicles, the second set of bars shows that the fraction folded is reduced to different extents. OmpA_{325} folding efficiency at 2 μM is reduced to 0.93 (±0.03) while folding at 5 μM is reduced to 0.72 (±0.03). These values correspond almost exactly to the fraction monomer measured under the same conditions by sedimentation velocity (shown in Figure 2.6A). OmpA_{171} folding efficiencies are reduced to 0.85 (±0.03) at 2 μM and 0.66 (±0.07) at 5 μM, which are also similar to the fraction monomer values measured by SV (Figure 2.6B), and are consistently lower than the corresponding fraction folded for OmpA_{325}. Altogether these data indicate that OmpA barrel folding successfully competes against oligomerization when membranes are present from the time of dilution into low urea, but when given the time to self-associate, oligomeric unfolded protein is unable to dissociate and fold within three hours and thus there is a reduction in folding efficiency. We also failed to observe an effect on the fraction folded when either OmpA_{325} or OmpA_{171} were incubated with equimolar concentrations of OmpA_{Per} before the addition of vesicles, as shown by the third set of bars in Figure 2.12. This result is consistent with the sedimentation velocity
results described above, and further demonstrates that the periplasmic domain acts in cis but not in trans to affect OmpA self-association.

_The folding competent unfolded conformations of OmpA_{171} and OmpA_{325} are expanded_

The conformation of the unfolded state is the reference point for the development of kinetic folding models. Importantly, it is the “folding competent” non-native state that would be populated under folding conditions that is the relevant conformation, not the conformation of the denatured state ensemble observed at high (>6 M) urea concentrations. Because membrane proteins require phospholipid vesicles to fold, we are uniquely suited to populate the relevant unfolded state by simply leaving out vesicles from the reaction. Because U_{AQ} states self-associate so strongly (shown above), we cannot directly visualize monomeric unfolded states in the absence of urea. However, we are able to tune the urea concentration to populate monomer and determine the sedimentation coefficient as a function of urea. At 2 μM total protein, we found that the sedimentation coefficients exhibit a linear dependence on urea concentration, which means we can extrapolate the values to the absence of denaturant. These data are shown in Figure 2.13, and the s*20,w values for OmpA_{325} and OmpA_{171} extrapolate to 2.34 S and 1.65 S, respectively. These values correspond to f/f₀ values of 1.68 and 1.60, respectively. We used Sednterp to interpret these data in terms of the simple molecular envelope of a prolate ellipsoid of revolution. Our results suggest expanded conformations for both of these proteins, with axial ratios of 8.3:1 and 7.3:1, respectively for OmpA_{325} and OmpA_{171}.
2.5 Discussion

*Unfolded conformations as a reference point for kinetic folding models*

Despite the extensive studies on OmpA folding, there are still kinetic phases and conformations that are not well understood. In particular, the self-association propensity of aqueous, unfolded OmpA has not previously been extensively investigated. We find that both the OmpA_{171} barrel and the full-length OmpA_{325} protein show significant propensities to form very large oligomeric structures when there is a delay in the addition of lipids. As previously shown for OmpT (Ebie Tan *et al.*, 2010), the weight average sedimentation coefficients of OmpA are in the 20-60 S range, comparable to the size of ribosomal subunits. In contrast, we expect that the thermodynamic stabilities of these OmpA oligomers must be less than the stabilities for OmpT U_{AQ} oligomers, as it requires much lower concentrations of urea to melt OmpA oligomers into monomers as compared to the previous studies on OmpT.

We have further shown that the unfolded conformation of the OmpA barrel has no regular structure. The interpretation of our CD data is in contrast to early studies implying that membrane-embedded regions of OmpA possessed a mixed \(\alpha/\beta\) secondary structure in aqueous solution. Rather, these secondary structure features of OmpA U_{AQ} arise from folding of its soluble periplasmic domain, which is stable independent of the OmpA transmembrane \(\beta\)-barrel. The “folding-competent” non-native state also has a CD signature quite distinct from that of the OmpA oligomers that form at higher concentrations over a time period of hours; the latter state has a broad negative peak in the beta region of the CD spectrum and lacks the 230 nm peak found in the native OmpA \(\beta\)-barrel. Combined with the applicability of a nucleated growth polymerization model,
the ability of the oligomers to specifically bind the dye Thioflavin T, and the morphology observed by TEM, it can be surmised that these are cross-β fibrils reminiscent of the structures that amyloidal proteins can form. This is perhaps not too surprising, as an increasing number of proteins not associated with protein aggregation diseases are being identified that can form amyloid fibrils, in some cases for functional purposes (Chapman et al., 2002; Chiti & Dobson, 2006). In addition, it has been shown that polypeptide sequences with a high propensity to form β-sheets, especially with alternating hydrophobic and hydrophilic residues (as are found in OMP β-barrels) are more prone to form amyloid fibrils (West et al., 1999).

A third aspect of the OmpA unfolded conformation that we investigated relates to its level of compactness. It had previously been proposed that OmpA325 forms a collapsed aqueous state that then folds into membranes (Kleinschmidt & Tamm, 1996; Surrey & Jähnig, 1995). However, this concept raises the question of how this conformation can then “uncollapse” to be able to partition onto membrane surfaces in a folding-competent state. Our sedimentation velocity data show, in fact, that the monomeric UAQ forms of both OmpA325 and OmpA171 adopt expanded conformations. This starting point for folding rationalizes the need for these proteins to effectively partition onto membranes in a productive structure.

*Competition between UAQ self-association and folding kinetics*

Another aspect of the in vitro folding pathway that remains to be investigated is the thermodynamic potential for partitioning of UAQ states onto the membrane surface. Certainly the data show that this reaction competes quite effectively because the proteins
do fold when exposed to membranes. Therefore, the consequences of $U_{AQ}$ self-associa-
tion will strongly depend on how quickly the proteins partition onto membranes
and whether or not all membrane binding occurs in folding-competent conformations.
Even if some membrane binding conformations are not productive for folding, membrane
binding will reduce the total aqueous OmpA concentration, which should diminish
oligomer formation. Moreover, early kinetic events that may be affected by $U_{AQ}$ self-
association will be essential to investigate in a systematic manner. Already, we observed
that the rate of OmpA$_{325}$ self-association occurs faster than the secondary structure
change that is ultimately indicative of the large oligomer formation. Multiple orthogonal
methods must be applied to fully dissect the kinetic details of these two processes. While
we did not observe any loss of folding efficiency over a three-hour period if membranes
were immediately available to $U_{AQ}$ OmpA$_{325}$ and OmpA$_{171}$, our data do not exclude an
effect on folding kinetics. It could be that $U_{AQ}$ oligomers slow the folding of both
OmpA$_{325}$ and OmpA$_{171}$. A future comparison of folding rate constants will be necessary
to address these questions.

*The OmpA periplasmic domain possesses chaperone activity when covalently attached to
the OmpA β-barrel*

Both OmpA$_{325}$ and OmpA$_{171}$ can fold *in vitro* upon the addition of synthetic
phospholipid vesicles, so it is evident that their sequences contain all the molecular
information needed to fold these proteins into their native conformations. However, along
with other OMPs, OmpA needs to remain in a folding-competent state in the cell until it
is sorted to the outer membrane in bacteria. Elaborate cellular machinery exists to ensure
proper membrane compartmentalization of these proteins (Selkrig et al., 2013). It is clear from the work in this study that self-association of the OmpA unfolded state represents a reaction that competes with folding. Although there is no loss in folding efficiency when membranes are present from the time of dilution to low urea, there is a decrease in folding when the proteins are incubated at low urea and self-association is allowed to occur before the addition of membranes. Indeed, in the cell the proteins encounter an environment more like the second situation as they must traverse the aqueous periplasm in the unfolded state before encountering the outer membrane. Self-association in vitro competes with folding even when folding is allowed to proceed for three hours, which is already a time period much longer than the doubling time of E. coli. Therefore the prevention of self-association while the unfolded OMPs proceed to the outer membrane is probably a major role for periplasmic chaperones (Denoncin et al., 2012).

In this study we discovered a novel chaperone function of the periplasmic domain of OmpA whereby it can act to increase folding efficiency by reducing self-association when present in cis (i.e., as part of OmpA_{325}). However, OmpA_{Per} is unable to reduce self-association when added separately to the barrel domain. We observed neither an increase in fraction monomer nor any indication of a stable complex formed between OmpA’s transmembrane and periplasmic domains when OmpA_{Per} is added in micromolar concentrations. Therefore the interaction between OmpA_{Per} and the OmpA barrel domain must be a relatively weak one. Nevertheless, the periplasmic domain does have an effect when covalently attached: the full-length OmpA_{325} self-associates less and displays less of a loss in folding efficiency than OmpA_{171} upon incubation without vesicles.
Even though the full-length OmpA<sub>325</sub> self-associates less than just the transmembrane β-barrel alone, both proteins form large oligomeric species at micromolar concentrations in low concentrations of urea. Therefore, we speculate that additional cellular factors must be involved in preventing OmpA self-interactions in vivo. Extrapolated to the absence of urea, the midpoint of oligomer formation of full-length OmpA is ~200 nM, which means that such assembly factors must bind with nanomolar affinities to thermodynamically compete with oligomerization. A likely candidate is the Skp protein, which has been shown to stoichiometrically bind to a number of outer membrane proteins at sub-micromolar concentrations, including OmpA (Bulieris et al., 2003; Walton et al., 2009). Recent work by the Fleming lab in fact found that Skp binds the OMPs OmpLA, OmpW, and PagP with dissociation constants ~11 nM (Moon et al., 2013). The cell therefore uses significant free energy to sequester OMPs into Skp:OMP complexes. Our data presented here on OmpA U<sub>AQ</sub> self-interactions rationalizes why the free energy of Skp binding needs to be so favorable.

2.6 Acknowledgements

This work was supported by grants from the NSF (MCB 0919868) and the NIH (R01 GM079440 and T32 GM008403). EJD is a recipient of an NSF graduate research fellowship. We thank Professor Blake Hill for the TEV expression plasmid and Mr. Michael Genualdi for cloning and purifying OmpA<sub>Per</sub>. We also acknowledge Michael McCaffery and Erin Pryce of the Johns Hopkins Integrated Imaging Center for assistance with electron microscopy.
2.7 Tables and Figures

Table 2.1 Primers used for cloning

<table>
<thead>
<tr>
<th>Protein</th>
<th>N-terminal (5’ → 3’)</th>
<th>C-terminal (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OmpA\textsubscript{325}\textsuperscript{a}</td>
<td>GGCATCCCATATGGCTCCG AAAGATAACACCTG</td>
<td>CGGGATCTTAAGCCTGCGG CTGAGTTAC</td>
</tr>
<tr>
<td>OmpA\textsubscript{171}\textsuperscript{b}</td>
<td>GGCATCCCATATGGCTCCG AAAGATAACACCTG</td>
<td>CGGGATCTCAACCAAAAAC GGTAGGAACACC</td>
</tr>
<tr>
<td>OmpA\textsubscript{Per}</td>
<td>GCCATATGCAGGGCGAAGC AGCTCCAG</td>
<td>GCCTCGAGCCCGGGTTAAGC CTGCAGGTGTTAC</td>
</tr>
</tbody>
</table>

\textsuperscript{a} The OmpA\textsubscript{325} nucleotide sequence was also mutated via QuikChange to remove an internal BamHI cut site. The nucleotide change (T687C, counting the A in the insert’s ATG start codon as nucleotide number 1) had no effect on the amino acid sequence. These were the primers used:

Forward: 5’-CTGAGCAACCTGGACCCGAAAGACGGTTC-3’
Reverse: 5’-GAACCGTCTTTTCGGGTCAGTTGCTCAG-3’

This construct has been published previously (Burgess et al., 2008).

\textsuperscript{b} These primers were published previously (Pautsch et al., 1999).
Table 2.2 Molecular weights, partial specific volumes, and extinction coefficients calculated in Sednterp

<table>
<thead>
<tr>
<th>Protein</th>
<th>Molecular Weight (Da)</th>
<th>Partial Specific Volume (ml/g)</th>
<th>Extinction Coefficient (M⁻¹ cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OmpA₃₂₅</td>
<td>35,304</td>
<td>0.7294 0.7273</td>
<td>50,210 a</td>
</tr>
<tr>
<td>OmpA₁₇₁</td>
<td>18,875</td>
<td>0.7241 0.7219</td>
<td>45,090 a</td>
</tr>
<tr>
<td>OmpA₇₆₆</td>
<td>16,772</td>
<td>0.7348 0.7327</td>
<td>5,960 b</td>
</tr>
</tbody>
</table>

a Sednterp calculates the extinction coefficient for unfolded protein using the method of Gill and von Hippel and the extinction coefficients of model amino acid chromophores in 6 M GdnHCl determined by Edelhoch (Edelhoch, 1967; Gill & von Hippel, 1989). We used this value for OmpA₃₂₅ and OmpA₁₇₁ because they were purified unfolded in 8 M urea.

b Sednterp calculates the extinction coefficient for folded protein in water using the values determined by Pace and co-workers (Pace et al., 1995). We used this value for OmpA₇₆₆ because it was purified folded in aqueous buffer.
Figure 2.1  Structural models of OmpA constructs

(A) An overlay is shown of the aligned RmpM structure (PDB id 1R1M, turquoise) and our OmpA Per homology model (blue). One region does not exactly overlay because RmpM has an extra turn of helix both preceding and following a turn, shown at the bottom of the structure. (B) Model of full-length OmpA (OmpA325) created by combining the homology model of OmpA Per (blue) and the crystal structure of OmpA171 (PDB id 1QJP, red). Tryptophan residues are shown as spheres. The membrane is shown in grey. The relative orientations of the two domains are unknown and arbitrarily placed. Structure images were made using MacPyMOL.
Figure 2.2  CD wavelength spectra of OmpA constructs

All spectra were measured at 25 ºC and are plotted as molar ellipticity. (A) OmpA\textsubscript{Per} under folding (solid blue) and unfolding (dotted blue) conditions. Folded OmpA\textsubscript{Per} was measured at a concentration of 40 µM in 20 mM Tris, pH 8, with a 1 mm path length. Unfolded OmpA\textsubscript{Per} was measured at a concentration of 20 µM in 6.1 M urea, 20 mM Tris, pH 8, with a 1 mm path length. (B) Spectra of folded proteins: OmpA\textsubscript{171} (red), OmpA\textsubscript{Per} (blue), and OmpA\textsubscript{325} (green) were incubated in β-barrel folding conditions for 6 h (1 µM protein, 800 µM diC\textsubscript{10}PC LUVs, 600 mM urea, 20 mM Tris, pH 8, 25 ºC). Data were collected with a 1 cm path length. The spectrum for an LUV-only sample was used.
for background subtraction. The mathematical sum of OmpA_{171} and OmpA_{Per} is shown as a purple dashed line. (C) and (D) Spectra for unfolded OmpA_{171} and folded OmpA_{Per}. In (C), protein samples were prepared at 1 µM in 600 mM urea, 20 mM Tris, pH 8 (monomer conditions) and measured with a 1 cm path length. In (D), protein samples were prepared at 10 µM in 600 mM urea, 20 mM Tris, pH 8 (self-associating conditions) and allowed to incubate at room temperature overnight before measurements were made with a 1 mm path length. Curves are colored as in (B).
Figure 2.3  Unfolding titration of OmpA<sub>Per</sub> measured by CD

Folded OmpA<sub>Per</sub> at a starting concentration of 8 µM in 20 mM Tris, pH 8 was titrated with 8 M urea in sequential steps. The CD signal was monitored at 220 nm and corrected for protein dilution. The data (open circles) were fit to a two-state model of unfolding (black line) and the linear extrapolation method used to determine the energy for folding in the absence of denaturant. Four titrations were performed and gave an average $\Delta G_f^0$ equal to -6.2 (±0.1) kcal mol<sup>-1</sup> with an m-value of -1.44 (±0.04) kcal mol<sup>-1</sup> M urea<sup>-1</sup>. 
Figure 2.4  Typical primary sedimentation velocity scans

Data are shown for 3 µM OmpA_{325} in 450 mM urea, 2 mM TCEP, 20 mM Tris, pH 8. The sample was sedimented at 50,000 rpm and 25 °C. Scans are colored by time with the earliest scans being purple and the later scans being red. Two boundaries are evident, demonstrating a fast moving species and a slower moving species, which are easily distinguishable by g(s*) and c(s*) analysis (see Figure 2.5). This figure was made using DCDT+.
Figure 2.5  Typical sedimentation velocity data analyzed by g(s*) and c(s*)

Sedimentation velocity data from Figure 2.4 (3 µM OmpA_{325} in 450 mM urea, 2 mM TCEP, 20 mM Tris, pH 8) were analyzed by (A) the time derivative g(s*) method using DCDT+, and (B) the continuous c(s*) method using Sedfit. Inset panels show a zoom-in of the high s region. Integration of both curves gives 75% monomer (s-value 2.35).
Figure 2.6 OmpA self-association as a function of concentration and urea

Varying concentrations of OmpA$_{325}$ (A) and OmpA$_{171}$ (B) were prepared at 300 mM (red triangles), 450 mM (orange circles), or 600 mM (gold diamonds) urea. Samples were incubated at 25 °C for 30 min before beginning centrifugation. Sedimentation velocity data were analyzed by the time derivative method and the g(s*) curves integrated to determine the fraction of monomeric protein at each concentration. Solid lines show fits to the Hill equation. (C) The midpoints of the Hill fits are plotted as a function of urea, fitted to a line and extrapolated to the absence of urea. These intercepts are 0.2 µM and -1 µM for OmpA$_{325}$ and OmpA$_{171}$, respectively.
Figure 2.7  The periplasmic and barrel domains do not interact \textit{in trans} (A) and (C) $g(s^*)$ curves for 20 $\mu$M OmpA$_{Per}$ and 5 $\mu$M OmpA$_{325}$ (A) or 5 $\mu$M OmpA$_{171}$ (C) measured individually (in 600 mM urea, 20 mM Tris, pH 8). Absorbance data were collected at 280 nm and analyzed in DCDT+. (B) and (D) $g(s^*)$ curves for mixtures of 10 $\mu$M OmpA$_{Per}$ and 5 $\mu$M OmpA$_{325}$ (B) or 5 $\mu$M OmpA$_{171}$ (D) overlaid with calculated curves for non-interacting mixtures (the sums of the scaled sedimentation coefficient distributions from (A) or (C)). In both cases the measured and calculated curves agree well, showing that OmpA$_{Per}$ is unable to form a stable complex with the full-length or barrel-only forms of OmpA. Experiments were also performed with higher ratios of OmpA$_{Per}$ (up to a four-fold molar excess) and the same result was obtained.
Figure 2.8  Time-dependence of OmpA self-association

Fraction monomer is plotted as a function of time for various total protein concentrations of OmpA$_{325}$ (A) and OmpA$_{171}$ (B). All samples were measured in 600 mM urea, 20 mM Tris, pH 8, and 25 °C. Open diamonds correspond to values measured by sedimentation velocity while solid lines show the CD signal at 218 nm converted to fraction monomer. Concentrations are denoted by color and are listed in the figure key. Dashed lines in (A) are shown only to guide the eye along the SV data.
Figure 2.9  Analysis of OmpA$_{171}$ CD kinetics

(A) Double logarithmic plot of the initial rate of monomer disappearance as a function of initial monomer concentration. A linear fit of the data gives a slope of 3.2, which is interpreted as the order of the reaction. (B) Analysis of association kinetics by a nucleated growth polymerization model. The slope of the double logarithmic plot corresponds to $n^*+2$, where $n^*$ is the critical nucleus size. This was determined to be 4.6, giving a critical nucleus of 2.6 OmpA molecules.
Figure 2.10 Thioflavin T binding indicates cross-β structure

(A) Excitation spectra for 3 μM Thioflavin T in the presence of unfolded OmpA_{325} (top, green) or OmpA_{171} (bottom, red). Solid lines correspond to conditions that promote self-association for both constructs (3 μM protein, 300 mM urea, 16 h incubation) and dashed lines correspond to conditions that maintain the monomeric state for both constructs (1 μM protein, 600 mM urea, measured immediately). The spectrum for ThT alone is shown in turquoise. Emission was collected at 482 nm. (B) Emission spectra for the same mixtures (excited at 450 nm).
Figure 2.11 TEM images of oligomeric OmpA$_{171}$

Transmission electron microscopy images of unfolded OmpA$_{171}$ under self-associating conditions (3 µM protein, 300 mM urea, 7 h incubation). The sample was adsorbed to a freshly ionized formvar/carbon-coated grid for 5 min and negatively stained with 2% uranyl acetate for 1 min. Images were captured using an FEI Tecnai 12 electron microscope equipped with an SIS Megaview III camera. Black bars in both images represent 100 nm.
Figure 2.12 Delayed folding of OmpA$_{325}$ and OmpA$_{171}$

Fraction folded is plotted for all samples after a delay of 0 or 30 min followed by folding for 3 h. Values are the average of 3-6 separate folding samples and error bars are the calculated standard deviation. OmpA$_{325}$ (A) and OmpA$_{171}$ (B) samples were prepared at 2 µM or 5 µM in 600 mM urea and incubated at 25 °C. After the appropriate delay, diC$_{16}$PC LUVs were added at a final lipid to protein ratio of 800:1 and folding was allowed to occur for 3 h with gentle stirring before quenching by the addition of 5X SDS-PAGE loading buffer to a final concentration of 1X. Fraction folded was determined by densitometry of folded and unfolded bands. Data labeled “+Per” indicate samples were incubated during the 30 min delay with equimolar concentrations of OmpA$_{Per}$ (2 µM or 5 µM).
Figure 2.13 Sedimentation coefficients extrapolated to the absence of urea

$s_{20,w}^*$ values from single-species fits of $g(s^*)$ curves for OmpA$_{325}$ and OmpA$_{171}$ are plotted as a function of urea concentration. OmpA$_{171}$ (red squares) was monitored in urea concentrations ranging from 1 M to 8 M. OmpA$_{325}$ (green triangles) was monitored in urea concentrations of 1 M to 2.5 M, in order to measure the sedimentation coefficient in the region below the unfolding transition of the periplasmic domain. The data were fit to linear functions and the sedimentation coefficients extrapolated to a urea concentration of zero. These values were determined to be 2.34 S for OmpA$_{325}$ and 1.65 S for OmpA$_{171}$.
Chapter 3

The OmpA transmembrane β-barrel folds via a multi-step mechanism with a partially inserted intermediate state containing extensive β-sheet structure

3.1 Abstract

We have examined the intrinsic folding process for the outer membrane β-barrel protein OmpA from *E. coli*. Numerous β-barrel proteins have been shown to be capable of spontaneously folding and inserting into lipid bilayers in the absence of folding chaperones or machinery, indicating that there is an inherent pathway by which these proteins fold, dictated by the amino acid sequence (as for soluble proteins) and the membrane environment. Previous studies on OmpA have identified several membrane-associated intermediate states along the folding pathway and a final concerted translocation mechanism across the bilayer. However, some aspects of the folding process are still unclear, such as the influence of the soluble periplasmic domain on the observed signal, the extent of secondary structure formation in the intermediate states, and the mechanism that gives rise to multiple exponential phases in the folding kinetics. We have addressed these questions by investigating the folding kinetics of the OmpA transmembrane β-barrel domain in the absence of the periplasmic domain, into a range of lipids. By varying bilayer thickness we were able to observe different regions of the folding pathway, with the fastest folding into the thinnest bilayers providing information on the later stages of the process, and the slowest folding into thicker bilayers revealing
the early kinetic steps. Folding was monitored using SDS-PAGE and CD spectroscopy, which provide complementary information about tertiary and secondary structure formation. We globally fit the folding data to kinetic schemes and found that in all lipid conditions the same core pathway was followed. Based on our data and previous work, we propose a detailed multi-step folding mechanism for the OmpA β-barrel that includes unstructured surface-adsorbed states converting through a partially inserted state with substantial β-sheet structure to the final natively inserted barrel. Utilization of bilayer defects and the initiation of β-hairpin formation are thought to be key facilitators of the folding and insertion process. We have also identified several off-pathway misfolded intermediate states and determined that the population of such species contributes to the complexity of the observed multi-exponential kinetics. These off-pathway states are expected to be suppressed in vivo where the BAM folding machinery accelerates the productive folding pathway enough to be the dominant route for substrate OMPs.

3.2 Introduction

Membrane proteins serve a variety of essential functions in biological systems, such as transport of metabolites into and out of cells, transduction of signals, and enzymatic reactions. The importance of these proteins is demonstrated by the fact that 20-30% of all genes in many organisms’ genomes encode membrane proteins (Krogh et al., 2001), and more than 50% of medical drugs currently on the market target membrane proteins (Overington et al., 2006). In addition, many diseases have been linked to mutations that result in the misfolding or incorrect targeting of membrane proteins.
(Sanders & Myers, 2004). Therefore, understanding the process by which these proteins fold to the correct tertiary structure and insert into their native bilayer environment is of utmost importance.

The need to satisfy all backbone H-bonds within the water-excluding bilayer environment has restricted transmembrane proteins to two types of structures: α-helical and β-barrel. Whereas α-helical membrane proteins are ubiquitous in most biological membranes, β-barrel membrane proteins are found only in the outer membranes of Gram-negative bacteria (Koebnik et al., 2000), mitochondria (Paschen et al., 2005), and chloroplasts (Schleiff, Eichacker, et al., 2003). These outer membrane proteins (OMPs) often serve as porins for small molecules but can have many other functions such as membrane biogenesis and maintenance, bacterial virulence, and antibiotic resistance (Koronakis et al., 2004; Tamm et al., 2004; Wimley, 2003). Additionally, dysfunction of mitochondrial OMPs has been implicated in several human diseases, including diabetes and Parkinson’s disease (Bender et al., 2013; Sasaki et al., 2012).

With their myriad functions and connections to disease, it is of significant interest to understand the folding pathway for β-barrel membrane proteins. In bacteria, OMPs are known to interact with a number of periplasmic chaperones en route to the outer membrane (Denoncin et al., 2012), and assembly into the membrane is catalyzed by the β-barrel assembly machinery, or BAM complex (T. Wu et al., 2005), but the details of how these proteins facilitate OMP folding are unclear. OMPs are also capable of spontaneously folding to the native conformation in lipid bilayers in the absence of any folding chaperones (Surrey & Jähnig, 1992), but on timescales too slow to be compatible with bacterial growth rates (Kleinschmidt & Tamm, 1996; Surrey & Jähnig, 1995),
especially in the presence of lipids native to the bacterial outer membrane (Gessmann et al., 2014). Therefore it appears that the BAM complex does not provide conformational instructions to substrate OMPs, but rather lowers the kinetic barrier to intrinsic OMP folding, possibly by inducing defects in the bilayer that OMPs utilize to insert and fold (Gessmann et al., 2014).

Great progress has been made in elucidating the intrinsic mechanism of OMP folding based on studies of the structural protein OmpA from *E. coli*. Monitoring OmpA folding into unilamellar vesicles using a variety of techniques has revealed that the protein folds via a multi-step mechanism with several membrane-associated intermediate states (Kleinschmidt & Tamm, 1996; Surrey & Jähnig, 1995). Additionally, it has been shown that β-barrel formation occurs in concert with membrane insertion, with the β-hairpins translocating the bilayer at the same time (Kleinschmidt, den Blaauwen, et al., 1999). However, there are still several aspects of the folding mechanism that have not been determined. For example, the conformation and interactions of the unfolded state for OmpA were not well defined, with the proposal of a partially folded aqueous state being complicated by the presence of the soluble periplasmic domain. We addressed this issue in Chapter 2, demonstrating that the periplasmic domain is an independently folding globular protein domain and that the transmembrane β-barrel domain of OmpA has no regular structure and is expanded in the absence of a membrane (Danoff & Fleming, 2011).

In addition, the conformations of the membrane-bound folding intermediates are unclear, particularly the extent of β-sheet formation. Early work on OmpA identified a trapped membrane-adsorbed state in the presence of gel phase lipid bilayers with the
same amount of β-structure as the native state (Rodionova et al., 1995; Surrey & Jähnig, 1992), but again these measurements were obscured by the strong secondary structure signal from the folded periplasmic domain. Kinetics measurements have indicated that β-sheet structure (measured by CD) and tertiary structure (measured by SDS-PAGE, see below) form at the same rate during OmpA folding (Kleinschmidt & Tamm, 2002a). This finding is consistent with the concerted mechanism of barrel formation, but does not address the presence of β-structure in intermediate states that lack tertiary structure.

Finally, it is uncertain whether parallel pathways occur in the OMP folding mechanism, as has been proposed for the OMPs FomA and PagP (Huysmans et al., 2012; Pocanschi et al., 2006). Multi-exponential kinetics were observed for the formation of these two membrane proteins, which can be indicative of independent unrelated reaction pathways. However, the presence of off-pathway intermediate states (i.e., optional misfolding) can also account for such behavior (Krishna & Englander, 2007).

To address these questions we have investigated the folding kinetics of OmpA into bilayers of different thicknesses and monitored the reaction using CD spectroscopy or SDS-PAGE. We conducted our measurements on OmpA171, which comprises the N-terminal β-barrel domain of OmpA, to eliminate contributions to the CD signal from the periplasmic domain. This is possible because we and others have shown that the two domains fold independently of one another (Danoff & Fleming, 2011; Pautsch & Schulz, 1998; Surrey & Jähnig, 1992). Additionally, through our studies in Chapter 2 of the oligomerization of the aqueous unfolded state (Danoff & Fleming, 2011), we determined conditions under which the barrel domain does not self-associate. This has enabled us to examine the folding reaction without the influence of the competing off-pathway
oligomerization reaction — a condition that more closely resembles the in vivo situation where periplasmic chaperones presumably suppress self-association prior to membrane insertion.

Interestingly, our SDS-PAGE measurements have led us to identify a novel conformation(s) for the OmpA β-barrel that migrates anomalously by gel, and which has not been discussed in the literature previously. We included this state in comprehensive kinetic modeling of the folding pathway based on global fitting of SDS-PAGE and CD kinetics data.

Through the modeling process we have identified three membrane-associated intermediate states the protein progresses through on the way to the native state and determined that the third intermediate state possesses substantial β-sheet structure. We also determined that the observed kinetics are consistent with the population of several off-pathway, presumably misfolded, intermediate states, and the presence of parallel pathways is not required. This result is in agreement with the protein folding theory of predetermined pathways with optional errors (PPOE) proposed previously (Krishna & Englander, 2007). We conclude this chapter by interpreting our kinetic model from a structural perspective. Building upon the previous model for OmpA folding, our data and modeling enable a more comprehensive description of the multi-step β-barrel formation and the driving forces for membrane insertion.
3.3 Materials and Methods

Preparation of OmpA171

The cloning and expression of the transmembrane barrel domain of OmpA (OmpA171) was described in Chapter 2. The protein was purified in Urea Buffer (8 M urea, 20 mM Tris, pH 8) as described in Chapter 2, except Zeba Spin Desalting Columns (Thermo Scientific) were used to desalt the protein instead of gel filtration. The protein stock concentration was determined by measuring the absorbance at 280 nm and using a calculated extinction coefficient based on the values determined by Pace and co-workers for model amino acid chromophores in 8 M urea (Pace et al., 1995). The extinction coefficient for OmpA171 (which has 5 tryptophans and 13 tyrosines) in 8 M urea was calculated to be 45,075 M$^{-1}$ cm$^{-1}$. The final protein concentration was typically ~100 µM. Purified protein was divided into 100 µl aliquots and stored at -80 ºC until use.

Determination of OmpA171 extinction coefficient in 1 M urea

The molar extinction coefficient for OmpA171 in 1 M urea was determined using the Edelhoch method (Edelhoch, 1967; Gill & von Hippel, 1989) and the calculated molar extinction coefficient in 8 M urea. In brief, OmpA171 was diluted into 8 M and 1 M urea (20 mM Tris, pH 8) at identical concentrations and the absorbance of both samples measured at 280 nm. The extinction coefficient for each buffer condition is related to the absorbance and concentration according to the Beer-Lambert Law:

\[ A_{8M} = \varepsilon_{8M}cl \]

\[ A_{1M} = \varepsilon_{1M}cl \]

Equation 3.1
where $A$ is absorbance, $\varepsilon$ is extinction coefficient, $c$ is concentration, and $l$ is path length.

For equal concentrations and path lengths, as was the case for the two dilutions made, the extinction coefficients are related by the respective absorbance values:

$$\frac{A_{8M}}{\varepsilon_{8M}} = \frac{A_{1M}}{\varepsilon_{1M}}$$  \hspace{1cm} \text{Equation 3.2}

This can be rearranged to give an expression for the extinction coefficient in 1 M urea:

$$\varepsilon_{1M} = \frac{A_{1M}}{A_{8M}} \varepsilon_{8M}$$  \hspace{1cm} \text{Equation 3.3}

The absorbance at 280 nm was measured for multiple dilutions in each buffer condition (four samples in 8 M urea and eight samples in 1 M urea) and the average values used along with the calculated $\varepsilon_{8M} = 45,075$ to calculate $\varepsilon_{1M} = 42,900 \text{ M}^{-1} \text{ cm}^{-1}$.

**Vesicle preparation**

Phosphatidylcholines of various chain lengths (diC$_9$PC, diC$_{10}$PC, diC$_{11}$PC, diC$_{12}$PC, diC$_{13}$PC, and diC$_{14}$PC) were purchased from Avanti Polar Lipids dissolved in chloroform. Lipids were dried to a thin film in glass vials (5 mg per vial) under a gentle stream of nitrogen gas. The lipid films were evacuated overnight to remove residual solvent and stored at -20 °C until use. For vesicle preparation, lipid films were reconstituted in 1 M urea, 20 mM Tris, pH 8 at a concentration of 10 mg ml$^{-1}$ (exact urea concentration determined by refractometry (Warren & Gordon, 1966)) and incubated for at least 30 min, with periodic vortexing. Large unilamellar vesicles (LUVs) were made by extruding reconstituted lipids 35 times through a 0.1 µm filter using a mini-extruder (Avanti)(Hope et al., 1985). For diC$_{14}$PC, lipids were incubated and extruded at 35 °C in order to be well above the phase transition temperature at 24 °C (see Figure 3.5).
**Transmission electron microscopy of vesicles**

LUVs of each lipid type were prepared as described above. Samples were diluted to 0.5 mg ml\(^{-1}\) in 1 M urea, 20 mM Tris, pH 8. Negatively stained samples were prepared for transmission electron microscopy (TEM) using freshly ionized formvar/carbon-coated copper grids. Diluted LUV samples were adsorbed to grids for 5 min, then washed six times with water. The grids were stained with 2\% uranyl acetate for 1 min before removing excess stain. The samples were observed in an FEI Tecnai 12 TWIN electron microscope operating at 100 kV. Images were captured using an SIS Megaview III camera (Olympus).

**Differential scanning calorimetry of diC\(_{14}\)PC**

Differential scanning calorimetry (DSC) was conducted using a VP-DSC microcalorimeter (MicroCal). LUVs of diC\(_{14}\)PC were prepared in 1 M urea as described above, and diluted to a concentration of 800 \(\mu\)M. DSC scans on the lipid sample were measured from 10 °C to 50 °C at a rate of 90 °C/h with a pre-scan thermostat time of 15 min. 1 M urea (in 20 mM Tris, pH 8) was used as the reference buffer. Multiple samples were measured and each sample was scanned multiple times, with no change in the observed data. DSC data were buffer-corrected and converted to heat capacity using Origin software.

**Folding kinetics measured by SDS-PAGE**

In all experiments, OmpA\(_{171}\) was folded into pre-formed LUVs in the standard buffer condition of 1 M urea, 20 mM Tris, pH 8 (Folding Buffer). The final protein
concentration was 1 µM and the final lipid concentration 800 µM (for an 800:1 lipid to protein ratio), with a final volume of 2.2 ml. To measure folding kinetics, OmpA\textsubscript{171} was rapidly diluted from Urea Buffer to Folding Buffer (8 M urea to 1 M urea) with constant stirring, and then LUVs were added to initiate folding. Folding samples were incubated in a custom-built stirring incubator (Aviv Biomedical) at 25 ºC with continual stirring for the duration of the experiment. For folding into diC\textsubscript{14}PC, measurements were also conducted at 24 ºC and 26 ºC. Aliquots were removed at specific time points after initiation (typically 5, 12, 30, 90, 180, 360, 720, 1800, 3600, 7200, 10800, 14400, and 18000 s) and folding quenched by mixing with 4X SDS gel-loading buffer to a final concentration of 1X (loading buffer did not contain reducing agent)(Sambrook et al., 1989). A duplicate sample was taken at the last time point and boiled at 95-100 ºC in SDS loading buffer for 5 min. All other time points remained at room temperature after mixing with SDS loading buffer.

The fraction folded was determined for each time point by subjecting the samples to SDS-PAGE (Laemmli, 1970) at 4 ºC using pre-cast 12% acrylamide gels (BioRad) and staining with Coomassie Blue R-350 (GE Healthcare). Gels were scanned at 1200 dpi using an Epson 4490 scanner in positive film mode. Densitometry was performed using ImageJ software.

Folding samples were maintained in the stirring incubator for 24 h and then transferred to microcentrifuge tubes at room temperature and longer-term time points collected over the course of 7 days. All samples were analyzed by SDS-PAGE the same day they were collected.
Calculating fraction folded, fraction unfolded, and fraction gone

OMPs have long been known to exhibit differential migration of the folded and unfolded forms when undergoing SDS-PAGE (Nakamura & Mizushima, 1976). Densitometry was conducted on the OmpA_{171} folding kinetics gels to determine the intensities of the “folded” and “unfolded” bands at each time point (see Figure 3.1). In the field in the past, fraction folded ($f_F$) and fraction unfolded ($f_U$) have typically been calculated as the intensity of the folded or unfolded band divided by the sum of the intensities of the folded and unfolded bands (Kleinschmidt & Tamm, 1996, 2002a):

$$f_F = \frac{F}{F + U} \quad \text{Equation 3.4}$$

$$f_U = \frac{U}{F + U} \quad \text{Equation 3.5}$$

where $F$ and $U$ denote intensities of the respective bands measured by densitometry.

Dividing by the sum of the folded and unfolded bands makes the assumption that this quantity represents the total amount of protein present in the sample and that it is not changing over time. However, we observed in our experiments that in many cases “F+U” decreased over time, indicating the population of another state that does not migrate with the typical folded and unfolded bands by SDS-PAGE (see Figure 3.1). For this reason we calculated fractional quantities by dividing by the intensity of a boiled sample, which more accurately represents the total amount of protein (upon boiling, all protein migrates with the unfolded band):

$$f_F = \frac{F}{B} \quad \text{Equation 3.6}$$

$$f_U = \frac{U}{B} \quad \text{Equation 3.7}$$
The anomalously migrating form of the protein that “disappears” from the folded and unfolded bands was termed the “gone” state and was quantified as the difference between the boiled sample and the sum of the folded and unfolded bands. The fraction gone ($f_G$) was calculated as:

$$f_G = \frac{B - (F + U)}{B} = 1 - \frac{(F + U)}{B}$$  \hspace{1cm} \text{Equation 3.8}

For all folding conditions, final fractional data are the averaged values from at least three independent folding reactions. Reported errors are the standard deviation of the averaged values.

**Circular dichroism**

CD measurements were conducted using an Aviv Circular Dichroism Spectrometer, Model 410 (Aviv Biomedical), with a custom inset detector to reduce the effects of light scattering. The temperature was maintained at 25 ºC for all measurements, except where noted otherwise. Samples were temperature-equilibrated with gentle stirring for 10-15 min before each set of measurements.

**Determination of mean residue ellipticity for unfolded OmpA_{171}**

CD is defined as the difference in absorbance of left-handed ($A_l$) and right-handed ($A_r$) circularly polarized light ($\Delta A$):

$$\Delta A = A_l - A_r$$  \hspace{1cm} \text{Equation 3.9}

The molar CD ($\Delta \varepsilon$) is defined similarly to the molar extinction coefficient and is essentially the CD normalized for concentration and path length. It can also be thought of
as the difference in extinction coefficients for left-handed and right-handed circularly polarized light (Woody, 1995):

\[ \Delta \varepsilon = \frac{\Delta A}{cl} = \varepsilon_l - \varepsilon_r \quad \text{Equation 3.10} \]

The Aviv CD instrument reports ellipticity (θ) in millidegrees, which is related to ΔA and Δε as follows:

\[ \theta = 32982(\Delta A) = 32982(\Delta \varepsilon)cl \quad \text{Equation 3.11} \]

Similar to molar CD, molar ellipticity ([θ]) is normalized for concentration and path length. It is typically converted to units of deg cm² dmol⁻¹ and has the following form:

\[ [\theta] = \frac{\theta}{10cl} = 3298.2(\Delta \varepsilon) \quad \text{Equation 3.12} \]

In the protein folding field, mean residue ellipticity is used ([θ]), which is the molar ellipticity divided by the number of residues in the protein (n):

\[ [\theta] = \frac{\theta}{10cln} = \frac{3298.2(\Delta \varepsilon)}{n} \quad \text{Equation 3.13} \]

Mean residue ellipticity has units of deg cm² dmol⁻¹ res⁻¹.

The mean residue ellipticity spectrum for unfolded OmpA₁₇₁ was determined by measuring the CD spectrum for OmpA₁₇₁ in 1 M urea (20 mM Tris, pH 8) and converting it to mean residue ellipticity using a concentration value directly measured through absorbance at 280 nm, with the concentration calculated using the extinction coefficient for OmpA₁₇₁ in 1 M urea described above. The number of residues, n, for OmpA₁₇₁ is 172, including the N-terminal methionine (see Chapter 2 for cloning details). These measurements were made for four separately prepared and measured samples. The full
wavelength spectrum for unfolded OmpA_{171} is shown in Figure 3.7 and the values at the wavelengths 216 nm and 230 nm are reported in Table 3.1.

**CD: wavelength spectra**

Wavelength spectra were recorded between 205 and 280 nm in 1 nm increments, with an averaging time of 5 s. For each sample, 3-5 scans were recorded and averaged. Hellma cuvettes with a path length of 1 cm were used. Spectra of cuvettes containing only buffer were subtracted from sample spectra to correct for background signal. For samples containing LUVs, spectra of LUV-only mixtures were used for background subtraction.

For samples where the wavelength spectrum needed to be measured as quickly as possible to capture the initial conformation, the data were collected in 10 nm sections. Fresh folding mixtures were prepared for each wavelength section and the signal recorded in 1 nm increments with a 2 s averaging time. Data were averaged for 4 or 5 separate folding mixtures for each wavelength section, and the final data combined to produce the full spectrum.

**CD: folding kinetics**

OmpA_{171} folding kinetics were measured by CD under the same conditions as for SDS-PAGE: 1 μM protein and 800 μM lipid, in Folding Buffer (1 M urea, 20 mM Tris, pH 8). Samples were stirred continuously during measurements. The CD signal was monitored at either 216 nm or 230 nm, with an interval between data points of 10 s and a time constant of 1 s. For each folding reaction, a baseline signal was measured for 3 min.
of Folding Buffer alone before the addition of protein or LUVs. The protein was then added (and rapidly diluted from 8 M urea to 1 M urea) and the signal measured for 2 min. LUVs were then added to initiate folding. The LUVs were prepared in 1 M urea so there was no further urea dilution concurrent with the initiation of folding. The folding kinetics were then monitored until completion (sample raw data are shown in the top panels of Figure 3.10). The reported kinetics data at 216 nm are the average values of measurements on 2 or 3 separate samples.

The average buffer-alone signal was used for background correction of the subsequent folding data and the LUV contribution to the folding signal was subtracted using a value measured for a separate LUV-only sample in Folding Buffer (data that have been background-corrected are shown in the middle panels of Figure 3.10). The average protein-alone signal was used to calculate the exact protein concentration of the sample, using the previously determined mean residue ellipticity for that wavelength:

\[
c = \frac{\theta}{10[\Theta]ln}
\]

Equation 3.14

where \(\theta\) is the measured signal, \([\Theta]\) is the mean residue ellipticity for that wavelength, \(l\) is the path length (1 cm), and \(n\) is the number of residues in OmpA_{171} (172). The concentration was corrected for the subsequent dilution caused by the addition of LUVs, and used to convert the folding data to mean residue ellipticity using Equation 3.13 (data that have been converted to mean residue ellipticity are shown in the bottom panels of Figure 3.10). The time-values for the kinetics data were also corrected by subtracting the time of LUV addition to obtain a curve with folding beginning at time zero.
After the completion of each kinetics run, a wavelength spectrum was collected for the folding sample. The spectrum was LUV-corrected and converted to mean residue ellipticity as described above.

**Kinetic modeling**

For each set of kinetics data for folding into a particular lipid, a kinetic mechanism was developed that was consistent with the gel and CD data. This was accomplished through simulation of various kinetic schemes and manual adjustment of microscopic rate constants until the best fit to the data was achieved. Complex kinetic schemes were simulated using Euler’s method of numerical integration as described below.

The rate laws for a particular reaction scheme composed of unimolecular steps can be written as a series of coupled linear differential equations (Nölting, 2005). For example, the simple case of a single-step, reversible reaction with the forward and reverse rate constants $k_1$ and $k_{-1}$, respectively:

$$\begin{align*}
A & \overset{k_1}{\underset{k_{-1}}{\rightleftharpoons}} B \\
\end{align*}$$

Equation 3.15

The rate laws for this mechanism are expressed by the following differential equations:

$$\frac{d[A]_t}{dt} = -k_1[A]_t + k_{-1}[B]_t$$

Equation 3.16

$$\frac{d[B]_t}{dt} = k_1[A]_t - k_{-1}[B]_t$$

where $[A]_t$ and $[B]_t$ are the concentrations of species A and B at time $t$, respectively.

This set of equations can be represented in matrix form:
\[
\frac{dY_t}{dt} = \mathcal{C}Y_t
\]

where \(Y_t\), the concentration vector, and \(\mathcal{C}\), the rate matrix, are defined as:

\[
Y_t = \begin{pmatrix} [A]_t \\ [B]_t \end{pmatrix}
\]

\[
\mathcal{C} = \begin{pmatrix} -k_1 & k_{-1} \\ k_1 & -k_{-1} \end{pmatrix}
\]

The solutions to the differential equations above can be obtained by using linear algebra to find the eigenvalues and eigenvectors of the rate matrix (Berberan-Santos & Martinho, 1990). For this simple mechanism, the following expressions are obtained for the concentration of each species as a function of time:

\[
[A]_t = \frac{([A]_0 + [B]_0)k_{-1}}{k_1 + k_{-1}} + \left(\frac{[A]_0k_1 - [B]_0k_{-1}}{k_1 + k_{-1}}\right)e^{-(k_1+k_{-1})t}
\]

\[
[B]_t = \frac{([A]_0 + [B]_0)k_1}{k_1 + k_{-1}} + \left(\frac{[B]_0k_{-1} - [A]_0k_1}{k_1 + k_{-1}}\right)e^{-(k_1+k_{-1})t}
\]

where \([A]_0\) and \([B]_0\) are the starting concentrations of species A and B, respectively.

If the kinetic mechanism is composed entirely of unimolecular steps, then there always exists an analytical solution. However, for more complex reaction mechanisms, it becomes impossible to obtain the analytical solution due to the inability to extract the roots of higher order polynomials, which is necessary when determining the eigenvalues.

Instead, the method of numerical integration can be used to obtain the time-dependence of all species. The simplest type of numerical integration is Euler’s method, which makes use of the following approximation (Berberan-Santos & Martinho, 1990):

\[
\frac{dY_t}{dt} \approx \frac{\Delta Y_t}{\Delta t}
\]

where \(\Delta Y_t\) is the change in the concentration vector over some small time interval, \(\Delta t\):

\[
\frac{dY_t}{dt} = \mathcal{C}Y_t
\]
\[ \Delta Y_t = Y_{t+\Delta t} - Y_t \]  

Equation 3.22

Combining Equation 3.21 and Equation 3.22 with Equation 3.17, we obtain the following expression:

\[ \frac{\Delta Y_t}{\Delta t} = \frac{Y_{t+\Delta t} - Y_t}{\Delta t} = CY_t \]  

Equation 3.23

which can be rearranged to:

\[ Y_{t+\Delta t} = \Delta t (CY_t) + Y_t \]  

Equation 3.24

\[ Y_{t+\Delta t} = [\Delta t C + I] Y_t \]

where \( I \) is the identity matrix. This equation can be used to calculate the concentrations of all species over time, in steps of \( \Delta t \). An appropriate \( \Delta t \) must be chosen based on the magnitudes of the rate constants.

For the single-step, reversible mechanism examined above, the following expressions are obtained using Euler’s method:

\[ [A]_{t+\Delta t} = (-k_1\Delta t + 1)[A]_t + k_{-1}\Delta t[B]_t \]  

Equation 3.25

\[ [B]_{t+\Delta t} = k_1\Delta t[A]_t + (-k_{-1}\Delta t + 1)[B]_t \]

Such expressions were derived for mechanisms of increasing complexity in order to find the most appropriate mechanism and corresponding rate constants for each kinetics dataset (see Appendix for equations). Due to the nature of numerical integration, it was not possible to explicitly fit the equations to the data using a nonlinear least-squares method. Rather, kinetics data were simulated for a given mechanism and the rate constants manually adjusted until the best fit to the data was obtained, as judged by visual inspection and minimization of the chi-squared value. Chi-squared was calculated as the sum of the squares of the residuals:
where $y_{obs}$ is the measured value and $y_{calc}$ is the value obtained from simulation.

Simulated curves were adjusted to fit all data for folding into a particular lipid type (i.e., the time-dependence of all species measured by gel and the CD signal kinetics measured at 216 nm and 230 nm). The kinetics data measured by gel are direct measurements of the folded, unfolded, and gone species, and therefore no transformation of Equation 3.24 was necessary to simulate progress curves to match the data. Although the measured gel data are fractional quantities, these values equate to the absolute concentration of each species in µM because the total protein concentration was 1 µM; therefore the numerical integration equations in terms of concentrations could be used to simulate data for fractional quantities.

It was necessary to transform the simulated populations to fit the CD data because the observed CD signal is a linear combination of the signals for all species present. The observed signal $([\theta]_{obs})$ at a particular wavelength at a certain time can be expressed by the following equation:

$$[\theta]_{obs} = [\theta]_A f_A + [\theta]_B f_B + \cdots$$  \hspace{1cm} \text{Equation 3.27}$$

where $[\theta]_A$ is the signal of 100% species A at that wavelength, and $f_A$ is the fraction of A present at that time (likewise for B and any other species present). The expressions for the time-dependence of each species from Equation 3.24 were substituted into Equation 3.27 to obtain the simulated CD curves at 216 and 230 nm. Again, these expressions are in concentration units, but could be used as fractional quantities because the total protein concentration was 1 µM. The simulated CD curve for the single-step mechanism discussed above would have the following form:
Equation 3.28

\[
[\Theta]_{obs,t+\Delta t} = [\Theta]_A \{ (-k_1\Delta t + 1) [A]_t + k_{-1}\Delta t [B]_t \} \\
+ [\Theta]_B \{ k_1\Delta t [A]_t + (-k_{-1}\Delta t + 1) [B]_t \}
\]

3.4 Results

*OmpA*$_{171}$ populates a lipid-mediated “gone” state that migrates anomalously by SDS-PAGE

OMPs have long been known to exhibit a “heat-modifiability” when undergoing SDS-PAGE (Nakamura & Mizushima, 1976). In the absence of heating, native β-barrel membrane proteins remain folded in the presence of SDS and migrate at an apparent molecular weight different than the actual molecular weight. This shift in position is thought to be due to differences in SDS binding and compactness between the two conformations (Ohnishi et al., 1998; Reithmeier & Bragg, 1977). Upon boiling in SDS, the protein becomes unfolded and migrates at the actual molecular weight. The identity of the shifted band as the folded form of the protein has been corroborated by spectroscopic measurements, protease protection, and phage receptor activity (Nakamura & Mizushima, 1976; Schweizer et al., 1978).

Under conditions where the protein population is only partially folded (as in folding studies), mixing the sample with SDS serves to quench the folding reaction and maintain the folded and unfolded conformations while undergoing electrophoresis, thereby leading to the appearance of both bands on the gel. Densitometry of the folded and unfolded bands can be used to quantify the extent of folding and this technique has been utilized extensively to measure OMP folding kinetics and stability (Burgess et al.,
We used this technique to study the details of the OmpA transmembrane β-barrel folding pathway. Towards this end, we measured OmpA$_{171}$ \textit{in vitro} folding kinetics into lipid bilayers of increasing acyl chain length and quantified the band intensities over time. However, it became apparent that the protein was populating an additional state that did not migrate with the canonical folded and unfolded bands. Figure 3.1 shows the densitometry results for folding into large unilamellar vesicles (LUVs) composed of the lipids diC$_{10}$PC (A) or diC$_{12}$PC (B) as well as a buffer condition lacking LUVs (C). In both lipids, the unfolded band (indicated by “U” and upside-down red triangles) disappears over time while the folded band (“F”; blue triangles) appears at a higher apparent molecular weight, although with different kinetics (discussed below). Also plotted is the sum of the folded and unfolded band intensities (green diamonds), which noticeably decreases in diC$_{12}$PC. Because the protein appears to be “gone” from the gel, we have termed this alternative conformation as the gone state. Boiling the sample after mixing with SDS causes the reappearance of all of the protein in the unfolded band, as can be observed in the final lanes of the gels in Figure 3.1A and B and the corresponding densitometry values (filled green diamonds). In diC$_{10}$PC, the initial decrease in “F+U” is too fast to be observed but it is clear from the higher intensity of the boiled sample that there is some gone state population.

Interestingly, formation of gone state does not occur in the absence of LUVs, as is shown in Figure 3.1C. When the protein is incubated under the same buffer conditions but without bilayers present, there is no decrease in the intensity of the unfolded band.
over time and boiling the sample produces a band of the same intensity, further indicating that no protein has populated the gone state.

One simple explanation for the gone state initially considered was that it is not actually a separate state but rather is due to a difference in the amount of Coomassie dye bound by the folded and unfolded states. If the folded conformation were to bind only a fraction of the dye molecules that the unfolded conformation binds, then we would expect to see a loss of dye intensity as unfolded protein converts to folded. This would also be consistent with the lack of gone state in the absence of bilayers, because the protein remains in the unfolded state. However, this explanation is not consistent with the data for several reasons. First, if the folded state were binding a specific fraction of the dye bound by the unfolded state, then it should be possible to multiply the folded band intensities by some scalar value and recover a constant “F+U” value for all time points. This is not possible for the diC\textsubscript{10}PC or diC\textsubscript{12}PC data, or any of the other lipids tested; no matter what value the folded band intensities are multiplied by, the sum of those values and the unfolded band intensities still changes over time (data not shown). In addition, we would expect the difference in dye binding to be the same no matter what lipid the protein was folding into because it presumably is folding to the same native conformation, and therefore the same scalar should work to correct all datasets. This is also not the case, as evidenced by the large difference in the final “F+U” values for diC\textsubscript{10}PC and diC\textsubscript{12}PC, despite both having close to zero unfolded population left; if there were no separate gone state then we would expect the folded state population to be the same and the total intensity to be reduced by the same fraction. For these reasons we
ruled out the possibility that the gone state is due to the folded state simply binding a lesser amount of dye.

Another strong indication that the gone state is a separate conformation is that as protein disappears from the folded and unfolded bands, a faint smear appears in the upper region of the gel. This smear disappears upon boiling the sample, concurrent with the full intensity reappearing in the unfolded band, and no smear is observed in the absence of bilayers, when there is no loss in “F+U” intensity. Qualitatively, the total intensity of the smear accounts for the loss of intensity from the folded and unfolded bands, but the smear is generally too faint to quantify reliably and the gone state is therefore defined in terms of the amount lost from “F+U”.

The heat-sensitivity of the gone state is further evidence that it is an alternative conformation of the protein. In addition, the “F+U” value increases slightly over time in diC₁₀PC, indicating that formation of the gone state is reversible under certain conditions. It should be noted that the gone state appears to comprise a range of species because it migrates over a range of positions on the gel, but we refer to it as a single state with the implication being that it includes an ensemble of lipid-mediated conformations that migrate anomalously by SDS-PAGE.

Another possibility is that the gone state is an aggregated species, which is supported by the fact that it migrates as a range of high molecular weight conformations in the upper region of the gel. However, the gone state is not the aqueous oligomeric state explored in Chapter 2 (Danoff & Fleming, 2011). Based on the concentration- and urea-dependence of OmpA₁₇₁ self-association measured by sedimentation velocity (Figure 2.6), conditions were chosen for these folding experiments where the protein remains
entirely monomeric in the absence of membranes (1 µM OmpA171 and 1 M urea). The lack of self-association under these exact experimental conditions was verified by sedimentation velocity (data not shown) and circular dichroism (see below). Therefore we conclude that the gone state is not a manifestation of the self-association of the aqueous unfolded state, but is some other conformation that forms in the presence of lipid bilayers.

Due to gel-to-gel variation in staining, it is essential for data comparison that folded and unfolded intensities be normalized to fractional quantities. Typically, the fractions folded and unfolded have been calculated by dividing by “F+U”, because this quantity is expected to represent the total protein population in each lane (see Equation 3.4 and Equation 3.5 in Materials and Methods). However, with the discovery of the gone state it has become clear that “F+U” is no longer a reliable measure of total protein and is itself variable with time. Therefore the intensity of the boiled band is used to normalize the values to fractional quantities (Equation 3.6 and Equation 3.7).

For conditions where substantial gone state is formed, this difference in calculation can have a major impact on the apparent kinetics of the folded and unfolded populations, as demonstrated in Figure 3.2. Panels A and B show the same lipid conditions as in Figure 3.1 converted to fractional quantities using either “F+U” (open symbols) or “B” (filled symbols). For folding into diC10PC (A), there is a marked difference in the values for the two methods of calculating fraction folded, with a slightly less pronounced difference in fraction unfolded. The differences are even larger for folding into diC12PC (B), due to the greater population of the gone state under this condition. In both lipids, the final value for fraction folded approaches one when
calculated using “F+U”. This is a consequence of the fact that the unfolded band almost completely disappears. However, the full population of the unfolded band does not reappear in the folded band but rather converts to the gone state, as shown in Figure 3.1, so the actual final fraction folded (represented by F/B) is less than one (drastically so for diC12PC). It will be discussed below what structural properties the gone state has and to what extent it resembles the canonical folded and unfolded states, but it is clear that the actual population in the definitively folded conformation (that migrates at the folded band position) is misrepresented by normalization with “F+U”. While not as noticeable at long times, fraction unfolded is also artificially inflated when calculated using “F+U”, due to the fact that this value decreases as the gone state is populated.

The data in Figure 3.2 were fitted to sums of exponentials to further demonstrate the difference in apparent kinetics for the two methods of calculation. Although a simple sum of exponential terms (each with its own amplitude and rate constant) does not correspond to a specific mechanism, it does give a sense of the observed rate constants and allows for a comparison between datasets. For the two methods of normalization, the fitted rate constants were found to be moderately different while the fitted amplitudes (which define the final folding efficiency) were very different (as is evident by eye). For an actual kinetic mechanism, both the observed rate constants and the amplitudes of the exponential terms are functions of the microscopic rate constants for all forward and reverse reactions (see Equation 3.20 for example), so these differences demonstrate that calculation of fractional species using “F+U” leads to completely incorrect kinetic parameters. For this reason we have normalized all of our folding data by a boiled band intensity measured with each dataset. The fraction gone is calculated as the difference
between “B” and “F+U” at each time point (Equation 3.8). Fitting these data to kinetic mechanisms and obtaining microscopic rate constants are discussed below.

In summary, when folding into LUVs in vitro, OmpA<sub>171</sub> populates a third state in addition to the well-established folded and unfolded forms. We have termed this conformation as the gone state due to the fact that it is manifested as a decrease in the sum of the folded and unfolded band intensities. A substantial amount of protein populates this state under all folding conditions (see below) and therefore it is essential that the gone state be taken into account when quantifying the degree of folding, and that it be included in modeling of the folding pathway.

**Folding kinetics and gone state formation are strongly dependent on bilayer thickness**

To more thoroughly elucidate the folding mechanism of the OmpA transmembrane domain, we measured folding kinetics into bilayers of varying lipid acyl chain length. Bilayer thickness is known to have profound effects on apparent folding kinetics (Burgess et al., 2008; Kleinschmidt & Tamm, 2002a), and thus we aimed to utilize this variable to probe different regions of the folding pathway. As the gone state has never previously been addressed in the literature, we also sought to examine the effects of bilayer thickness on gone state population and kinetics.

To achieve a range of bilayer thicknesses, we used phospholipids with acyl chain lengths ranging from 9 carbons (diC<sub>9</sub>PC) to 14 carbons (diC<sub>14</sub>PC, also known as DMPC). It was previously determined by <sup>31</sup>P-NMR that C<sub>9</sub> is the minimal chain length for lipids with phosphocholine headgroups that will assemble into a bilayer structure (Kleinschmidt & Tamm, 2002b). The authors also utilized the theory of amphiphile self-assembly
developed by Israelachvili to demonstrate that the geometry of diC₉PC and longer chain phospholipids is incompatible with micellar structures (either spherical or cylindrical) based on packing constraints, and therefore these lipids must form bilayers (Israelachvili et al., 1976; Kleinschmidt & Tamm, 2002b). To further verify that the phospholipids were forming bilayers under our experimental conditions, we observed each of the lipids, after extruding to form LUVs, by transmission electron microscopy. Representative images are shown in Figure 3.3, and it is evident that in all cases there are circular structures present indicative of unilamellar vesicles of the correct size (~100 nm in diameter).

Figure 3.4 shows the measured fractions folded (A), unfolded (B), and gone (C) for OmpA₁₇₁ folding into LUVs composed of diC₉PC – diC₁₄PC. Consistent with previous data, folding is generally faster into thinner bilayers (Burgess et al., 2008; Kleinschmidt & Tamm, 2002a), which is thought to be due to a greater incidence of membrane defects and pores in thinner bilayers (Bennett et al., 2014). There are two exceptions: diC₉PC (light blue) induces slightly slower formation of the folded state than diC₁₀PC (dark blue), but reaches a higher fraction folded at longer times; in addition, diC₁₄PC (gold) produces folding kinetics much faster than expected (faster than in diC₁₃PC (purple) and even slightly faster than in diC₁₂PC (pink)). This is probably due to the experiment being conducted near the lipid phase transition temperature for diC₁₄PC, and is discussed further below.

For all three gel species, the kinetics are highly complex and require multiple exponential phases to fit. In addition, a lag phase appears for folding into the longer chain lipids (diC₁₂PC, diC₁₃PC, and diC₁₄PC). Altogether, this kinetic complexity implies a
multi-step mechanism for the folding process, which is explored further and modeled in conjunction with CD data below.

It is apparent that the kinetics of gone state formation are dependent on bilayer thickness as well (Figure 3.4C). As observed for the folded state, the gone state is formed slower in longer chain lipids. However, the final fraction of gone state formed is larger in longer chain lipids. The increased fraction gone formed at long times with increasing chain length is coupled to a decrease in final fraction folded, because for all lipids the fraction unfolded approaches zero at long times (longer-term time points can be seen in Figures 3.18-24). The lipid-dependence of the gone state kinetics provides further evidence that the gone state is an alternative lipid-induced conformation of the protein and the negative correlation with fraction folded implies that it is formed by an off-pathway reaction that competes with folding. Modeling the relationship between the folded, unfolded, and gone states is discussed below.

*Folding into diC_{14}PC is accelerated at the lipid phase transition temperature*

Lipid bilayers can exist in different phases with varying degrees of order in the lipid constituents. The two most common lamellar phases are the solid-ordered, or gel phase (L_{β'}), and the liquid-disordered, or fluid phase (L_{α}). As the name implies, the gel phase is highly ordered, with the lipid chains in all-trans configurations and the head groups packed in a regular lattice. In contrast, the fluid phase has highly disordered lipid chains and no lattice arrangement (Heimburg, 2007). Some lipids can also exist in a “ripple” phase (P_{β'}), which has an intermediate level of disorder and contains areas of lipids in both the gel and fluid configurations (Heimburg, 2000; Janiak et al., 1979). The
ripple name stems from the appearance of periodic one-dimensional ripples formed by linear defects of disordered lipids. Depending on the physical and chemical properties of the lipids, transitions between these phases occur at characteristic temperatures. The main phase transition, or “melting,” refers to the formation of the fluid phase and the temperature at which this occurs is the $T_m$. Lipids that form a ripple phase undergo a “pre-transition” from the gel phase at a temperature several degrees below the $T_m$ (Heimburg, 2007).

For membranes that contain mixtures of lipids, as is the case for all biological membranes, the $T_m$ depends on the lipid composition. Interestingly, cells adapt their membrane composition depending on the growth temperature, so that the melting temperature occurs 10-15 ºC below the growth temperature. For example, when shifted to lower temperatures, *E. coli* cells increase the proportion of unsaturated lipids in their membranes, thus decreasing the phase transition temperature (Marr & Ingraham, 1962). This phenomenon has also been observed in *Bacillus subtilis*, amoebae, and trout liver cells (Heimburg, 2007). Evidently it is necessary for biological function that the membrane be in the fluid phase, so cells have evolved mechanisms of altering their membranes to adapt to changing environment temperature. In line with these observations, it has been shown that OMPs cannot assemble into membranes in the gel phase both *in vivo* and *in vitro* (DiRienzo & Inouye, 1979; Surrey & Jähnig, 1992). In fact, *in vitro* folding experiments have utilized gel phase bilayers formed at low temperatures to trap and characterize membrane-adsorbed intermediate states of OmpA (Rodionova *et al.*, 1995). However, OMP folding has not been investigated with the lipid bilayer in the transition region where the fluid and gel phases coexist.
For most of the lipids in this work, the experimental temperature of 25 °C is well above the T_m reported in the literature: diC_{13}PC has a T_m of 14 °C, and diC_{9}PC – diC_{12}PC have reported T_m values below 0 °C (Lewis et al., 1987; Silvius, 1982). The reported T_m for diC_{14}PC is near 25 °C, so we hypothesized that the unexpectedly fast folding kinetics observed for OmpA_{171} in this lipid were due to the bilayer being in the transition region at the experimental temperature.

The measured T_m for diC_{14}PC ranges between 20 and 30 °C in the Lipid Thermodynamic Database (LIPIDAT), with the bulk of the data being between 22 and 24 °C (Caffrey & Hogan, 1992; Caffrey et al., 1991). To determine the exact transition temperature under our experimental conditions, we conducted differential scanning calorimetry (DSC) measurements on diC_{14}PC LUVs in 1 M urea, 20 mM Tris, pH 8. Lipid chain melting is highly cooperative and exhibits a sharp endothermic heat capacity maximum at the transition temperature. Figure 3.5 shows a representative DSC curve for diC_{14}PC, plotting heat capacity as a function of temperature, and a sharp peak is observed at 24 °C. Therefore, we concluded that the T_m for diC_{14}PC under our folding conditions is 24 °C (which is consistent with the data in LIPIDAT).

To verify whether being near the transition temperature causes the faster folding kinetics, we conducted folding experiments at the transition temperature of 24 °C, as well as at a temperature farther away from the transition, 26 °C. Figure 3.6 shows the resulting timecourses of fraction folded (A), unfolded (B), and gone (C) for OmpA_{171} folding into diC_{14}PC at these temperatures. The data measured at 25 °C are also plotted for comparison. It can clearly be observed that the formation of F and G and the disappearance of U are all slightly faster at 24 °C (green sideways triangles) than at 25 °C.
(gold diamonds), and substantially slower at 26 °C (maroon x’s), although all three conditions result in the same final fractions of folded, unfolded, and gone states (see Figures 3.24-26 for longer term time points). Therefore we conclude that the fastest folding kinetics occur at the phase transition temperature for the bilayer, and as the temperature is increased, the kinetics are slowed down. A more detailed analysis of the kinetic mechanism for folding in diC14PC and how it is affected by temperature are discussed below.

The observation that folding slows down in diC14PC as temperature is increased may seem counter-intuitive at first; after all, reaction kinetics generally become faster as temperature goes up due to increased thermal motion. In addition, as the temperature is increased above the Tm we would expect the bilayer to become more uniformly liquid-disordered phase and exhibit more fluid-like properties. However, the kinetics are also a function of bilayer thickness, and as can be seen from the diC12PC and diC13PC data, folding is slowed down as the bilayer becomes thicker. In fact, the kinetics measured for diC14PC at 26 °C are more in line with the expected kinetics based on the trend observed for the other lipids, with folding in diC14PC at 26 °C being slower than folding into diC13PC.

The acceleration of folding (and gone state formation) at the transition temperature probably stems from the physical nature of the bilayer at this temperature. Fully hydrated diC14PC is known to undergo a pre-transition from the gel phase to the ripple phase around 14 °C (Caffrey & Hogan, 1992; Janiak et al., 1979), so the main transition at 24 °C would be between the partially-gel/partially-fluid ripple phase and the fully fluid phase. At the Tm there is extensive coexistence of the two lipid conformations,
and Monte-Carlo simulations and AFM have shown that this results in the formation of many separate lipid domains of each phase (Heimburg, 2007; Nielsen et al., 2000; Tokumasu et al., 2002). Due to the different geometries and dynamics of the lipids, gel phase and fluid phase domains have significantly different dimensions: the fluid phase is ~20% thinner and covers ~20% more area than the gel phase (Mouritsen et al., 1995). These differences cause incompatibilities in molecular packing and hydrophobic matching at interfaces between domains, and therefore greater lateral compressibility, bilayer fluctuations, and an increased incidence of defect/pore formation at domain interfaces (Heimburg, 2007). Experimental support for such behavior comes from the observation that lipid bilayers exhibit the greatest permeability to small solutes during the main phase transition (Mouritsen et al., 1995; Papahadjopoulos et al., 1973), and lipid membranes have been shown to exhibit ion conductivity in the transition region but not at other temperatures (Antonov et al., 1980). We propose that OmpA utilizes the defects at the boundaries between phase domains to initiate membrane insertion and folding, thus resulting in the fastest folding at the $T_m$, where there is the highest degree of lipid domain interfaces. This proposal is also consistent with the observations of faster folding in thinner and more curved membranes, which are also more prone to bilayer defects (Bennett et al., 2014; Kleinschmidt & Tamm, 2002a; Surrey & Jähnig, 1992).

**CD of unfolded OmpA$_{171}$ in 1 M urea provides a reference point for folding kinetics data**

CD spectroscopy is frequently used to determine the type and extent of secondary structure elements in biological macromolecules because different structure types (e.g., $\alpha$-helix and $\beta$-sheet) exhibit unique CD spectra (Woody, 1995). Measuring the evolution
of this signal over time is a further means of determining how and when structure is formed in a macromolecule. We sought to gain a more structurally detailed understanding of the OmpA folding pathway by measuring the kinetics of secondary structure formation for the protein under different lipid conditions.

To monitor the changes in a signal, it is necessary to have a precise measurement of the starting value. For a protein, this corresponds to the signal for the unfolded state in the same buffer condition used for folding. For soluble proteins this is generally difficult to obtain because the protein is maintained in the unfolded state in high denaturant and folding begins immediately after dilution to low denaturant. An unfolded protein’s CD signal is usually dependent on the concentration of denaturant, so measuring the signal in high denaturant is inadequate for determining the starting value for the unfolded conformation under folding conditions. One way of solving this issue is to measure the signal of the unfolded protein over a range of denaturant concentrations (where folding is not favored), which typically has a linear dependence, and then extrapolate to the denaturant condition where folding is measured, but this of course has error associated with it. In contrast, the unfolded signal of a membrane protein can be directly measured in low denaturant, provided there is no aggregation, because folding does not begin until lipid bilayers are added. These folding experiments were conducted under conditions that inhibit self-association of the OmpA$_{171}$ aqueous unfolded state, so we can precisely measure the CD signal of the unfolded conformation.

Panel A of Figure 3.7 shows the CD spectra for 1 μM OmpA$_{171}$ in 8 M urea (green dotted line) or 1 M urea (red solid line), in 20 mM Tris, pH 8. The spectrum in 8 M urea is less intense than that in 1 M urea, indicating that there is a urea-dependence for
the CD signal, but both spectra are indicative of no regular secondary structure. Panel B shows kinetic traces for the sample in 1 M urea at 216 nm (purple) and 230 nm (blue) over 4 h, and it is evident that there is no signal change and thus no structure is formed over 4 h under this condition. As discussed in Chapter 2, the aqueous self-associated form has a CD spectrum indicative of β-sheet structure (Figure 2.2D), so the lack of a β-trough in the wavelength spectrum (panel A) and the unchanging CD signals over time (panel B) are strong evidence that there is no self-association of the unfolded state under our folding conditions.

As described in Materials and Methods, the CD signal for the unfolded conformation in 1 M urea was converted to mean residue ellipticity using the concentration determined by absorbance at 280 nm. The values at the two wavelengths of interest, 216 nm and 230 nm, are -3860 and -1216 deg cm$^2$ dmol$^{-1}$ res$^{-1}$, respectively (Table 3.1). As the CD signal for unfolded OmpA$_{171}$ in 1 M urea corresponds to the starting value for signal changes upon folding, these values are used as reference points when analyzing CD kinetics below.

*CD of folded OmpA$_{171}$ exhibits β-sheet structure and an aromatic exciton signal*

OmpA$_{171}$ folding into LUVs composed of diC$_9$PC – diC$_{14}$PC was monitored by CD under the same conditions as the SDS-PAGE experiments. However, there are several challenges associated with measuring CD in the presence of lipid vesicles. The first is that the scattering of light by LUVs in the wavelength region of interest (<230 nm) reduces the signal intensity and thus contributes to a higher noise level in the measurements. This is compounded by the fact that the buffer contains 1 M urea
(necessary to prevent protein self-association), which absorbs in that region of the spectrum, further reducing the signal-to-noise ratio. We have dealt with this issue by using a custom inset detector (Aviv Biomedical) that reduces the effects of light scattering, working with low protein and LUV concentrations (1 µM and 800 µM, respectively), and collecting data with longer time-averaging and averaging over multiple sets of measurements. Even with the combination of these factors, we can only collect reliable wavelength data down to 212 nm, but this range is completely adequate for assessing the secondary structure and monitoring the kinetics at the two wavelengths of interest, 216 nm and 230 nm.

Another caveat of CD measurements in the presence of LUVs is that the lipids contribute to the CD signal in the region of interest due to differential scattering of left-handed and right-handed circularly polarized light (Bustamante et al., 1983). Figure 3.8 shows wavelength spectra for LUVs of diC\textsubscript{9}PC – diC\textsubscript{14}PC (under the same conditions used for folding) and it is evident that the lipids exhibit a peak around 215 nm. However, it has been demonstrated that ellipticity due to differential light scattering of liposomes can simply be subtracted from the sample signal as background (Chakraborty & Lentz, 2012; Ladokhin et al., 2010). We have verified this by conducting measurements with unfolded protein and LUVs in separate, side-by-side 1 mm cuvettes, and determined that the signals from the two species are additive (data not shown). Therefore we correct all of the experimental data for the lipid contribution by subtracting the corresponding LUV spectrum before conversion to mean residue ellipticity.

Before investigating the CD signal kinetics for OmpA\textsubscript{171} folding into lipid bilayers, we examined the overall spectrum of the folded conformation. Figure 3.9A
demonstrates the differences between the spectra of the unfolded protein (red dashed line) and after folding into diC\textsubscript{10}PC (blue solid line), and two distinct changes are evident (indicated by arrows): a trough has appeared at 216 nm, indicative of $\beta$-sheet structure, and a peak has appeared at 230-231 nm. The $\beta$-sheet signal is highly consistent with all previous CD data on OMPs (Dekker et al., 1995; Eisele & Rosenbusch, 1990; Huysmans et al., 2007; Khan et al., 2007; Kleinschmidt & Tamm, 2002a; Kramer et al., 2000; Nakamura & Mizushima, 1976; Sugawara et al., 1996; Surrey & Jähnig, 1992, 1995) and with the crystal structure of OmpA\textsubscript{171}, which verified it was a $\beta$-barrel (Pautsch & Schulz, 2000). CD bands in the 230 nm-region of the spectrum have previously been observed in a number of soluble proteins such as dihydrofolate reductase and bovine pancreatic trypsin inhibitor, and have been attributed to aromatic residues that interact to give rise to exciton couplets in the spectra (Grishina & Woody, 1994; Kuwajima et al., 1991; Manning & Woody, 1989). A peak at 232 nm also occurs in the OMP PagP, and a specific pair of aromatic residues was identified as responsible for the signal, which are brought into close proximity across the interior of the folded $\beta$-barrel (Khan et al., 2007). OmpA\textsubscript{171} contains a large number of aromatic residues so we hypothesize that a similar interaction is occurring in our protein. To verify this and pinpoint the residues responsible for the signal, mutagenesis studies would be required.

Figure 3.9B shows the wavelength spectra for OmpA\textsubscript{171} after folding into lipids with increasing acyl chain lengths (spectra were all measured at the conclusion of kinetics measurements, which are discussed below). Interestingly, in all conditions the trough at 216 nm overlays, indicating the same amount of $\beta$-sheet structure in all lipids. In contrast, the peak at 230 nm is reduced in intensity with increasing acyl chain length.
As shown in Figure 3.4, there are different final fractions folded by gel for different lipids, so this difference in folding efficiency could account for the change in signal at 230 nm if the exciton interaction is only formed in the native state. The reduced amounts of fraction folded correspond to increasing amounts of gone state as measured by gel, so the CD spectra also imply that the gone state possesses the same amount of β-sheet structure as the native state but not the interaction responsible for the signal at 230 nm. The CD characteristics of the gel species are further discussed in the kinetic modeling section below.

**CD kinetics are dependent on bilayer thickness and reveal an intermediate state with high β-sheet content**

The wavelength spectra for folded OmpA_{171} indicate that the largest changes in ellipticity from the unfolded spectrum occur at 216 nm and 230 nm, so we chose to monitor kinetics at these two wavelengths. Measuring the evolution of the CD signal at 216 nm gives us information about the appearance of β-sheet structure in the protein, while measuring at 230 nm indicates the appearance of the aromatic exciton interaction, and perhaps therefore the native barrel structure. It should be noted that the signal changes at the two wavelengths could be coupled, with either one or both signals influencing the change at the other wavelength due to overlapping of the full spectral signals. However, the observation that the signal at 216 nm is the same in all lipid conditions while the signal at 230 nm changes, is a strong indication that the 216-signal is unaffected by the 230-signal. It is still possible that the β-trough at 216 nm affects the signal at 230 nm. This possibility is addressed through the analysis discussed below.
Figure 3.10 demonstrates how the raw CD kinetics data were processed, using data for folding into diC\textsubscript{10}PC as an example. The procedure included measuring background signals for subtraction and using the signal of the protein alone to determine the sample concentration and convert the folding data to mean residue ellipticity (see Materials and Methods). Final kinetics data for folding into all lipids are shown in Figure 3.11. It is clear that the data follow the same dependence on bilayer thickness as observed in the SDS-PAGE data, with the signal changes at 216 nm and 230 nm being slower in longer acyl chain lipids and a lag phase appearing in the longest chain lipids. The same exceptions occur as well, with the diC\textsubscript{9}PC data (light blue) being slightly slower than the diC\textsubscript{10}PC data (dark blue) at both wavelengths, and the diC\textsubscript{14}PC data (gold) being as fast as the diC\textsubscript{12}PC data (pink) at 216 nm.

Measurements were also conducted for folding into diC\textsubscript{14}PC at 24 °C and 26 °C, as for the gel kinetics. These data are shown in Figure 3.12, and it is again evident that the kinetics are fastest at the phase transition temperature of 24 °C (green) and slowest at 26 °C (maroon).

The CD kinetics follow the same trends as the gel kinetics, but there is more complexity in the kinetics observed by CD due to the spectroscopic signal being a combination of the signals from all species present rather than a direct measure of the population of each species. Of particular note are the kinetics behavior for folding into the shortest chain lipids, diC\textsubscript{9}PC and diC\textsubscript{10}PC. Under these conditions, the data measured at 216 nm (panel A in Figure 3.11) begin at a more negative ellipticity than the expected unfolded protein signal (shown as a red dashed line). Although not as large a difference,

\*\* Wavelength scans were also collected of diC\textsubscript{14}PC LUVs at 24 °C and 26 °C for background correction, and at the conclusion of the folding kinetics, but these spectra are omitted from the plots in Figure 3.8 and Figure 3.9 because they overlay with the diC\textsubscript{14}PC data at 25 °C.
the data at 230 nm (panel B) also begin below the unfolded value. This drop in the mixing time to a more negative ellipticity before increasing to the final value could indicate the population of an intermediate species with a higher content of β-sheet structure than the final native state. For diC₁₁PC, the drop is slow enough to be observed at 216 nm and the signal only dips slightly below the final value before increasing back to it, indicating that the intermediate is less populated under this condition. Similarly, at 230 nm the signal dips slightly below the unfolded signal before increasing to the final value. In the longer chain lipids diC₁₂PC and diC₁₃PC, the measured data at 216 nm and 230 nm begin at the unfolded value and decrease (at 216 nm) or increase (at 230 nm) to the final value with only a slight indication of the more negative signal (see Figure 3.18 and Figure 3.20 for clearer individual plots). Although this could indicate that the β-structured intermediate is not an obligate intermediate for folding in diC₁₂PC and diC₁₃PC, it is also possible that the protein passes through the intermediate but it is not populated enough to make a strong impact on the overall CD signal. In support of this, the β-structured intermediate was observed in diC₁₃PC with the lipids below the phase transition temperature (see below).

The data for diC₁₄PC also provide an indication that the β-structured intermediate is populated in the longer chain lipids, because the kinetics at 24 ºC are observed to decrease to a more negative value before increasing to the final ellipticity at 216 nm and 230 nm (Figure 3.12). In contrast, the kinetics at 25 ºC and 26 ºC do not decrease below the native value at 216 nm and only reach a slightly more negative signal at 230 nm before increasing to the final signal. Such behavior is consistent with faster folding at 24 ºC leading to a higher population of the β-structured intermediate and thus a more
negative signal in the observed kinetics at early times. This will be discussed further in the modeling section below.

To verify the secondary structure type of this hypothesized intermediate, it was necessary to measure the full wavelength spectrum. As the signal reaches the most negative ellipticity in diC₉PC, we chose to capture the spectrum in this lipid condition. Figure 3.13A shows the wavelength spectrum exhibited by OmpA₁₇₁ immediately after mixing with diC₉PC LUVs, measured in 10 nm sections at a time and then combined (dotted light blue line). The signal at 216 nm is consistent with the initial kinetics value in Figure 3.11 (-6700 deg cm² dmol⁻¹ res⁻¹), and the shape of the spectrum is indicative of a β-trough, verifying that the intermediate contains a high content of β-sheet structure. Additionally, the peak at 230 nm associated with the exciton signal is completely absent from the spectrum, demonstrating that the aromatic interaction responsible for this signal has not formed yet in this intermediate state.

The greater negative intensity of the β-trough in the spectrum could indicate that this intermediate conformation has more β-sheet structure than the final native state, but another possibility is that the positive peak at 230 nm extends into the 216 nm-region and increases the signal when it is present. In its absence, the more negative signal at 216 nm could be the actual ellipticity value for the native amount of β-structure. However, this is unlikely because the final signal at 216 nm is the same in all lipid conditions, when the signal at 230 nm varies (Figure 3.9B). If the peak at 230 nm were influencing the value at 216 nm then we would expect the trough at 216 nm to be at increasing values as the 230-signal increases. Because the final value at 216 nm is always the same, we conclude that this region is independent of the peak at 230 nm, and the more intense signal of the
intermediate indicates it has more β-signal than the native state. Conversely, the β-trough at 216 nm does appear to influence the ellipticity at 230 nm, because the value at 230 nm for the β-structured intermediate is more negative than in the unfolded spectrum. This is accounted for in the kinetic modeling section below.

Contrasting with the spectrum measured in diC₅PC, the spectrum measured immediately after mixing with diC₁₃PC LUVs (Figure 3.13B; dotted purple line) overlays well with the unfolded spectrum, indicating that the protein does not immediately adopt the β-sheet conformation under this condition, which is consistent with the kinetics data for diC₁₃PC beginning at the unfolded values.

The β-structured intermediate can be captured in diC₁₃PC bilayers at 4 ºC

To determine whether the β-structured intermediate is populated in diC₁₃PC, we conducted measurements with more ordered bilayers as a means of trapping the intermediate. This strategy was based on previous experiments in which a partially-inserted intermediate form of OmpA was observed in gel phase diC₁₄PC bilayers presumably because the protein could not fully insert into the membrane when it was in the more ordered phase (Rodionova et al., 1995; Surrey & Jähnig, 1992). The intermediate appeared to have similar β-sheet content to the native state, so we hypothesized that this partially-inserted intermediate corresponded to our β-structured conformation. The main phase transition for diC₁₃PC occurs at 14 ºC, with the pre-transition to the ripple phase occurring at 0 ºC (Lewis et al., 1987). We conducted measurements in diC₁₃PC at 4 ºC, and therefore utilized bilayers in the partially-gel/partially-fluid ripple phase.
In the course of these measurements, we determined that the CD signal of OmpA_{171} exhibits a temperature-dependence, as demonstrated in Figure 3.14A for the unfolded form of the protein. It is evident that the ellipticity at 216 nm for OmpA_{171} in 1 M urea becomes more negative in a linear fashion as the temperature is raised. A linear dependence of CD signal on temperature is commonly observed for soluble proteins in the absence of a major conformational change (i.e., in the baseline regions of a thermal denaturation curve), in a similar manner to the frequently observed linear dependence on denaturant concentration (Street et al., 2008). We determined that the values at 216 nm at 4 °C and 25 °C differ by a factor of 1.25, so we reasoned that the measured data at the lower temperature could be scaled by this factor to make it comparable to the data at 25 °C. In support of this idea, we observed that the folded form of OmpA_{171} exhibited the same temperature dependence; panel B of Figure 3.14 shows that the wavelength spectrum for OmpA_{171} folded into diC_{13}PC at 25 °C and cooled to 4 °C decreases in negative intensity (dash-dot pink line), but upon scaling by a factor of 1.25 the spectrum overlays well with that measured at 25 °C (solid pink and purple lines). Therefore we conclude that the unfolded and folded conformations of OmpA_{171} exhibit the same linear dependence of CD signal on temperature, and this dependence can be corrected for by multiplying the data by a scalar factor (1.25 in the case of comparing 4 °C data to 25 °C data).

Panel C of Figure 3.14 shows the spectrum that was obtained upon incubating OmpA_{171} with diC_{13}PC LUVs at 4 °C (dash-dot orange line). Note that this spectrum was collected at the conclusion of kinetics measurements, discussed below, so it is the final conformation formed under this condition. Strikingly, scaling the spectrum by 1.25 (solid
orange line) produces a curve that matches very closely to the spectrum of the β-structured intermediate obtained immediately upon mixing with diC₉PC (dotted light blue line). Slight differences are observed in the region of 230 nm, which could indicate a small degree of aromatic exciton formation in diC₉PC, or simply error in the diC₉PC spectrum (which had to be collected over a very narrow time range). Still, the agreement of the signal at 216 nm is a strong indication that the same β-structured intermediate is formed in diC₁₃PC as in diC₉PC, and it can be trapped when the lipid bilayer is in a more ordered phase.

The kinetics in diC₁₃PC at 4 ºC were also measured and it can be observed in Figure 3.15A that the appearance of β-structure is actually much faster at 4 ºC (orange) than at 25 ºC (purple). Identical kinetics were also observed at 230 nm, with the signal decreasing to the more intensely negative value of the β-structured intermediate (data not shown). The reason for the more rapid kinetics could again be due to bilayer defects. In the ripple phase, the membrane contains a mixture of ordered and disordered lipids, which results in periodic bilayer curvature and packing defects (Heimburg, 2000). These defects could promote the formation of a partially-inserted intermediate state. However, the ordered components of the bilayer apparently prevent complete insertion because there is no formation of the native state, as evidenced by the lack of both the native β-trough signal at 216 nm and the aromatic exciton signal at 230 nm. We also verified by SDS-PAGE that there is no formation of the folded band under these conditions (data not shown).

We further determined that the β-structured intermediate trapped in diC₁₃PC at 4 ºC is an on-pathway intermediate by allowing this conformation to form and then raising
the temperature to 25 °C. Figure 3.15B shows the kinetics at 230 nm upon raising the
temperature (light green), and it can be observed that the CD signal increases to the same
final ellipticity at 230 nm obtained for folding into diC_{13}PC directly at 25 °C (purple).
The final wavelength spectrum also overlays well with the spectrum for OmpA_{171} folded
in diC_{13}PC at 25 °C and the same fractions folded, unfolded, and gone are obtained by
SDS-PAGE (data not shown), indicating that the same equilibrium of conformations is
reached upon raising the temperature to 25 °C. Interestingly, the kinetics of the β-
structured intermediate converting to the native state are much faster than the kinetics
obtained entirely at 25 °C of the unfolded state folding to the native state. This is further
evidence that the β-structured intermediate is an on-pathway state, because having it
already completely formed induces a much faster appearance of the native state than
when folding is initiated from the unfolded state. Evidently, forming this intermediate is
rather slow at 25 °C, because there is a substantial lag phase in the kinetics at 25 °C that is
not observed when folding is initiated from the intermediate state.

In summary, we have determined that OmpA_{171} passes through an intermediate
state with substantial β-sheet structure before forming the native β-barrel. In shorter
chain lipids, this intermediate is formed within the mixing time of the experiment; in
longer chain lipids, the formation of this state is slowed down so that it is only populated
transiently before the protein continues on to the next steps in the pathway, although this
conformation can be trapped when the lipid bilayers are in a more ordered phase. These
results demonstrate that folding into different bilayer thicknesses allows us to access
different parts of the folding pathway that are kinetically silent under other conditions. In
the next section we perform kinetic modeling and assess the data for each lipid condition to determine further details of the folding pathway.

*A complex multi-step model is required to describe the diC12PC gel data*

The ultimate goal of this project was to determine a kinetic mechanism consistent with all of the kinetics data for OmpA171 folding into lipid bilayers. The kinetics measured by SDS-PAGE and CD as a function of bilayer thickness are highly complex and provide a wealth of information. Our strategy for reconciling all of these data into one overarching mechanism was to develop a model for one lipid condition and then extend this model to the other lipid data and adjust it as needed. For each lipid condition, we started with the SDS-PAGE data and developed the mechanism to adequately represent the time-dependence of the species observed by gel (folded, unfolded, and gone); the CD kinetics data were then used to assign CD signal values to the species in the mechanism and extend the model as necessary. We begin here by presenting the mechanism obtained for folding into diC12PC. The model is subsequently extended to the other lipids, and finally a general mechanism for OmpA β-barrel folding is described.

The SDS-PAGE kinetics data comprise progress curves for three species: folded (F), unfolded (U), and gone (G) states. We began the modeling process by considering simple reaction schemes relating these three species. Euler’s method of numerical integration was used to simulate data for each scheme (see Materials and Methods and Appendix for equations) and the rate constants adjusted manually to obtain curves that overlaid with the measured data. For simplicity, each reaction step was assumed to be first order and reversible in all mechanisms examined. In addition, it was assumed that all
of the population begins in U, and F and G are not populated at the beginning of the reaction (which is consistent with all of the protein migrating as unfolded in the absence of membranes).

The first mechanism considered for fitting the diC12PC gel data was a sequential two-step reaction with U converting to F, which subsequently converts to G. In other words, U converts directly to F while G is an off-pathway state that forms from F:

Scheme 3.1

\[ U \xrightleftharpoons{\text{k}_1}{\text{k}_1^{-1}} F \xrightleftharpoons{\text{k}_2}{\text{k}_2^{-1}} G \]

An alternative form of this model was also considered, where G is an on-pathway intermediate state between U and F:

Scheme 3.2

\[ U \xrightleftharpoons{\text{k}_1}{\text{k}_1^{-1}} G \xrightleftharpoons{\text{k}_2}{\text{k}_2^{-1}} F \]

However, both of these mechanisms were ruled out based on the observed gel data. As shown in Figure 3.4, there is a lag phase in the appearance of F for the diC12PC data, which typically indicates one or more reaction steps must occur before the final step that produces F (Wallace & Matthews, 2002). Because Scheme 3.1 has U converting to F in a single step, this mechanism cannot adequately produce the lag phase behavior of the folded species. In addition, if G were formed subsequently to F, then we would expect F to appear faster than G, when in fact the opposite is observed in the diC12PC data. Figure 3.16 demonstrates the shortcomings of Scheme 3.1, as illustrated by the poor agreement between the simulated progress curves for the mechanism and the measured data. Note that the data are plotted over three time ranges to provide a comprehensive view of the fits. Scheme 3.1 is depicted in the middle plot with the species color-coded to match the data, and residuals for the fits are shown above each plot with the total chi-squared value.
(defined as the sum of the squares of the residuals for all three species) shown in the rightmost plot. The rate constants used to simulate the mechanism and obtain the best fit are listed in Table 3.2. It is evident from the residuals and the high chi-squared value that Scheme 3.1 is inadequate for fitting the data, particularly at early times, so this model was eliminated from consideration.

Scheme 3.2 places G before F in the two-step reaction, and would therefore be consistent with G appearing faster than F, and a lag phase in F (fits to diC_{12}PC data not shown). However, it was also observed that F forms faster than G in the shorter chain lipids, which is not consistent with Scheme 3.2. In fact, the gel data for all lipids indicate an inverse relationship between the final amounts of F and G formed; i.e., when more F is formed there is less G and *vice versa* (Figure 3.4). As our goal was to find a mechanism that was consistent with all of the data, we therefore ruled out Scheme 3.2.

The negative correlation between the final fraction folded and fraction gone implies that the two states are formed by competing reactions, so we next considered a mechanism with F and G forming from U by separate reactions steps:

\[
\begin{align*}
U & \xrightleftharpoons[k_1]{k_{-1}} F \\
& \xrightleftharpoons[k_2]{k_{-2}} G
\end{align*}
\]

Scheme 3.3

The simulated progress curves for this mechanism fit to the diC_{12}PC data are plotted in panel A of Figure 3.17. More complex models to be discussed below are presented in panels B-D. The rate constants obtained for each mechanism are listed in Table 3.2.

It is evident from panel A that Scheme 3.3 better describes the data than Scheme 3.1, with the simulated curves for the folded and gone species more closely matching the
observed data at early times. However, a lag in the appearance of F is still lacking, U
decreases to the final value much too fast, and as a result, G overshoots the data and then
decreases to the final value at long times. To accommodate the lag in F, we adjusted the
model to include an intermediate state between U and F:

\[ \begin{align*}
A & \quad \overset{k_1}{\rightleftharpoons} \quad B \quad \overset{k_2}{\rightleftharpoons} \quad F \\
& \quad \overset{k_3}{\downarrow} \quad \overset{k_3}{\uparrow}
\end{align*} \]

Scheme 3.4

The fit to this model is shown in panel B of Figure 3.17. Because only three
species are observed by gel, we assume that the intermediate also migrates as unfolded.
This is a reasonable assumption because an intermediate state would likely not have the
fully formed, stable non-covalent interactions and the compact shape that the native state
possesses, and would therefore be denatured in the presence of SDS and migrate with the
fully unfolded state. However, it is possible that the intermediate state contains some
structure, as will be discussed below based on the CD data, so we chose to refer to the
initial and intermediate states as A and B, respectively, rather than U₁ and U₂. To fit the
gel data, the simulated progress curve for the observed U state was produced by summing
the simulated curves for the A and B species. The letters A and B are both colored red in
the model depicted in the figure to reflect this. Note that all of the protein population is
assumed to begin in state A.

As can be observed in the shortest time range of panel B, Scheme 3.4 adequately
produces the lag phase in F (also reflected in the improved residuals). However, at longer
times the U curve still decreases to the final value too fast, and the G curve still
overshoots the final value (although, as reflected in the decreased chi-squared value of
0.11 compared to 0.18, the overall fit is better than for Scheme 3.3).
An alternative form of Scheme 3.4 was briefly considered, in which the G state is formed from B rather than A:

\[
\begin{align*}
A & \underset{k_1}{\overset{k_3}{\rightleftharpoons}} B \overset{k_2}{\rightarrow} F \\
& \underset{k_4}{\overset{k_3}{\rightleftharpoons}} G
\end{align*}
\]

Scheme 3.5

However, this scheme was eliminated because it would produce a lag phase in G as well as F, which is not observed in the data.

It is evident from the middle plots of Figure 3.17A and B and the large corresponding residual values that there is a slow phase in the disappearance of U that is not described by Schemes 3.3 and 3.4. To account for this slow phase, we introduced an optional misfolding step into the mechanism, via an off-pathway intermediate C:

\[
\begin{align*}
A & \underset{k_1}{\overset{k_4}{\rightleftharpoons}} B \overset{k_2}{\rightarrow} F \\
& \underset{k_4}{\overset{k_3}{\rightleftharpoons}} C
\end{align*}
\]

Scheme 3.6

The simulated progress curves for Scheme 3.6 fit to the diC12PC data are plotted in panel C of Figure 3.17. Like species B, C is assumed to migrate as unfolded by gel and is colored red to denote this in the scheme depiction. The simulated curve to fit the U data is the sum of the progress curves for A, B, and C. It is clearly shown in the middle plot of panel C that the slow phase in U is fit well by the simulated curve, and the residuals are greatly decreased. This is the result of a population of protein following the misfolding reaction from B to C (governed by the relative rates of B to C and B to F), and the reaction from C back to B (error-correction) being slow and rate-limiting for that
population of protein. This type of probabilistic misfolding has been proposed previously as part of a “predetermined pathway with optional errors” (PPOE) view of protein folding, and was shown to be consistent with a variety of kinetic folding data (Bédard et al., 2008; Krishna & Englander, 2007).

The rate constants used to simulate the best fit to Scheme 3.6 are listed in Table 3.2. It should be noted that the reverse rate constants $k_{-1}$ and $k_{-4}$ (corresponding to the reaction steps from B to A and G to A, respectively) are reported as zero. This can be interpreted as indicating that these reaction steps are irreversible. However, it is also possible that the reverse rates are non-zero but are so small that they have a negligible influence on the apparent kinetic behavior. Because these rate constants can be varied over a range of small values with no effect on the quality of the fit, we conclude that $k_{-1}$ and $k_{-4}$ are not well-defined and therefore assign them a value of zero. This behavior will be explored in more detail below and in Figure 3.19.

The greatly improved fit to Scheme 3.6 is reflected in the decreased chi-squared value of 0.024. However, this model less adequately describes the lag phase in F at early times. We therefore introduced another on-pathway intermediate state (that migrates as U) into the pathway:

$$
\begin{array}{c}
A \xleftrightarrow{k_1} B \xleftrightarrow{k_2} C \xleftrightarrow{k_3} F \\
\downarrow k_5 \quad \downarrow k_5 \quad \downarrow k_5 \\
D \quad G \\
\end{array}
$$

Panel D of Figure 3.17 shows the fit that was obtained with this mechanism, resulting in a chi-squared value of 0.017 and small, randomly distributed residuals.
Although the fit is only modestly improved over that for Scheme 3.6 (noticeable in the first 800 s; leftmost plots of panels C and D), we determined that having the additional intermediate state is also necessary for fitting the CD data, which will be discussed below. Note that the reverse rate constants $k_{-1}$, $k_{-2}$, and $k_{-5}$ are reported as zero for the reasons discussed for Scheme 3.6.

*Extending the kinetic model to describe the diC$_{12}$PC CD data*

With a model now developed to describe the gel data, the next step was to expand the mechanism to be consistent with the CD kinetics data. First we considered the data collected at 216 nm, which monitors the amount of secondary structure formed. The measured ellipticity at a particular wavelength is a linear combination of the signals from all species present, with the overall signal being defined by Equation 3.27. Therefore our strategy was to assign CD signals to each species in Scheme 3.7 and simulate the resultant CD progress curve using the rate constants determined for the gel data. We then compared this simulated curve to the measured data and adjusted the values and the mechanism as needed.

We determined the signal for the unfolded form of the protein (in the absence of a bilayer) at 216 nm to be -3860 deg cm$^2$ dmol$^{-1}$ res$^{-1}$ (see above). As shown in Figure 3.11A, the kinetics at 216 nm for folding into diC$_{12}$PC (pink) appear to begin at this value (red dashed line), so we assigned this signal to the initial species, A, in our model. Similar to the gel data, there is also a lag in the change in CD signal at early times, which would be consistent with a species subsequent to A in the pathway having the same signal; therefore we assigned the unfolded signal value to B as well.
The 216 nm kinetics end at a value of -5000 deg cm$^2$ dmol$^{-1}$ res$^{-1}$, so we assigned this signal to the F state. Interestingly, the observed signal at 216 nm reaches the value of -5000 by 2 h (7200 s) and does not change further, even after several days (data not shown). In contrast, the fraction unfolded by gel continues to slowly decrease up to 2 days (~200000 s) after initiation, accompanied by a slow increase in fraction folded (Figure 3.17), as a result of the slow error-correction step from D to C. Because this slow phase is not present in the CD data at 216 nm, we hypothesized that D has the same ellipticity as the native state at 216 nm (-5000) and therefore the slow conversion from D to C is not observable at this wavelength.

Based on the rapid drop in the CD signal at 216 nm when folding into shorter chain lipids, we determined that OmpA$_{171}$ populates an intermediate species with a high degree of β-content during folding. Although not as pronounced when folding into diC$_{12}$PC, the signal does dip slightly below -5000 before increasing to the final value under these conditions. Therefore, we predicted that the remaining intermediate species in the folding pathway, C, corresponds to the more highly β-structured form, and assigned it an ellipticity of -6700 at 216 nm (the value at 216 nm for OmpA$_{171}$ immediately after mixing with diC$_9$PC LUVs and in diC$_{13}$PC at 4 ºC).

The assignment of CD signals at 216 nm to each of the “U” species in Scheme 3.7 also demonstrates the necessity of two on-pathway intermediate states between the initial state and the final state. The second intermediate was introduced into the mechanism to adequately fit the lag in the gel data, but it is also necessary to account for all of the features of the measured 216 nm kinetics: species A and B must both have the unfolded signal to produce the lag in the CD kinetics, species C must have the β-structured
intermediate signal to cause the signal to dip below the final value, and species D and F must both have the native signal for the kinetics to reach the final value of -5000 by 2 h.

The only remaining species to assign a 216 nm CD signal to was the gone state, G. We observed that the CD signal at 216 nm reaches the same final value of -5000 in all lipid conditions measured, despite having differing final fractions folded and differing CD signals at 230 nm. Thus, the gone state must have the same ellipticity at 216 nm as the native state, and therefore the same amount of β-sheet structure. However, when we assigned species G in Scheme 3.7 to have a signal of -5000 at 216 nm, the simulated CD progress curve could not be fit to the observed data. This was due to the fact that species A converts directly to G in Scheme 3.7; when G has a more negative signal than A, the kinetics exhibit an exponential decrease and it is impossible to obtain a lag phase. To produce a lag in the observed CD signal, G must have the same signal as A (just as B must have the same signal as A). However, if G has the signal of the unfolded species A (-3860), then the final CD signal is not -5000, but somewhere between -3860 and -5000, depending on the relative amounts of fraction folded and fraction gone, which is not consistent with the data.

To reconcile this problem, we introduced multiple gone states into the mechanism, which all migrate anomalously by gel but have differing CD signals (indicating conformational changes). Through simulation and adjustments, we determined that a minimum of four G species were required to adequately fit the data, resulting in an expanded mechanism with a parallel gone state pathway:
Interestingly, each gone state resembles a state in the “main” pathway, as we found that the data were fit well when we assigned G₁ and G₂ to have the unfolded signal (-3860), G₃ to have the same signal as the β-structured intermediate (-6700), and G₄ to have the same signal as the native state (-5000). This indicates that the population of protein that migrates as gone by gel actually undergoes conformational changes akin to the main folding pathway in terms of the secondary structure composition.

Figure 3.18 shows the final fits that were obtained for this model. The corresponding rate constants are listed in Table 3.3, along with the assigned CD signals for all species in the mechanism. Panel A of Figure 3.18 depicts Scheme 3.8 with the species color-coded to match the gel data. The fit to the gel data in panel B is essentially the same as the fit to Scheme 3.7 in Figure 3.17D because the additional G states all migrate as gone and do not affect the behavior of the U and F states. Similarly, the rate constants obtained for this model are the same as those determined for Scheme 3.7, with the addition of the rate constants for the G state conversions. The reverse rate constants $k_{-6}$ and $k_{-7}$ as well as $k_{-1}$, $k_{-2}$, and $k_{-5}$ are reported as zero because they are not well-defined. This is discussed further below.

Figure 3.18C shows that the assigned CD signals for each species allowed for an excellent fit to the observed kinetics at 216 nm. This is also confirmed by the residuals for the fit being small and evenly distributed above and below the x-axis. Note that the
more negative signals of the intermediate species C and G₃ are essential for fitting the slight dip below -5000 around 2000 s, and the fact that D, F, and G₄ all have a signal of -5000 is what causes the curve to reach -5000 around 6000 s and stay at that value for the rest of the kinetics measurements.

Scheme 3.8 was next applied to the CD data measured at 230 nm. Species A, B, G₁, and G₂ were assigned the signal for unfolded protein at 230 nm, -1216, which is consistent with where the measured data start and with the observed lag in the data at early times (see Panel D of Figure 3.18).

Based on the wavelength spectra of folded OmpA₁₇₁ (Figure 3.9), the native state F is proposed to contain an aromatic exciton interaction that gives rise to a positive peak at 230 nm. We assigned species F an ellipticity of +700 at 230 nm for this exciton signal based on the data for diC₉PC and diC₁₀PC; these lipids result in the highest amount of final fraction folded and have a final CD signal at 230 nm around +700 (see below). It should be noted that we have taken into account the influence of the signal at 216 nm on the 230 nm-region because the value of +700 is the resultant ellipticity in the presence of the β-signal. The actual CD signal due to the exciton interaction is probably much higher but the overall signal is decreased to +700 because of cancellation from the shoulder of the negative β-sheet signal at 216 nm.

We assigned the β-structured intermediate state, C, a signal of -2700 at 230 nm based on the spectrum measured for this conformation in diC₁₃PC at 4 ºC. Although the spectrum of the intermediate formed in diC₉PC had a slightly less negative value at this wavelength (-2200), fitting of the simulated curve to the data yielded the best fit for a value of -2700 so we believe that the discrepancy with the diC₉PC data is either due to
measurement error or a spectral contribution at 230 nm unique to the diC₉PC environment.

The off-pathway intermediate D has the same β-signal at 216 nm as F, but does not migrate as folded by gel and therefore most likely does not possess the exciton signal at 230 nm. In addition, the signal for D at this wavelength must be more negative than the value for the unfolded species at 230 nm (-1216) due to the β-trough at 216 nm extending into the 230 nm-region of the spectrum. Simulation and fitting led us to assign the signal for species D at 230 nm to be -2200. Based on the attributes of D, this value corresponds to the ellipticity at 230 nm for the native amount of β-signal and in the absence of the exciton signal. In essence, the signal at 230 nm of the β-structured intermediate (-2700) is increased (to -2200) by the β-trough converting to the less negative value in the native state (-6700 changing to -5000 at 216 nm). Therefore, assigning this value to species D takes into account the influence of the β-signal at 216 nm on the signal at 230 nm. It should also be noted that the slow phase in the gel kinetics at long times, which arises from the slow conversion of D back to C, is also observed in the CD kinetics at 230 nm and is thus accounted for by D not having the exciton signal at 230 nm.

The species G₃ and G₄ were less straightforward to assign CD signals to at 230 nm. Although G₄ has the native signal at 216 nm (-5000), it cannot have the native signal at 230 nm (+700) because the final ellipticity observed at 230 nm is far below +700 in diC₁₂PC. Furthermore, the signal at 230 nm diminishes in the other lipid conditions as the gone state is more populated, indicating that the gone state does not possess the same exciton interaction as the native state. G₃ has the same signal at 216 nm as the β-structured intermediate, C, so we tried assigning it a signal of -2700 at 230 nm as for C;
similarly, we assigned a signal of -2200 to G_4, making it parallel to species D. However, the resultant simulated curve was totally inconsistent with the observed data, being drastically too negative. Therefore we hypothesized that G_3 and G_4 could have signals at 230 nm between those for the intermediates and the native state, implying some degree of exciton signal. Trial and error eventually led to an excellent fit of the data, shown in Figure 3.18D, with G_3 and G_4 having signals of -700 and -400, respectively (using the same rate constants as determined from the gel and 216 nm data). Evidently these conformations possess a small positive CD band that increases the ellipticity at 230 nm from the baseline value of -2700 or -2200 to -700 or -400, respectively, but not as large a positive band as in the native state, which increases the ellipticity at 230 nm to +700. It is perhaps not too surprising that these gone state forms have a non-zero CD band in the 230 nm-region. The exact conformations of the gone states are unknown, but it is possible that they contain aromatic residues brought into proximities that give rise to weak exciton interactions. The interacting residues could be the same as those that cause the signal in the native state, but in a slightly different orientation so as to make the signal weaker, or they could be residues from separate protein molecules that are brought together by the multimeric nature of the gone state. Verification and identification of the interacting residues would require mutagenesis studies and analysis of the protein variants’ spectra.

The final rate constants obtained from fitting the diC_{12}PC kinetics data to Scheme 3.8 are listed in Table 3.3, but in the process of adjusting the rate constants to fit the data we determined that some of the rate constants are less well-defined than others. The nature of the mechanism combined with the magnitudes of the fitted rate constants causes
some rate constants to have a large effect on the fit when they are varied by a small amount, resulting in their being narrowly defined. Other rate constants can be varied by a greater factor or even set to zero (as reported in Table 3.3) without much effect on the goodness of fit. To further explore the sensitivity of each rate constant in the mechanism, we varied the rate constants individually over several orders of magnitude and monitored the chi-squared values for the fits to the gel data and CD data at both wavelengths. These values are plotted in Figure 3.19 for all of the forward and reverse rate constants of Scheme 3.8.

The rate constant values are plotted on a log scale to display a large range of values, with the distance between tick marks representing a factor of 10. Although the graphs cover different regions of the scale (reflecting the differing magnitudes of the fitted rate constants), the scaling of the x-axes is kept constant throughout. The leftmost plots show the variation in the chi-squared value for the fit to the gel data (red circles), while the middle and rightmost plots show the chi-squared values for the fits to the CD data at 216 nm (purple triangles) and 230 nm (light blue diamonds), respectively. The filled symbol in each plot denotes the rate constant value used in the final fits in Figure 3.18 and listed in Table 3.3.

Inspection of Figure 3.19 reveals that the chi-squared values for $k_1$, $k_3$, $k_4$, $k_4$, and $k_5$ exhibit a trough for the three sets of data (except $k_3$, $k_4$, and $k_4$ do not exhibit a trough in the CD data at 216 nm), with the final value used in the fit always occurring at the minimum of the trough. Therefore the values listed in Table 3.3 for these rate constants are well-defined by the data. The rate constants $k_1$ and $k_5$, in particular, have very sharp troughs for the gel data, and thus are very well-defined. This is due to the fact that these
are the forward rate constants for the two initial reaction steps from the starting state, A, and thus strongly influence the initial apparent kinetics and determine the relative populations that follow the “main” pathway and the “gone” pathway. The other forward rate constant in the main pathway, $k_2$, exhibits less pronounced troughs for the three datasets: the chi-squared values significantly increase at low values but at higher values increase and level off at values only slightly above the minima. But though the effect of varying the rate constant is more subtle, the final values used in the fits still correspond to the minimum chi-squared values, and therefore this rate constant is also fairly well-defined.

Varying the forward rate constants in the gone state pathway ($k_6$, $k_7$, and $k_8$) has no effect on the gel data chi-squared value because the different gone state species cannot be distinguished in the gel data and thus this part of the pathway has no effect on the apparent gel kinetics. The chi-squared values for the CD data, though, follow similar trends to $k_2$, and exhibit asymmetrical troughs that level off at values slightly higher than the minimal values used in the fits. Interestingly, $k_6$ and $k_8$ have better defined minima than $k_7$.

The reverse rate constants are generally not as well-defined. With the exception of $k_{-4}$ (which does exhibit a clear trough in the gel data chi-squared plot) the reverse rate-constants exhibit chi-squared plots that increase at larger values but are essentially flat at lower values. This indicates that there are upper limits for the reverse rate constant values but the lower limits are poorly defined. The rate constants $k_{-3}$ and $k_{-8}$ do exhibit slight increases in chi-squared at lower rate constant values and therefore are reported as the values that result in a minimal chi-squared. But $k_{-1}$, $k_{-2}$, $k_{-5}$, $k_{-6}$, and $k_{-7}$ can be set to zero.
with no effect on the chi-squared values for all three datasets, indicating that these rate 
constants need only be below a certain value before they become negligible in affecting 
the apparent kinetics. For this reason we report these rate constants as zero in Table 3.3. 
The filled symbols in the chi-squared plots denoting the final rate constant values used 
for fitting are plotted on the y-axes for these rate constants to represent the zero-value on 
the log scale.

Because the rate constants listed as zero in Table 3.3 could in fact have non-zero 
values, we maintain the reverse reaction arrows in Scheme 3.8 (and the previous 
mechanisms) as an indication that the reverse reactions are possible. In addition, we show 
below that the reverse reactions are not negligible under other lipid conditions.

In summary, analysis of OmpA_{171} folding kinetics in diC_{12}PC lipid bilayers has 
led us to develop a complex mechanism to describe the folding pathway (Scheme 3.8). The model includes an off-pathway series of “gone” states that migrate anomalously by 
gel but undergo similar conformational changes in the secondary structure as the 
“normally” migrating species (but do not attain the native state). In both the main 
pathway and the gone state pathway the protein transforms through multiple states, 
including an intermediate with a high degree of β-sheet content, before attaining the 
native amount of β-structure. The main pathway also includes an optional misfolding step 
to an off-pathway intermediate. Interestingly, this state exhibits the native CD signal at 
216 nm, indicating the same amount of β-sheet structure as the folded state, but it does 
not migrate as folded by gel nor contain the aromatic exciton signal at 230 nm, indicating 
it lacks the native tertiary structure. The slow conversion of this state back to C in the 
main folding pathway gives rise to a slow phase in the observed folding kinetics by gel
and CD at 230 nm. The rate constants obtained from fitting the data to this mechanism are fairly well-defined with the exception of several of the reverse rate constants, which are reported as zero since they have a negligible effect on the apparent kinetics in \( \text{diC}_{12}\text{PC} \).

*Similar models to Scheme 3.8 describe Omp\( \text{A}_{171} \) folding into \( \text{diC}_{13}\text{PC}, \text{diC}_{11}\text{PC}, \text{diC}_{9}\text{PC}, \text{and diC}_{10}\text{PC} \).*

As shown in Figure 3.4 and Figure 3.11, the folding of Omp\( \text{A}_{171} \) into \( \text{diC}_{13}\text{PC} \) is much slower than into \( \text{diC}_{12}\text{PC} \), and there is a more pronounced lag phase. However, other aspects of the data are similar, such as the slow phase in the long-term gel data, and the CD kinetics ending at the same values as in \( \text{diC}_{12}\text{PC} \) at 216 and 230 nm. For this reason we predicted that the kinetic model consistent with these data would be very similar to that developed for \( \text{diC}_{12}\text{PC} \), but with reduced rate constants. In fitting the data to Scheme 3.8 though, we found that this mechanism could not adequately describe the gel and CD kinetics data, even with substantially slower rate constants. We noted that there appears to be a lag phase in the gone state data for \( \text{diC}_{13}\text{PC} \), which is not present in \( \text{diC}_{12}\text{PC} \); to accommodate this and the more pronounced lag in the folding data, we introduced another state, \( A' \), kinetically before the initial state \( A \) in the mechanism:

![Scheme 3.9](image_url)
Figure 3.20 shows that this model describes the data well. The CD kinetics were fit using the same ellipticity values for each species as for diC_{12}PC (A' being assigned the unfolded species’ signals). Additionally, the fitted rate constants (listed in Table 3.4) are generally an order of magnitude smaller than the corresponding rate constants determined for diC_{12}PC, as predicted (again, with the non-well-determined reverse rate constants reported as zero). Therefore, a combination of slower rate constants and an additional state at the beginning of the pathway were necessary to fit the data for folding in diC_{13}PC. We interpret this additional state as being an earlier unfolded (but distinct) conformation of the protein that is invisible when folding into diC_{12}PC because the conversion of A' to A is faster in the shorter chain lipid.

Folding into the thinner bilayer of diC_{11}PC was observed to be substantially faster than into diC_{12}PC, and so it was not too surprising that the data could not be fit to Scheme 3.8. However, there is still a slow phase in the disappearance of the U species at long times and the CD kinetics dip slightly to a more negative value before reaching the native signal. Therefore we predicted that the same model could be applied but the early steps could be kinetically silent, similar to the first step in Scheme 3.9 being kinetically silent in diC_{12}PC. First we considered a truncated form of Scheme 3.8 where all of the protein begins in state B instead of state A (i.e., the reaction from A to B is too fast to be observed). In addition, we incorporated a reaction step from B directly to G_2, due to the need for a path to form the gone state (with A converting rapidly to B, we assumed that no G_1 is formed):
Although this mechanism allowed us to fit the fast initial changes and the very slow phase at long times in the gel data, it did not account for the intermediate phase observed (see Figure 3.21B, 100-800 s). This phase required another reaction step feeding protein into the pathway at an intermediate rate, so we incorporated another optional misfolding step into the pathway, this time stemming from state B:

In this mechanism, some of the protein population converts to the off-pathway state B' instead of continuing on the main pathway or the gone state pathway. The slow correction of this misfolding to convert back to B and continue on the productive folding pathway gives rise to the intermediate phase in the observed kinetics. Panel B of Figure 3.21 shows the excellent fit that was obtained to the gel data using this model. The faster kinetics are a result of the protein population beginning in state B instead of A, as well as the rate constants (listed in Table 3.5) being 2-4 times larger than the corresponding values determined for diC\textsubscript{12}PC. It should also be noted that although the main pathway reverse rate constants \(k_2\) and \(k_3\) are not well-determined and are therefore reported as
zero, the gone pathway reverse rate constants $k_{-5}$, $k_{-7}$, and $k_{-8}$ are non-zero values. The observation that the G state slowly converts back to the F species at long times (see the rightmost plot of Figure 3.21B) requires the reverse reactions to occur and therefore the rate constant values are non-zero. Such behavior was not observed in diC$_{12}$PC and diC$_{13}$PC, with the gone species population remaining constant at long times, so the reverse reactions were not necessary to describe the kinetic behavior. However, in the shorter chain lipid diC$_{11}$PC it is apparent that the reverse reactions do occur, providing further support for maintaining the reverse reaction arrows in the general mechanism descriptions.

The CD kinetics measured at 216 nm for diC$_{11}$PC were fit well by Scheme 3.11 (panel C of Figure 3.21), using the same ellipticity values for the species as for diC$_{12}$PC and diC$_{13}$PC (B' being assigned the unfolded species’ signals). The 230 nm data were also fit using the same CD signals as the longer chain lipids (panel D), with the exception of G$_3$ and G$_4$ having signals of -1500 and -50, respectively. Interestingly, it appears that G$_3$ has a more negative signal at 230 nm than in diC$_{12}$PC while G$_4$ has a more positive signal, somehow stemming from the nature of the gone state in the presence of diC$_{11}$PC bilayers. However, the consistency of the signals of the species along the main pathway provides evidence that the protein follows the same folding mechanism in different lipid bilayers. Note also that the slight dips in the CD kinetics at early times are a result of the more negative signals of the $\beta$-structured states C and G$_3$.

Folding was observed to be fastest in the thinnest bilayers of diC$_9$PC and diC$_{10}$PC. In both of these lipids, the gel data exhibited substantial burst phases in the folded and unfolded species’ kinetics, and the CD data began at a more negative ellipticity, which
led to the discovery of the β-structured intermediate state. Because the observed CD signal for both lipids had already decreased to the more negative values at 216 and 230 nm within the mixing time of the experiment, we concluded that all of the protein begins in the intermediate C state (i.e., the conversions through the unfolded A and B states are too fast to be seen with our experimental set-up). Similarly, G₁ and G₂ are not populated and C must be able to convert directly to G₃. Additionally, to account for the multiple phases in the gel data (including the burst phase), it was necessary to include an optional misfolding step from species C to C' as well as the off-pathway intermediate D included in the other mechanisms:

\[
\begin{align*}
D & \quad \xleftarrow{k_4} \quad \xrightarrow{k_4} \\
C' & \quad \xleftarrow{k_3} \quad \xrightarrow{k_3} \quad C \quad \xleftarrow{k_5} \quad \xrightarrow{k_3} \quad F \\
G_3 & \quad \xleftarrow{k_8} \quad \xrightarrow{k_8} \quad G_4
\end{align*}
\]

Scheme 3.12

Figure 3.22B shows this model fit to the diC₉PC gel data (rate constants listed in Table 3.6). The burst phase in the appearance of F is due to some of the protein population converting directly from C to F, while other populations convert to C', D, or G₃. The error correction from C' to C causes the intermediate phase in the observed kinetics while the slower correction from D to C gives rise to the slow phase. In addition, there is a slow flux of gone state back to the main pathway (similar to diC₁₁PC), resulting in a gradual conversion of species G to F at long times. These features of the gel data resulted in all of the rate constants being fairly well-determined and none of the reverse rate constants being reported as zero because they are all necessary to define the observed kinetic behavior.
Scheme 3.12 also fits the diC\textsubscript{10}PC gel data well, as shown in Figure 3.23B (rate constants listed in Table 3.7). Again, the burst phase is a result of the fast conversion from C directly to F, and the slower phases are due to error correction from the off-pathway states C' and D. We determined that the fitted rate constants for these steps are slightly larger than those obtained for the diC\textsubscript{9}PC data, which results in the observed kinetics for folding in diC\textsubscript{10}PC being slightly faster than in diC\textsubscript{9}PC. The reason for the rates being slower in the thinner bilayer condition (the opposite of the trend observed for the other lipids) is unclear but is undoubtedly related to the physical properties of the lipid bilayers.

Scheme 3.12 and the respective fitted rate constants are also consistent with the multi-phasic CD kinetics data for diC\textsubscript{9}PC and diC\textsubscript{10}PC. The 216 nm kinetics for both lipids were fit well using the same ellipticity values for each species as determined in the longer chain lipids (-6700 for C, C', and G\textsubscript{3}, and -5000 for D, F, and G\textsubscript{4}), again showing the consistency of the secondary structure formation pathway. The kinetics at 230 nm were also fit to Scheme 3.12 using the previously determined values for the C, D, and F species (-2700 for C and C', -2200 for D, and +700 for F), but as for diC\textsubscript{11}PC, it was necessary to assign different signals to G\textsubscript{3} and G\textsubscript{4} than in the other lipids to fit the data. Interestingly, we determined that G\textsubscript{4} must have the same signal at 230 nm as the F state, +700. This is based on the fact that the CD signal at 230 nm reaches a final value of +700 in both diC\textsubscript{9}PC and diC\textsubscript{10}PC, even though there is a slightly higher final fraction folded in diC\textsubscript{9}PC than in diC\textsubscript{10}PC. Because the final CD signal is the same, the final folded and gone states must have the same signal of +700 at 230 nm. The F state having a signal of +700 is also highly consistent with the kinetic models developed for the other lipids, as
discussed above. In addition, the G₄ state having a signal of +700 follows the trend of its signal at 230 nm increasing with thinner bilayers (from -400 in diC₁₂PC and diC₁₃PC to -50 in diC₁₁PC to +700 in diC₉PC and diC₁₀PC). Thus it appears that either the exciton interaction responsible for the peak at 230 nm in the native state forms more completely in the final gone state as the bilayers become thinner, or some other aromatic residues are interacting more in the gone state in thinner bilayers to produce a signal at 230 nm.

We also determined from fitting the CD kinetics that G₃ has a signal of -900 at 230 nm in diC₉PC and diC₁₀PC. Interestingly, this value is similar to the signal obtained for G₃ in diC₁₂PC and diC₁₃PC, indicating that the states could be similar between the lipids. However, this value is very different from that determined for G₃ in diC₁₁PC, -1500, and it is unclear why this is the case.

*OmpA₁₇₁ folding into diC₁₄PC at and above the phase transition temperature follows the same kinetic mechanism determined for the shorter chain lipids*

The folding kinetics measured by gel for diC₁₄PC at 25 °C were observed to be as fast as folding in diC₁₂PC, presumably due to being near the phase transition temperature of this lipid. Therefore we applied the model developed for diC₁₂PC (Scheme 3.8) to the diC₁₄PC data and were able to obtain a good fit (shown in Figure 3.24B). The fitted rate constants (listed in Table 3.8) are slightly larger than the corresponding rate constants for diC₁₂PC, which is consistent with the kinetics being slightly faster in diC₁₄PC. The reverse rate constants $k_{-2}$, $k_{-3}$, $k_{-5}$, $k_{-6}$, $k_{-7}$, and $k_{-8}$ are reported as zero due to their negligible effect on the observed data.
In fitting the CD data, we determined that the 216 nm kinetics were well described using the same CD signals for the species determined for the other lipids, with A, B, G1, and G2 having the unfolded signal, C and G3 having the more negative signal of the β-structured intermediate, and D, F, and G4 having the native β-signal at 216 nm (panel C). For the 230 nm kinetics, we determined that the data could be fit using the same signals for the species A, B, C, D, G1, and G2 as in the other lipids, but different signals for the G3, G4, and F species.

As shown in panel D of Figure 3.24, the data at 230 nm dip below the unfolded value for ~3000 s before increasing to the final ellipticity, which was not observed in diC12PC. To fit this dip, we found that G3 must have a signal of -2700 at 230 nm, similar to species C. This contrasts with the models for diC12PC and diC13PC in which G3 has an elevated signal of -700, indicating the formation of some aromatic interactions that give a positive signal at this wavelength.

In addition, a major difference between the 230 nm kinetics for diC14PC and diC12PC is that the final signal is at a more negative ellipticity (see Figure 3.11B), even though the final fractions folded and gone measured by gel are the same for both lipids (Figure 3.4). This implies that the final states, F and G4, have more negative signals at 230 nm in diC14PC. Indeed, we determined that the best fit to the data was achieved when G4 has a signal of -850 (as opposed to -400 in diC12PC) and F has a signal of -300 (as opposed to +700 in all the other lipids). Therefore it appears that G3, G4, and F all have more negative signals at 230 nm in diC14PC, indicating there is less of a positive contribution from aromatic exciton signals in these species under this lipid condition. The fact that the native F state has a greatly reduced signal at 230 nm, when it consistently has
a signal of +700 in all the other lipids examined, indicates that the thicker bilayer plays a role in influencing the aromatic interaction that gives rise to the exciton signal in the native state. Being near the lipid phase transition temperature for diC\textsubscript{14}PC could also be having an effect.

As described above, we also conducted folding experiments in diC\textsubscript{14}PC at 24 ºC and 26 ºC to explore the effect of the lipid phase transition on the folding kinetics. The kinetics at 24 ºC were observed to be moderately faster by gel than at 25 ºC, and exhibited a significantly faster decrease in the CD kinetics to a more negative ellipticity before increasing to the final signal at 216 and 230 nm. In fitting the 24 ºC data, we determined that Scheme 3.8 could not adequately produce the initial decrease in the CD data but the truncated mechanism, Scheme 3.10, allowed for an excellent fit of all data (shown in Figure 3.25; rate constants listed in Table 3.9). In this model, all of the protein population begins in state B because the conversion from A to B is too fast to be observed. Interestingly, no optional error reaction from B to B\textsuperscript{'} was required, which was introduced to fully describe the folding behavior in diC\textsubscript{11}PC. Similar to diC\textsubscript{11}PC, a slow conversion of G to F was observed at long times, resulting in the reverse rate constants for the gone state pathway having non-zero values.

The 24 ºC kinetics at 216 nm were fit using the same signals for the species determined for the other lipids; the combination of the population starting in B and converting directly to the \(\beta\)-structured intermediate C (or through G\textsubscript{2} to the \(\beta\)-structured state G\textsubscript{3}) and generally faster rate constants causes the observed drop in the CD signal at 216 nm. The kinetics at 230 nm display similar behavior, and the data were fit using the same signal values as in diC\textsubscript{14}PC at 25 ºC (including F having a signal of -300 instead of
with the exception of G\(_3\) and G\(_4\) having signals of -2000 and -1000, respectively
(though these species’ signals are still very similar to those determined at 25 °C).

The kinetics in diC\(_{14}\)PC at 26 °C were slower than at 25 °C and more on par with
the kinetics measured in diC\(_{13}\)PC, due to being further away from the phase transition
region. However, we determined that the data were fit well by Scheme 3.8 (shown in
Figure 3.26; rate constants listed in Table 3.10). The unfolded state A’, introduced in
Scheme 3.9 to fit the diC\(_{13}\)PC data, was unnecessary. This is evident by the fact that there
is no lag in the appearance of gone state, which was observed in diC\(_{13}\)PC and was
accommodated by incorporating an additional unfolded state before the state that
converts to G\(_1\). The much slower appearance of the F state at 26 °C appears to be a result
of the rate constants for the reactions B to C and C to F being two orders of magnitude
smaller than the corresponding rate constants at 25 °C. Similar to diC\(_{13}\)PC and diC\(_{14}\)PC at
25 °C, the gone state population remains constant at long times, resulting in the reverse
reactions having negligible flux; thus the rate constants \(k_{-6}\), \(k_{-7}\), and \(k_{-8}\) are reported as
zero.

The 26 °C kinetics at 216 nm were fit well using the same CD signal values for
the species as for all the other lipids, again indicating the consistency of the secondary
structure formation process. The 230 nm kinetics were fit using the same signal values as
in diC\(_{14}\)PC at 25 °C, with the exception of G\(_3\) having a signal of -2000 instead of -2700.
Still, this value is the same as the signal determined for G\(_3\) in diC\(_{14}\)PC at 24 °C, indicating
a degree of consistency between the temperatures.

In summary, OmpA\(_{171}\) folding into diC\(_{14}\)PC appears to follow the same general
pathway as in shorter chain lipids, but with faster rate constants near the lipid phase
transition temperature, resulting in faster apparent folding kinetics. The protein transforms through the same intermediate, partially-structured states, based on the secondary structure formation kinetics; the only major difference is the CD attributes of the native F state, as it was consistently found to have a reduced ellipticity at 230 nm of -300 in diC$_{14}$PC, compared to +700 in the shorter chain lipids. Because the native state in diC$_{14}$PC migrates at the same shifted position when undergoing SDS-PAGE as in the other lipids, it is likely that the difference in CD signal is the result of a local environmental difference around the aromatic pair responsible for the positive peak at 230 nm, and not a major structural difference in the β-barrel.

A general mechanism for the folding of the OmpA β-barrel

The previous sections have described the kinetic mechanisms developed to fit the SDS-PAGE and CD kinetics data measured for OmpA$_{171}$ folding into lipid bilayers of diC$_9$PC – diC$_{14}$PC. We argue that the protein follows the same general mechanism under all these lipid conditions, and the differences in apparent kinetics are a result of differing rate constants, parts of the mechanism being kinetically invisible under different conditions, and differing populations of off-pathway misfolded states. The major evidence for Schemes 3.8-12 being variations on the same general model is the remarkable consistency of the CD signals assigned to the species for each lipid condition. The constancy of the signals at 216 nm is a particularly strong indication that the protein follows the same secondary structure formation pathway in the different lipids. The signals of the species at 230 nm are also quite consistent, with a few exceptions most likely due to subtle structural differences causing spectral changes. Figure 3.27
demonstrates the uniformity of the species’ CD signals by plotting the assigned ellipticity at 216 and 230 nm for each species across all lipid conditions. The purple bars represent the signal at 216 nm, and for each species the value is essentially the same in all lipids. The light blue bars, representing the signal at 230 nm, are the same in all lipid conditions except for the species F, G3, and G4, and the F state has the same signal in all lipids except diC14PC.

To summarize: When present in the mechanism, the states A', A, B, B', G1, and G2 (panel A) were consistently assigned signals corresponding to the unfolded conformation at 216 and 230 nm (these signals being determined by separate measurements in the absence of bilayers; Figure 3.7). Species C (panel B) was determined to be an intermediate state with substantial β-sheet structure in all lipid conditions, and was uniformly assigned signals of -6700 and -2700 at 216 and 230 nm, respectively (also measured separately, captured immediately upon mixing with diC9PC and in diC13PC at 4 °C). Note that the species C' was also assigned these signals, although this state was only necessary for the scheme describing the diC9PC and diC10PC data. The ubiquitous off-pathway intermediate D (panel C) was consistently found to have the native signal at 216 nm, -5000 (i.e., the same amount of β-structure as the native state), and a signal of -2200 at 230 nm. The native folded state, F (panel D), uniformly had a signal of -5000 at 216 nm, and a signal of +700 at 230 nm in all lipid conditions except diC14PC, in which F was consistently assigned a signal of -300. This difference in ellipticity at 230 nm is most likely a result of the thicker bilayer influencing the aromatic interaction responsible for the positive peak at 230 nm.
The gone state was expanded into four species to be consistent with the CD kinetics data, and it was found that this anomalously migrating species undergoes conformational changes akin to the “main” folding pathway. While the first two gone states, \( G_1 \) and \( G_2 \), were assigned the unfolded signals (panel A), \( G_3 \) was consistently found to have the same more intense \( \beta \)-signal at 216 nm as state C (-6700) and \( G_4 \) was always assigned the native \( \beta \)-signal at 216 nm, as for D and F (-5000)(panels E and F, respectively). The signals of the last two gone states at 230 nm, however, were less uniform across lipid conditions, and appeared to indicate some degree of exciton signal in these non-native states. \( G_3 \) (panel E) was found to have a signal ranging from -1500 to -700 in diC\(_9\)PC – diC\(_{13}\)PC, which is elevated from the expected 230 nm signal of -2700 corresponding to the more intense signal at 216 nm (see species C). The signal was closer to the expected value in diC\(_{14}\)PC, where it ranged from -2700 to -2000 at the different temperatures measured. Therefore, it appears that the conformation of species \( G_3 \) has some degree of aromatic exciton formation that increases the signal at 230 nm, but this signal is diminished in the thicker bilayer of diC\(_{14}\)PC. \( G_4 \) (panel F) was found to have a signal at 230 nm ranging from +700 in diC\(_9\)PC and diC\(_{10}\)PC to -1000 in diC\(_{14}\)PC (at 24 \(^\circ\)C). For this species there appears to be a trend of the signal decreasing with increasing lipid acyl chain length. With the same signal at 230 nm as the native state (+700) in the shortest chain lipids, it appears that \( G_4 \) has substantial formation of the exciton interaction under this lipid condition, but the signal is diminished in thicker bilayers. However, the signal in diC\(_{14}\)PC at all three temperatures (between -1000 and -850) is still less negative than the expected ellipticity due to the \( \beta \)-trough in the absence of the exciton signal.
(-2200, as seen for species D). Therefore, some aromatic interaction still occurs in the final gone state in the thicker bilayers.

Evidently, with the exceptions of G 3 and G 4, the CD signals of the species are highly consistent between lipid conditions, providing strong evidence that these are common species within the same general pathway for folding of the β-barrel into a lipid bilayer. Thus we propose the following complete mechanism to describe the folding pathway of the OmpA β-barrel, constructed by combining the schemes developed for each lipid condition:

This mechanism contains all of the observed unfolded, intermediate, gone, and native species discussed above. The protein population begins in state A' and proceeds along the folding pathway through states A, B, C, and F. Population of off-pathway misfolded states B', C', and D can also occur, but B' and C' are only observed under certain conditions. State D is populated in all lipid conditions and therefore represents a prevalent misfolded species in vitro that gives rise to the slow phase of folding. The protein can also divert into the “gone state” pathway, which most likely consists of a protein/lipid aggregate that can undergo secondary structure formation. Reaction arrows are included between states B and G 2, and C and G 3, because these reactions were observed in the thinner bilayers of diC 9PC, diC 10PC, and diC 11PC, as well as diC 14PC at 24 °C. It appears these cross-reactions are not necessary when state A is populated and
able to access the gone pathway, as is the case for folding in diC_{12}PC, diC_{13}PC, and diC_{14}PC at 25 and 26 °C, because such reactions were not required for modeling the gone state kinetics in these conditions. However, when the early states are passed through very rapidly and therefore not observed (as in the thinner bilayers), the latter states feed into the gone pathway and these become important reactions. Similarly, all steps of Scheme 3.13 include reverse reaction arrows to indicate the possibility of reverse reactions taking place. We determined that in the longer chain lipids certain reverse rate constants had a negligible influence on the apparent kinetics and thus could be set to zero; however, in the shorter chain lipids the reverse reactions were observed, as indicated by the slow conversion of the gone species to the folded state at long times. Therefore all reaction steps are considered reversible in the general mechanism.

3.5 Discussion

*Structural mechanism proposed for OmpA β-barrel folding*

We have proposed a complex kinetic pathway for OmpA β-barrel folding (Scheme 3.13) with numerous off-pathway species that strongly impact the observed kinetics. However, the main productive folding pathway in our scheme is highly consistent with the previously proposed model of Kleinschmidt and colleagues (Kleinschmidt et al., 2011; Kleinschmidt, den Blaauwen, et al., 1999). In addition, our detailed knowledge of the CD and SDS-PAGE kinetics allows further refinement of this pathway.
We previously determined that the unfolded aqueous form of OmpA (UAQ) has no regular secondary structure and is expanded (Danoff & Fleming, 2011). The first step in the folding pathway is for this state to associate with a membrane. Membrane binding has been shown to occur rapidly by tryptophan (Trp) fluorescence kinetics and to happen before the appearance of secondary or tertiary structure (Kleinschmidt & Tamm, 1996; Surrey & Jähnig, 1995). Trp quenching experiments have further been used to identify three membrane-associated intermediate states during OmpA β-barrel folding, distinguished by the degree of Trp penetration into the lipid bilayer. The first membrane-adsorbed state (IM1) has minimal Trp penetration, the second state (IM2) has Trps penetrating to the interfacial region of the bilayer, and the third state (IM3) exhibits Trps that are deeply buried in the membrane, indicating a partially inserted conformation (Kleinschmidt, den Blaauwen, et al., 1999). We interpret the three on-pathway intermediate states in our mechanism as corresponding to these three membrane-associated states. Therefore, the surface-adsorbed and interfacial states IM1 and IM2 correspond to species A and B, respectively, and the partially inserted penultimate state IM3 corresponds to our β-structured intermediate C. It should be noted that it is unclear whether species A' in our mechanism is membrane-associated. It is possible that A' is a conformation similar to A but interacting slightly differently with the membrane, or that A' is the UAQ state and converts to A upon binding to the membrane. Further experiments examining the rate of OmpA171 association to diC13PC lipids bilayers are needed to distinguish between these possibilities, as this is the only condition under which A' is populated to a noticeable extent.
Our results substantiate Kleinschmidt and colleagues’ multi-step pathway for OmpA β-barrel assembly because we have identified these states based on fitting of the data directly to kinetic schemes, which, to our knowledge, has not been done before for a β-barrel protein folding into membranes. We will now further describe the structural features of the folding process subsequent to membrane association, based on our data and previous results. We illustrate the relevant on-pathway folding species in Figure 3.28 and summarize the known attributes of each state in Table 3.11. First, species A/IM1 becomes adsorbed to the membrane surface, based on Trp fluorescence, but with no regular secondary structure, as shown by our CD measurements. This state converts to species B/IM2, which has slightly penetrated into the membrane, based on the Trps being located in the interfacial region of the bilayer. Our CD data indicate that this species also has no regular secondary structure. However, this state appears to contain some approximation of the circular arrangement of β-strands, based on site-directed Trp quenching experiments that examined the association of adjacent strands (Kleinschmidt et al., 2011). It was found that the trans ends of adjacent strands (the ends that translocate the bilayer and are located on the extracellular side in the native state) are in closer proximity than the cis (periplasmic) ends, and that the first and last strands that close the barrel are in some proximity in the membrane-adsorbed state. We envision this as a clover-like conformation parallel to the membrane surface, with the trans ends of the strands located at the center of the clover. Figure 3.28 illustrates this configuration for species B with the experimentally observed associating residues on the trans and cis portions of strands indicated by red and green circles, respectively. This conformation is also consistent with the Trp residues penetrating to the interfacial region of the bilayer,
because the Trps and the other (hydrophobic) externally facing residues would be facing the membrane surface in this conformation, and thus could penetrate into the bilayer.

This loosely organized but unstructured intermediate then converts to the partially inserted, $\beta$-structured intermediate $C/I_{M3}$. Previous work has shown that the tryptophans in this state are deeply buried in the membrane, but that the protein is still accessible to proteolysis and does not migrate at the shifted position by SDS-PAGE, indicating it lacks full membrane protection and the compact barrel structure (Kleinschmidt, den Blaauwen, et al., 1999; Rodionova et al., 1995; Surrey & Jähnig, 1992). It was also observed by CD and FTIR spectroscopy that this state contains substantial $\beta$-sheet structure (Rodionova et al., 1995; Surrey & Jähnig, 1992), so it was proposed that a portion of the protein partially inserts into the membrane and adopts a $\beta$-sheet configuration in a closed barrel to satisfy hydrogen-bonds within the hydrophobic bilayer (see Figure 3.28, species C). Based on the site-directed quenching experiments, the $trans$ ends of the $\beta$-strands insert into the bilayer first (i.e., the center of the clover)(Kleinschmidt et al., 2011), which is consistent with the Trps in these locations being buried in the membrane (again indicated by red circles in the figure). In the next step of converting to the native conformation, the $\beta$-strands translocate the rest of the way across the bilayer in concert to form the fully inserted, folded state (Kleinschmidt, den Blaauwen, et al., 1999).

Our CD measurements also indicated $\beta$-sheet structure for intermediate state C, but we have unambiguously determined that the $\beta$-signal is actually more intense than that of the native barrel, indicating that it contains more $\beta$-sheet structure. We hypothesize that this could be due to the uninserted sections of the protein engaging in $\beta$-structure in addition to the membrane-inserted portion. In other words, the portion of
each strand still located on the membrane surface adopts a β-strand conformation and H-bonds with the neighboring strand attached at the cis end (see Figure 3.28). The splayed geometry of the uninserted β-strands prevents full H-bonding between β-hairpins, which could explain why this state is sensitive to SDS denaturation (i.e., migrates as unfolded by SDS-PAGE). In addition, the extracellular loops, which are presumably buried in the membrane in this state, could also be taking part in the β-sheet structure, thus explaining why the β-signal is more intense for this state. Upon emerging into the extracellular environment where there is no longer a driving force to satisfy H-bonds, the loops would unfold, thus causing the native state to have a slightly lower β-content than the partially inserted intermediate.

The other spectroscopic feature we monitored in our studies was an aromatic exciton interaction that induces a positive peak in the CD signal at 230 nm. In fitting the kinetics data, we determined that only the native barrel (in the main folding pathway) possesses this signal. Because it is unknown exactly which residues of the 5 tryptophans and 13 tyrosines in OmpA171 are responsible for the signal, we cannot specify the structural feature that forms in the native state and gives rise to the interaction. In PagP, the aromatic pair identified as responsible for an exciton signal are positioned on either side of the interior of the β-barrel, and thus only interact in the native state (Khan et al., 2007). Similarly, there are two tyrosine residues on the interior of the OmpA β-barrel in the crystal structure (Pautsch & Schulz, 2000), so these residues could be forming an exciton interaction. However, there are also several pairs of aromatic residues brought into close proximity on adjacent strands on the exterior of the β-barrel. There is some evidence that the interacting aromatic residues in OmpA171 could be externally located.
because the native state exhibits a weaker CD contribution at 230 nm in the thicker bilayers of diC_{14}PC than in thinner bilayers, possibly indicating that the lipid bilayer environment influences the exciton interaction. It is difficult to predict significant aromatic exciton formation based on side chain geometry and distance (Grishina & Woody, 1994), so experimental mutagenesis studies are needed to explicitly identify the interacting residues. Still, it is clear that the CD band at 230 nm is only formed in the fully inserted native state, so this signal provides an orthogonal method to SDS-PAGE for monitoring β-barrel formation. The presence of the exciton signal in the folded state is indicated in a non-specific location on the β-barrel by a yellow star in Figure 3.28.

Intermediate C might have been expected to exhibit the aromatic exciton signal as well, because it is presumably partially inserted with a partly formed β-barrel structure. However, the potentially interacting residue pairs identified from the OmpA crystal structure are all located on the periplasmic half of the barrel (both interior and exterior facing pairs), and therefore would not be in close proximity in intermediate C, for which the cis ends of the barrel are proposed to be β-hairpins splayed out on the membrane surface. Therefore, our CD measurements are consistent with the proposed conformation of the β-structured intermediate C, and the lack of the exciton interaction in C is further indication that it provides a sensitive probe of complete β-barrel formation.

*OmpA is subject to folding errors*

Our extensive kinetic modeling led us to identify several off-pathway intermediate states in the OmpA folding mechanism that give rise to slower phases of folding. Species D was necessary for fitting the apparent kinetics in all lipid conditions,
indicating that it is a ubiquitous intermediate state. This conformation is formed by state C in competition with F, and must convert back to C to reach the native state. The slower rate of “error correction” is what causes the slow continued appearance of F at long times. Under some lipid conditions we observed that species B and C can also populate the off-pathway states B' and C', respectively, which contribute intermediate kinetic phases to the data. These reaction steps are representative of the “predetermined pathway with optional errors” (PPOE) model for protein folding proposed by the Englander group as an alternative to unrelated parallel pathways (Krishna & Englander, 2007). This theory suggests that rather than folding via multiple separate pathways within a funnel-shaped free energy landscape, proteins follow predetermined paths of folding based on the sequential incorporation of structural elements, or “foldons.” Kinetic heterogeneity can be caused by the population of off-pathway intermediate states and does not require the presence of parallel folding routes. As it is difficult to envision a β-barrel folding by a substantially different mechanism than the one we have described, there is no direct experimental evidence for the population of significantly different intermediate states, and the presence of off-pathway states is supported by the fitting of our data, we argue that OMPs also fold by a dominant pathway with the correction of optional errors causing apparent multi-exponential kinetics.

Globally fitting our kinetics data from orthogonal methods has revealed some structural information about these off-pathway states, although the details of the conformations are unknown. Species B' and C' were determined to have identical CD spectra to B and C, respectively, and to migrate as unfolded by SDS-PAGE. Therefore, B' is an unstructured state, but could have some loose arrangement as for species B. Because
B’ is a misfolded state, it is possible that this species has incorrect formation of the clover shape and thus cannot proceed with further insertion and formation of intermediate C. Similarly, species C’ has the more intense β-signal of C, indicating it possesses high β-sheet content. The misfolded nature of the state could stem from incorrect H-bonding of the non-inserted β-strands. Interestingly, we determined that although species D also migrates as unfolded by SDS-PAGE and does not possess the aromatic exciton signal indicative of the native β-barrel, it contains β-sheet content equivalent to the native state. Therefore intermediate D must have a different conformation from C’, possibly closer to the native form but still lacking the complete β-barrel structure. One possibility is that the β-hairpins have translocated across the bilayer, and the loops have unfolded (resulting in lower β-content as in the native state), but the β-strands are incorrectly H-bonded and out of register, so the periplasmic side of the barrel is misaligned and unstable. However, it is unknown whether species D is in fact membrane-inserted, so further experiments examining the degree of Trp burial and protease accessibility of this state would be useful.

Another major off-pathway species identified in these studies was the gone state. More likely an ensemble of conformations, the gone state denotes the population of OmpA_{171} that migrates anomalously by SDS-PAGE. CD measurements revealed that the gone state gains secondary structure similar to the species on the main folding pathway, resulting in the model being expanded to four distinct gone states. It was also observed that the latter gone states exhibit a degree of (possibly non-native) aromatic exciton interaction, as evidenced by the elevated CD signals at 230 nm for the species G_3 and G_4. The molecular details of these conformations are unknown, but none of the gone states
possess the full exciton signal of the native \( \beta \)-barrel, nor do they migrate at the folded position by gel, so it can be assumed that these are non-native states that compete with folding and constitute additional off-pathway states in a PPOE model. It is unknown whether the gone states are populated in the \textit{in vivo} pathway, but these competing reactions strongly influence the apparent folding kinetics \textit{in vitro} (Figure 3.2) and therefore must be accounted for in modeling the pathway. Although never addressed previously in the literature, we have observed many other OMPs also exhibit the formation of gone state when folding into synthetic bilayers, so its existence is of broad importance for OMP folding studies. Interestingly, different OMPs have differing propensities for forming gone state (unpublished results), indicating sequence-specific effects similar to those that induce varying folding behavior between OMPs (Burgess \textit{et al.}, 2008). Hence, modeling of the gone state pathway in conjunction with the main folding pathway must be conducted on a case-by-case basis for a given OMP.

\textit{Bilayer defects and secondary structure formation drive \( \beta \)-barrel folding and insertion}

It was previously determined that membrane properties strongly influence OMP folding kinetics, with thinner and more curved bilayers promoting faster folding (Burgess \textit{et al.}, 2008; Kleinschmidt & Tamm, 2002a). We utilized the dependence on membrane thickness to access different regions of the folding pathway and thereby gained a more complete picture of the mechanism. Based on our data and previous observations, we argue that the physical properties of the lipid bilayer, specifically membrane defects, are key in driving the folding and insertion process, in combination with secondary structure formation.
Fast folding of OmpA into the highly curved membranes of small unilamellar vesicles (SUVs) was attributed to the higher incidence of defects in the strained bilayers (Surrey & Jähnig, 1992; Tamm et al., 2004). It was subsequently discovered that OmpA can fold rapidly into low-curvature membranes (i.e., LUVs) when the bilayers are thin (12 carbons or less in the lipid acyl chains§§), and folding is fastest in the thinnest bilayers (Kleinschmidt & Tamm, 2002a). It has also long been known that thinner bilayers are more permeable to solutes (Blok et al., 1975; Paula et al., 1996), and it has been demonstrated in molecular dynamics simulations that spontaneous pore formation is more frequent in thinner bilayers (Bennett et al., 2014). This suggests that bilayer defects are important for OMP insertion and folding, and our kinetics data strongly support this idea.

Not only did we observe a clear trend of faster folding into thinner bilayers, we also discovered that approaching the lipid phase transition temperature leads to accelerated folding of OmpA in bilayers composed of diC\textsubscript{14}PC. When lipid bilayers undergo a phase transition, multiple separate domains of the two phases coexist and increased permeability due to packing defects has been observed at the interfaces between these domains (Heimburg, 2007; Mouritsen et al., 1995). Such defects could play the same role that the more prevalent defects of thinner bilayers play in promoting OMP folding.

Furthermore, having developed a kinetic scheme for OmpA folding and structural models for the intermediate states, we can examine in more detail the role of the bilayer environment and defects on the folding process. In comparing the models and fitted rate constants for the different lipid conditions, it is apparent that the faster folding observed

§§ We have shown in this work that OmpA can also fold into LUVs with lipid chains of 13 and 14 carbons. The previous lack of folding into diC\textsubscript{14}PC LUVs was probably due to differences in experimental conditions and monitoring folding for too short of a time.
in thinner bilayers is due to the early steps being so fast that they are kinetically invisible, in addition to the later steps also being faster. For example, we determined that for folding into diC₉PC and diC₁₀PC, the protein converts to the partially folded intermediate state C within the mixing time of the experiment. In the next longest chain lipid, diC₁₁PC (and in diC₁₄PC at the phase transition temperature, 24 °C), the conversion from B to C is observed, but prior steps are still faster than can be measured without a stopped-flow apparatus. Following this trend, in diC₁₂PC (and diC₁₄PC at 25 and 26 °C) the pathway is observed beginning with state A, and in diC₁₃PC an additional reaction from A' to A is manifested in the observed kinetics. In addition to the early steps becoming too fast to be measured, the final reaction step of C to F and the error correction step from D to C are also faster in thinner bilayers (see fitted rate constants in Tables 3.3-10). Thus all of the main pathway steps are accelerated by the lipid bilayer becoming thinner (or having more lipid phase domain interfaces), with the earliest steps affected the most. Considering the structural features of the intermediate states (Figure 3.28), it is clear that progressive bilayer insertion is an important aspect of the transformation of each conformation to the next one. Therefore it makes sense that these steps would occur faster in bilayers that have more defects. The conversion from state A to B involves insertion of some or all of the membrane-facing residues into the interfacial region of the bilayer, B to C induces β-hairpin insertion into the hydrophobic core of the bilayer, and formation of the native barrel requires the complete penetration of the strands through the membrane. In bilayers that have packing defects and occasional pore formation, the energy barrier to such insertion steps would be greatly decreased and the protein would thus achieve the transition much faster. The earlier insertion events are probably more accelerated by
increasing bilayer defects because the barrier to initial membrane penetration (A to B and B to C) would be higher than continued translocation across the bilayer from a partially inserted state (C to F).

Inserting into the membrane is not the only requirement for β-barrel folding; the proper secondary structure must also be formed to progress along the pathway. As discussed above, species B is predicted to have a loose arrangement of the β-hairpins in a clover conformation on the membrane surface. Previous work has demonstrated that the OmpA_{171} β-barrel inserts unidirectionally into the membrane (Surrey & Jähnig, 1992), so this conformation must be arranged so that only the trans ends of the β-strands are at the center of the clover shape and insert across the bilayer. In addition, the correct polypeptide segments need to be positioned next to each other so that the correct H-bonds can form and the β-strands will be in register in the final barrel structure. We propose that the formation of the periplasmic turns is what initiates proper β-hairpin arrangement and subsequent strand formation. Folding of the turn regions would force the polypeptide into the clover conformation (with the turns at the tips of the clover “leaves”) and the trans ends toward the center where they could penetrate into the bilayer in a concerted fashion. In conjunction with insertion, the β-strands would “zip up” along the inserted and surface-located regions of the chain to form the high β-sheet content observed experimentally. In support of this model, it has been shown that mutations in turn residues result in impaired OmpA assembly in vivo (Koebnik & Krämer, 1995). Having now established a detailed picture of the in vitro folding process, it would be of great interest to more closely examine the role of turn formation in OMP folding.
Thus it appears that the two key facilitators of β-barrel membrane protein folding are membrane defects and the propensity to form the correct secondary structure. Bilayer defects promote insertion of the protein into the membrane, which occurs in conjunction with β-hairpin formation and eventually complete barrel translocation. Still, folding into the thin bilayers examined here results in OmpA folding kinetics on a timescale of hours, with slow conversion from off-pathway intermediate states continuing for days. Clearly OMP folding in the cell must be much faster, to be compatible with bacterial doubling times. Therefore the role of the BAM complex (T. Wu et al., 2005) in vivo may be to further accelerate the intrinsic OMP folding process. The information required for proper secondary structure formation is contained within the OMP polypeptide sequence, so it is likely that the BAM complex serves to facilitate membrane insertion by creating bilayer defects. This idea is supported by the observation that the subunit of the BAM complex capable of catalyzing OMP folding in vitro, BamA, is located in the outer membrane, and molecular dynamics simulations of this protein have indicated thinning of the bilayer adjacent to the protein surface (Gessmann et al., 2014; Noinaj et al., 2013). Acceleration of membrane insertion would also help reduce the population of off-pathway intermediate states and thereby further improve the efficiency of OMP folding.
3.6 Acknowledgements

This work was supported by grants from the NSF (MCB 0919868) and the NIH (R01 GM079440 and T32 GM008403). EJD is a recipient of an NSF graduate research fellowship. We gratefully acknowledge Michael McCaffery and Erin Pryce of the Johns Hopkins Integrated Imaging Center for assistance with electron microscopy. We also thank members of the Fleming lab and Doug Barrick for many helpful discussions.
3.7 Tables and Figures

Table 3.1 CD signal values for unfolded OmpA_{171}

<table>
<thead>
<tr>
<th>Wavelength (nm)</th>
<th>[Θ]^a (deg cm^2 dmol^{-1} res^{-1})</th>
<th>Δε^b (M^{-1} cm^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>-3860</td>
<td>-201.3</td>
</tr>
<tr>
<td>230</td>
<td>-1216</td>
<td>-63.4</td>
</tr>
</tbody>
</table>

^a Mean residue ellipticity for unfolded OmpA_{171} in 1 M urea, 20 mM Tris, pH 8, calculated using Equation 3.13 in Materials and Methods.

^b Molar circular dichroism, calculated from mean residue ellipticity using Equation 3.13.
Table 3.2  Kinetic rate constants for diC$_{12}$PC model development

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheme 3.1:</td>
<td></td>
</tr>
<tr>
<td>$U \xrightleftharpoons[k_1]{k_1} F \xrightleftharpoons[k_2]{k_2} G$</td>
<td>$k_1$ 0.0006, $k_{-1}$ 0.00001, $k_2$ 0.008, $k_{-2}$ 0.005</td>
</tr>
<tr>
<td>Scheme 3.3:</td>
<td></td>
</tr>
<tr>
<td>$U \xrightleftharpoons[k_1]{k_1} F \xrightleftharpoons[k_2]{k_2} G$</td>
<td>$k_1$ 0.000025, $k_{-1}$ 0.000011, $k_2$ 0.0007, $k_{-2}$ 0.00002</td>
</tr>
<tr>
<td>Scheme 3.4:</td>
<td></td>
</tr>
<tr>
<td>$A \xrightleftharpoons[k_1]{k_1} B \xrightleftharpoons[k_2]{k_2} F$</td>
<td>$k_1$ 0.0004, $k_{-1}$ 0.0005, $k_2$ 0.003, $k_{-2}$ 0.000055, $k_3$ 0.0007, $k_{-3}$ 0.00001</td>
</tr>
<tr>
<td>Scheme 3.6:</td>
<td></td>
</tr>
<tr>
<td>$C \xrightleftharpoons[k_1]{k_1} A \xrightleftharpoons[k_2]{k_2} B \xrightleftharpoons[k_3]{k_3} F$</td>
<td>$k_1$ 0.0005, $k_{-1}$ 0, $k_2$ 0.01, $k_{-2}$ 1 x 10$^{-6}$, $k_3$ 0.005, $k_{-3}$ 0.000025, $k_4$ 0.0007, $k_{-4}$ 0</td>
</tr>
<tr>
<td>Scheme 3.7:</td>
<td></td>
</tr>
<tr>
<td>$D \xrightleftharpoons[k_1]{k_1} A \xrightleftharpoons[k_2]{k_2} B \xrightleftharpoons[k_3]{k_3} C \xrightleftharpoons[k_4]{k_4} F$</td>
<td>$k_1$ 0.0005, $k_{-1}$ 0, $k_2$ 0.006, $k_{-2}$ 0, $k_3$ 0.01, $k_{-3}$ 1 x 10$^{-6}$, $k_4$ 0.005, $k_{-4}$ 0.000025, $k_5$ 0.0007, $k_{-5}$ 0</td>
</tr>
</tbody>
</table>
Table 3.3  Kinetic rate constants and CD signals for final diC$_{12}$PC model

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants (s$^{-1}$) $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>0.0005</td>
</tr>
<tr>
<td>$k_1$</td>
<td>0</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.006</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0</td>
</tr>
<tr>
<td>$k_3$</td>
<td>0.01</td>
</tr>
<tr>
<td>$k_3$</td>
<td>1 x 10$^{-6}$</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.005</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.000025</td>
</tr>
<tr>
<td>$k_5$</td>
<td>0.0007</td>
</tr>
<tr>
<td>$k_5$</td>
<td>0</td>
</tr>
<tr>
<td>$k_6$</td>
<td>0.003</td>
</tr>
<tr>
<td>$k_6$</td>
<td>0</td>
</tr>
<tr>
<td>$k_7$</td>
<td>0.007</td>
</tr>
<tr>
<td>$k_7$</td>
<td>0</td>
</tr>
<tr>
<td>$k_8$</td>
<td>0.001</td>
</tr>
<tr>
<td>$k_8$</td>
<td>0.00005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>CD signals of species (deg cm$^2$ dmol$^{-1}$ res$^{-1}$)</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td></td>
<td>-3860</td>
<td>-3860</td>
<td>-6700</td>
<td>-5000</td>
<td>-5000</td>
</tr>
<tr>
<td>230</td>
<td></td>
<td>-1216</td>
<td>-1216</td>
<td>-2700</td>
<td>-2200</td>
<td>+700</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G$1$</th>
<th>G$2$</th>
<th>G$3$</th>
<th>G$4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>-3860</td>
<td>-3860</td>
<td>-6700</td>
</tr>
<tr>
<td>230</td>
<td>-1216</td>
<td>-1216</td>
<td>-700</td>
</tr>
</tbody>
</table>

$^a$ Rate constants were determined by simulation of fitting curves and manual adjustment of values to minimize chi-squared. See Figure 3.19 for illustration of how well-determined each rate constant is.
Table 3.4 Kinetic rate constants and CD signals for diC$_{13}$PC model

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_1'$</td>
</tr>
<tr>
<td></td>
<td>$k_{-1}'$</td>
</tr>
<tr>
<td></td>
<td>$k_1$</td>
</tr>
<tr>
<td></td>
<td>$k_{-1}$</td>
</tr>
<tr>
<td></td>
<td>$k_2$</td>
</tr>
<tr>
<td></td>
<td>$k_{-2}$</td>
</tr>
<tr>
<td></td>
<td>$k_3$</td>
</tr>
<tr>
<td></td>
<td>$k_{-3}$</td>
</tr>
<tr>
<td></td>
<td>$k_4$</td>
</tr>
<tr>
<td></td>
<td>$k_{-4}$</td>
</tr>
<tr>
<td></td>
<td>$k_5$</td>
</tr>
<tr>
<td></td>
<td>$k_{-5}$</td>
</tr>
<tr>
<td></td>
<td>$k_6$</td>
</tr>
<tr>
<td></td>
<td>$k_{-6}$</td>
</tr>
<tr>
<td></td>
<td>$k_7$</td>
</tr>
<tr>
<td></td>
<td>$k_{-7}$</td>
</tr>
<tr>
<td></td>
<td>$k_8$</td>
</tr>
<tr>
<td></td>
<td>$k_{-8}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>CD signals of species (deg cm$^2$ dmol$^{-1}$ res$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>A'$</strong></td>
</tr>
<tr>
<td>216</td>
<td>-3860</td>
</tr>
<tr>
<td>230</td>
<td>-1216</td>
</tr>
<tr>
<td></td>
<td><strong>F</strong></td>
</tr>
<tr>
<td>216</td>
<td>-5000</td>
</tr>
<tr>
<td>230</td>
<td>+700</td>
</tr>
</tbody>
</table>
Table 3.5  Kinetic rate constants and CD signals for diC$_{11}$PC model

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_2'$ 0.018</td>
</tr>
<tr>
<td></td>
<td>$k_{2'} 0.002$</td>
</tr>
<tr>
<td></td>
<td>$k_2 0.028$</td>
</tr>
<tr>
<td></td>
<td>$k_3 0$</td>
</tr>
<tr>
<td></td>
<td>$k_4 0.02$</td>
</tr>
<tr>
<td></td>
<td>$k_5 0$</td>
</tr>
<tr>
<td></td>
<td>$k_{4} 0.00008$</td>
</tr>
<tr>
<td></td>
<td>$k_{5} 0.023$</td>
</tr>
<tr>
<td></td>
<td>$k_{4} 0.002$</td>
</tr>
<tr>
<td></td>
<td>$k_7 0.03$</td>
</tr>
<tr>
<td></td>
<td>$k_8 0.0001$</td>
</tr>
<tr>
<td></td>
<td>$k_8 0.0025$</td>
</tr>
<tr>
<td></td>
<td>$k_8 0.00007$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>CD signals of species (deg cm$^2$ dmol$^{-1}$ res$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>216</td>
<td>-3860</td>
</tr>
<tr>
<td>230</td>
<td>-1216</td>
</tr>
<tr>
<td></td>
<td>G$_2$</td>
</tr>
<tr>
<td>216</td>
<td>-3860</td>
</tr>
<tr>
<td>230</td>
<td>-1216</td>
</tr>
</tbody>
</table>
Table 3.6  Kinetic rate constants and CD signals for diC₉PC model

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants (s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( k_{3'} ) 0.07</td>
</tr>
<tr>
<td></td>
<td>( k_{3'} ) 0.005</td>
</tr>
<tr>
<td></td>
<td>( k_3 ) 0.045</td>
</tr>
<tr>
<td></td>
<td>( k_3 ) 3 \times 10^{-6}</td>
</tr>
<tr>
<td></td>
<td>( k_4 ) 0.02</td>
</tr>
<tr>
<td></td>
<td>( k_4 ) 0.0004</td>
</tr>
<tr>
<td></td>
<td>( k_5 ) 0.02</td>
</tr>
<tr>
<td></td>
<td>( k_5 ) 0.00018</td>
</tr>
<tr>
<td></td>
<td>( k_8 ) 0.0003</td>
</tr>
<tr>
<td></td>
<td>( k_8 ) 0.00003</td>
</tr>
</tbody>
</table>

Wavelength | CD signals of species (deg cm² dmol⁻¹ res⁻¹)
---|---
<table>
<thead>
<tr>
<th>C</th>
<th>C'</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>-6700</td>
<td>-6700</td>
<td>-5000</td>
</tr>
<tr>
<td>230</td>
<td>-2700</td>
<td>-2700</td>
<td>-2200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C₃</th>
<th>C₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>-6700</td>
</tr>
<tr>
<td>230</td>
<td>-900</td>
</tr>
</tbody>
</table>
Table 3.7  Kinetic rate constants and CD signals for diC\textsubscript{10}PC model

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants (s\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_{3}')</td>
<td>0.035</td>
</tr>
<tr>
<td>(k_{3}')</td>
<td>0.01</td>
</tr>
<tr>
<td>(k_3)</td>
<td>0.05</td>
</tr>
<tr>
<td>(k_3)</td>
<td>(3 \times 10^{-6})</td>
</tr>
<tr>
<td>(k_{4})</td>
<td>0.018</td>
</tr>
<tr>
<td>(k_{4})</td>
<td>0.0008</td>
</tr>
<tr>
<td>(k_5)</td>
<td>0.025</td>
</tr>
<tr>
<td>(k_5)</td>
<td>0.0008</td>
</tr>
<tr>
<td>(k_8)</td>
<td>0.002</td>
</tr>
<tr>
<td>(k_8)</td>
<td>0.00003</td>
</tr>
</tbody>
</table>

Wavelength | CD signals of species (deg cm\textsuperscript{2} dmol\textsuperscript{-1} res\textsuperscript{-1})
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>-6700</td>
<td>-6700</td>
<td>-5000</td>
<td>-5000</td>
</tr>
<tr>
<td>230</td>
<td>-2700</td>
<td>-2700</td>
<td>-2200</td>
<td>+700</td>
</tr>
</tbody>
</table>

\[\text{G}_3 \leftrightarrow \text{G}_4\]

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>CD signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>-6700</td>
</tr>
<tr>
<td>230</td>
<td>-900</td>
</tr>
</tbody>
</table>
Table 3.8  Kinetic rate constants and CD signals for diC₁₄PC at 25 ºC model

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants (s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_1$</td>
<td>0.001</td>
</tr>
<tr>
<td>$k_1$</td>
<td>0.003</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.01</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0</td>
</tr>
<tr>
<td>$k_3$</td>
<td>0.015</td>
</tr>
<tr>
<td>$k_3$</td>
<td>0</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.015</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.0001</td>
</tr>
<tr>
<td>$k_5$</td>
<td>0.0012</td>
</tr>
<tr>
<td>$k_5$</td>
<td>0</td>
</tr>
<tr>
<td>$k_6$</td>
<td>0.01</td>
</tr>
<tr>
<td>$k_6$</td>
<td>0</td>
</tr>
<tr>
<td>$k_7$</td>
<td>0.0008</td>
</tr>
<tr>
<td>$k_7$</td>
<td>0</td>
</tr>
<tr>
<td>$k_8$</td>
<td>0.0015</td>
</tr>
<tr>
<td>$k_8$</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>CD signals of species (deg cm⁻² dmol⁻¹ res⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>216</td>
<td>-3800</td>
</tr>
<tr>
<td>230</td>
<td>-1200</td>
</tr>
<tr>
<td>G₁</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>-3800</td>
</tr>
<tr>
<td>230</td>
<td>-1200</td>
</tr>
</tbody>
</table>
Table 3.9  Kinetic rate constants and CD signals for diC$_{14}$PC at 24 °C model

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants ($s^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_2$</td>
<td>0.016</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.013</td>
</tr>
<tr>
<td>$k_3$</td>
<td>0.0025</td>
</tr>
<tr>
<td>$k_3$</td>
<td>0</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.0018</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.00012</td>
</tr>
<tr>
<td>$k_5$</td>
<td>0.007</td>
</tr>
<tr>
<td>$k_5$</td>
<td>0.015</td>
</tr>
<tr>
<td>$k_7$</td>
<td>0.025</td>
</tr>
<tr>
<td>$k_7$</td>
<td>0.00005</td>
</tr>
<tr>
<td>$k_8$</td>
<td>0.0008</td>
</tr>
<tr>
<td>$k_8$</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>CD signals of species (deg cm$^2$ dmol$^{-1}$ res$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>B -3860, C -6700, D -5000, F -5000</td>
</tr>
<tr>
<td>230</td>
<td>B -1216, C -2700, D -2400, F -300</td>
</tr>
<tr>
<td>216</td>
<td>G$_2$ -3860, G$_3$ -6700, G$_4$ -5000</td>
</tr>
<tr>
<td>230</td>
<td>G$_2$ -1216, G$_3$ -2000, G$_4$ -1000</td>
</tr>
</tbody>
</table>
Table 3.10  Kinetic rate constants and CD signals for diC$_{14}$PC at 26 °C model

<table>
<thead>
<tr>
<th>Model</th>
<th>Rate constants (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>0.004</td>
</tr>
<tr>
<td>$k_{-1}$</td>
<td>0.00045</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.0001</td>
</tr>
<tr>
<td>$k_{-2}$</td>
<td>0.0003</td>
</tr>
<tr>
<td>$k_3$</td>
<td>0.0004</td>
</tr>
<tr>
<td>$k_{-3}$</td>
<td>0</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.0003</td>
</tr>
<tr>
<td>$k_{-4}$</td>
<td>0.000012</td>
</tr>
<tr>
<td>$k_5$</td>
<td>0.00072</td>
</tr>
<tr>
<td>$k_{-5}$</td>
<td>0.00001</td>
</tr>
<tr>
<td>$k_6$</td>
<td>0.0004</td>
</tr>
<tr>
<td>$k_{-6}$</td>
<td>0</td>
</tr>
<tr>
<td>$k_7$</td>
<td>0.0003</td>
</tr>
<tr>
<td>$k_{-7}$</td>
<td>0</td>
</tr>
<tr>
<td>$k_8$</td>
<td>0.0002</td>
</tr>
<tr>
<td>$k_{-8}$</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>CD signals of species (deg cm$^2$ dmol$^{-1}$ res$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>A -3800, B -3800, C -6700, D -5000, F -5000</td>
</tr>
<tr>
<td>230</td>
<td>A -1200, B -1200, C -2700, D -2200, F -300</td>
</tr>
<tr>
<td>216</td>
<td>G$_1$ -3800, G$_2$ -3800, G$_3$ -6700, G$_4$ -5000</td>
</tr>
<tr>
<td>230</td>
<td>G$_1$ -1200, G$_2$ -1200, G$_3$ -2000, G$_4$ -900</td>
</tr>
</tbody>
</table>
Table 3.11  Attributes of OmpA main pathway folding species

<table>
<thead>
<tr>
<th>Species</th>
<th>SDS-PAGE migration</th>
<th>Protease accessible</th>
<th>Trp location</th>
<th>Secondary structure</th>
<th>Exciton signal at 230 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>U_{AQ}</td>
<td>U</td>
<td>Yes</td>
<td>Solvent exposed</td>
<td>None</td>
<td>No</td>
</tr>
<tr>
<td>A</td>
<td>U</td>
<td>Yes</td>
<td>Membrane surface</td>
<td>None</td>
<td>No</td>
</tr>
<tr>
<td>B</td>
<td>U</td>
<td>Yes</td>
<td>Membrane interface</td>
<td>None</td>
<td>No</td>
</tr>
<tr>
<td>C</td>
<td>U</td>
<td>Yes</td>
<td>Membrane interior</td>
<td>β-sheet [\Theta]_{216} = -6700</td>
<td>No</td>
</tr>
<tr>
<td>D</td>
<td>U</td>
<td>?</td>
<td>?</td>
<td>β-sheet [\Theta]_{216} = -5000</td>
<td>No</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>No</td>
<td>Membrane interface</td>
<td>β-sheet [\Theta]_{216} = -5000</td>
<td>Yes</td>
</tr>
</tbody>
</table>

\[a\] Determined by this work and consistent with (Kleinschmidt & Tamm, 1996; Rodionova et al., 1995; Surrey & Jähnig, 1992).

\[b\] Determined by (Rodionova et al., 1995; Surrey & Jähnig, 1992, 1995).

\[c\] Determined by (Kleinschmidt, den Blaauwen, et al., 1999; Rodionova et al., 1995).

\[d\] Determined by this work and consistent with (Kleinschmidt & Tamm, 2002a; Rodionova et al., 1995; Surrey & Jähnig, 1992).

\[e\] Determined by this work.
Figure 3.1 Densitometry of folding gels demonstrates the “gone” state
Representative gels and corresponding densitometry values for OmpA_{171} folding into (A) diC_{10}PC, (B) diC_{12}PC, and (C) in the absence of LUVs. Folded and unfolded bands are indicated by “F” and “U”, respectively. The last lane on each gel contains a duplicate sample of the last time point that has been boiled at 95-100 °C for 5 min. Gels have been artificially darkened for better clarity but densitometry analysis was performed on unaltered scans. Folded band intensities are shown as blue triangles, unfolded band intensities as upside-down red triangles, and the sum of the folded and unfolded intensities as green diamonds. The boiled band intensity is shown as a filled green diamond. The difference between the “F+U” data and the boiled point indicates the amount of “gone” state formed. In (C) there is no folded band formed in the absence of LUVs, so only the unfolded band intensities and the boiled band intensity are plotted. The boiled intensity matches the unfolded band intensity, indicating that no gone state is formed under this condition.
Figure 3.2 Dividing by the boiled band intensity is a more accurate method of calculating fractional values

Fractions folded and unfolded for OmpA$_{171}$ folding into LUVs of (A) diC$_{10}$PC, and (B) diC$_{12}$PC. Fractional quantities were calculated by dividing by the sum of the folded and unfolded band intensities (“F+U”) or by the boiled band intensity (“B”). F/(F+U) values are plotted as light blue triangles, F/B as filled dark blue triangles, U/(F+U) as upside-down orange triangles, and U/B as filled upside-down red triangles. All values are averaged from three independent folding reactions with error bars denoting the standard
deviation. Data are fitted to simple sums of exponential terms. To show more kinetic
detail, data are plotted in two time ranges: 3600 s (1 h) and 172800 s (2 days) for the
diC₁₀PC data, or 7200 s (2 h) and 432000 s (5 days) for the diC₁₂PC data.
Figure 3.3  Electron microscopy images of LUVs for all lipids

Transmission electron microscopy images of LUVs composed of (A) \( \text{diC}_9\text{PC} \), (B) \( \text{diC}_{10}\text{PC} \), (C) \( \text{diC}_{11}\text{PC} \), (D) \( \text{diC}_{12}\text{PC} \), (E) \( \text{diC}_{13}\text{PC} \), and (F) \( \text{diC}_{14}\text{PC} \). LUVs were prepared by extrusion at a concentration of 10 mg ml\(^{-1}\) lipid in 1 M urea, 20 mM Tris, pH 8, and diluted to 0.5 mg ml\(^{-1}\) in the same buffer. Samples were adsorbed to freshly ionized formvar/carbon-coated grids for 5 min and negatively stained with 2% uranyl acetate for 1 min. Images were captured using an FEI Tecnai 12 electron microscope equipped with an SIS Megaview III camera. Variation in vesicle darkness between images is due to variability in staining. Black bars in all panels represent 200 nm.
Figure 3.4 The kinetics of the folded, unfolded, and gone species are highly dependent on bilayer thickness
Time-dependence of species measured by SDS-PAGE for OmpA$_{171}$ folding into LUVs composed of diC$_9$PC – diC$_{14}$PC, at 25 °C: (A) fraction folded, (B) fraction unfolded, and (C) fraction gone. Densitometry values were converted to fractional quantities by dividing by the boiled band intensity. All values are averaged from at least three independent folding reactions with error bars denoting the standard deviation. Data colors and symbols are indicated in the legends. Solid lines are fits to kinetic models discussed in the text and shown in Figures 3.18-24. To show more kinetic detail, data are plotted in two time ranges: 1800 s (30 min) and 18000 s (5 h). Longer-term time points can be seen in individual lipid plots (Figures 3.18-24).
Figure 3.5  

DSC of diC<sub>14</sub>PC LUVs reveals a phase transition at 24 °C

Differential scanning calorimetry was conducted on diC<sub>14</sub>PC LUVs at a concentration of 800 µM lipid in 1 M urea, 20 mM Tris, pH 8. The scan rate was 90 °C/h and the sample was scanned from 10 °C to 50 °C multiple times (one representative curve is shown). Data were converted to heat capacity in Origin. The peak occurs at 24 °C.
Figure 3.6  Folding kinetics in diC₁₄PC are fastest at the transition temperature
Time-dependence of species measured by SDS-PAGE for OmpA_{171} folding into diC_{14}PC at 24, 25, or 26 °C: (A) fraction folded, (B) fraction unfolded, and (C) fraction gone. All values are averaged from at least three independent folding reactions with error bars denoting the standard deviation. Data colors and symbols are indicated in the legends. Solid lines are fits to kinetic models discussed in the text and shown in Figures 3.24-26. Data are plotted in two time ranges: 1800 s (30 min) and 18000 s (5 h). Longer-term time points can be seen in individual lipid plots (Figures 3.24-26). The fastest kinetics for this lipid condition occur at 24 °C, which was determined to be the lipid phase transition temperature by DSC (Figure 3.5).
Figure 3.7  Unfolded OmpA$_{171}$ has no regular structure in the absence of a bilayer

(A) CD spectra, converted to mean residue ellipticity, for 1 µM OmpA$_{171}$ in 8 M (green dotted line) or 1 M urea (red solid line), 20 mM Tris, pH 8. In the absence of a bilayer, the protein is unfolded and shows no regular secondary structure at both urea concentrations. Purple and blue filled circles indicate the values at 216 nm and 230 nm, respectively. (B) Kinetic traces for OmpA$_{171}$ in 1 M urea, under the same conditions as in (A), measured at 216 nm (purple) or 230 nm (blue). Black lines are plotted at the values measured in the wavelength spectrum (-3860 at 216 nm and -1216 at 230 nm) to show that the signal does not change at either wavelength over the course of 4 h. This lack of change indicates that the protein does not undergo any conformational changes in 1 M urea in the absence of a bilayer.
Figure 3.8  LUVs exhibit a CD signal

CD wavelength spectra of large unilamellar vesicles (LUVs) composed of different lipids (diC₉PC – diC₁₄PC), measured at a concentration of 800 µM, in 1 M urea, 20 mM Tris, pH 8, in a 1 cm cuvette at 25 °C. Each spectrum is the average of 3-5 separately prepared and measured samples. The CD signal is reported as ellipticity, in mdeg.
Figure 3.9 CD spectra of folded OmpA<sub>171</sub> indicate β-sheet structure and an aromatic exciton interaction

CD wavelength spectra of OmpA<sub>171</sub> in various lipids measured after kinetics. Each spectrum is the average of 3-5 separately prepared and measured samples. Curves were corrected with the appropriate LUV spectrum, shown in Figure 3.8, and converted to mean residue ellipticity using the concentration measured before the folding kinetics (see Materials and Methods). Each plot also contains the spectrum for unfolded OmpA<sub>171</sub> in 1 M urea from Figure 3.7A, as a red dashed line. (A) Spectrum after folding into LUVs of diC<sub>10</sub>PC for 5 h (blue). Black arrows indicate the change in signal at 216 nm and 230 nm from the unfolded state. (B) Spectra after folding into LUVs of diC<sub>9</sub>PC – diC<sub>14</sub>PC for 5 h, measured at 25 ºC. The trough at 216 nm is indicative of β-sheet structure and the peak at 230 nm is thought to be due to an exciton interaction between aromatic residues in the native state.
Figure 3.10  Processing of raw CD kinetics data

CD kinetics data at (A) 216 nm, or (B) 230 nm for OmpA_{171} folding into LUVs of diC_{10}PC at final concentrations of 1 μM protein and 800 μM lipid, in Folding Buffer (1 M urea, 20 mM Tris, pH 8). Measurements were conducted at 25 °C with a 1 cm path.
length and constant stirring. The interval between data points was 10 s and the time constant was 1 s. The top panels show the raw data, reported by the instrument in mdeg. For each experiment, the signal for Folding Buffer alone was measured first for 3 min followed by the addition of protein (indicated by an arrow), which was then measured for 2 min. LUVs were then added (indicated by an arrow) to initiate folding, and the kinetics were measured to completion (data only shown to 2000 s). The middle panels show data that have been corrected by subtracting the cuvette/buffer background and the LUV signal at the corresponding wavelength. The cuvette/buffer background was determined by averaging the kinetics data measured for Folding Buffer alone. This value was subtracted from the entire dataset, which is apparent from the initial signal being shifted to zero. The LUV signal was subtracted from the data measured after the addition of LUVs, and was determined from measurements on a separately prepared LUV-only sample. The bottom panels show data that have been converted to mean residue ellipticity. This was accomplished using the protein concentration determined from the signal of protein alone and the previously measured mean residue ellipticity value for unfolded protein (see Materials and Methods for equations). The mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted in each panel as a red dashed line and indicates what the signal should be before folding is initiated. The data have also been shifted so that the folding initiation (i.e., LUV addition) corresponds to time zero.
Figure 3.11 CD kinetics at 216 nm and 230 nm are complex and dependent on bilayer thickness

CD signal kinetics measured at (A) 216 nm, and (B) 230 nm, for OmpA_{171} folding into LUVs composed of diC_{9}PC – diC_{14}PC, under the same conditions as SDS-PAGE kinetics (Figure 3.4). Data were converted to mean residue ellipticity using the concentration measured before initiation of folding. Colors are indicated in the legends. The mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted in each panel as a red dashed line. Traces measured at 216 nm are the average of 2 or 3 separate
folding experiments. To show more kinetic detail, data are plotted in two time ranges: 1800 s (30 min) and 14400 s (4 h) for the 216 nm data, or 1800 s (30 min) and 18000 s (5 h) for the 230 nm data.
Figure 3.12 CD kinetics in diC₁₄PC are fastest at the transition temperature

CD signal kinetics measured at (A) 216 nm, and (B) 230 nm, for OmpA₁₇₁ folding into diC₁₄PC at 24, 25, or 26 °C. Data were converted to mean residue ellipticity using the concentration measured before initiation of folding. Colors are indicated in the legends. The mean residue ellipticity for unfolded OmpA₁₇₁ at the appropriate wavelength is plotted in each panel as a red dashed line. Traces measured at 216 nm are the average of two separate folding experiments. To show more kinetic detail, data are plotted in two
time ranges: 1800 s (30 min) and 14400 s (4 h) for the 216 nm data, or 1800 s (30 min) and 18000 s (5 h) for the 230 nm data. The inset in (B) shows the 230 nm data over the course of 38000 s (10.5 h).
Figure 3.13 The initial wavelength spectrum of OmpA<sub>171</sub> in diC<sub>9</sub>PC reveals an intermediate with high β-sheet content

CD wavelength spectra of OmpA<sub>171</sub> immediately after initiating folding into LUVs of (A) diC<sub>9</sub>PC, or (B) diC<sub>13</sub>PC (dotted lines) at 25 ºC. Data were collected in 10 nm sections to capture the signal at the earliest times possible, and then combined into a full spectrum. Each wavelength section was averaged from 4 or 5 separately prepared folding samples. Data were converted to mean residue ellipticity using the concentration measured before adding LUVs. Also included in each plot for comparison are the spectrum for unfolded OmpA<sub>171</sub> from Figure 3.7A (red dashed line), and the spectrum measured after folding was complete in the corresponding lipid (solid line), from Figure 3.9B. In diC<sub>9</sub>PC, the initial spectrum is clearly different from the unfolded spectrum, indicating a fast conformational change takes place within the mixing time of the experiment. This fast change is not observed when folding into diC<sub>13</sub>PC at 25 ºC, as evidenced by the initial spectrum overlaying well with the unfolded spectrum.
Figure 3.14  OmpA\textsubscript{171} adopts the same β-structured conformation in diC\textsubscript{13}PC at 4 °C as was observed in diC\textsubscript{9}PC

(A) Mean residue ellipticity at 216 nm of unfolded OmpA\textsubscript{171} in 1 M urea between 4 and 30 °C. The data are fit to a line to demonstrate the linear dependence of the signal on temperature. The values at 4 °C and 25 °C differ by a factor of 1.25. (B) The CD wavelength spectrum of OmpA\textsubscript{171} folded into diC\textsubscript{13}PC LUVs at 25 °C and cooled to 4 °C (dash-dot pink line). Scaling the data by a factor of 1.25 produces a spectrum (solid pink line) that matches well with the spectrum obtained at 25 °C (solid purple line). (C) The
CD wavelength spectrum of OmpA$_{171}$ incubated with diC$_{13}$PC at 4 °C (dash-dot orange line), and scaled as in (B) (solid orange line). Plotted for comparison are the wavelength spectrum of OmpA$_{171}$ folded in diC$_{13}$PC at 25 °C (solid purple line) and the spectrum of the β-structured intermediate obtained immediately after mixing with diC$_9$PC at 25 °C (dotted light blue line).
Figure 3.15  The β-structured intermediate forms quickly in diC₁₃PC at 4 °C and converts rapidly to the native state at 25 °C

(A) CD kinetics at 216 nm for OmpA₁₇₁ in the presence of diC₁₃PC LUVs at 4 °C (orange), scaled by a factor of 1.25 to correct for temperature dependence. Plotted data are the average of two separate experiments. Also shown are the kinetics at 216 nm for OmpA₁₇₁ folding into diC₁₃PC at 25 °C (purple) and the unfolded value at 216 nm at 25 °C (red dashed line). The signal at 4 °C reaches a more negative ellipticity than at 25 °C because the protein is forming the more highly β-structured intermediate state. (B) CD
kinetics at 230 nm for OmpA_{171} folding into diC_{13}PC at 25 °C after incubating at 4 °C for 5 h (light green). The signal starts at a more negative ellipticity than the unfolded signal (red dashed line) due to the protein already being in the β-structured intermediate conformation. The slight lag at the beginning of the kinetics is likely due to temperature equilibration occurring for the first ~250 s. Plotted for comparison are the kinetics at 230 nm for OmpA_{171} folding into diC_{13}PC directly at 25 °C (purple). All data were converted to mean residue ellipticity using the concentration measured before initiation of kinetics. To show more kinetic detail, data are plotted in two time ranges: 1800 s (30 min) and 18000 s (5 h).
Figure 3.16  A two-step sequential mechanism is inadequate to fit the SDS-PAGE data for OmpA$_{171}$ folding into diC$_{12}$PC

SDS-PAGE data for OmpA$_{171}$ folding into diC$_{12}$PC, fitted to Scheme 3.1. The three plots depict the same data over three time ranges: 800 s (13 min), 18000 s (5 h), and 432000 s (5 days). Measured data are denoted by symbols and fitted curves are shown as solid lines. The two-step kinetic mechanism used to fit the data is displayed in the middle plot with the states color-coded according to which gel species they correspond to (folded species in blue, unfolded species in red, and gone species in green). Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot. Rate constants obtained for the fit are listed in Table 3.2.
Figure 3.17 Models of increasing complexity better describe the diC_{12}PC gel data
SDS-PAGE data for OmpA\textsubscript{171} folding into diC\textsubscript{12}PC, fitted to kinetic mechanisms of increasing complexity: (A) Scheme 3.3, (B) Scheme 3.4, (C) Scheme 3.6, and (D) Scheme 3.7. Data in each panel are plotted over three time ranges: 800 s (13 min), 18000 s (5 h), and 432000 s (5 days). Measured data are denoted by symbols and fitted curves are shown as solid lines. The kinetic mechanism used to fit the data is displayed in each middle plot with the states color-coded according to which gel species they correspond to (folded species in blue, unfolded species in red, and gone species in green). Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel. Rate constants obtained for each fit are listed in Table 3.2.
Figure 3.18  Final model and fitted kinetic data for folding into diC$_{12}$PC
(A) The mechanism used to fit the kinetics data, Scheme 3.8, with the states color-coded to match the gel data. The protein population is assumed to all begin in state A. (B) Fitted SDS-PAGE data, shown over three time ranges: 800 s (13 min), 18000 s (5 h), and 432000 s (5 days). Measured data are denoted by symbols and fitted curves are shown as solid lines. (C) CD kinetics measured at 216 nm (purple line), plotted over two time ranges: 1200 s (20 min) and 14400 s (4 h). Fitted curve shown as a black line. (D) CD kinetics measured at 230 nm (light blue line), plotted over two time ranges: 1200 s (20 min) and 18000 s (5 h). Fitted curve shown as a black line. In (C) and (D) the mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted as a red dashed line. The fitted rate constants and the CD signals for all species used to fit the CD data are listed in Table 3.3. Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel.
(continued on next page)
Figure 3.19 Variability of rate constants determined for diC\textsubscript{12}PC model

To explore how well-defined the rates constants of Scheme 3.8 were after fitting the diC\textsubscript{12}PC kinetics data, each rate constant was varied independently (holding the other rate constants at the final values) and the chi-squared value monitored for the fits to the gel and CD data. Gel data chi-squared values are shown in the leftmost plots (red circles), and CD data chi-squared values are shown in the middle (216 nm, purple triangles) and rightmost plots (230 nm, light blue diamonds). Rate constants: (A) \( k_1 \), (B) \( k_{-1} \), (C) \( k_2 \), (D) \( k_{-2} \), (E) \( k_3 \), (F) \( k_{-3} \), (G) \( k_4 \), (H) \( k_{-4} \), (I) \( k_5 \), (J) \( k_{-5} \), (K) \( k_6 \), (L) \( k_{-6} \), (M) \( k_7 \), (N) \( k_{-7} \), (O) \( k_8 \), (P) \( k_{-8} \). Rate constant values are plotted on a log scale, with the distance between tick marks representing a factor of 10. The filled symbol in each plot indicates the value used in the final fitting and listed in Table 3.3. For the rate constants \( k_{-1}, k_{-2}, k_{-5}, k_{-6}, \) and \( k_{-7} \), the final value of zero is plotted on the y-axis to represent this value on the log scale.
Figure 3.20  Model and fitted kinetic data for folding into diC\textsubscript{13}PC
(A) The mechanism used to fit the kinetics data, Scheme 3.9, with the states color-coded to match the gel data. The protein population is assumed to all begin in state A'. (B) Fitted SDS-PAGE data, shown over three time ranges: 3600 s (1 h), 18000 s (5 h), and 432000 s (5 days). Measured data are denoted by symbols and fitted curves are shown as solid lines. (C) CD kinetics measured at 216 nm (purple line), plotted over two time ranges: 5400 s (1.5 h) and 18000 s (5 h). Fitted curve shown as a black line. (D) CD kinetics measured at 230 nm (light blue line), plotted over two time ranges: 5400 s (1.5 h) and 18000 s (5 h). Fitted curve shown as a black line. In (C) and (D) the mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted as a red dashed line. The fitted rate constants and the CD signals for all species used to fit the CD data are listed in Table 3.4. Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel.
Figure 3.21 Model and fitted kinetic data for folding into diC$_{11}$PC
(A) The mechanism used to fit the kinetics data, Scheme 3.11, with the states color-coded to match the gel data. The protein population is assumed to all begin in state B. (B) Fitted SDS-PAGE data, shown over three time ranges: 800 s (13 min), 18000 s (5 h), and 604800 s (7 days). Fitted curves are shown as solid lines. (C) CD kinetics measured at 216 nm (purple line), plotted over two time ranges: 800 s (13 min) and 14400 s (4 h). Fitted curve shown as a black line. (D) CD kinetics measured at 230 nm (light blue line), plotted over two time ranges: 800 s (13 min) and 18000 s (5 h). Fitted curve shown as a black line. In (C) and (D) the mean residue ellipticity for unfolded OmpA\textsubscript{171} at the appropriate wavelength is plotted as a red dashed line. The fitted rate constants and the CD signals for all species used to fit the CD data are listed in Table 3.5. Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel.
Figure 3.22 Model and fitted kinetic data for folding into diC$_9$PC
(A) The mechanism used to fit the kinetics data, Scheme 3.12, with the states color-coded to match the gel data. The protein population is assumed to all begin in state C. (B) Fitted SDS-PAGE data, shown over three time ranges: 800 s (13 min), 18000 s (5 h), and 432000 s (5 days). Fitted curves are shown as solid lines. (C) CD kinetics measured at 216 nm (purple line), plotted over two time ranges: 800 s (13 min) and 14400 s (4 h). Fitted curve shown as a black line. (D) CD kinetics measured at 230 nm (light blue line), plotted over two time ranges: 800 s (13 min) and 18000 s (5 h). Fitted curve shown as a black line. In (C) and (D) the mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted as a red dashed line. The fitted rate constants and the CD signals for all species used to fit the CD data are listed in Table 3.6. Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel.
Figure 3.23 Model and fitted kinetic data for folding into diC\textsubscript{10}PC
(A) The mechanism used to fit the kinetics data, Scheme 3.12, with the states color-coded to match the gel data. The protein population is assumed to all begin in state C. (B) Fitted SDS-PAGE data, shown over three time ranges: 800 s (13 min), 18000 s (5 h), and 432000 s (5 days). Fitted curves are shown as solid lines. (C) CD kinetics measured at 216 nm (purple line), plotted over two time ranges: 800 s (13 min) and 14400 s (4 h). Fitted curve shown as a black line. (D) CD kinetics measured at 230 nm (light blue line), plotted over two time ranges: 800 s (13 min) and 18000 s (5 h). Fitted curve shown as a black line. In (C) and (D) the mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted as a red dashed line. The fitted rate constants and the CD signals for all species used to fit the CD data are listed in Table 3.7. Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel.
Figure 3.24 Model and fitted kinetic data for folding in diC$_{14}$PC at 25 ºC
(A) The mechanism used to fit the kinetics data, Scheme 3.8, with the states color-coded to match the gel data. The protein population is assumed to all begin in state A. (B) Fitted SDS-PAGE data, shown over three time ranges: 800 s (13 min), 18000 s (5 h), and 604800 s (7 days). Measured data are denoted by symbols and fitted curves are shown as solid lines. (C) CD kinetics measured at 216 nm (purple line), plotted over two time ranges: 3600 s (1 h) and 14400 s (4 h). Fitted curve shown as a black line. (D) CD kinetics measured at 230 nm (light blue line), plotted over two time ranges: 3600 s (1 h) and 18000 s (5 h). Fitted curve shown as a black line. In (C) and (D) the mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted as a red dashed line. The fitted rate constants and the CD signals for all species used to fit the CD data are listed in Table 3.8. Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel.
Figure 3.25 Model and fitted kinetic data for folding in diC$_{14}$PC at 24 °C
(A) The mechanism used to fit the kinetics data, Scheme 3.10, with the states color-coded to match the gel data. The protein population is assumed to all begin in state B. (B) Fitted SDS-PAGE data, shown over three time ranges: 800 s (13 min), 18000 s (5 h), and 604800 s (7 days). Measured data are denoted by symbols and fitted curves are shown as solid lines. (C) CD kinetics measured at 216 nm (purple line), plotted over two time ranges: 1800 s (30 min) and 14400 s (4 h). Fitted curve shown as a black line. (D) CD kinetics measured at 230 nm (light blue line), plotted over two time ranges: 1800 s (30 min) and 25200 s (7 h). Fitted curve shown as a black line. In (C) and (D) the mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted as a red dashed line. The fitted rate constants and the CD signals for all species used to fit the CD data are listed in Table 3.9. Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel.
Figure 3.26 Model and fitted kinetic data for folding in diC_{14}PC at 26 °C
(A) The mechanism used to fit the kinetics data, Scheme 3.8, with the states color-coded to match the gel data. The protein population is assumed to all begin in state A. (B) Fitted SDS-PAGE data, shown over three time ranges: 800 s (13 min), 18000 s (5 h), and 604800 s (7 days). Measured data are denoted by symbols and fitted curves are shown as solid lines. (C) CD kinetics measured at 216 nm (purple line), plotted over two time ranges: 5400 s (1.5 h) and 21000 s (6 h). Fitted curve shown as a black line. (D) CD kinetics measured at 230 nm (light blue line), plotted over two time ranges: 5400 s (1.5 h) and 38000 s (10.5 h). Fitted curve shown as a black line. In (C) and (D) the mean residue ellipticity for unfolded OmpA_{171} at the appropriate wavelength is plotted as a red dashed line. The fitted rate constants and the CD signals for all species used to fit the CD data are listed in Table 3.10. Residuals for the fits are shown above each plot and the total chi-squared value is shown in the rightmost plot of each panel.
Figure 3.27 CD signals of mechanism species are consistent across all lipid conditions
The CD signals at 216 nm (purple bars) and 230 nm (light blue bars) determined for each species in the kinetic models developed for each lipid condition. (A) The signals assigned to species $A'$, $A$, $B$, $B'$, $G_1$, and $G_2$. These are grouped into one plot because these species all have the same signals at 216 and 230 nm, that of unfolded protein, when they are present in the model. Asterisks indicate that these species were not observed, and thus not part of the mechanism, for diC$_9$PC and diC$_{10}$PC. $A'$ is only present in the model for diC$_{13}$PC and $B'$ is only present in the model for diC$_{11}$PC. (B) The signals assigned to species $C$ and $C'$. $C'$ is only present in the model for diC$_9$PC and diC$_{10}$PC. (C) The signals assigned to species $D$. (D) The signals assigned to species $F$. (E) The signals assigned to species $G_3$. (F) The signals assigned to species $G_4$. 
Figure 3.28 Structural mechanism proposed for OmpA β-barrel folding

Illustration of the proposed on-pathway intermediate states in the folding mechanism of the OmpA β-barrel. The attributes of each species are summarized in Table 3.11. $U_{AQ}$ is the folding competent, non-membrane-associated state that was discussed in Chapter 2. Upon membrane association, $U_{AQ}$ converts to species A, which contains no regular structure. A then converts to species B, which has no regular secondary structure but has a loose arrangement of β-hairpins in a “clover-like” shape, oriented parallel to the membrane surface. Membrane-facing residues penetrate to the interfacial region of the bilayer. Site-directed quenching studies have revealed that residues on the translocating ($trans$) portions of adjacent strands (red circles) are in closer proximity in this state than residues on the periplasmic ($cis$) portions of strands (green circles)(Kleinschmidt et al., 2011). B converts to species C, which is partially inserted and contains a higher β-sheet
content than the native state. For this reason, the portions of the strands remaining on the surface as well as the membrane-embedded extracellular loops are proposed to be engaged in H-bonded β-strand conformations in addition to the membrane-inserted portion of the β-barrel. The closer-proximity trans residues are again indicated by red circles and the cis residues as green circles. Full insertion of intermediate C leads to formation of the native β-barrel structure, F (OmpA_{171} PDB id 1BXW)(Pautsch & Schulz, 1998). This state exhibits a shifted SDS-PAGE migration and an aromatic exciton signal in the CD spectrum at 230 nm. The interacting residues responsible for this signal are unknown, but the interaction only occurs in the native state and is depicted by a yellow star. The trans and cis residues used in site-directed quenching experiments are again indicated by red and green circles, respectively.
Chapter 4

Concluding Remarks

We have presented a detailed kinetic analysis of the folding pathway for the OmpA transmembrane β-barrel. Through kinetic modeling and global fitting of folding data we determined that the protein transforms through three membrane-associated intermediate states with increasing degrees of membrane insertion, and is prone to populating off-pathway misfolded states. We have proposed the initiation of secondary structure via turn formation and the presence of bilayer defects to drive β-barrel folding and insertion, thus suggesting that a role for the β-barrel assembly machinery in vivo is to create bilayer defects.

Our folding model for OmpA is built upon the previous extensive work that has been conducted on this protein, reviewed in Chapter 1. However, several major aspects of the pathway are novel to our work. It was previously suggested that the unfolded state was a collapsed and partially folded conformation, but we determined that the unfolded form of the transmembrane domain (UAQ) is expanded and has no regular structure (Chapter 2). This conformation is a more logical starting point for folding as it has always been unclear how the proposed compact aqueous state could “uncollapse” and partition onto the membrane in a folding-competent form. We also discovered that in the absence of membranes (and denaturant), the transmembrane domain undergoes extensive self-association to form amyloid-like oligomers. Interestingly, the periplasmic domain present in full-length OmpA reduces self-association of UAQ, therefore demonstrating a novel chaperone activity possibly important for folding in vivo. The presence of denaturant also
serves to inhibit oligomerization of the aqueous state, and by systematically quantifying the extent of self-association as a function of denaturant and protein concentration, we identified conditions under which oligomerization is suppressed. This information was essential for our folding studies because it enabled us to eliminate the competing $U_{AQ}$ self-association reaction, which can strongly influence the observed folding kinetics.

Another major contribution to the $\beta$-barrel folding pathway from this work was the elucidation of secondary structure formation during folding (Chapter 3). It was previously shown that a folding intermediate of OmpA contains $\beta$-sheet structure, but with the identification of multiple intermediate states it became unclear how much secondary structure each conformation contains. The issue was additionally complicated by the proposal that secondary and tertiary structure (i.e., barrel) formation are synchronized, implicitly suggesting that the intermediate states lack secondary structure (Kleinschmidt & Tamm, 2002a). Our studies have resolved the question of when secondary structure forms by directly fitting the observed CD data to kinetic mechanisms and determining the CD attributes of each intermediate state. We have demonstrated that the early membrane-associated intermediates have no regular secondary structure and that the penultimate conformation contains a higher content of $\beta$-structure than the native state. The transformation of this state to the native amount of $\beta$-structure occurs in conjunction with complete barrel formation and membrane insertion, which we demonstrated by an SDS-PAGE gel-shift and the formation of an aromatic exciton interaction.

Lastly, kinetic modeling of the data led to the identification of several off-pathway misfolded states, including the anomalously migrating “gone” state (Chapter 3).
To our knowledge, these states have never been discussed previously in the literature. The presence of parallel pathways has been suggested for several OMPs based on multi-exponential kinetics. However, a dominant pathway with optional misfolding steps, as presented here and previously in the literature for soluble protein folding, provides an alternative model of structure formation more consistent with all experimental evidence.

Historically, OmpA has been the most studied OMP β-barrel due to its high natural abundance in *E. coli* and its lower propensity for aggregation compared to other OMPs. With a few exceptions, the folding kinetics of other OMPs have not been studied in as much detail. Therefore, an important future issue to address will be the applicability of the OmpA folding model to other bacterial OMPs, as well as the eukaryotic OMPs of mitochondria and chloroplasts.

A recent study compared the *in vitro* folding of nine OMPs from *E. coli* and found that the proteins exhibited a range of folding behavior under the same conditions (Burgess et al., 2008). Because these proteins have evolved to reside in the same biological environment, this result indicates that there are sequence-specific differences between OMPs that govern their folding. However, similar trends were observed in the proteins’ folding behaviors under varying conditions, such as slowed folding into thicker bilayers and low-curvature membranes. This suggests that the OMPs follow similar processes for folding. For instance, lag phases appeared in the folding kinetics for many of the OMPs in the thickest bilayers, implicating the same type of multi-step pathway we have developed for OmpA. We observed that population of off-pathway misfolded states strongly influence the apparent folding kinetics of OmpA, so it is likely that similar off-pathway states occur for other OMPs and have a varying effect on the apparent kinetics.
Therefore, we hypothesize that all β-barrel membrane proteins follow the same general folding pathway and that the sequence-dependent propensity to misfold is what causes the differences in observed folding behavior. With the mechanism for OmpA now understood in such detail, similar kinetic modeling of other OMPs would help verify whether the pathway is universal for all β-barrels.

It would also be highly useful to determine if the β-barrels of mitochondrial outer membranes fold by the same mechanism, as these proteins are of critical importance to cellular processes and have been implicated in human pathologies such as diabetes and Parkinson’s disease. One challenge associated with studying mitochondrial OMPs is that these proteins do not exhibit a gel-shift by SDS-PAGE as for bacterial OMPs, possibly due to lower stability. However, other techniques can be used to monitor β-barrel folding, such as protease protection, CD, and fluorescence spectroscopy. In addition, a homologous folding machinery to the BAM complex has been identified in mitochondria, so it would be interesting to determine if the complex performs the same function in facilitating mitochondrial OMP folding as occurs in bacteria.
References


Bustamante, C., Tinoco, I., Jr., & Maestre, M.F. (1983) Circular differential scattering can be an important part of the circular dichroism of macromolecules. *Proc Natl Acad Sci USA, 80*, 3568-72.


281


by periplasmic lipid head groups and BamA. *Proc Natl Acad Sci USA*, **111**, 5878-83.


Appendix: Fitting Equations

This section contains all of the equations used to fit data to kinetic schemes in Chapter 3. The equations were derived based on Euler’s method of numerical integration, which calculates the populations of all species in steps of \( \Delta t \) (see Materials and Methods for Chapter 3).

**Scheme 3.1**

\[
U \xrightleftharpoons[k_1]{k_{-1}} F \xrightarrow{k_2} G
\]

\[
[U]_{t+\Delta t} = (-k_1 \Delta t + 1)[U]_t + k_{-1} \Delta t[F]_t
\]

\[
[F]_{t+\Delta t} = k_1 \Delta t[U]_t + (-k_{-1} + k_2) \Delta t + 1)[F]_t + k_{-2} \Delta t[G]_t
\]

\[
[G]_{t+\Delta t} = k_2 \Delta t[F]_t + (-k_{-2} \Delta t + 1)[G]_t
\]

**Scheme 3.3**

\[
U \xrightarrow{k_1} F \\
\xleftarrow{k_2} \\
F \xrightarrow{k_2} G \\
\]

\[
[U]_{t+\Delta t} = (-k_1 + k_2) \Delta t + 1)[U]_t + k_{-1} \Delta t[F]_t + k_{-2} \Delta t[G]_t
\]

\[
[F]_{t+\Delta t} = k_1 \Delta t[U]_t + (-k_{-1} \Delta t + 1)[F]_t
\]

\[
[G]_{t+\Delta t} = k_2 \Delta t[U]_t + (-k_{-2} \Delta t + 1)[G]_t
\]
Scheme 3.4

\[
\begin{align*}
\text{A} & \xrightleftharpoons[k_1]{k_2} \text{B} \xrightleftharpoons[k_3]{k_4} \text{F} \\
\end{align*}
\]

\[\begin{align*}
[A]_{t+\Delta t} &= -(k_1 + k_3)\Delta t + 1)[A]_t + k_{-1}\Delta t[B]_t + k_{-3}\Delta t[G]_t \\
[B]_{t+\Delta t} &= k_1\Delta t[A]_t + -(k_{-1} + k_2)\Delta t + 1)[B]_t + k_{-2}\Delta t[F]_t \\
[F]_{t+\Delta t} &= k_2\Delta t[B]_t + -(k_{-2}\Delta t + 1)[F]_t \\
[G]_{t+\Delta t} &= k_3\Delta t[A]_t + -(k_{-3}\Delta t + 1)[G]_t
\end{align*}\]

Scheme 3.6

\[
\begin{align*}
\text{A} & \xrightleftharpoons[k_1]{k_3} \text{B} \xrightleftharpoons[k_2]{k_4} \text{F} \\
\end{align*}
\]

\[\begin{align*}
[A]_{t+\Delta t} &= -(k_1 + k_4)\Delta t + 1)[A]_t + k_{-1}\Delta t[B]_t + k_{-4}\Delta t[G]_t \\
[B]_{t+\Delta t} &= k_1\Delta t[A]_t + -(k_{-1} + k_2 + k_3)\Delta t + 1)[B]_t + k_{-2}\Delta t[F]_t + k_{-3}\Delta t[C]_t \\
[C]_{t+\Delta t} &= k_3\Delta t[B]_t + -(k_{-3}\Delta t + 1)[C]_t \\
[F]_{t+\Delta t} &= k_2\Delta t[B]_t + -(k_{-2}\Delta t + 1)[F]_t \\
[G]_{t+\Delta t} &= k_4\Delta t[A]_t + -(k_{-4}\Delta t + 1)[G]_t
\end{align*}\]
Scheme 3.7

\[ A \xleftrightarrow[k_5]{k_1} B \xleftrightarrow[k_2]{k_6} C \xleftrightarrow[k_3]{k_7} F \]

\[
[A]_{t+\Delta t} = (-(k_1 + k_5)\Delta t + 1)[A]_t + k_{-1}\Delta t[B]_t + k_{-5}\Delta t[G]_t \\
[B]_{t+\Delta t} = k_1\Delta t[A]_t + (-(k_{-1} + k_2)\Delta t + 1)[B]_t + k_{-2}\Delta t[C]_t \\
[C]_{t+\Delta t} = k_2\Delta t[B]_t + (-(k_{-2} + k_3 + k_4)\Delta t + 1)[C]_t + k_{-3}\Delta t[F]_t + k_{-4}\Delta t[D]_t \\
[D]_{t+\Delta t} = k_4\Delta t[C]_t + (-(k_{-4}\Delta t + 1)[D]_t \\
[F]_{t+\Delta t} = k_3\Delta t[C]_t + (-(k_{-3}\Delta t + 1)[F]_t \\
[G]_{t+\Delta t} = k_5\Delta t[A]_t + (-(k_{-5}\Delta t + 1)[G]_t \\

Scheme 3.8

\[ A \xleftrightarrow[k_5]{k_1} B \xleftrightarrow[k_2]{k_6} C \xleftrightarrow[k_3]{k_7} F \]

\[ G_1 \xleftrightarrow[k_6]{k_5} G_2 \xleftrightarrow[k_7]{k_4} G_3 \xleftrightarrow[k_8]{k_4} G_4 \]

\[
[A]_{t+\Delta t} = (-(k_1 + k_5)\Delta t + 1)[A]_t + k_{-1}\Delta t[B]_t + k_{-5}\Delta t[G_1]_t \\
[B]_{t+\Delta t} = k_1\Delta t[A]_t + (-(k_{-1} + k_2)\Delta t + 1)[B]_t + k_{-2}\Delta t[C]_t \\
[C]_{t+\Delta t} = k_2\Delta t[B]_t + (-(k_{-2} + k_3 + k_4)\Delta t + 1)[C]_t + k_{-3}\Delta t[F]_t + k_{-4}\Delta t[D]_t \\
[D]_{t+\Delta t} = k_4\Delta t[C]_t + (-(k_{-4}\Delta t + 1)[D]_t \\
[F]_{t+\Delta t} = k_3\Delta t[C]_t + (-(k_{-3}\Delta t + 1)[F]_t \\
[G_1]_{t+\Delta t} = k_5\Delta t[A]_t + (-(k_{-5} + k_6)\Delta t + 1)[G_1]_t + k_{-6}\Delta t[G_2]_t \\

313
\[ [G_2]_{t+\Delta t} = k_6 \Delta t [G_1]_t + (-k_{-6} + k_7) \Delta t + 1)[G_2]_t + k_{-7} \Delta t [G_3]_t \]
\[ [G_3]_{t+\Delta t} = k_7 \Delta t [G_2]_t + (-k_{-7} + k_8) \Delta t + 1)[G_3]_t + k_{-8} \Delta t [G_4]_t \]
\[ [G_4]_{t+\Delta t} = k_8 \Delta t [G_3]_t + (-k_{-8} \Delta t + 1)[G_4]_t \]

Scheme 3.9

\[
\begin{array}{cccccc}
A' & \xleftarrow{k_{-1}'} & A & \xrightarrow{k_1} & B & \xleftarrow{k_2} C & \xrightarrow{k_3} F \\
& \xleftarrow{k_5} & & \xleftarrow{k_4} & & \\
G_1 & \xleftarrow{k_{-6}} & G_2 & \xleftarrow{k_{-7}} & G_3 & \xleftarrow{k_{-8}} G_4
\end{array}
\]

\[ [A']_{t+\Delta t} = (-k_{1}' \Delta t + 1)[A']_t + k_{-1}' \Delta t [A]_t \]
\[ [A]_{t+\Delta t} = k_1 \Delta t [A']_t + (-k_{-1}' + k_1 + k_5) \Delta t + 1)[A]_t + k_{-1} \Delta t [B]_t + k_{-5} \Delta t [G_1]_t \]
\[ [B]_{t+\Delta t} = k_1 \Delta t [A]_t + (-k_{-1} + k_2) \Delta t + 1)[B]_t + k_{-2} \Delta t [C]_t \]
\[ [C]_{t+\Delta t} = k_2 \Delta t [B]_t + (-k_{-2} + k_3 + k_4) \Delta t + 1)[C]_t + k_{-3} \Delta t [F]_t + k_{-4} \Delta t [D]_t \]
\[ [D]_{t+\Delta t} = k_4 \Delta t [C]_t + (-k_{-4} \Delta t + 1)[D]_t \]
\[ [F]_{t+\Delta t} = k_3 \Delta t [C]_t + (-k_{-3} \Delta t + 1)[F]_t \]
\[ [G_1]_{t+\Delta t} = k_5 \Delta t [A]_t + (-k_{-5} + k_6) \Delta t + 1)[G_1]_t + k_{-6} \Delta t [G_2]_t \]
\[ [G_2]_{t+\Delta t} = k_6 \Delta t [G_1]_t + (-k_{-6} + k_7) \Delta t + 1)[G_2]_t + k_{-7} \Delta t [G_3]_t \]
\[ [G_3]_{t+\Delta t} = k_7 \Delta t [G_2]_t + (-k_{-7} + k_8) \Delta t + 1)[G_3]_t + k_{-8} \Delta t [G_4]_t \]
\[ [G_4]_{t+\Delta t} = k_8 \Delta t [G_3]_t + (-k_{-8} \Delta t + 1)[G_4]_t \]
Scheme 3.10

\[
\begin{align*}
[B]_{t+\Delta t} &= (-k_2 + k_5)\Delta t + 1)[B]_t + k_{-2}\Delta t[C]_t + k_{-5}\Delta t[G_2]_t \\
[C]_{t+\Delta t} &= k_2\Delta t[B]_t + (-(k_{-2} + k_3 + k_4)\Delta t + 1)[C]_t + k_{-3}\Delta t[F]_t + k_{-4}\Delta t[D]_t \\
[D]_{t+\Delta t} &= k_4\Delta t[C]_t + (-k_{-4}\Delta t + 1)[D]_t \\
[F]_{t+\Delta t} &= k_3\Delta t[C]_t + (-k_{-3}\Delta t + 1)[F]_t \\
[G_2]_{t+\Delta t} &= k_5\Delta t[B]_t + (-(k_{-5} + k_7)\Delta t + 1)[G_2]_t + k_{-7}\Delta t[G_3]_t \\
[G_3]_{t+\Delta t} &= k_7\Delta t[G_2]_t + (-(k_{-7} + k_8)\Delta t + 1)[G_3]_t + k_{-8}\Delta t[G_4]_t \\
[G_4]_{t+\Delta t} &= k_8\Delta t[G_3]_t + (-k_{-8}\Delta t + 1)[G_4]_t
\end{align*}
\]

Scheme 3.11

\[
\begin{align*}
[B]_{t+\Delta t} &= (-k_2 + k_2')\Delta t + 1)[B]_t + k_{-2}'\Delta t[B']_t + k_{-2}\Delta t[C]_t + k_{-5}\Delta t[G_2]_t \\
[B']_{t+\Delta t} &= k_2'\Delta t[B]_t + (-k_{-2}'\Delta t + 1)[B']_t \\
[C]_{t+\Delta t} &= k_2\Delta t[B]_t + (-(k_{-2} + k_3 + k_4)\Delta t + 1)[C]_t + k_{-3}\Delta t[F]_t + k_{-4}\Delta t[D]_t \\
[D]_{t+\Delta t} &= k_4\Delta t[C]_t + (-k_{-4}\Delta t + 1)[D]_t
\end{align*}
\]
\[
[F]_{t+\Delta t} = k_3 \Delta t [C]_t + (-k_{-3} \Delta t + 1) [F]_t
\]

\[
[G_2]_{t+\Delta t} = k_5 \Delta t [B]_t + (-k_{-5} + k_7) \Delta t + 1) [G_2]_t + k_{-7} \Delta t [G_3]_t
\]

\[
[G_3]_{t+\Delta t} = k_7 \Delta t [G_2]_t + (-k_{-7} + k_5) \Delta t + 1) [G_3]_t + k_{-8} \Delta t [G_4]_t
\]

\[
[G_4]_{t+\Delta t} = k_8 \Delta t [G_3]_t + (-k_{-8} \Delta t + 1) [G_4]_t
\]

**Scheme 3.12**

\[
\begin{array}{c}
\text{D} \\
\downarrow k_4 \\
\text{C'} \xrightarrow{k_3'} \text{C} \\
\downarrow k_3 \\
\text{F} \xleftarrow{k_3} \\
\downarrow k_5 \\
\text{G}_3 \xleftarrow{k_8} \text{G}_4
\end{array}
\]

\[
[C]_{t+\Delta t} = (-k_3' + k_3 + k_4 + k_5) \Delta t + 1) [C]_t + k_{-3}' \Delta t [C']_t + k_{-3} \Delta t [F]_t
\]

\[
+ k_{-4} \Delta t [D]_t + k_{-5} \Delta t [G_3]_t
\]

\[
[C']_{t+\Delta t} = (-k_{-3}' \Delta t + 1) [C']_t + k_{-3}' \Delta t [C]_t
\]

\[
[D]_{t+\Delta t} = k_4 \Delta t [C]_t + (-k_{-4} \Delta t + 1) [D]_t
\]

\[
[F]_{t+\Delta t} = k_3 \Delta t [C]_t + (-k_{-3} \Delta t + 1) [F]_t
\]

\[
[G_3]_{t+\Delta t} = k_5 \Delta t [C]_t + (-k_{-5} + k_8) \Delta t + 1) [G_3]_t + k_{-8} \Delta t [G_4]_t
\]

\[
[G_4]_{t+\Delta t} = k_8 \Delta t [G_3]_t + (-k_{-8} \Delta t + 1) [G_4]_t
\]
| **Number of figures/tables/illustrations** | 1 |
| **Format** | both print and electronic |
| **Are you the author of this Elsevier article?** | No |
| **Will you be translating?** | No |
| **Title of your thesis/dissertation** | Elucidating the folding pathway of outer membrane protein A |
| **Expected completion date** | Jul 2014 |
| **Estimated size (number of pages)** | 200 |
| **Elsevier VAT number** | GB 494 6272 12 |
| **Permissions price** | 0.00 USD |
| **VAT/Local Sales Tax** | 0.00 USD / 0.00 GBP |
| **Total** | 0.00 USD |
Dear Emily,

We hereby grant you permission to reprint the material below at no charge in your thesis subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgment to the source must be made, either as a footnote or in a reference list at the end of your publication.

3. Your thesis may be submitted to your institution in either print or electronic form.

4. Reproduction of this material is confined to the purpose for which permission is hereby given.

5. This permission is granted for non-exclusive world English rights only. For other languages please reapply separately for each one required. Permission excludes use in an electronic form other than submission. Should you have a specific electronic project in mind please reapply for permission.

6. This includes permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Regards,

Lakshmi Priya
Global Rights Department
Elsevier
(A division of Reed Elsevier India Pvt. Ltd.)


Emily J. Danoff
T.C. Jenkins Department of Biophysics
Johns Hopkins University
3400 North Charles Street
Baltimore, MD 21218
Phone: 410-516-5452
E-mail: edanoff@jhu.edu

Personal

Born December 11, 1984 in Washington, DC

Education

Ph.D., Molecular Biophysics, Johns Hopkins University, 2014
  Dissertation: Elucidating the Folding Pathway of Outer Membrane Protein A

B.S., Chemistry, University of Maryland, 2007
  Graduated Summa Cum Laude
  University honors citation
  Departmental high honors (thesis: Studies of the high-efficiency capture and separation of charged organic solutes by catanionic surfactant vesicles)

Research Experience

2008-2014: Dissertation research project in JHU biophysics department lab
  Faculty Advisor: Dr. Karen Fleming
  Elucidating the folding pathway of outer membrane beta-barrel proteins, using in vitro folding conditions, SDS-PAGE, and circular dichroism spectroscopy.

2007-2008: Graduate school lab rotations
  1st Rotation Advisor: Dr. Sarah Woodson
  Studied folding and protein binding of 16S rRNA in the presence of a leader sequence, using PCR, in vitro transcription, and native gel electrophoresis.
  2nd Rotation Advisor: Dr. Karen Fleming
  Studied folding energetics of outer membrane proteins containing ionizable groups, using site-directed mutagenesis, in vitro folding into synthetic vesicles, and SDS-PAGE.
3rd Rotation Advisor: Dr. Jon Lorsch
Studied eukaryotic translation initiation and the effect of initiation factor 2 on complex formation kinetics, using fluorescence emission spectroscopy.

2004-2007: Undergraduate researcher in UMD chemistry department lab
Faculty Advisor: Dr. Douglas English
Studied the ability of catanionic surfactant vesicles to encapsulate charged solutes, using dynamic light scattering, column chromatography, and UV-visible spectroscopy.

Summer 2005: Intern in NCI Lab of Cell Biology at NIH
Principle Investigator: Dr. Michael Maurizi
Studied structure and function of the intracellular protease ClpXP, using an activity assay, SDS-PAGE, and Western blots.

Summer 2004: Intern in NIMH Lab of Neuroendocrine Immunology and Behavior at NIH
Principle Investigator: Dr. Esther Sternberg; Mentor: Dr. Cherie Butts
Studied the effects of hormones on dendritic cells from rat bone marrow. Duties included culturing and dosing cells, making slides, and performing FACS and ELISA analyses.

Research Skills

- Recombinant DNA technology: primer design, colony PCR, Quikchange mutagenesis, plasmid DNA preparation, restriction digestion, gel extraction, transformation.
- Protein purification: expression in E. coli, extraction from cells, purification via ion exchange and size-exclusion chromatography (gravity and high-pressure systems).
- Protein characterization: SDS-PAGE, analytical ultracentrifugation, spectroscopy (UV-visible, circular dichroism, fluorescence emission), differential scanning calorimetry, transmission electron microscopy.
- Lipid handling and formation of unilamellar vesicles by extrusion.
- Software: Microsoft Office, EndNote, ImageJ, IGOR Pro, MacPyMOL.
- Administrative direction of lab: maintenance of equipment, ordering lab supplies, organizing lab space and materials, and maintaining lab files.

Publications and Patents


**Patent Applications**


**Presentations**


**Danoff, E.J. & Fleming, K.G.** Secondary and tertiary structure formation in outer membrane beta-barrel proteins. Poster presentation. 24th Annual Gibbs


Scholarships and Fellowships

National Science Foundation Graduate Research Fellowship, 2009-2012
Alpha Lambda Delta Adele Hagner Stamp Fellowship for graduate study, 2007-2008
Howard Hughes Medical Institute Undergraduate Research Fellowship, 2006-2007
University of Maryland President's Scholarship recipient, 2003-2007
Maryland Distinguished Scholar Finalist and Scholarship recipient, 2003-2007

Awards

College Chemistry Achievement Award from the Chemical Society of Washington, 2006
1st place in Chemical Sciences Group at UMBC Undergraduate Research Symposium, 2006
Hypercube Scholar Award for Excellence in Chemistry, 2006
American Chemical Society Undergraduate Award in Analytical Chemistry, 2006

Professional Associations

Biophysical Society member, 2010-2013
American Chemical Society member, 2003-2008
President of UMD American Chemical Society Student Affiliates chapter, 2006-2007