HIDDEN DYNAMICS IN THE UNFOLDING OF A MEMBRANE

PROTEIN REVEALED BY ULTRAFAST CANTILEVERS

by

Matthew G. W. Siewny

B.A., Kenyon College, 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Physics

2018

ProQuest Number: 10843147

ProQuest 10843147

This thesis entitled:

Hidden Dynamics in the Unfolding of a Membrane Protein Revealed by Ultrafast Cantilevers

written by Matthew G.W. Siewny

has been approved for the Department of Physics

_____

Thomas T. Perkins

_____

Michael Stowell

Date_____

The final copy of this thesis has been examined by the signatories, and we

find that both the content and the form meet acceptable presentation standards

of scholarly work in the above mentioned discipline.

Siewny, Matthew G.W. (Ph.D. Physics)

Hidden Dynamics in the Unfolding of a Membrane Protein Revealed by Ultrafast Cantilevers

Thesis directed by Prof. Thomas T. Perkins

Single molecule force spectroscopy (SMFS) is a technique by which a biomolecule is unfolded by the application of force. Analysis of the forces applied to a protein and the distance the protein consequently extends reveal the configurations the protein adopts during the unfolding process, known as intermediates. By looking at a single molecule at a time, SFMS offers the ability to resolve rarely or transiently occupied intermediates that are obscured by averaging in ensemble techniques. However, detection of such intermediates has been limited by the spatiotemporal resolution of the force probe. Historically, atomic force microscopy-based SMFS has used long cantilevers (>100 µm in length). We utilized modified ultrashort cantilevers (9 µm in length), which were optimized for 1-microsecond temporal resolution and improved spatial resolution, to reexamine the unfolding of the model membrane protein Bacteriorhodopsin within its native lipid bilayer. Numerous new intermediates were detected, with many spaced by as little as two amino acids. The pathways by which the protein unfolded exhibited complex dynamics, including frequent unfolding and refolding, and intermediate occupancies shorter than 10 µs. For a particular fast folding transition, we deduced the folding free-energy landscape. Further, by unfolding the protein from both the C-terminal and N-terminal ends, we obtained complementary sets of intermediates that help identify some of the interactions that fold and stabilize the protein. Lastly, we removed the retinal cofactor and examined the change in the unfolding behavior in an attempt to discern the role of the retinal in the stability of the protein. Surprisingly, for this last experiment, we observed no significant change, although more study is needed to confirm this conclusion. These results sharpen the picture of the mechanical unfolding of membrane proteins and provide details into the interactions that are responsible for the folded structure of membrane proteins. These ultrashort cantilevers can be applied to SMFS studies on other biomolecules, to reveal dynamics previously obscured by a lack of spatiotemporal resolution of the force probe.

Dedication

To my fiancée Kelsey, whose encouragement and support were vital to this endeavor.

To my parents, for inspiring my curiosity and nurturing the instinct to ask questions.

Acknowledgements

The person I want to thank first is my advisor and committee chair, Tom Perkins, for both overseeing my PhD work and for mentoring me through matters both inside and outside of the lab. Working for him taught me how an experimentalist thinks, and his foresight steered the research from a small outlier into a high impact publication. His indulgence of tangents in my work lead to many dead ends, also the development of many skills that I would otherwise never would have exposed. Under Tom's tutelage, I went from a mere learning enthusiast to a focused scientist.

Many other members of the lab contributed to the project. Most significant is Hao Yu, who responsible for the equilibrium energy landscape reconstruction analysis, as well as an equal partner in the collection of all data that went into the Science paper, as well as assisting on intermediate determination. He taught me how to craft a message, and conversation with him have probably played the largest role in how I think about protein and their dynamics. Rob Wader was developed and adapted the protocols for the specific functionalization used in later experiments. His good-natured role as a skeptic of my ideas and conclusions sharpened them into their final forms. Devin Edwards played a crucial role in developing and providing FIB-modified cantilevers, as well as improvements to the AFM (especially the custom small spot size). Jaevyn Faulk also produced FIB-modified cantilevers for a significant length of time, especially during the C-terminal experiments. Recently, David Jacobson has become the main producer of FIB cantilevers, including most of them used in the N-terminal experiments, as well as my intellectual partner in continuing bR studies. Lyle Uyetake was the responsible for the growth, as well as functionalization of all of the bacteriorhodopsin in this study. Violent Roskens also performed a similar role in many of the initial studies on bR that preceded this work. I also wish to thank the junior members of the lab (Ayush Adhikari, William Van Patten, Guy Margalit and Ty Miller), who helped functionalize, and perform vital chemistry for the production of the substrates and cantilevers.

While not contributing directly to the work published in this thesis, Allison Churnside also played a significant role in helping me learn how a scientist conducts an experiment, how atomic force

beyond support received from Asylum Research enabled me to run the instruments exactly as desired. I also wish to thank Aric Sanders, who helped optimize the FIB modification process for high-spatiotemporal resolution.

I cannot properly thank my fiancée, Kelsey Coolidge her support both emotionally and domestically, to enabled me to focus on this work. I also owe mom and dad, Gregory and Jeanne Siewny, a large debt of gratitude for helping me keep things in perspective over the course of my graduate work. I also cannot thank all my friends and family enough for providing me with encouragement and wisdom relating during both this dissertation, as well as graduate school.

Contents

List of Tables

List of Figures

List of Abbreviations & Acronyms

AA/aa: Amino Acid

AFM: Atomic Force Microscopy

APDMES: Aminopropyl-dimethoxysilane

bO: Bacterioopsin

bR: Bacteriorhodopsin

DFS: Dynamic Force Spectroscopy

DI: Deionized (Water)

DBCO: Dibenzocyclooctyl

FEC: Force Extension Curve

FIB: Focused Ion Beam

FRINL: Folded Remains in Native Location

GPCR: G-protein coupled receptor

invOLS: Inverse Optical Lever Sensitivity

IPA: Isopropyl Alcohol

μm: Micrometer

MD: Molecular Dynamics (simulation)

nm: Nanometer

nN: Nanonewton

Occu. : Occupancy

PBS: Phosphate buffer solution

PEG: Polyethylene Glycol

PM: Purple Membrane

pN: Piconewton

S.E.M.: Standard Error of the Mean

SEM: Scanning Electron Microscope

SD/S.D.: Standard Deviation

SMFS: Single-Molecule Force Spectroscopy

SPA: Silane-PEG-Azide

CHAPTER I: Introduction and Literature Review

1.0) Motivation

Membrane proteins are an important class of proteins as they perform critical life-sustaining functions, including nutrient transport, signal transduction, and cellular adhesion. While they comprise 20-30% of all proteins encoded within a typical genome (*1*), many of their functional mechanisms, including their folding behaviors, remain poorly characterized compared to globular proteins. Understanding protein folding, the process by which unstructured polypeptide chain adopts its functional confirmation is immensely important and is one of the most active areas of current research: 60% of drugs currently in development target membrane proteins (*2*). Significant advances in the understanding of membrane protein folding hold the potential to elucidate the causes and cures of protein misfolding diseases, such as cystic fibrosis (*3*), and revolutionize targeted drug delivery. Despite this significant potential, the membrane protein folding problem still remains a fundamental research question. Part of this delay is caused by difficulties in experimental determination of membrane protein structure (which still represent less than 2% of solved structures within the protein data bank) (*4*). Membrane proteins are in general difficult to study due to the fragility of the native lipid membranes in which they are embedded. Many techniques that aid in solvating these proteins (required for techniques such as X-ray crystallography and NMR) disrupt this lipid environment, and can place the proteins in non-native conditions, and accordingly, non-native conformations (*5*). Furthermore, even if the protein can reconstituted into its native state, many chemical and thermal denaturants that are used study the folding of globular and fibrous proteins can unfold membrane proteins incompletely, with significant secondary structure still intact in the unfolded state, making interpretation of the results difficult (*6*). Thus, novel experimental assays are needed for studying of membrane protein folding.

This is not to say that no progress has been made on the membrane protein folding problem: General properties of the membrane protein folding process, such as the two-stage model, have been elucidated (*7*), and advances have been made in computer aided prediction (*8*), yet erroneous predictions

of protein structure are still common (*9*) especially for 'free modeling' (i.e. structures resulting from amino acid sequences dissimilar to those with known structure (*10*)). To improve the accuracy of protein folding prediction depends, in part, on the ability to predict folding intermediates and pathways from an amino acid sequence. These intermediates may include configurations that are short-lived, and therefore difficult to detect, that can have significant impact on the folding process. Thus, observing all configurations the protein adopts during the folding process is necessary to construct predictive models of protein folding. A useful concept for discussing protein folding is known as the energy landscape. The energy landscape models the folding process as diffusion from high energy unfolded configurations to low energy fully folded configuration, regulated by diffusion between partially folded configurations (known as folding intermediates). Theoretically, if an energy landscape is well known, then the folding process can be accurately modeled by computer simulations (*11*). Measuring the energy landscape in high detail aids in predicting protein structures.

One such set of experimental techniques that has proven quite adept in the study of membrane protein folding (and specifically for measuring energy landscapes) is force spectroscopy. Force spectroscopy includes any method that measures the response of molecules or molecular bonds after applying a known force. It was first achieved in 1988 by measuring the tensile strength of actin filaments using microneedles (*12*). The focus of this thesis, single-molecule force spectroscopy (SMFS), involves physically attaching a single biomolecule to a force probe, then exerting force on that molecule until a conformational change occurs: e.g. a bond ruptures (*13*, *14*), a nucleic acid unwinds (*15*), or a protein unfolds. The first SMFS protein unfolding experiment was first published in 1997 on the globular protein (*16*), followed by the first forced unfolding of a membrane protein in 2000 (*17*). Both of these papers utilized atomic force microscopy (AFM). These two studies, amongst others, spurred a lot of excitement for the future of AFM SMFS specifically, and force spectroscopy as a whole (*18*). Since then, SMFS studies have been performed on additional membrane proteins, including:

- Other Rhodopsins: bovine rhodopsin & murine rhodopsin (*19–21*), Halorhodopsin (*22*), Proteorhodopsin (*23*)

- Alpha helical proteins: NhaA (*24*), SteT (*25*), Aac3p (*26*), human aquaporin-1 (*27*), KpOmpA (*28*), BetP (*29*), MjNhaP1 (*30*), LacY (*31*, *32*), GlpG (*33*), GPCR β$_2$ Adrenergic Receptor (*34*)

- Beta barrel proteins: OmpG (*35*), FhuA (*36*), LamB (*37*)

- Other miscellaneous proteins: Light Harvesting Complex 2 (*38*, *39*), Notch proteins (*40*), as well as a few non-native short peptides (*41*, *42*)

A significant majority of these studies have been conducted either by, or in collaboration with the lab of Daniel Müller, and their influence in the field of membrane protein SMFS is immense. As a result of this influence, there is a significant amount of methodological overlap. For example, all (sans one) membrane protein SMFS studies were performed on an AFM, utilizing the same set of cantilevers, despite the availability of other force inducing instruments, and other cantilevers.

AFM is a conceptually simple instrument: a microscopic small cantilever with a measurable spring constant mounted on the end of a precision piezoelectric motor. By measuring the deflection of the cantilever (Δz in Fig 1-1) yields knowledge of the forces applied to the cantilever (*43*). Additionally, the cantilevers are capped with a nanoscopically sharp tip that can be attached to a single molecule at a time. Once a molecule is attached to the cantilever, it can be stretched from the surface by retracting the base of the cantilever with a piezo-electric motor. This motion applies an increasing force to the molecule until it partially unfolds.

The AFM is the most commonly used instrumentation for membrane proteins SMFS for a few key reasons. First, it has a geometry that is inherently surface coupled, making it well suited to study 2D interfaces such as a lipid bilayer. More stable membranes can even be deposited intact onto the substrate, allowing the embedded proteins to be studied in their native lipid environment. Another advantage of AFM for protein unfolding experiments is the ability to apply and measure forces from single piconewtons (pN) to multiple nanonewtons (nN). Due to the hydrophobic coupling to the lipid membrane, many membrane proteins need > 100 pN applied to induce unfolding, which is not easily attainable with other common SMFS techniques such as optical tweezers or magnetic tweezers (with maintenance of temporal resolution) (*44*).
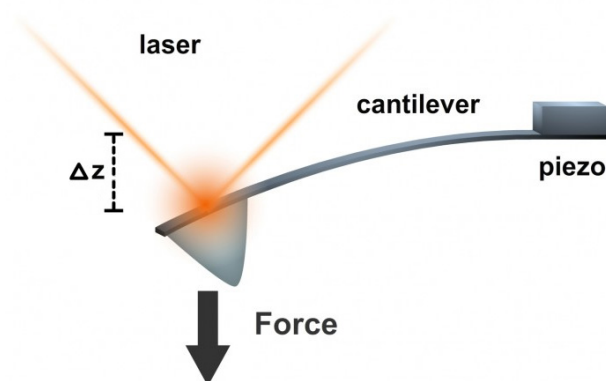
Figure 1-1: Diagram of AFM operating principles. A cantilever (under force) is depicted, with the deflection (Δz) being measured by the deflection laser (red) (Credit: Greg Kuebler).

One persistent criticism that vexes the SMFS literature is the biological relevance of the results that emerge from the studies. While the forced unfolding does reveal subdomain architecture of a protein (the local bonds and interactions that maintain the protein in its native form), they are an imperfect model to study refolding as the forced unfolding process does not model the conditions of folding in reverse. When bacteriorhodopsin and most alpha-helical transmembrane proteins are folded, the insertion and orientation of the polypeptide is usually performed by the ribosome directly into the translocon, a membrane protein complex (Fig 1-2). In general, force spectroscopy studies have produced results difficult to corroborate with other chemical/thermal based assays (*45*). One goal of this work was to discern interhelical bonds previously observed in a mutation/SDS denaturation study (*46*) and crystallographic experiments (*47*).

First, a general introduction of AFM force spectroscopy, and with improvements in the resolution of SMFS AFM studies achieved in the Perkins lab, is discussed. Next, I will have a brief overview of bacteriorhodopsin (bR), the protein that is the focus of this thesis, both in its properties as well as its history within SMFS literature. Lastly, I will include a general outline of the remainder of this thesis.
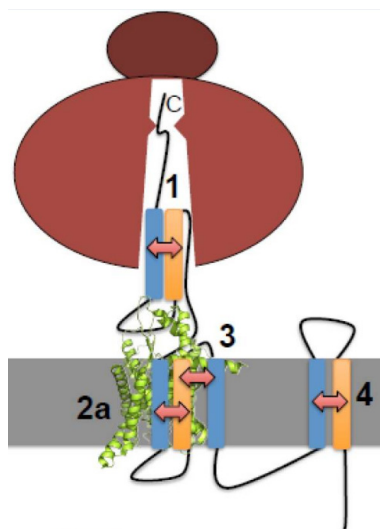
Figure 1-2: Native integral membrane protein folding. A depiction of folding and insertion of a membrane protein. The ribosome (maroon), translocon channel (green), and helical segments being inserted (blue and orange) are shown. The general consensus is that protein forms into its helical structure, and is inserted into the membrane via the translocon cannel, and partitions into the membrane from the translocon. Reprinted from J. Mol. Biol. 427, Cymer, F. *et al*. Mechanisms of Integral Membrane Protein Insertion and Folding. 999-1022. Copyright 2015, with permission from Elsevier. (*48*).

1.1) AFM Force spectroscopy overview

Force spectroscopy is a single-molecule method, which are more sensitive to transient or sparsely populated events than ensemble methods. The utilization of force to unfold the protein, as opposed to chemical or heat denaturants, also enables the localization of unfolding to specific segments of the protein based on the change in contour length upon unfolding. Unlike globular proteins, membrane proteins are stabilized by the coupling of the helices to the hydrophobic membrane. This makes it likely that the protein will unfold before being extracted from the membrane, which imposes a sequentially in the unfolding of the protein (*49*).

SMFS experiments monitor the force applied to a cantilever as well as the distance that the base of the cantilever has been moved. This latter quantity can be combined with the measured deflection of the cantilever to derive a quantity known as the tip-sample separation (or extension) that forms the (quasi) independent variable of SMFS experiments, with the force forming the dependent variable. The ability to detect short-lived and closely spaced intermediates in an unfolding record depends on the performance of the cantilever. Achieving high spatial precision requires low force/deflection noise which is achieved by

minimizing the inherent fluctuations of the cantilevers in liquid due to underdamping (quality factor $Q >$ 0.5). Achieving high responsiveness, however, is multifaceted and requires some exploration of its limiting factors.

The ability of a cantilever to respond quickly to a change in force applied to it depends on the mechanical properties of the cantilever and the damping forces that oppose motion. Mechanically, a cantilever can be modeled as a damped harmonic oscillator. Accordingly, the resonance frequency can be understood to represent the responsiveness of the cantilever and can be increased by stiffening the cantilever. This, unfortunately, increases force noise which is not generally an acceptable tradeoff, as the new short-lived states will likely be buried in the increased force noise. The more productive strategy then is to reduce the damping forces that oppose cantilever motion. These forces arise mainly from the hydrodynamic drag of the liquid on the cantilever. Hydrodynamic drag is dominated by largest dimension of the cantilever, which is the length. Therefore, using a shorter cantilever is the principal way to reduce hydrodynamic drag (*50*).

In the past, with the focus placed on sufficiently soft cantilever, all AFM SMFS studies on membrane proteins outside the Perkin's lab used only 4 different cantilevers[1], the most recent of which is located in Fig 1-3A). These cantilevers share features with all biological SMFS-focused cantilevers in their relatively low stiffness (<100 pN/nm), which aids in the detection of small forces. Historically, cantilevers at this stiffness were only available at lengths greater than 60 μm, and as a result these cantilevers have temporal resolutions on the order of hundreds of microseconds, at best. These cantilevers act as low-pass filters to folding and unfolding steps of proteins that are hypothesized to occur on the 0.1 μs to 1 μs timescale (*51*). Thus, many details of molecular unfolding trajectories are obscured by this temporal filtering. Furthermore, these cantilevers were not optimized to have low force noise, such that even if short lived states were resolvable, they would still be difficult to distinguish with an individual unfolding record. This fact led the lab to begin to explore using smaller cantilevers. First with the

---

[1] The four cantilevers being the Olympus OMCL-TR400PSA, the Olympus BL-RC150VB (Biolever), the Bruker DNP-S (NPS) and the Olympus OMCL-RC800PSA

Olympus Biolever Mini (l=38 μm), and then the Olympus Biolever Fast (l=9 μm, See Fig 1-3B). This last

cantilever is over an order of magnitude shorter than many of the listed cantilevers, and accordingly has

single microsecond resolution. Predictably, one problem with these shorter cantilevers was their high

stiffness, which made them challenging to apply due to their increased force noise. In order to reduce

their stiffness, a focused ion beam was used to mill and thin the cantilevers and render then 4-10x softer

[work originally done by Matthew Bull (*52*) and improved upon by Devin Edwards (*53*)]. In that

modification process, a large amount of material was removed from the base of the cantilever and the

remaining legs were thinned. The net result was a cantilever softer than the previous literature (~ 20-40

pN/nm), and with marginally reduced hydrodynamic drag, thus maintaining its microsecond temporal

resolution. Additionally, the cantilevers have an order of magnitude improvement in force precision over

previous cantilevers. Thus, the Perkins Lab has produced a superior cantilever for all purposes that do not

require long term stability, or need to be exceedingly (<10 pN/nm) soft.
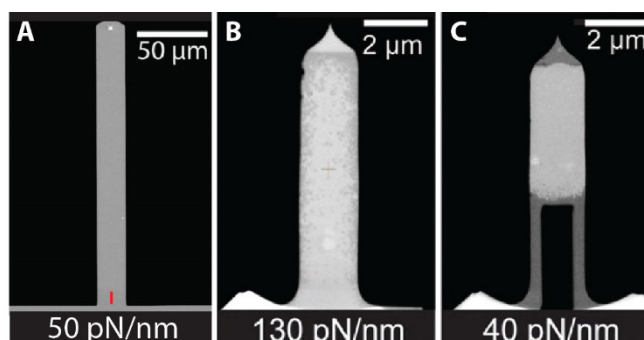
Figure 1-3: SEM images of 3 different cantilevers used for bR SMFS. (A) Olympus RC-800PSA, the
most recent cantilever used by the Müller lab in a bR SMFS study. Nominally l=200 μm, w=20 μm. A
rough estimate of the size of the cantilever depicted in the next panel (AC10DS), on the panel A scale is
shown (red bar) (photo credit: David Jacobson) (B) Olympus AC10DS Biolever Fast (credit: Devin
Edwards), designed for imaging, used in initial studies of bR unfolding in Perkins lab. Nominally l=9 μm
w=2 μm (C) Focused Ion Beam (FIB) modified AC10DS. Same dimensions as (B). This style cantilever,
with superior spatiotemporal resolution, is what will be used for the studies within this thesis (credit:
Devin Edwards) Average stiffness listed on bottom of each panel. (B) and (C) Reprinted from Nano Lett.
15, Edwards, D.T. *et al*. Optimizing 1-μs-Resolution Single-Molecule Force Spectroscopy on a
Commercial Atomic Force Microscope. 7091-7098. Copyright 2015, with permission from Elsevier. (*53*).

1.2) Bacteriorhodopsin

1.2.1) General overview

Bacteriorhodopsin is a rhodopsin, a group of α-helical transmembrane proteins with a covalently bound retinal molecule. They act as photoactive proton pumps, which use light to power the pumping of protons from within the cell, and then (elsewhere in the cell) using the resulting chemical gradient to create ATP. Bacteriorhodopsin is comprised of 7 transmembrane alpha helices (Fig 1-4), labeled by the letters A-G with A being closest to the N-terminus and G-being closest to the C-terminus.
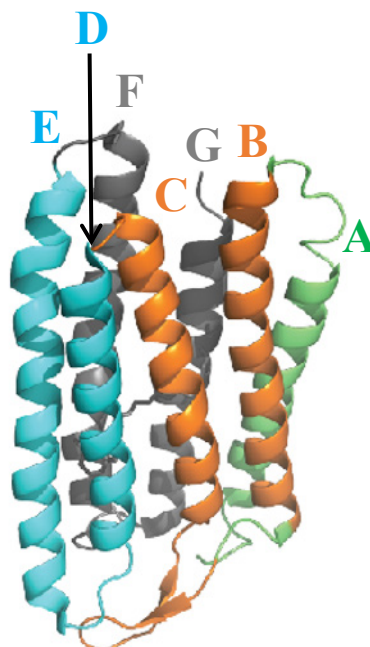
Figure 1-4: 3D structural model of bR. The 7 helices are colored by C-terminal forced unfolding pairs, and labeled. Model constructed from PDB 1FBB (*54*) within PyMOL.

Quite similar in structure to G-protein coupled receptor (GPCR), bR has been the model membrane protein for a variety of breakthrough studies and assay demonstrations (*55–58*). Purified from Halobacteria Salinarum, an archaebacteria found in environments with extremely high salt concentrations [4.3 M NaCl, optimally (*59*)], its status as a model membrane protein is due to a few useful features. First, the protein is naturally expressed in high levels, and is packed extremely densely in H. Salinarum membrane, occupying up to 50% of the membrane surface (known specifically as Purple Membrane, or PM). Secondly, the proteins naturally form into a trigonal lattice, allowing it to be examined by crystallographic methods (Fig 1-5). Thirdly, the protein is highly stable, able to remain in their native
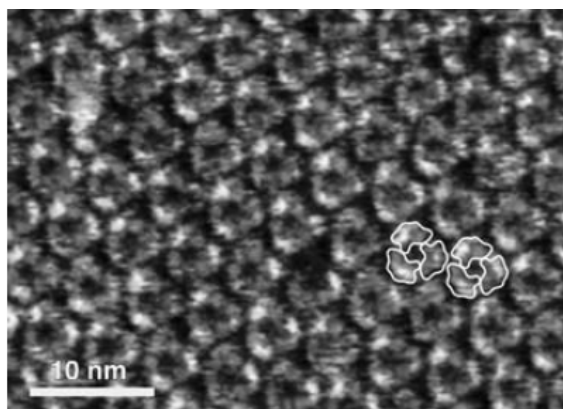
Figure 1-5: Purple membrane topograph. AFM topograph of the cytoplasmic side of bR within purple membrane. Each trigonal shape is a bR trimer, with 6 individual monomers outlined in white. Reprinted from J. Mol. Biol. 355, Sapra, K.T. *et al*. Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy. 640-650. Copyright 2006, with permission from Elsevier. (*64*).

trimeric structure until 80° C and not fully unfold until 100° C (*60*). It is also (relatively) chemically

stable, resisting unfolding in urea concentrations up to 5 M (*61*), and in guanidinium chloride up to 8 M

(*62*). Lastly, it is fluorescent and its spectroscopic properties are sensitive to the physical state of the

protein, which can serve as a proxy for examining the functionality of a membrane protein before the

development of non-spectroscopic methods (*60, 63*).

1.2.2) A selection of non-SMFS unfolding studies of Bacteriorhodopsin in literature

In general, membrane protein folding has proven difficult to explore via traditional assays used

on globular proteins. Whereas globular protein folding are commonly explored by heat or chemical

denaturants, both of these methods have proven problematic for bR. bR, as previously stated, is quite

stable thermally, maintaining some tertiary structure up to 100° C (thus is not a useful means to induce

unfolding). Chemically, bR has proven easier to unfold, especially by SDS denaturation, but the

unfolding state tends to be ill defined. A recent study showed that the SDS denatured state poorly

resembles the native prefolded state of the protein, with much of the secondary structure partially intact

(*62*), making analysis difficult. Secondly, SDS denaturation has shown to progress along different

unfolding and refolding pathways, raising questions about the relevance of the observed folding, and

unfolding behavior (*65*). That is not to say that SDS denaturation experiments of bR do not produce

important information about protein folding trends. For example, bR was instrumental in the studies that established the two-stage folding model for membrane proteins (*7*). It does suggest that non-chemical denaturation would be helpful to study bR in higher detail.

1.2.3) Comprehensive overview of Müller bR SMFS studies

The Daniel Müller lab, as previously discussed has extensively studied the forced unfolding of bR. Below is a brief overview of each important paper in this literature, including pertinent conditions and experimental parameters to which bR was subjected. They are ordered chronologically. Note that all of the following studies are performed exclusively by pulling on the C-terminus/cytoplasmic side of the protein:

A. Müller *et al.* Surface structures of native bacteriorhodopsin depend on the molecular packing arrangement in the membrane. J. Mol. Biol. 1999 (*66*)

A purely topographic study using contact mode topography to image both the extracellular and cytoplasmic sides of the bR trimer. By imaging at forces of both 100 pN and 200 pN, and comparing the relative deformations, the authors were able to make conclusions about the rigidity of the interhelical loops in the proteins.

B. Oesterhelt *et al.* Unfolding Pathways of Individual Bacteriorhodopsins. Science 2000 (*17*)

The first SMFS study of a membrane protein, this paper also marked the first observation of the pairwise helical unfolding behavior seen in alpha helical integral membrane proteins. Also, they cleaved the interhelical loop connecting the F & E helices, and measured enhanced unfolding force of the remaining unfolding peaks. This indicates that the presence of the G and F helices stabilize the remaining helices. The unfolding length differences remained the same, which indicates that the stable folding segments are dictated by intrahelical bonds.

C. Müller *et al.* Stability of Bacteriorhodopsin α-Helices and Loops Analyzed by Single-Molecule Force Spectroscopy. Biophys. J. 2002 (*67*)

In this paper, the authors first observed of non-obligate unfolding intermediates (also known as side peaks) that sometimes occur between the previously seen obligate intermediates (referred to

here as major states). In addition, the authors established the invariability of the location of these side peaks by observing no significant change at different buffer pH, in a single point mutation that locks bR into one of its functional configurations, with the retinal removed, or by changing the substrate. For retinal removed bR (known as bacterioopsin, or bO), there was an increase in the non-obligate intermediate occurrence percentages (indicating that the G & F helices are likely stabilized by contacts with the retinal).

D. Janovjak *et al.* Unfolding pathways of native bacteriorhodopsin depend on temperature. Embo J. 2003 (*68*)

   The authors performed forced unfolding at temperatures varying from 8° C to 52° C. Once again, the location of the obligate and non-obligate intermediates did not change, but at higher temperatures the rupture force for most intermediates decreased. Furthermore, the percentage of non-obligate intermediate occurrence went down with increasing temperature . With prior studies showing that hydrophobic effects grow stronger at higher temperatures (*69*), this reinforced the idea that unfolding intermediates are largely caused by intrahelical and interhelical bonds, as opposed to hydrophobic interactions.

E. Janovjak *et al.* Probing the Energy Landscape of the Membrane Protein Bacteriorhodopsin. Structure 2004 (*70*)

   Here, the authors performed a dynamic force spectroscopy (DFS), which is a SMFS assay where they vary the pulling speed to determine how the rupture force changes at different pulling speeds. This can be used to measure the energy landscape of the molecule. The location of the intermediates did not change depending on pulling velocity. In general, the occurrence of non-obligate intermediates did increase at higher pulling speeds (an unexpected result). The resulting measurement of $x^{\ddagger}$ (the width of the barrier), being 3.2 Å – 8.6 Å, does reinforce the idea that unfolding intermediates are stabilized largely by interhelical bonds, as opposed to hydrophobic interactions (which should have $x^{\ddagger} \sim 30$ Å).

F. Kessler *et al*. Bacteriorhodopsin Folds into the Membrane against an External Force. J. Mol. Biol. 2006 (*71*)

By lowering the force after partially unfolding bR, the authors were successfully able to observe refolding, albeit rarely. This was an unexpected result, as it wasn't clear if bR could refold without the presence of the translocon, and at forces as high as 12 pN. Interestingly, non-native folding (possibly misfolding) was observed at a higher frequency than native refolding, as evidenced the appearance of intermediates not observed in normal unfolding experiments.

G. Sapra *et al.* Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy. J. Mol. Biol. 2006 (*64*)

The oligomeric state of bR, which natively occurs as a trimer, was varied as bR was unfolded in a trimer, dimer and monomer configurations. Once again, the intermediates did not change location, but the average rupture forces did decrease with decreasing oligomerization state. The occurrence rate of non-obligate intermediates did increase with decreasing oligomerization state. These facts once again suggest that while intramolecular bonds create the stable folding units for bR, intermolecular interactions (be it bonds to other monomer units or hydrophobic interactions) help dictate the intermediate stability.

H. Sapra *et al.* Point Mutations in Membrane Proteins Reshape Energy Landscape and Populate Different Unfolding Pathways. J. Mol. Biol. 2008 (*72*)

5 bR mutants were made by mutating individual residues to alanine (P50A, M56A, Y57A, P91A, P186A), and DFS was conducted on each mutant. Once again, no change in the unfolding intermediates locations was observed. In terms of forces, the ED helices saw no change in force, while intermediates within CB and A helices saw small increases in force. In general, the occurrence rate of the non-obligate unfolding intermediates became smaller with mutations. This indicates that while individual residues influence state stability, their locations are likely caused by collective interactions.

I. Sapra *et al.* Role of extracellular glutamic acids in the stability and energy landscape of bacteriorhodopsin. Biophys. J. 2008 (*73*)

       Another mutation study, this time unfolding a triple mutant (E9Q, E194Q, E204Q), each residue being involved in the proton release pathway, and are known to affect photocycle kinetics. This tested the hypothesis that changes in the function are coupled with changes in the mechanical properties of the protein. Once again, no change was seen in the location of the intermediates, but a new unfolding pathway was observed. In a reversal of the conclusion in the last paper (*72*), these mutations led to increases in the occurrence rate of non-obligate intermediates.

J. Zocher *et al.* Single-Molecule Force Spectroscopy from Nanodiscs: An Assay to Quantify Folding, Stability, and Interactions of Native Membrane Proteins. ACS Nano 2012 (*74*)

       Here, the lab reconstituted bR within lipid nanodiscs to determine if the native folding was present in the mimic lipid environment (which allows one to treat bR like a globular protein in a variety of assays). Other than the significant decrease in attachment rates, no significant changes in state location, force and occurrence occurred with respect to native purple membrane, thus demonstrating it as a successful reconstitution, and that the intermediate forces do not depend on the specific lipid composition of the membrane.

K. Petrosyan *et al.* Single-Molecule Force Spectroscopy of Membrane Proteins from Membranes Freely Spanning Across Nanoscopic Pores. Nano. Lett. 2015 (*75*)

       A simple force spectroscopy experiment on purple membrane patches suspended over nanoscopic pores to determine if physical support from the substrate altered the unfolding behavior. While a bit of new analysis had to be implemented to account for the elasticity of the unsupported membrane, no significant changes to intermediate location rupture forces or pathways were detected, indicating that the substrate support did not influence the established unfolding behavior beyond altering the application of force.

1.2.4) Non-Müller lab SMFS studies of bR

In addition to the aforementioned studies, a few important SMFS studies of bR have been conducted by other labs, which both pulled on both independently pulled on the C-terminal (cytoplasmic) and N-terminal (extracellular) ends of the protein.

L. Kessler & Gaub. Unfolding Barriers in Bacteriorhodopsin Probed from the Cytoplasmic and the Extracellular Side by AFM. Structure 2006 (*49*)

In 2 sets of experiments, the authors pulled from C-terminal, and N-terminal tails of the protein (the first time N-terminal forced unfolding was achieved on bR), and compared the location of the intermediates observed. In general, the location of the C-terminal unfolding intermediates of this paper were quite similar to those observed in the Müller literature (see Fig 1-6, labeled Kessler *et al*. vs. Sapra *et al*.), with the addition of intermediates identified in the ED loop, the CB loop, and the A helix. It is unclear if this change in the number of intermediates is from the differences in pulling speed (1.4 µm/s vs. 40 nm/s in (*67*)). Also, the authors hypothesize for the first time that the changing residue height has an influence on the measured contour length. Due to this theory, they assign their intermediates to different residues than in observed for the Müller literature. Also in the paper, they hypothesized that an intermediate observed from their C-terminal pulling experiments, and an intermediate observed from their N-terminal pulling experiments occurred close (though not within uncertainty), to a predicted interhelical hydrogen bond, and theorized that this is the cause of those forced unfolding intermediates. One weakness of this study was the poor statistics collected for N-terminal unfolding experiments (only 12 molecules), as achieving attachment to the shorter N-terminus proved to be an order of magnitude more difficult than for the C-terminus.
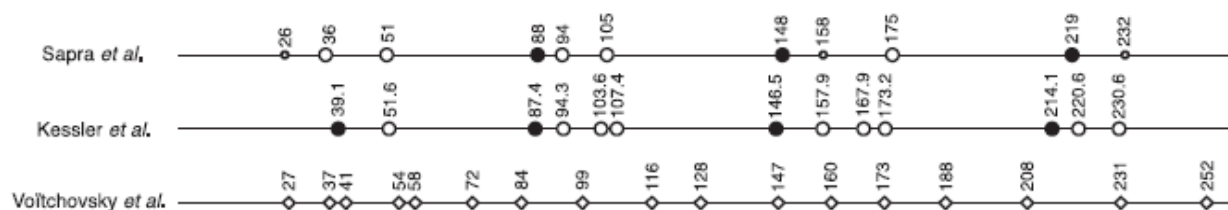
Figure 1-6: Comparison of C-terminal intermediate locations detected from prior experiments. This figure lists the locations of the c-terminal (cytoplasmic) forced unfolding intermediates observed for the Müller literature (here, labeled Sapra *et al.*), the Kessler-Gaub (L) observed intermediates, and the Voïtchiovsky (M) intermediates. The large, filled in circles represent intermediates that are observed in more than 90% of records, the large open circles represent intermediates that are observed between 30% and 90% of records, and the small open circles represents states that are observed in less than 30% of records. Reprinted from Biophys J. 100, Yamada, T. *et al*. Forced Unfolding Mechanism of Bacteriorhodopsin as Revealed by Coarse-Grained Molecular Dynamics. 2086-2098. Copyright 2016, with permission from Elsevier. (*76*).

M. Voïtchiovsky *et al.* Electrostatic and Steric Interactions Determine Bacteriorhodopsin Single-

Molecule Biomechanics. Biophys. J. 2007 (*77*)

Using significantly lower salt concentrations (20, 30 and 40 mM KCl vs. 150 mM KCl in

previous papers), in 2 sets of experiments the authors pulled from both termini of bR, looking to

explore the effects of electrostatic interactions on the stability of bR. For their cytoplasmic results,

they saw significant differences in the location of the intermediates (see Fig 1-4, labeled Sapra *et al.*

vs. Voïtchiovsky *et al.*). Some of this change is due to a difference in their amino acid assignment

strategy compared to the Müller literature. The number of intermediates observed increased

dramatically as the buffer concentration was increased. The rupture forces of each stayed about the

same on average with changing buffer concentrations, but interestingly, the variability of the rupture

forces from molecule to molecule significantly increased as salt concentration increased. The

strongest conclusion of this study was that all intermediates they detected correspond to steric

interactions of Trp residues with other charged residues (or prolines). This conclusion was based on

the persistence and strength of intermediates assigned to Trp residues at 20 mM KCl. Considering

that the salt concentration is significantly lower than the rest of the literature, direct comparison to the

rest of the literature is hard, other than concluding that electrostatic interactions play a significant role in the forced unfolding behavior of bR.

The most striking result all of these previously studies (with the exception of Voïtchiovsky, performed at significantly lower ionic concentrations) is that they observed approximately the same number and locations of native unfolding intermediates for the C-terminal unfolding experiments, despite the significant differences in experimental and environmental conditions (pH, pulling velocity, presence of retinal, temperature, oligomerization, lipid composition), as well as a few sets of significant point mutations.

1.2.5) Computational simulations of bR forced unfolding

A few computational studies of forced unfolding of bR have been conducted, with results ranging from rough agreement with experimental results, to predicting a significantly denser array of intermediates. Here is an overview of 3 significant papers:

N. Cieplak *et al.* Pulling single bacteriorhodopsin out of a membrane: Comparison of simulation and experiment. Biochim. Biophys. Acta.- Biomembranes. 2006 (*78*)

   The authors look at the unfolding behavior for bR when pulled at from both the cytoplasmic (C-terminal) and extracellular (N-terminal) side using a part all-atom, part coarse-grained simulation. They report that they observe the predicted negative relationship between temperatures and rupture forces. They also claim that they see the same pairwise unfolding behavior on the cytoplasmic side, with qualitatively similar unfolding forces to the Müller intermediates (given, the effective pulling speed is multiple orders-of-magnitude greater than for any experiment listed so far). Specific locations for those intermediates were not listed. Interestingly, they report only 3 unfolding peaks for the extracellular unfolding.

O. Kappel and Grubmüller. Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network. Biophys. J. 2011 (*79*)

An all-atom simulation of bR, again pulling from both sides of the protein. The pulling velocity (1 m/s) was orders-of-magnitude faster than either of the other two simulations, or any experiments, due to the immense computational overhead for an all-atom simulation of 4 bR trimers, their associated lipids, and waters. Additionally, they did not simulate a cantilever, so force peaks were not calculated. To determine intermediates, they looked at displacement from the native location, from which they determined the transition point between natively folded protein and the linearly displaced unfolded polypeptide. While this is more physically relevant to the underlying bonds that stabilize the protein, it makes comparisons to experiments difficult. Once again, they recapitulated the pairwise unfolding, marked by large stable folding intermediates at the top of the helix pairs. One significant difference was in the amount of intermediates observed: the simulation had a denser array unfolding intermediates, significantly higher than the amount seen in experimental studies. They roughly observed intermediates in proximity of the Müller intermediates locations. It is unclear if the difference in the number of intermediates is due to the differences in pulling speed (1 m/s at the slowest). Two concerns were raised about the interpretation of the experimental intermediates: 1) a breakdown in the underlying experimental assumption that folded protein remains in its native location until after the secondary structure is unwound, and 2) that non-native bonds and interactions transiently form during the unfolding process, indicating that unfolding behavior is not merely from the result of native structure.

P. Yamada *et al.* Forced Unfolding Mechanism of Bacteriorhodopsin as Revealed by Coarse-Grained Molecular Dynamics Biophys. J. 2016 (*76*)

A coarse-grained simulation which models each amino acid as a single particle, with hydrophobic properties adjusted according to the Roseman hydrophobicity scale (*80*). Significantly, they excluded interhelical interactions from their simulations, focusing on hydrophobicity forces between the residues and the membrane, and the intrahelical bonds. Their pulling velocity is 3 orders-of-magnitude slower than Kappel-Grubmüller (1 mm/s vs. 1 m/s), which, while still about three orders-of-magnitude faster than experimental pulling velocities, it is significantly closer, making the

comparison more appropriate. They also employed two different intermediate detection techniques: "Analysis of chain length frequency", which is very similar to the technique described in the Kappel-Grubmüller, as well as rupture peaks. In general, while the location of the intermediates did not perfectly align with the Müller literature, the number of intermediates is approximately the same (in both pulling directions). They also raise concerns once again about the validity of the underlying assumption that folded protein remains in its native location until after the secondary structure is unwound.

With the caveat that each of the simulations differs from each other in implementation and pulling velocity, they reflect a significantly different degrees of agreement with the previous experimental literature. Thus, there was need to resolve these differences.

1.3) Outline of thesis

This thesis centers on the use of cantilevers with improved force and temporal resolution to reexamine the forced unfolding behavior of the model membrane protein bR.

In chapter 2, I will outline experimental and analytical techniques used in the course of this thesis, especially relating to obtaining high enough throughput and removing artifacts from the data to make it interpretable. The closely related sets of experiments were all performed with the FIB modified Biolever Fast (AC10DS).

Next, in chapter 3. I will discuss work published in *Science* in 2017 that utilized these cantilevers in forced unfolding experiments on the C-terminus of bR (*81*). This paper reveals a significant number of new intermediates that were not observed in the Müller literature, and somewhat correlated with the intermediates predicted in Kappel-Grubmüller 2011(*79*). Some of these intermediates are assigned to residues a mere two amino acids apart from each other. Surprisingly, the improvements in temporal resolution also revealed rapid unfolding and refolding, indicating that the protein is experiencing near-equilibrium conditions. We were able to leverage this behavior by holding the cantilever at a constant distance from the surface to observe many back and forth transitions, building up sufficient statistics to reconstruct part of the energy landscape.

In chapter 4, I applied the concepts and experimental advances described in chapter 3 to perform forced unfolding from the N-terminal side of bR. This was not a trivial change, as the shorter length of the N-terminus compared to the C-terminus necessitated a change in the attachment chemistry. Once again, a significant increase in the number of intermediates (compare to the limited experimental literature) was observed. More excitingly, by comparing the intermediates observed in the C-terminal and N-terminal experiments, I show modest corroboration with previously observed and predicted hydrogen bonds. Lastly, I look at the location of intermediates with respect to the hydrophobic properties, and comment on trends observed with the location of intermediates with respect to transferring hydrophobic residues out of the membrane or hydrophilic residues into the membrane.

In chapter 5, I replicate the same experiment as chapter 4, but with protein that had had the retinal removed. This arrangement has been studied before, but never from the N-terminal side, which due to the location of its covalent bond to the protein, means that it should impact the unfolding behavior of the protein throughout the process, rather than just at the beginning (as seen in C-terminal unfolding experiments). Interestingly, no significant change in the unfolding behavior is seen in comparison to the chapter 4 results, (except for the appearance of few anomalous unfolding behaviors never before in the records) which prompts a discussion about possible interpretations of the results. This work is to be understood as in progress.

In chapter 6, I conclude the three studies covered in chapters 3, 4 and 5, and potential follow-up studies are discussed.

CHAPTER II: Experimental techniques and analytical methods

2.0) Introduction

This chapter is a technical section focusing on details of the experimental methods and analysis processes used in the experiments described in chapter 3, 4 and 5.

2.1) Substrates used and preparation

2.1.1) Mica

In the C-terminal SMFS discussed in chapter 3, freshly cleaved mica was utilized as the substrate on which we deposited the purple membrane (PM) patches. Mica offered a few advantages as a substrate for these experiments: 1) It had a slight negative surface charge as a result from the cleaving process which orients our PM patches in a predictable manner (extracellular side down) and 2) It is atomically flat, which simplified the analysis, as it removed the need to account for height variations in the fitting of the worm-like chain models (see section 2.5). Unfortunately, mica's flatness did increase the magnitude of the interference artifact (see section 2.6), as the atomically flat surface reflects the laser more coherently than glass. Concern might be raised about the impact of the charge on the unfolding behavior, but a previous paper observed no change in the unfolding patterns between mica, graphite and double-layered bR (*67*), indicating that the influence is negligible at traditional resolutions.

2.1.2) APDMES functionalized glass

In the chapter 4 and chapter 5 experiments, the substrate had to be positively charged due to the need to reverse the sidedness by which the PM patches deposits onto the substrate. As this is dictated by electrostatic interactions, we needed a positive charge on the surface. This was accomplished by depositing a layer of aminopropyl-dimethoxysilane (APDMES) (Gelest Inc. SIA0603.0) onto the surface. First, a KOH cleaned coverslide (Ted Pella Inc., borosilicate glass, Prod No. 26023) was created using techniques outlined in previous work (*82*). This was able to be done weeks ahead of the APDMES functionalization by storing the cleaned glass in parafilm lined pipet tip boxes. The night before, a Pyrex petri dish was placed into an oven at 110º C under vacuum, in order to bake-out any moisture from the

dish. Then, the glass coverslide was placed in custom-made a Teflon holder and placed into a UV-Ozone cleaner (Novascan Digital UV Ozone system) for 1 hour, to further clean the coverslide and activate the surface for the APDMES silane reaction. Then, immediately after removing the petri dish from the oven, a new 5g container of APDMES was opened and the entire contents were deposited into the petri dish. The coverslide was immediately placed into the solution. 30 minutes later the coverslide was removed from the APDMES solution and washed by immersing in at least 50 ml of reagent grade Toluene, Isopropyl alcohol, and deionized water (in that order) for 30 seconds each. At the end of the 30 seconds in each solution, each slide was gently raised and lowered through the liquid interface to aid in removal of excess APDMES. Nitrogen gas was then used to blow dry the coverslide. The glass coverslide was immediately deposited onto a metal AFM puck using optical wax on a hotplate set to 110º C. Attempts to use epoxy to fix the glass to the metal AFM puck failed, as placing the puck into a vacuum vaporized some of the epoxy, and deposited it onto the coverslide, as evidenced by AFM imaging.

2.2) Bacteriorhodopsin and Bacterioopsin preparation

2.2.1) Bacteriorhodopsin growth and purification

Bacteriorhodopsin (bR) in its native lipid bilayer (i.e., purple membrane) was purified from Halobacterium salinarum as previously described (*83*). Briefly, this process involved increasing the expression of bR in Purple Membrane by the use of plasmids on the bacteria. The bacteria was also mutated to make it resistant to the antibiotic Mevinolin, so that the growth solution could be treated with Mevinolin to reduce populations of the non-mutated bacteria. After a growth in the incubator under a strong lamp for a few weeks in 4M NaCl, the bR is purified out via a few steps of filtration and centrifugation, culminating in a 17 hour sucrose gradient centrifugation. Concentrated aliquots of PM (2.4 µg/ml) were stored at -80ºC until use, where they were thawed, and stored at 4ºC for up to a month.

2.2.2) Purple Membrane deposition onto substrate

Prior to deposition, both onto the Mica in chapter 3 and the APDMES-functionalized glass in chapters 4 and 5, the stock solution was diluted 1,200-fold into absorption buffer [10 mM Tris-HCl(pH 7.8), 300 mM KCl], sonicated for 3 minutes (Sonics & Materials Inc., VC 750), and then 50 – 100 µl was

deposited onto freshly cleaved mica, and placed into a humidity chamber (a pipet tip box flooded with water, and a soaked Kimwipe) for 1 hr. at room temperature. After this, the sample was rinsed at least five times with 100-150 µl imaging buffer [10 mM Tris-HCl (pH 7.8), 150 mM KCl]. The sample was then ready for experimentation.

2.3) Cantilever modification and functionalization

2.3.1) Cantilever modification and functionalization

The cantilevers used in all experiments were modified Biolever Fasts (Olympus AC-10DS). Prior to modification, they are nominally 9 µm x 2 µm (see Fig 1-3B). In work pioneered by Matt Bull (*52*), and improved upon by Aric Sander, Devin Edwards, and David Jacobson (*53*), they were modified by Focused Ion Beam (FIB) lithography. The process involves placing the cantilevers in an scanning electron microscope (with a FIB module installed), under vacuum, and directing a concentrated beam of gallium ions at the backside of the cantilever, milling a void near the base, as well as thinning the remaining legs (see Fig 1-3C). The stiffness ranged from 15 pN/nm – 45 pN/nm.

2.3.2) FIB contamination issues

During the course of both the C-terminal and N-terminal pulling experiments, there were periods of weeks where the cantilevers were not providing attachment to our biological molecules. These periods would start and end regardless of the bR aliquot being used. While the N-terminal experimental failings could have been caused by issues with the tip functionalization (section 2.3.4), the C-terminal experiments, not being functionalized, must have be the result of a different tip conditions. In fact, we saw a difference between the attachment rate between cantilevers from the same wafer, modified by two different FIB lithographs (One located in the engineering department at the University of Colorado, Boulder and the other at NIST-Boulder). Considering the stark contrast in attachment rate from temporally proximal experiments, it is hypothesized that the engineering department's FIB lithograph, being a communal instrument, is more susceptible to carbon contaminations. Interestingly, these differences in performance persisted, even after rinsing the modified cantilevers in toluene. Eventually,

the CU-Boulder FIB lithograph began yielding useable cantilevers again, but one should be weary of the possibility of contaminated tips, as this cost the author months of productivity.

2.3.3) Chapter 3 tip cleaning

Before the C-terminal pulling experiments, which utilized nonspecific attachment chemistry, we placed the cantilevers in a custom built tip-holder, and inserted into the plasma cleaner within the Keck lab (AXIC Inc. PlasmaSTAR). The tip-holder prevented damage to the tips from drafts within the plasma cleaner during operation. The tips were then exposed to ionized oxygen, with parameters of 100 W, 30 sec and 100 SCCM. Not all modified cantilevers were usable after this process (especially ones with spring constants < 20 pN/nm), assumedly because the plasma cleaning process removes a small amount of material, causing thinner tips to buckle. Ideally, the tips should be placed within the instrument and utilized within a half hour of plasma cleaning. Plasma cleaned tips exhibited a more than five-fold greater attachment rates than non-plasma cleaned tips. Cantilevers which were structurally intact after an experimental session could be reused by additional plasma cleanings.

2.3.4) Specific attachment tip functionalization protocol

Before the N-term and BO pulling experiments, we switched to using a specific attachment chemistry, (Dibenzocyclooctyne (DBCO) to Azide, i.e. copper-free click chemistry, see Fig 2-1). To achieve this, we ordered commercially manufactured Silane-PEG-Azide (SPA) constructs (Nanocs), either with a PEG 600 linker, or a PEG 3400 liker. While Hao Yu and David Jacobson regularly achieved acceptable attachment rates using the PEG 600 linker in C-terminal experiments, we could not reliably achieve an attachment rate greater than 0.1%. Thus, we opted to use the PEG 3400 linker for these experiments.

To functionalize the cantilevers, we first mixed the stock SPA (stored in -14º C, with desiccant) into 2 ml of reagent grade isopropyl alcohol initially, and added toluene until a concentration of 7.5 mg/ml was achieved. The cantilevers were placed into the UV-Ozone cleaner for 1 hour prior to insertion into the SPA-toluene solution to remove oxide layers on the cantilever. The SPA-toluene solution was deposited into petri dishes modified to accommodate a small stir rod. These petri dishes were modified by

fusing a 10 ml beaker cut in half horizontally with vertical slits were cut into the bottom, and fused to the bottom of a glass dish, so as to mix the SPA-Toluene solution during deposition, while confining the stir bar to prevent it damaging the fragile cantilevers. The petri dish was then placed onto a hot plate set at 98 ºC (to achieve a temperature solution of 60 ºC), and the cantilevers were deposited in the dish for 3 hours, with stir bar set to 600 rpm, with cover applied to prevent evaporation. After 3 hours, the cantilevers were removed, and washed in toluene, IPA, and deionized water, in that order, each for 30 seconds. The cantilever was then placed in phosphate buffer solution (PBS) in an airtight container, and refrigerated in a humidity chamber (a pipet tip holder filled with water) for up to 2 weeks prior to use.

2.4) Specific N-terminal attachment bR functionalization protocol

The other half of the functionalization protocol to the specific attachment chemistry (DBCO-Azide) first mentioned in section 2.3.4, which was necessary given the low attachment rate that was achievable with nonspecific attachment on the N-terminus of the protein. In work performed by Lyle Uyetake, the N-terminal glutamic acid was mutated into a cysteine. The bR was sonicated (to break up PM aggregates) before the N-terminal cysteine was reduced using TCEP, and then exposed to Maleimide-PEG-DBCO (Fig 2-1 right). The TCEP was diluted out of the solution, and the bR was concentrated via centrifugation. Initially, a PEG monomer was used in the construct, but significantly higher attachment chemistries were achieved after the PEG linker length was increased to a 4-mer. See section 4.7.2 for more details about concentration.
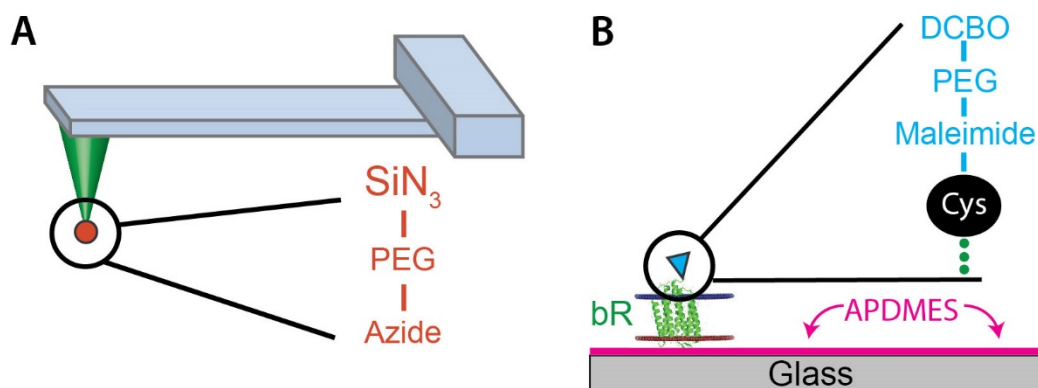
Figure 2-1: Specific attachment functionalization chemistry. Chemistries performed on the cantilever tip, and the N-terminus of the protein, to achieve a specific attachment include **(A)** Silane-PEG3400-Azide (orange) functionalization of the tip, and **(B)** E1C mutation of the bacteriorhodopsin, then reduced using TCEP and attached via a thiol bond to a Malemide-PEG4-DBCO (blue) molecules. The surface APDMES functionalization is also depicted (pink).

2.5) Worm-like chain model

The main form of data created by force spectroscopy experiments are force-extension curves (FEC), briefly described in chapter 1. Basically, they measure how unfolded polypeptide, modeled as a semi-flexible polymer, will stretch when force is applied. This behavior, in the force ranges that will be exhibited in this thesis, is modeled by the Worm-like chain, an extension of the Kratky-Porod model first proposed in the 1949 (*84*), the version we use was developed by David Marko and Eric Siggia developed (*85*) a simple empirical formula for WLC model as follows:

$$F = \frac{k_B T}{L_p} \left(\frac{1}{4} \left(1 - \frac{x - x_0}{L_c} \right)^{-2} - \frac{1}{4} + \frac{x - x_0}{L_c} \right) \qquad \text{(Eqn. 2-1)}$$

where F is force, x is the extension, x_0 is the extension offset, L_p is the persistence length (0.4 nm for unfolded polypeptide), L_c is the contour length (the length of the polymer at maximum extension, also sometimes signified as L_o), and $k_B T$ is the thermal energy. Conversion between force-extension and contour length will be utilized extensively in analysis, especially for alignment purposes (described in Section 3.10.1), as well as in state determination in chapter 5.

2.6) Deflection laser interference artifact and removal

2.6.0) Interference artifact causes

The main disadvantage to using ultrashort cantilevers comes from an optical interference artifact between the portion of the deflection laser that is reflected off of the cantilever with the portion of the deflection laser that reflects off of the sample surface. The interference artifact is present in the deflection signal as a sinusoid with slightly growing amplitude as the tip approaches the surface (Fig 2-2C). For most ultrashort cantilevers, it is usually 1-2 nm peak-to-peak in amplitude near the surface, which depending on the cantilever stiffness can mean that the artifact will be of comparable size to a typical bR rupture force (~ 40-90 pN). Thus, the removal of the interference artifact is paramount to correctly analyzing the unfolding traces.

It is not clear which pathway the laser light that interferes with our deflection signal takes. There are two (non-mutually exclusive) possibilities: either from light that transmits through the cantilever (which should be worse for a Biolever Fast than it is for a Biolever, considering their relative cantilever thickness: 0.13 μm and 0.18 μm, respectively), or moving around the edges of the cantilever (which should also be worse for a Biolever Fast than it is for a Biolever, considering the relative cantilever widths: 2 μm and 30 μm, respectively).

2.6.1) Interference artifact removal: experimental

While a "small spot size" laser module was provided by the manufacturer of our AFM (Fig 2-2A), its spot proved insufficiently focused (~9 x 3 μm^2) and thus a large interference artifact was observed.

A custom small spot size laser was installed in our AFM in work done by Devin Edwards (*53*). In addition to having a smaller laser spot (~3 x 3 μm^2) (Fig 2-2B) the custom laser also employed 2 different strategies (chronologically) to further reduce the magnitude of the interference artifact: first, we modulated the laser at radio frequency using an acousto-optical modulator (AOM) (results in Fig 2-2C), and later (for chapter 4 and 5) we swapped out the standard laser for a superluminescent diode. The latter has a broader linewidth, which reduces the magnitude of the interference artifact.
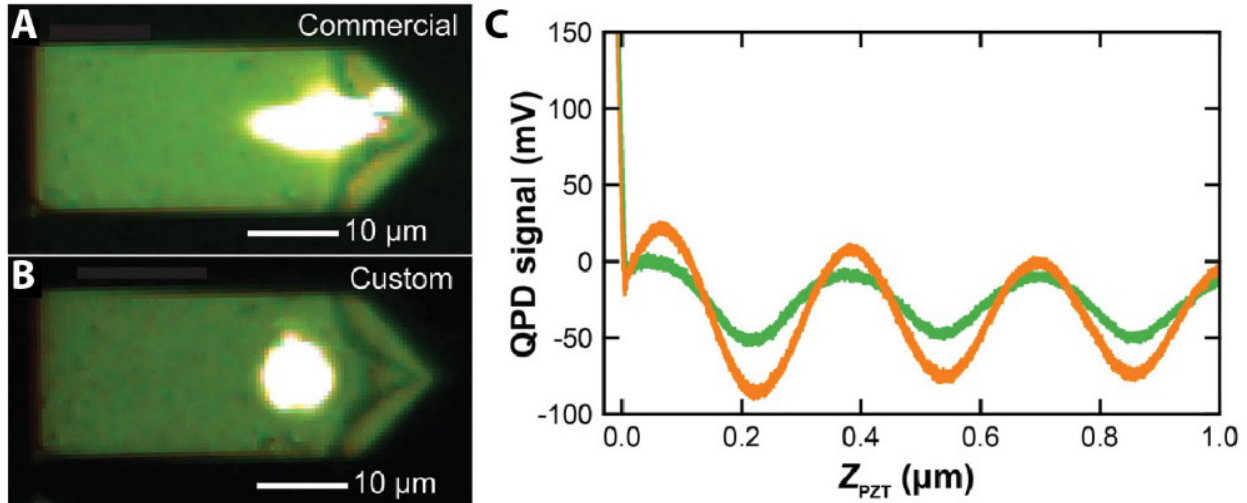
Figure 2-2: Comparison of spot size from commercial and custom laser and impact of modulation on interference artifact. **(A)** Biolever mini with commercial "small spot size" laser focused on the back. Spot size is ~ 9 x 3 μm^2. **(B)** Biolever mini with custom laser module. Spot size is ~ 3 x 3 μm^2. **(C)** Interference artifact recorded on a FIB modified Biolever Fast with (green) and without (orange) laser current modulation. Adapted from Nano Lett. 15, Edwards, D.T. *et al*. Optimizing 1-μs-Resolution Single-Molecule Force Spectroscopy on a Commercial Atomic Force Microscope. 7091-7098. Copyright 2015, with permission from Elsevier.

2.6.2) Interference artifact removal: analytical

The other means of removing the interference artifact from data was subtraction during analysis. The shape of the interference artifact is relatively constant, though the magnitude and phase are prone to drifting over the course of an experimental run, and vary slightly by location on the substrate. We developed a phenomenological equation that fit the behavior:

$$\Delta z_{interference} = w_1 + w_2 * x + (w_3 + w_4 * x) * \sin((w_5 + w_6 * x) * x + w_7) \qquad \text{(Eqn. 2-2)}$$

$$x = Z_{PZT} - \Delta z_{Cantilever}$$

$$\Delta z_{Cantilever} = \Delta z_{Measured} - \Delta z_{interference}$$

where $\Delta z_{Measured}$ is the measured deflection of the cantilever, $\Delta z_{Interference}$ is the contribution of the optical-interference artifact on $\Delta z_{Measured}$, $\Delta z_{Cantilever}$ is the actual deflection of the cantilever that gives rise to the applied force ($F_{Corrected} = k\Delta z_{Cantilever}$), Z_{PZT} is the distance the base of the cantilever has been retracted, x is the extension (*i.e.*, tip-sample separation), and w_1–w_7 are fitting parameters. The results of the initial fit are demonstrated in Fig 2-3. A closer demonstration of the residual interference

artifact can be observed in Fig 2-3B-C, with the initial interference artifact being smaller than 4pN

globally.

For N-terminal studies (Chapter 3 and Chapter 4), the fit of Eqn. 2-2 to the interference artifact proved insufficient for proper interference artifact removal. We suspect this is caused by the change of substrates from mica to glass, where the flatness of the substrate diminished leading to more irregularity of the interference artifact. A new algorithm was needed. A spline method, where the incoming (extension) wave is smoothed and used as a reference to reconstruct the interference artifact observed in the retraction wave, was implemented.

It was necessary to implement both of these processes in an iterative method, as the degree of interference artifact that was contributing was a function of the tip position, not of the *Zpzt* (base of the cantilever positions).Thus, with multiple iterations, an accurate calculation of $\Delta z_{\mathrm{Cantilever}}$ and x for the full retraction curve including the unfolding events was derived. We initially approximated the $\Delta z_{\mathrm{Interference}}$ artifact from Z_{PZT} alone (*i.e.*, $x = Z_{\mathrm{PZT}}$). We then deduced a first estimate for x using this approximate value for $\Delta z_{\mathrm{Interference}}$ and calculate a new value of $\Delta z_{\mathrm{Interference}}$. Only 5 iterations were necessary to reach convergence in both the phenomelogical equation method, and the spline method.
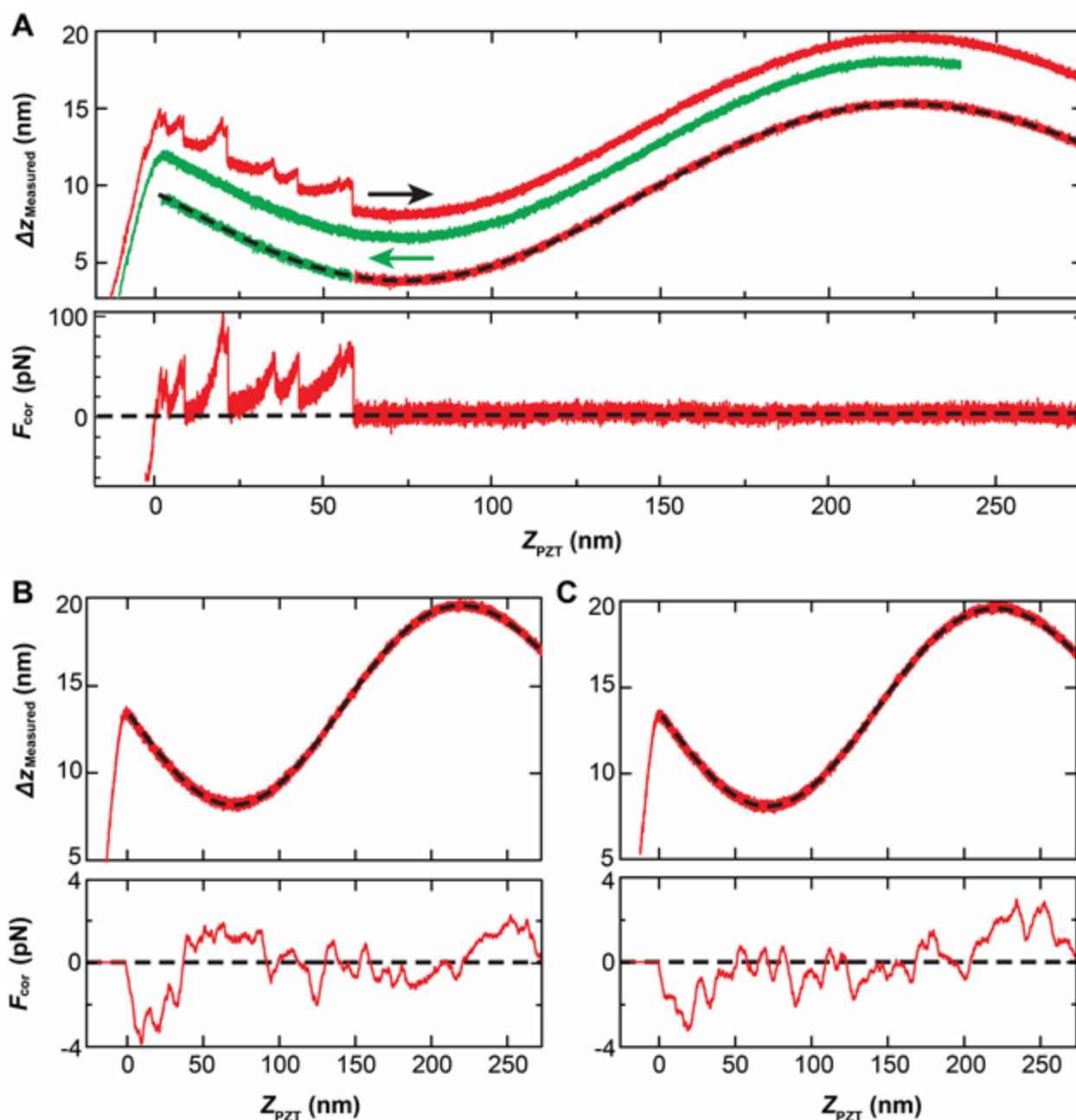
Figure 2-3: Process for removing optical-interference artifacts in analysis. **(A)** Measured deflection vs. cantilever motion (Z_{PZT}) during an approaching (green) and retraction (red) cycle shows optical-interference artifacts (top panel). The traces are offset for clarity. The bR unfolding signal is superimposed on top of these larger interference fringes. To remove the interference artifact, part of the retraction curve after the molecule was unfolded (red) and part of the approaching curve (green) were concatenated (red and green) and parameterized with Eqn. 2-2 (dashed black). The canonical bR unfolding pattern is seen in the corrected F (lower panel). **(B–C)** The corrected force-vs.-cantilever motion using the parameters used to correct the bR unfolding record is shown for the prior and subsequent retraction curve (left and right panels, respectively). The fitting residues were less than 4 pN, a small force error compared to the standards of AFM-based SMFS (*44*). All data traces were smoothed to 10 kHz for presentation.

2.7) Instrumental noise mitigation

2.7.1) Peltier cooler noise joined into close-loop measurement.

　　Midway through collecting the C-terminal unfolding data, we noticed an increased amount of low force short-lived transitions in our data. This was unusual relative to the force ranges over which transitions were observed in the initial dataset. After a search of different potential noise sources, we (with help from Jason Cleveland) determined that it was related to the Peltier cooler that maintained the sample temperature. It was coupling into the deflection signal indirectly, by adding electrical noise to the stage position measurement (Z sensor) channel on the AFM, which, when in closed loop positional feedback mode (where the Z sensor channel is monitored for deviation from an assigned route, and corrected in a proportional manner) would move the tip in the opposite direction to the noise (see Fig 2-4). Thus, this noise only coupled into our measurements when operating in closed loop mode. What made this especially difficult to determine was that the magnitude of the noise depended on the Peltier's duty cycle in a non-linear fashion (See Fig 2-5). As the temperature inside the noise isolation box would change over the course of the experiment, the duty cycle would change, and the noise level would increase and decrease. We speculate that the noise is being added onto the Z sensor channel at a much higher frequency, and was only coupling in when it was aliasing. Unfortunately, the equilibrium temperature of our instrument on an overnight session had the Peltier cooler approximately running at -18%, right near the major spike in the noise. This artifact forced us to exclude 45% of our molecular unfolding curves being used in our ED intermediates analysis, 29% in our CB intermediates analysis, and 12% in our A intermediates analysis. These percentages are different, as we did not remove all closed loop data vulnerable to this noise, but rather only the unfolding traces whose Z sensor records showed spikes of 1 nm (peak to peak) or greater, and the increased compliance of the linker when pulling on helix A compared to helix E effectively filtered the noise spikes from impacting our molecule. Note though that this problem has been avoided since the C-terminal unfolding experiment by working exclusively in open loop mode (where the feedback is disabled).
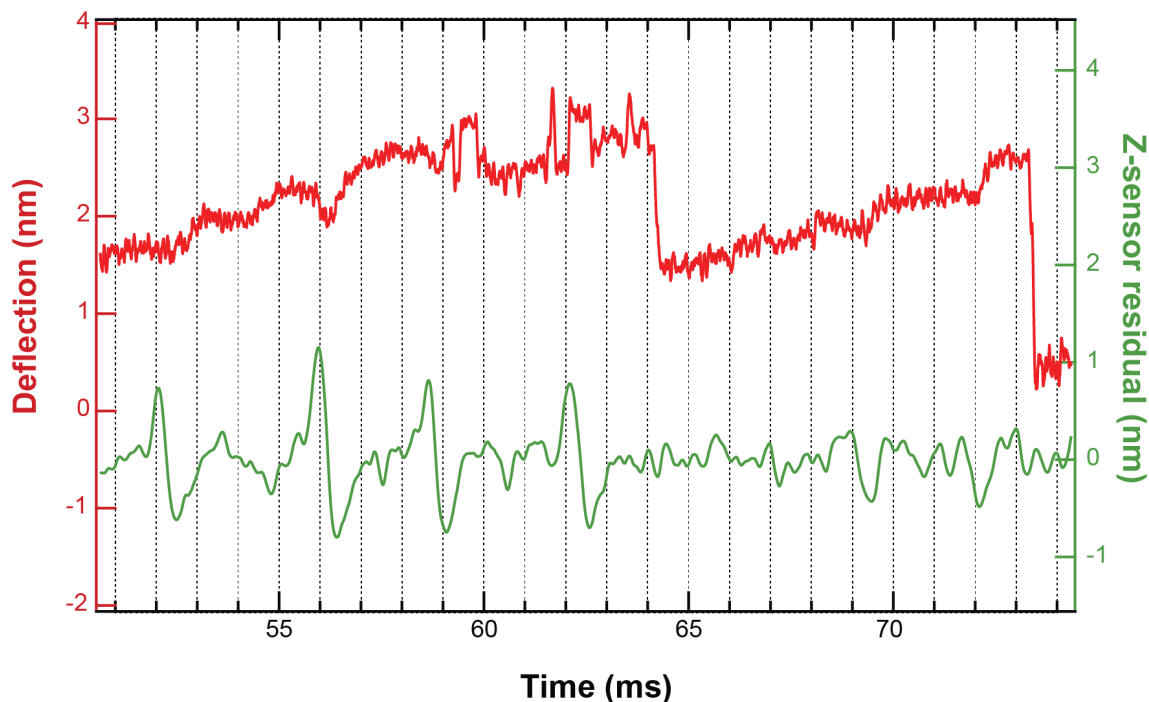
Figure 2-4: Electronic noise driving transitions in closed-loop mode operation. In the course of our experiments, we performed pulls in both open loop and closed loop mode. A transient noise around 8 kHz appeared in our capacitive measurement of our piezo position (Z-sensor). In open loop mode, these noise signals did not couple into the piezo motion, but did in closed loop. An example of this closed loop coupling is shown above. The red curve is a cantilever deflection-time plot for a single molecule, while the green curve is the z-sensor residual signal (linear change from constant velocity pull) removed. The noise spikes on the green channel were initially electronic, but it induced a motion of the Z-piezo, which could induce transitions, as evidenced here by the motion in the deflection signal within a millisecond after the noise spike appears in the z-sensor residual, especially at high force (when the compliance of the unfolded polypeptide linking the cantilever to the protein is less).
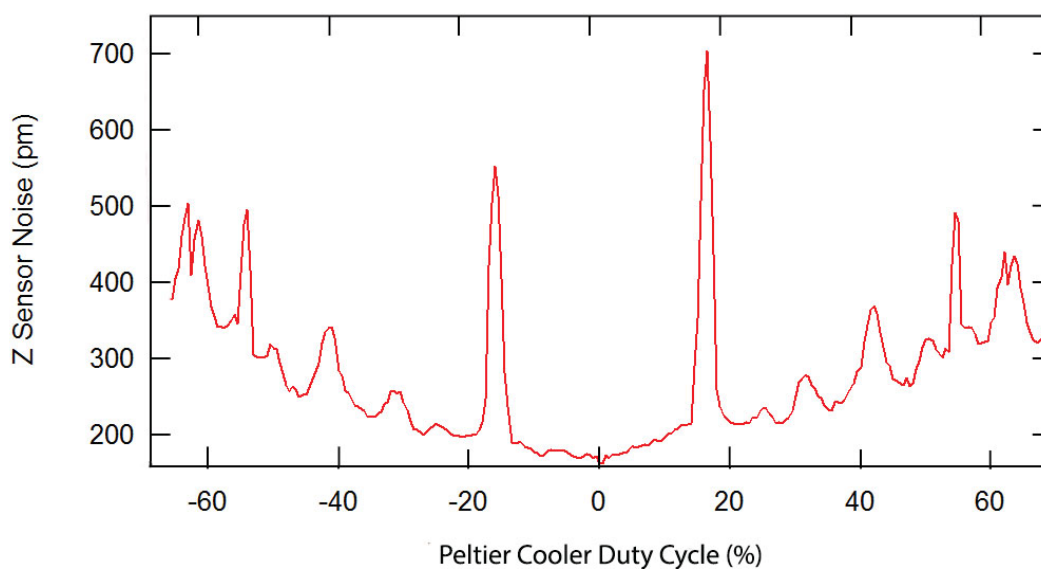
Figure 2-5: Peltier noise coupling into the Z piezo position. The horizontal axis corresponds to the duty cycle (which is effectively the percent of power applied) to the Peltier cooler on our AFM. The vertical axis here is the electronic noise measured on the Z sensor channel on the Cypher. Notice how there is both a gradual increase in the noise level both directions away from 0%, as well as short lived noise spikes (e.g. at -18%). This latter phenomenon is a reflection of the fact that the Z sensor has is being filtered in the course of its feedback cycle. Considering that during the course of an experiment, the Peltier duty cycle would increase in response to rising temperatures in our noise isolation box (See section 2.7.2), the noise level could vary highly over the course of an average force spectroscopy pulling session. Figure Credit: Jason Cleveland.

2.7.2) Ambient/HVAC noise

Early in the experimental process, we noticed a noise in our deflection records, especially in the 100-300 Hz range (Fig 2-6, red). After an extensive search for the source of the noise using accelerometers, it was determined to be the audible noise coming from the HVAC system within the room. We constructed a noise isolation box by repurposing previously built sound isolation panels. The reduction in the noise is shown in Fig 2-6, as the noise spectrum is shown with the box door open and the box door closed. A picture of the box, with a panel removed for visibility, is shown in Fig 2-7.
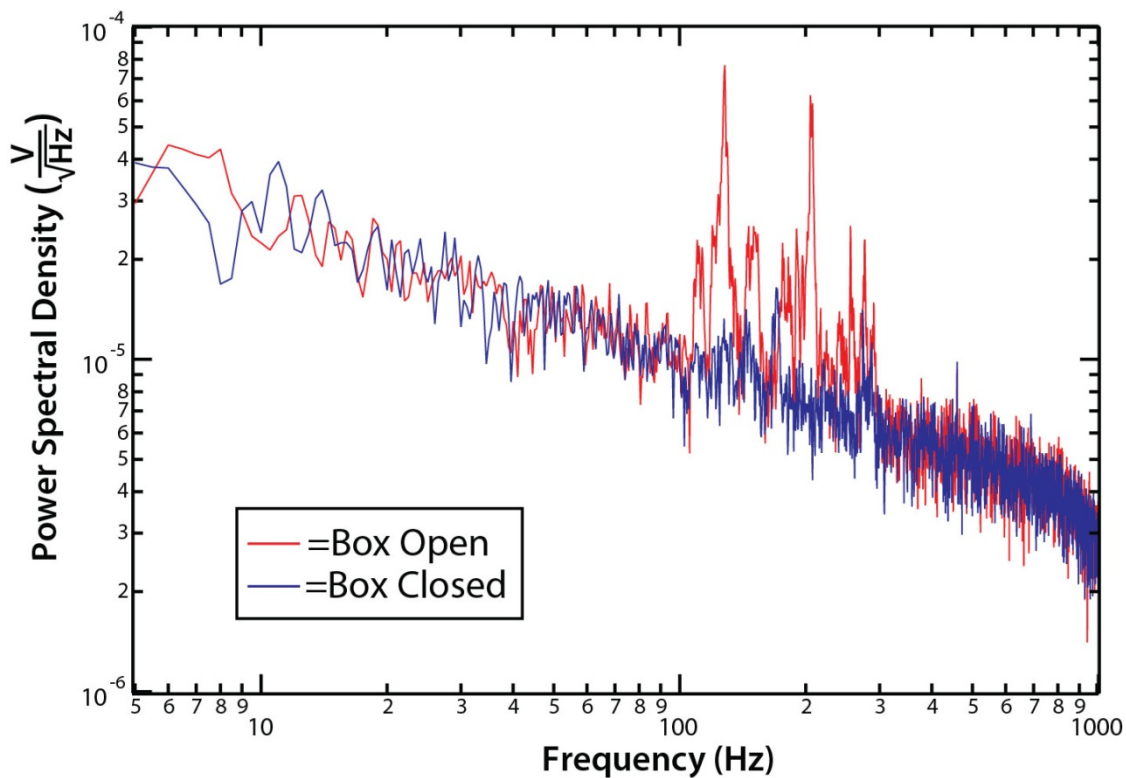
Figure 2-6: HVAC related noise. Power spectral density of the vibrations measured both with the noise isolation box door open (red), and the noise isolation box door closed (black), and the isolation box closed. Note that this was measured from the uncalibrated lateral channel of the AFM, thus the vertical axis units are arbitrary. The figure still qualitatively shows the improvement attained using the noise isolation box.

Figure 2-7: **Noise isolation box around the Cypher AFM, with the front panel removed.** The panels are constructed of different foams with contrasting densities, making then non-transmissive to noise. Quiet computer fans and foam tubing were installed to maintain reasonable temperatures within the box when closed. Photo adapted from (*86*).

2.8) Force spectroscopy parameters

2.8.1) Chapter 3 experiments

After the purple membrane patches are deposited on the substrate, topographic imaging was required to locate a patch, as well as determine if the patch is optimal for force spectroscopy. Optimal patches are greater than 300 nm in diameter, and 4-7 nm in height (membrane thickness), as well as a surface texture with 20-50 nm diameter protrusions (see Fig 2-8 for an example).

Once an appropriate patch is located, a set of points was assigned over a square area of the patch (what is known as a 'force map'). The map consists of somewhere between 25 and 100 points depending on the scale of the map (as each point was separated by ~2.5 nm). Then a pattern is started, first with FEC is collected at a different point, followed by a FEC taken off-patch, to check the cantilever sensitivity, as well as remove any proteins adhered to the tip from previous attempts. Each FEC is collected using parameters established in the previous literature specifically for non-specific attachment on bR (*67*). First,
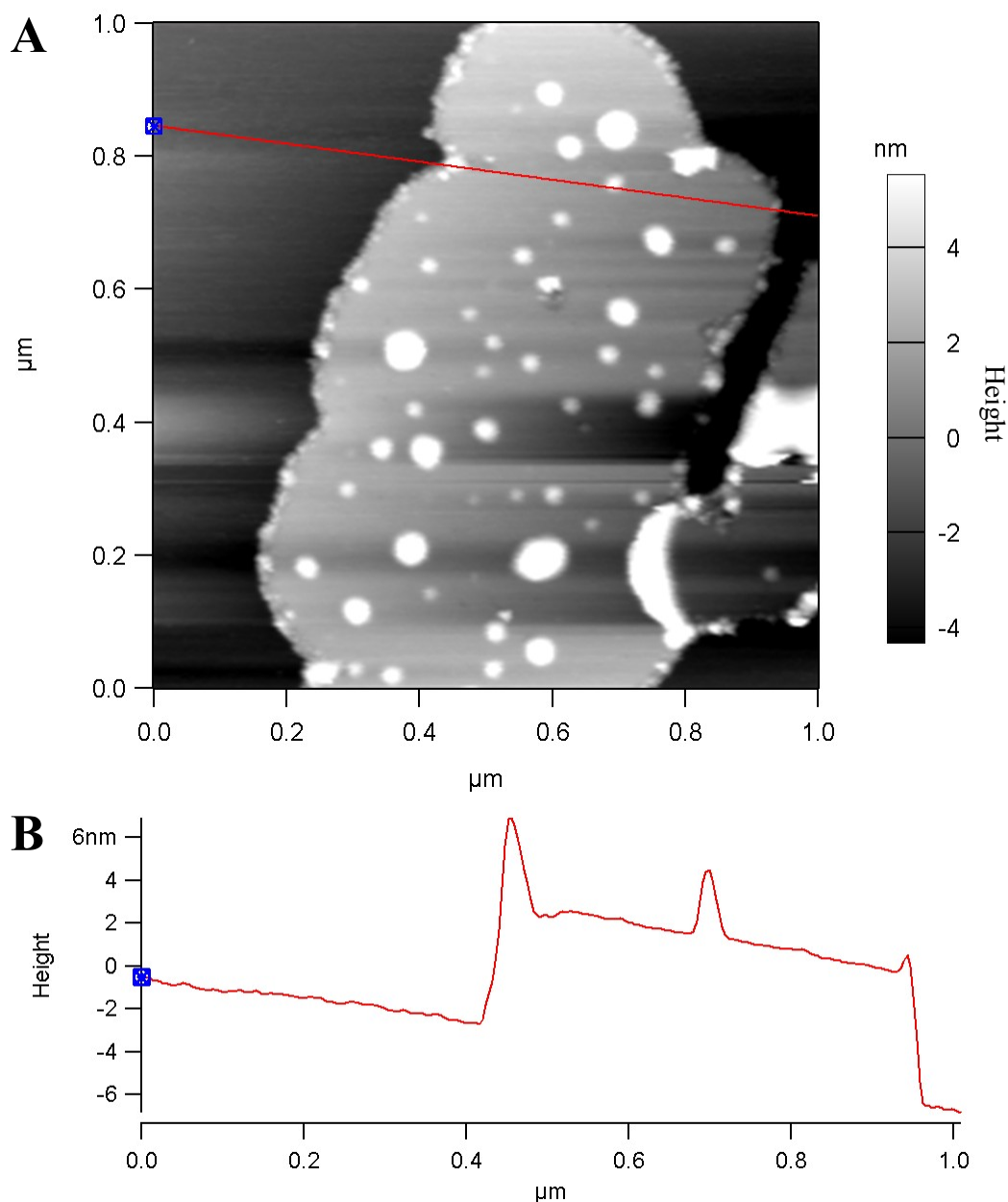
Figure 2-8: Purple membrane patch before a C-terminal force spectroscopy pull. (A) AFM topograph of Purple Membrane deposited on mica. The size of the patch (>300 nm in diameter) ensures that it is both much greater than the nominal tip radius (20 nm), precluding interactions with the substrate, as well as ensuring that the membrane patch will be well attached to the substrate, due to the collective strength of each proteins electrostatic attraction to the mica. (B) A line profile taken of the section highlighted by the red line on the topograph, demonstrating the height of the patch (~5 nm). This indicates that it is likely a single membrane patch, as opposed to multiple patches stacked on top of each other. The presence of small (< 50 nm bumps on top of the membrane patch), indicates that it is likely cytoplasmic side up, as those bumps (likely tiny patches of PM) only tend to accumulate on the cytoplasmic interface of bR.

the tip is extended into the PM at 300 nm/s into the patch until a force of 700-900 pN is reached (referred to as the contact force). The cantilever is then held at that position for ~1 second. The cantilever is then retracted from the surface at a constant velocity of 300 nm/s. The deflection is being measured during this process. It was important to keep the contact force below 1.4 nN risks, as exceeding this force risks puncturing the PM, which likely places that bR into a non-native configuration (*87*).

2.8.2) Chapter 4 and 5 experiments

For N-terminal pulling experiments, the first two criterion for the patches (horizontal size, as well as thickness) are also used, but the tip functionalization process added large polymers to the tip surface and made imaging the surface texture impossible (See Fig 2-9).

The force spectroscopy assay used here was similar, but the parameters of the surface contact force and surface contact dwell were lowered and raised, respectively, to account for the change in chemistry. The contact force was kept below 300 pN (normally 150 pN-200 pN) to reduce the formation of nonspecific attachments to the tip. The contact time was between 3-5 seconds, so as to promote the formation of DBCO-azide bonds. The extension and retraction speeds were kept at 300 nm/s.
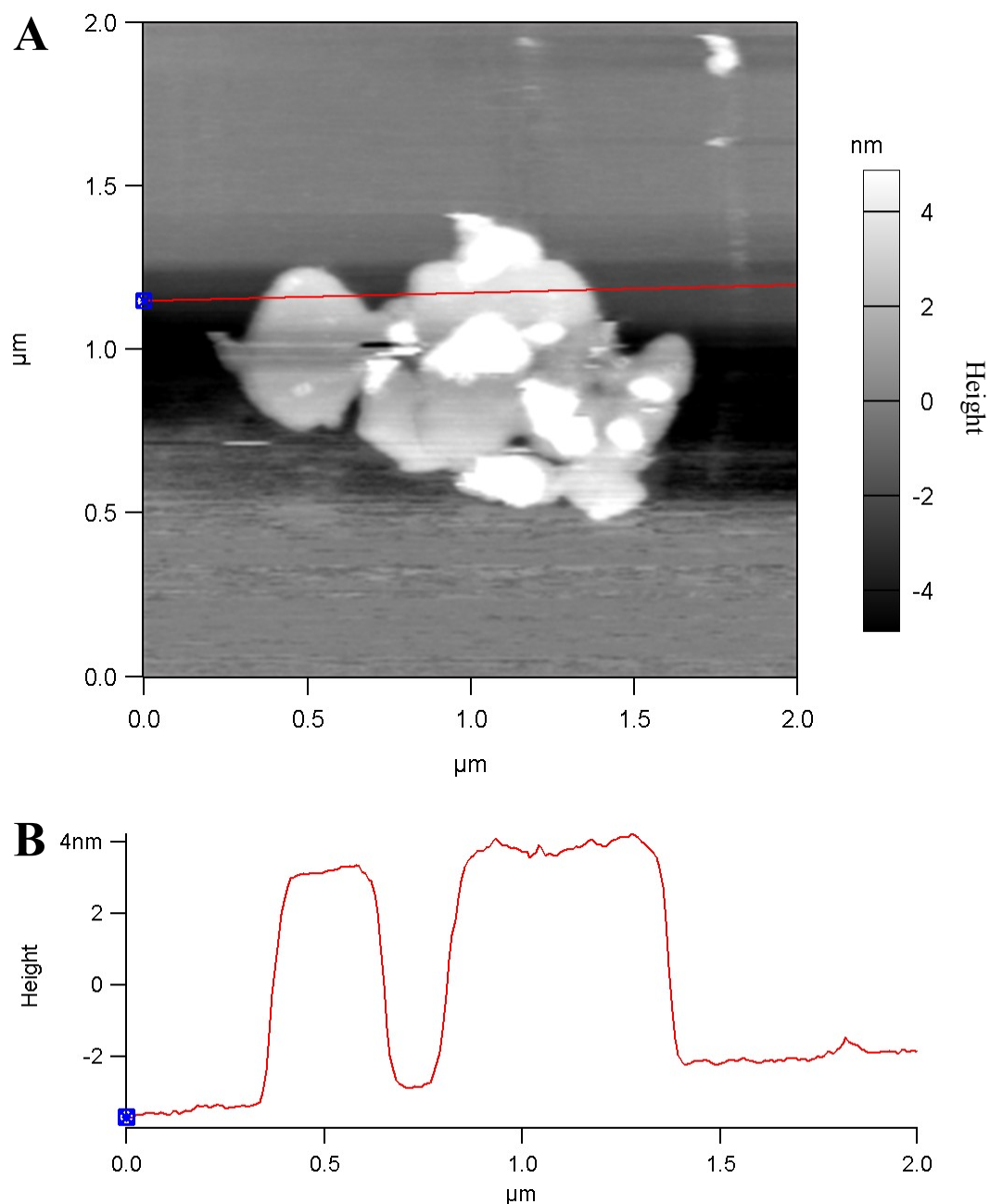
Figure 2-9: Purple membrane patch before a N-terminal force spectroscopy pull. (A) An AFM topography of Purple Membrane deposited on APDMES functionalized glass. Similar to the C-terminal patch (Fig 2-8A), it is large enough (>300 nm) and the appropriate thickness (4-7 nm), but the polymers on the tip preclude the ability to see surface texture at anywhere near the same resolution as in Fig 2-8A. Note that the white sections in the upper figure arise from PM sticking to the single layered patch. As those double-layered regions are of indeterminate sidedness, they are avoided. (B) A line profile taken of the red line overlaid on the topograph that demonstrates the patch thickness.

2.9) Filtering/selection criterion for bR force-extension curves (FECs)

2.9.1) Selection criterion for C-terminus (Chapter 3)

As is generally true for AFM-based SMFS, the non-specific attachment between the tip and the bR led to many FECs that were not interpretable (*e.g.*, tip attachment to bR between the E and F helix). Analogous to established metrics (*64*, *67*, *74*, *75*, *88*), we used the following criteria for selecting a valid FEC for further analysis: (*i*) the FEC must exhibit three major unfolding peaks, in addition to some measureable unfolding of the GF helix pair; (*ii*) the FEC must have the onset of E helix unfolding occur at an extension of ≥ 18 nm; (*iii*) the FEC must have the onset of A helix unfolding occur at extensions ≥ 58 nm; (*iv*) the FEC must show no extra adhesion events after A helix is fully removed (*i.e.*, no rupture events at extensions greater than ~ 75 nm.) Examples of force curves that fail each of these criterion are shown in Fig 2-10.

2.9.2) Selection criterion for N-terminus (Chapter 4 only)

For the N-terminal and BO experiments, specific attachment of DBCO-Azide was utilized (see section 2.4). As a result, attachment to the terminus of the protein is certain. As a consequence of this, large (> 60 pN) rupture forces for the first peak (in these experiments, corresponding to the AB helices) are consistently observed. Thus, condition (*i*) from the C-terminal experiments was expanded to require 4 rupture peaks. Condition (*ii*) was changed from 18 nm to 21 nm. In addition, another criterion is necessary: the tip approach curve must not diverge from its nominal form (no irregular force noise), due the difficulty that this adds to the removal of the interference artifact. Variability in where on the tip the SPA molecule is attached was still possible, meaning that the relative distance between the attachment point and the tip apex changes for each unfolding record. Thus horizontal alignment is still necessary.

2.9.3) Selection criterion for N-terminus bO

The selection criteria are more central to the analysis of chapter 5, and will be presented there (see section 5.2.4).
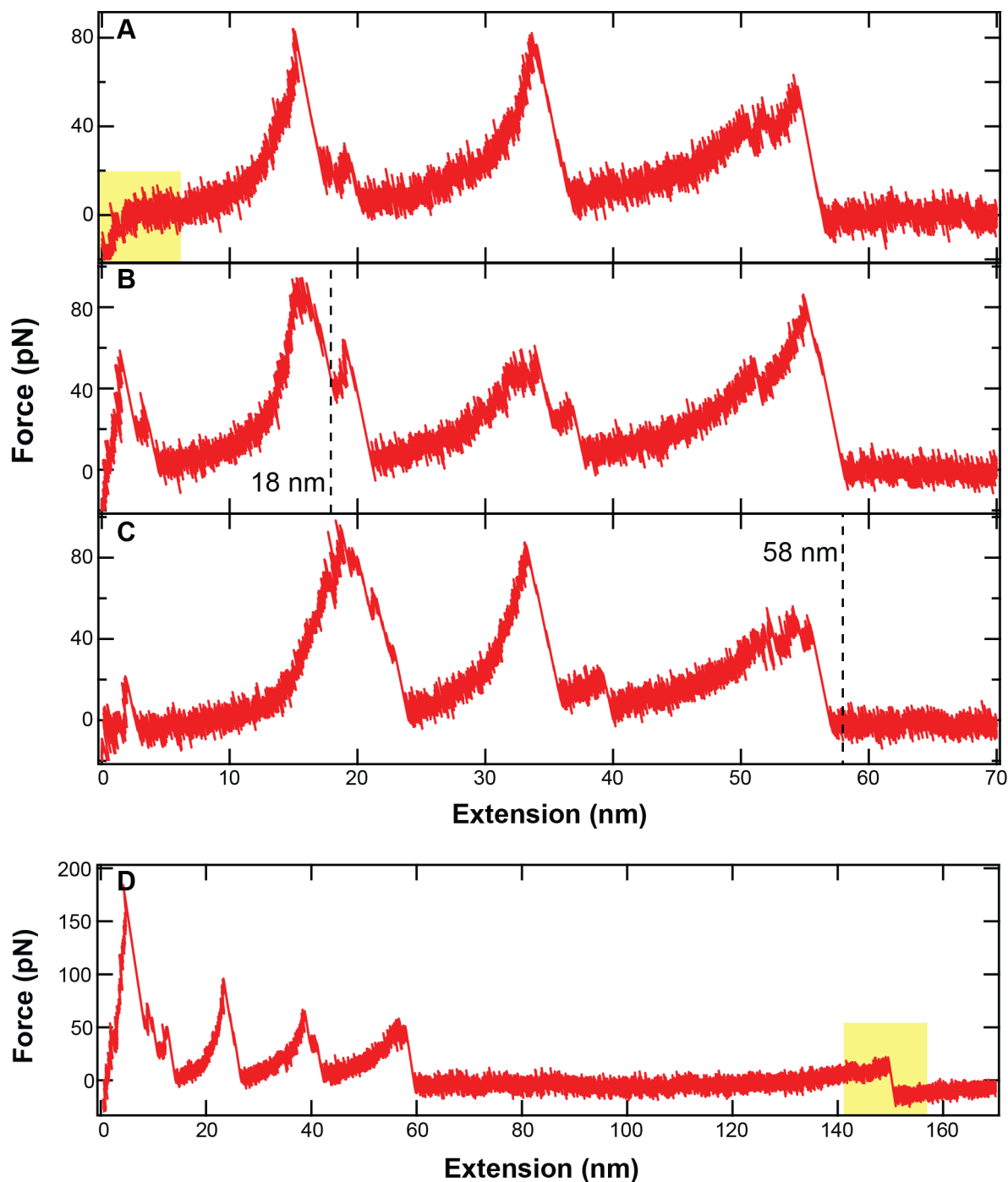
Figure 2-10: Examples of pulls that failed the extension criterion. Section 2.9.1 establishes four criterion for rejecting curves, in order to guarantee single molecular attachment at or near the C terminus of the BR molecule. The following molecules were rejected because they **(A)** failed to demonstrate any GF rupture event (yellow box is area of interest for this judgment), **(B)** the ED rupture event failed to reach past 18 nm (see dotted line), **(C)** failed to achieve an extension of 58 nm or more (see dotted line), and **(D)** failed to fully detach by 70 nm, showing a rupture event after the A helix (highlighted by the yellow box)

2.10) Data sampling rates & filtering/smoothing

Data from the AFM was collected at 5 MHz, but for the state determination, and pathway analysis the data was smoothed to improve the signal-to-noise. To remove the noise, but to minimally impact the transitions rate, a Savitzky-Golay filter was employed with a window size of 501.

In section 4.5, less filtered data was analyzed to determine if rapid unfolding and refolding transitions were being obscured by this standard filtering. There, the data was filtered with a window size of 51, an order-of-magnitude smaller than our standard filtering. It was discovered that while some short lived transitions were obscured by the standard 501 filtering, no specific pathway was completely obscured to the point where a given intermediate was completely missed, meaning that the occupancy rates were largely unaffected.

CHAPTER III: Cytoplasmic forced unfolding of Bacteriorhodopsin

3.0) Introduction

The process by which protein folds into its native structure is the subject of intense research across many fields. A few different frameworks have arisen to interpret this process (such as the energy landscape, Section 1.0). What is common to most of these frameworks is the necessity of capturing all the unfolding steps, including transient configurations and rarely observed pathways. Omission of any folding intermediates or pathways likely will result in inaccurate modeling of the folding process, and with the rise of simulations that leverage cooperative folding units for modeling, could lead to inaccurate predictions of the final folding configuration (*89*). For example, limited experimental precision can lead to two closely spaced states being misinterpreted as a single composite state exhibiting nonexponential lifetimes (*90*). Thus, measuring the folding process (and the underlying energy landscape) in high resolution should be a priority in any assay.

The membrane protein bacteriorhodopsin (bR) has been extensively studied in forced unfolded experiments for the past 18 years (see Section 1.2) (*17*, *64*, *67*, *68*, *70*, *72–75*). In that time, the number and location of states has remained constant (Fig 3-3A top left inset, 3-3B top left inset, 3-3C top left inset). This may be the result of the cantilevers used during this period, which all share similar temporal resolutions (>50 µs), as well as force precisions. We hypothesized that the lack of resolution may have been obscuring different unfolding intermediates during the forced unfolding process, which would result in missed details in the energy landscape reconstructions (Fig 3-1A). Inspired by previous work on globular proteins (*91*), we utilized modified ultrashort cantilevers (see Section 1.1) to perform SMFS on the cytoplasmic side of bR (Fig 3-1B). This chapter (including many of the figures herein) is based on work published in (*81*), with some text reproduced verbatim.

3.1) Brief review of bR forced unfolding behavior

During the forced unfolding process of a molecule, the length of the unstructured segment of the protein is measured by fitting the force-extension record with a worm-like chain model (see section 2.5).
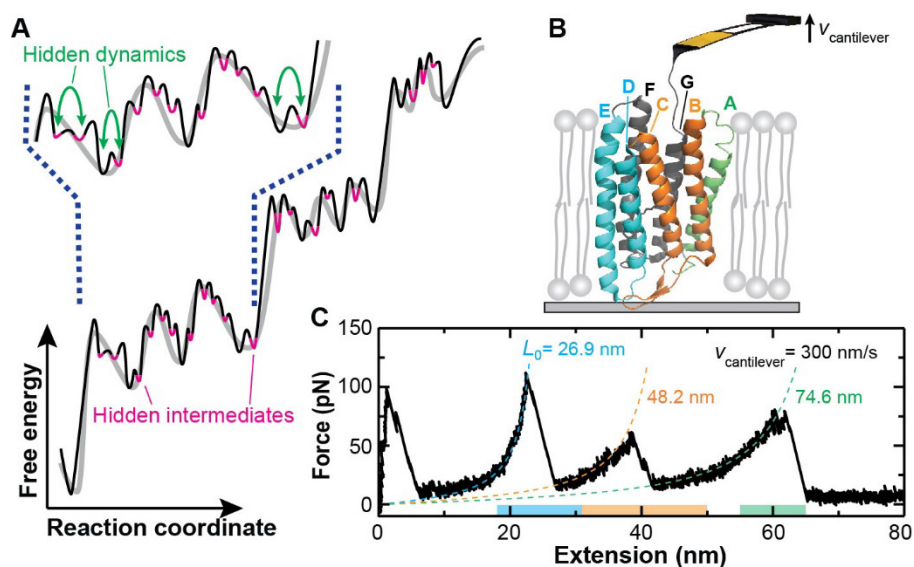
Figure 3-1: Single-molecule force spectroscopy of bacteriorhodopsin (bR) using ultrashort cantilever. (**A**) A conceptual sketch shows a low (grey) and a high-resolution (black) representation of the same free-energy landscape. Each free-energy valley represents an intermediate, with lower energy and therefore more fully folded states depicted on the left while higher energy, more unfolded states are shown on the right. Assays with improved sensitivity enable the detection of previously hidden folding intermediates (magenta) and protein dynamics between closely-spaced states separated by low barriers (green arrows). (**B**) A cartoon illustrates the unfolding of individual bR molecules by a modified ultrashort cantilever. Mechanical unfolding occurs by retracting the cantilever at a constant velocity. Each transmembrane helix is identified by its standard letter label. (**C**) A typical force-extension curve (FEC) using a modified ultrashort cantilever recapitulates the three previously detected major states corresponding to pulling on the top of E (cyan), C (orange) and A (green) helices. The FEC segments associated with these major states are well described by a worm-like chain model (colored dashed lines) and labeled with their associated contour lengths. The colored bars denote the extension range in panels D to F. From Yu, Siewny et al. Science 2017. Reprinted with permission from AAAS.

Integral membrane proteins unfold in sequential order, which means that a peak can be assigned to a specific configuration of the protein (*49*) (see section 3.3.1). When bR is forcibly unfolded, it nominally exhibits a 4 peak unfolding pattern, which corresponding to the sequential unfolding of the GF helices, ED helices, CD helices, and A helix (Fig 3-1C). These four peaks correspond to stable configurations observed at the top of the G, E, C and A helix. Often the G helix state is distorted or obscured due to nonspecific interactions between the tip and the surface, but the latter 3 must be observed undistorted in order for the unfolding record to be interpretable. For these reasons, in this chapter we will only be studying the unfolding of the last 3 peaks. This obligate nature has resulted in a distinct name for

these configurations: major states. In addition, there are other configurations that are occasionally

observed, known as intermediates.

While advances in a variety of techniques have seen the identification of increasing numbers of

unfolding intermediates, this is particularly the case in single molecule assays (FRET and Force

spectroscopy). By observing the folding and unfolding of single molecules, as opposed to an population

average folding observed in ensemble measurements, short-lived intermediates, normally obscured in

averaging, become detectable. These studies have identified the energy barriers between states and have

distinguished obligatory, nonobligatory, and off-pathway intermediates.

3.2) Improvements to the resolution demonstrated

3.2.1) Improvements in spatiotemporal resolution

The improvements to our results were immediately apparent: Most of the intermediates were

closely spaced and transiently populated, indicating that they would be difficult to detect using previous

cantilevers. Fig 3-2 depicts a force-time plot from a single bR record, and it has two remarkable features:

first, it demonstrates a dwell time as short as 8 μs (Fig 3-2B), a time commonly associated with the

transition path time between states (*92, 93*), rather than state occupancy time. Secondly, two of these

states (I_{ED}^4 & I_{ED}^5) are separated by only two amino acids, or half an α-helical turn. In pervious bR

literature, the smallest unfolding step observed was 6 aa (and did not comment on the state lifetimes, or

detection thresholds) (*67*), so resolving half a helical turn represents a significant advancement in

instrument precision.

3.2.2) Significant increases in the number of observed intermediates

In addition to resolving closely spaced states, the quantity of states that was resolved significantly

increased. For example, whereas prior studies over the past 16 years reported two nonobligate

intermediates when unfolding the ED helix pair (Fig 3-3A, top left inset) (*67*), we observed 14

intermediates (Fig 3-3A) denoted I_{ED}^1 to I_{ED}^{14}. Newly identified intermediates were detected throughout the

unfolding pathway. For the CB helix pair (Fig 3-3B), where prior studies reported two nonobligate

intermediates, (*67*) we observed 7 unfolding intermediates. Lastly, for helix A (Fig 3-3C), where previous

literature reports a single nonobligate intermediate, we observed 3 unfolding intermediates. All together,

we reported nearly a fivefold increase in the number of resolved intermediates as compared with the

consensus number of observed intermediates (*2-9*), with a seven-fold increase within an individual pair of

helices.
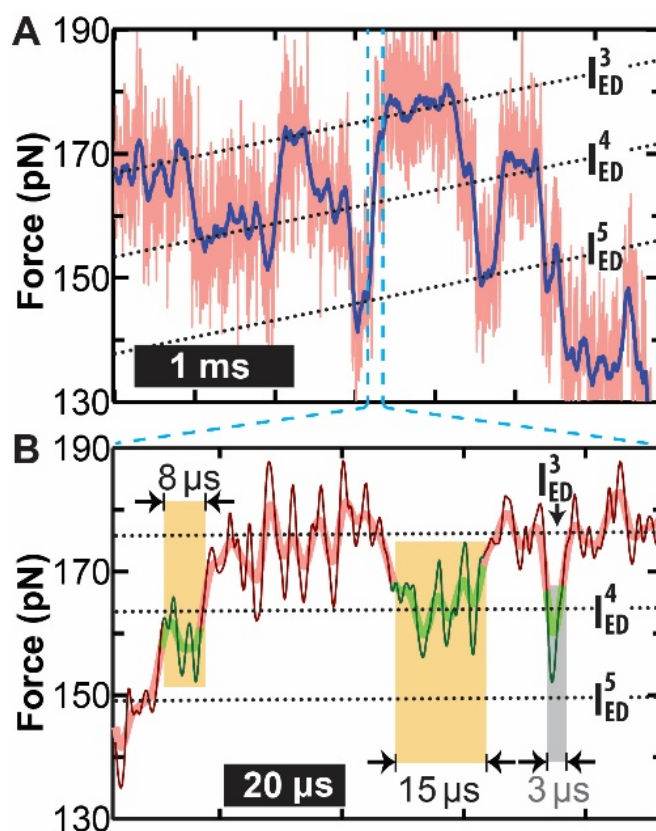
Figure 3-2: Improved spatiotemporal resolution details complex and rapid dynamics between close spaced states. **(A)** Force-versus-time trace shows rapid back-and-forth transitions between three states determined with hidden-Markov model analysis (black dotted lines) and correspond to I_{ED}^3, I_{ED}^4, and I_{ED}^5. Data were smoothed to 10 kHz (blue) and 200 kHz (pink), respectively. A highlighted portion of the trace (light blue) is shown in detail in (B). **(B)** High-resolution force-versus-time trace illustrates rapid dynamics between I_{ED}^3, I_{ED}^4 (green), and I_{ED}^5. Here, two state lifetimes of 15 and 8 µs are identified with a hidden-Markov-model analysis (orange highlights). A potentially even shorter state lifetime of 3 µs (gray) is seen but not identified as a state by the hidden-Markov-model analysis. Traces were smoothed to 100 kHz (light colors) and 830 kHz (dark colors). From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.
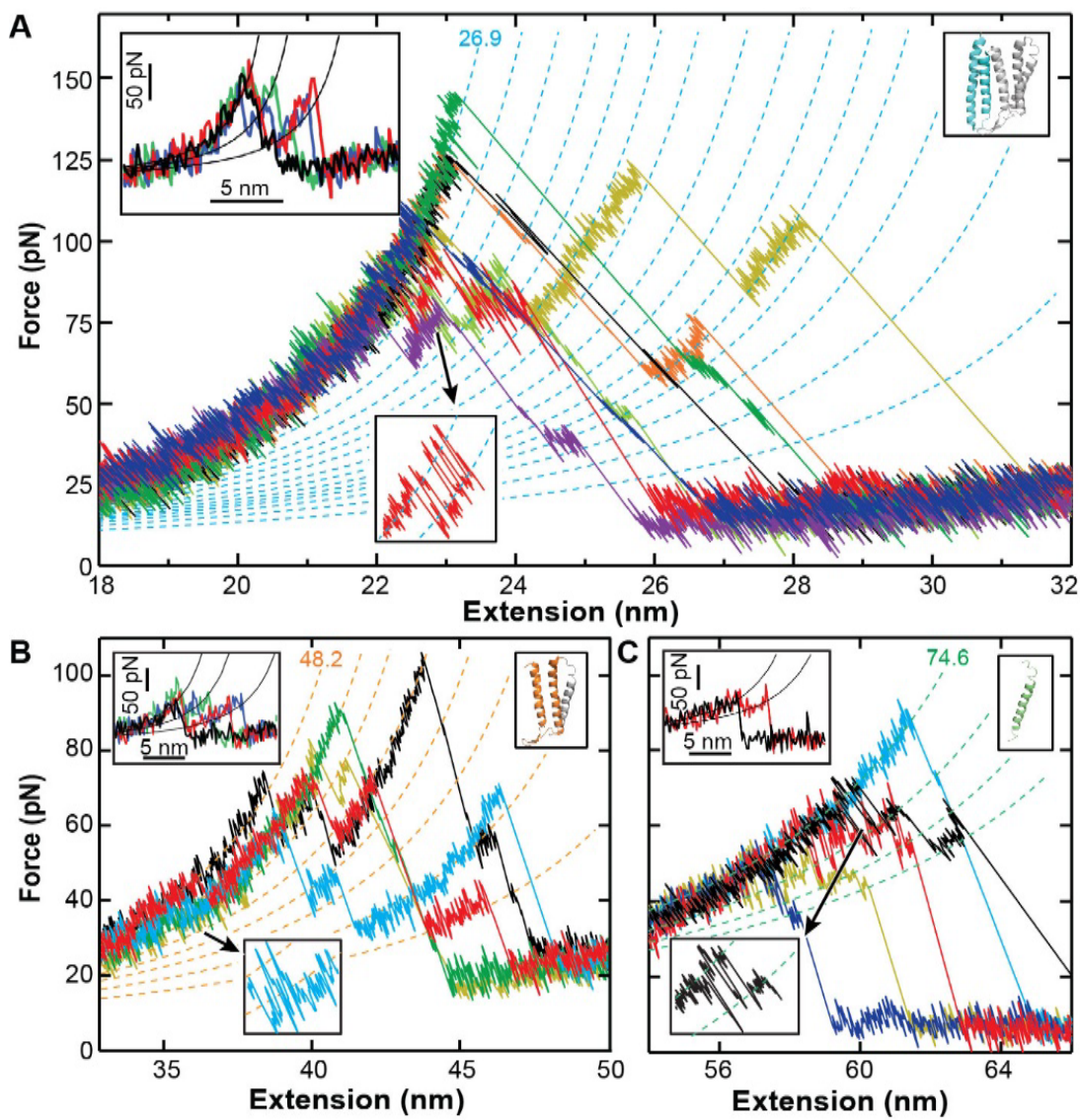
Figure 3-3: Significant increase in unfolding intermediates observed throughout bR unfolding. (**A**) Representative high-resolution FECs reveal 14 intermediates when unfolding the ED helix pair. In contrast, two intermediates were reported in prior studies (*67*) (upper left inset). This inset and the corresponding insets in panels E and F reprinted from ref (*64*) with permission from Elsevier. (**B**) FECs show 7 intermediates during the unfolding of the CB helix pair instead of 2 observed previously *(2-9)*. (**C**) FECs show 3 intermediates while unfolding helix A instead of 1 observed previously (*2-9*). Near-equilibrium fluctuations between multiple states were observed when stretching at 300 nm/s (lower inset, D to F; see Fig S5 for force-time curves). From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.

3.3) Intermediate location assignment

3.3.1) Amino acid assignment method

Once the intermediate contour lengths are determined, the next task was to assign them to specific residues. As is standard in the field (*17, 64, 67, 68, 70, 72–75*), we assumed the intact secondary structure remains stationary within the lipid bilayer as each unfolding event occurs (which we will refer to as the Folded Remains in Native Location model, or FRINL). Using this model, the number of amino acids unfolded (n_{aa}) was calculated based on: $n_{aa} = (\Delta L_0 + \Delta d)/L_0^{aa}$, where Δd is the vertical distance of the folded structure along the pulling axis in native bR (*49*) and $L_0^{aa} = 0.366$ nm is the contour length per amino acid (based on the distance between I_{ED}^0 to I_A^0). Note the sign of Δd depends on the pulling geometry; it is negative for helix E and positive for helix D. In other words, the measured change in contour length is smaller when unwinding a helix from the top (helix E) than from the bottom (helix D). This model assumes that that Δd alters the measured contour length, and not the measured horizontal offset of the worm-like chain (see section 2.5 for WLC explanation). Physically, this means that we are assuming that from the cantilever tip to the folded structure of the protein is one polypeptide chain with constant tension throughout. That indicates that there is no change in the environmental conditions that the protein chain is exposed to, *i.e.* that the void left by the unfolded protein is filled with water (*49*). This was chosen because this model empirically fits our data better than the alternative change in horizontal offset model (which implies the void is filled by lipid) (*49*).

3.3.2) Underlying assumptions in amino acid assignment method

As mentioned in section 3.1, one advantage of working with an integral membrane protein (such as bR) is that the coupling of the protein to the lipid environment dissipates forces applied to a given folded element, preventing unfolding of secondary structure elements later in the peptide. Consequently, the helices unfold in sequential order. Further, the energy change required to extract an α-helix intact is significantly greater than the energy change required to unfold an α-helix, which indicates an α-helix is significantly more likely to unwind prior to extraction (*42, 49*). This enables the use of the assumption that the protein goes through a discrete transition point that delineates the unfolded polypeptide chain

from the remaining folded protein. Thus, by measuring the length of the unfolded section (as outlined in

section 3.3.1), and subtracting that from the total length of the protein (248 aa, from literature), we can

determine the specific residue at the transition point. Inherent in this calculation is the assumption that the

folded part of the protein remains at its native location which might be a valid assumption for the

unfolding of the first few helices, but seems unlikely during the unfolding of the last helices in the

protein. Helix A, lacking stabilizing interhelical bonds from the now absent G though B helices, likely

reorients in response to the force being applied to it. Unfortunately, SMFS experiments lack information

that would enable a more sophisticated model, and consequentially, we will use FRINL assumption in all

the residue assignments in this thesis, as it has been in all SMFS experiments on integral membrane

proteins in the literature (*49*, *67*).

3.3.3) Amino acid assignment locations and comparison

Utilizing FRINL, along with the contour length changes between these curves, we assigned the

intermediates to specific residues within the structure protein, which is detailed both in the column

"Nearest residue number" in Table 3-1, and Fig 3-4. In additional, the frequency with which each state is

observed and the unfolding forces are also listed.

Table 3-1: Structural parameters of intermediates for C-terminal unfolding. The contour length change, structural position, probability, average unfolding force of each intermediate, and number of trajectories detected in each state are listed. The presented data were derived from 98, 129 and 159 molecules for the ED, CB and A helices, respectively.

State	ΔLc (nm) of previously identified intermediates	ΔLc (nm) ± s.d. from I_{ED}^0	Nearest residue number	Occupancy probability (%)	Mean unfolding force (pN) ± s.e.m.	Number of trajectories occupying each state	Description
I_{ED}^0	0	0 ± 0.1	160	100	94 ± 1	98	Top of helix E
I_{ED}^1		0.7 ± 0.1	157	86	87.3 ± 0.8	84	
I_{ED}^2		1.2 ± 0.1	154	83	89 ± 1	81	
I_{ED}^3	2.16	1.9 ± 0.1	151	81	85 ± 1	79	
I_{ED}^4		2.4 ± 0.1	148	68	85 ± 1	67	
I_{ED}^5		3.0 ± 0.1	146	58	80 ± 1	57	
I_{ED}^6		3.6 ± 0.2	143	38	74 ± 2	37	
I_{ED}^7		4.4 ± 0.2	139	40	71 ± 2	39	
I_{ED}^8		5.1 ± 0.2	136	46	68 ± 2	45	
I_{ED}^9		5.8 ± 0.1	132	54	63 ± 1	53	
I_{ED}^{10}	6.12	6.4 ± 0.1	130	50	58 ± 1	49	Bottom of helix E
I_{ED}^{11}		7.1 ± 0.2	129	45	51.4 ± 0.9	44	
I_{ED}^{12}		8.0 ± 0.2	127	48	46.3 ± 0.9	47	Top of helix D
I_{ED}^{13}		9.4 ± 0.3	124	37	42 ± 1	36	
I_{ED}^{14}		11.9 ± 0.4	119	23	31 ± 2	23	
I_{CB}^0	21.6	21.3 ± 0.3	101	47	49 ± 2	61	Top of helix C
I_{CB}^1		22.7 ± 0.1	96	89	58 ± 1	115	
I_{CB}^2		23.7 ± 0.1	91	66	60 ± 2	85	
I_{CB}^3	25.2	25.2 ± 0.1	83	81	55 ± 2	104	Bottom of helix C
I_{CB}^4		27.0 ± 0.1	77	72	50 ± 2	93	
I_{CB}^5		28.9 ± 0.1	71	56	46 ± 2	72	
I_{CB}^6	31.32	31.9 ± 0.2	63	90	43 ± 2	116	Top of helix B
I_{CB}^7		36.4 ± 0.3	54	62	31 ± 1	80	
I_A^0	47.16	47.7 ± 0.1	29	100	62.0 ± 0.6	159	Top of helix A
I_A^1		50.3 ± 0.1	16	90	60 ± 1	143	
I_A^2	51.84	52.4 ± 0.1	8	73	57 ± 1	116	Bottom of helix A
I_A^3		54.5 ± 0.2	2	27	50 ± 2	43	
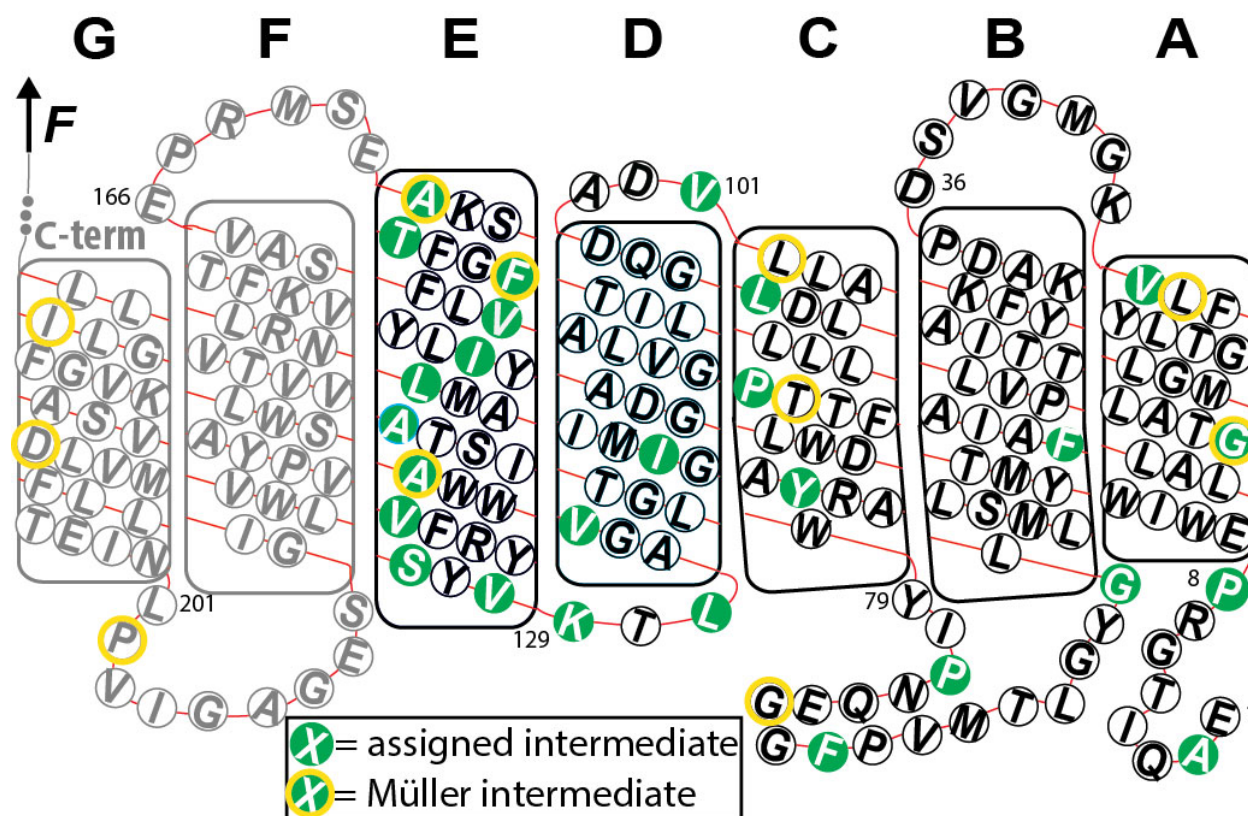
Figure 3-4: Residue assignment from C-terminal experiment. A cartoon depicting the primary (labeled residues) and secondary structure (helices) of bR. Circles filled in with green correspond to the residues mapped as intermediates from the unfolded intermediates. The gold circles around the residues represent the intermediates observed in the previous unfolding literature (*67*). A few residue numbers are displayed to aid interpretation.

With assignment uncertainties of 1 aa, this assignment (especially helix E) represents the highest density of folding intermediates in the unfolding of a transmembrane protein in the literature. Comparing the location of these unfolding intermediates to those identified within the Müller literature (*67*), we can see robust agreement, with only one residue (G73, located in the CB loop), falling outside the uncertainty of our assigned intermediates. It should be noted that the uncertainty of the Müller literature intermediates is unlisted in the literature, but given the consistency of their assignment over the past 18 years, over the course of multiple experiments and hundreds of unfolding curves, one can assume from the statistics scale that the uncertainty is small.

A reasonable concern about such a dense and previously unobserved state assignment is that we conflated noise with transitions, and split what are really single unfolding intermediates. In section 3.10

we will go into the methods we used to segment, assign, and verify the states. A qualitative demonstration of the distinctness of the intermediates is presented in Fig 3-5, which depicts transitions between every intermediate and its proximal intermediates within helices ED.

3.3.4) Discussion of relative densities of intermediates in each helix

By analyzing the distribution of locations within the secondary structure of the helices, one may note that there is a very high concentration of intermediates within the E helix, especially relative to the other 4 helices that we analyzed. This is likely the result of a few different experimental conditions, rather than a reflection of structural features of the protein. One likely reason stems from the relationship between the pulling direction and the protein orientation: three helices (E, C and A) will unwind antiparallel the pulling direction (downward), and two (D and B) will unwind parallel to the pulling direction. Unfolding steps within the downward unwinding helices lower the vertical location of the transition residue, making it necessary for some of the unfolded peptide to span this gap, reducing the increase in contour length. The opposite occurs for upward winding helices: unfolding steps raise the vertical position of the transition residue, making the change in the contour length greater. This difference means that the refolding barriers between states within downward winding helices (relative to the refolding barriers within upward unwinding helices) are larger, due to the larger entropic costs for the transition. This ultimately should not impact the probability of an unfolding event occurring, but can impact the probability of an unfolding event being detected, as refolding is likely common for all transitions. By decreasing the chance of a refolding event occurring, unfolding events are more likely to be detected. This means that intermediates within downward winding helices have longer lifetimes, and are more detectable compared to intermediates within upward winding helices.
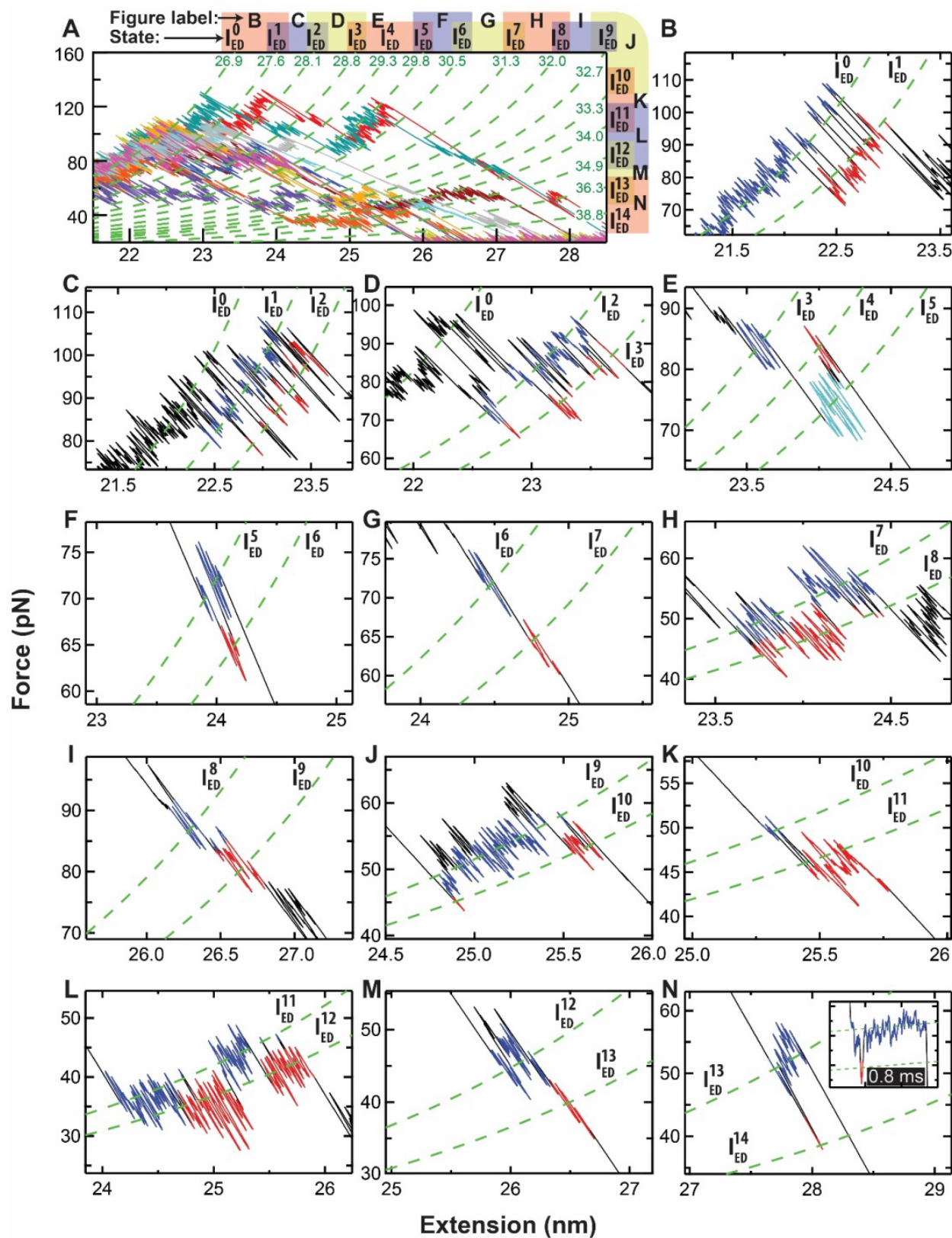
Figure 3-5: Demonstration of state resolution via proximal transitions. (A) The FECs shown in the subsequent panels illustrate resolution of closely spaced states. Green lines indicate WLC models of the

Figure 3-5 (cont.): with contour length (in nm) listed for each intermediate. (Top and right) Labels showing relevant panel that demonstrates selected intermediates, overlaid on top and right. These FECs link adjacent states within the unfolding pathway of the ED helix pair, with **(B)** I_{ED}^0 and I_{ED}^1; **(C)** I_{ED}^1 and I_{ED}^2; **(D)** I_{ED}^2 and I_{ED}^3; **(E)** I_{ED}^3, I_{ED}^4, and I_{ED}^5; **(F)** I_{ED}^5 and I_{ED}^6; **(G)** I_{ED}^6 and I_{ED}^7; **(H)** I_{ED}^7 and I_{ED}^8; **(I)** I_{ED}^8 and I_{ED}^9; **(J)** I_{ED}^9 and I_{ED}^{10}; **(K)** I_{ED}^{10} and I_{ED}^{11}; **(L)** I_{ED}^{11} and I_{ED}^{12}; **(M)** I_{ED}^{12} and I_{ED}^{13}; and **(N)** I_{ED}^{13} and I_{ED}^{14}. The portion of the trace shaded blue in each panel represents the more folded state, and the red trace represents the more unfolded state (except in **E**, where cyan represents the most unfolded state), with black indicating regions of the trace not assigned to the states of interest. Records smoothed to 10 kHz using a Savitzky-Golay filter. **(N,** Inset) Force-time plot of molecule record in force-extension plot filtered at 40 kHz. The dwell in I_{ED}^{14} is ~40 μs, well within the ~1-μs resolution of this assay. From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.

The unfolded polypeptide linker also plays into another probable cause of the lack of intermediate in later helices (C and A) compared to earlier helices (E). This results from the increase in the length of the unfolded polypeptide linker as additional helices are unfolded. By effectively making this linker longer, the compliance of the linker increases, and as a result the magnitude of the thermal fluctuations grows larger (Fig 3-6). As a result, the force noise, even at equivalent forces are greater for later helices than they are for earlier helices, making state detection of hypothetically closely spaced states more difficult. In additional, additional compliance of the linker corresponds to a reduction in the temporal resolution of the measurement system, making short lived stated harder to detect. Thus, a similar density of unfolding intermediates in the true unfolding behavior would measure as fewer intermediates. The impact of this additional noise is also paired with the diminishing likelihood that states even exist, as one of the theorized causes of unfolding intermediates, interhelical hydrogen bonds, rely on the location of both pair residues in their native locations. Near the end of the unfolding process, when most of the helices have been unfolded and extracted from the membrane, this is unlikely to still be valid.

3.4) Comparisons to computational simulations

The enhanced precision achieved in these records helps to resolve a long-standing discrepancy between theoretical and experimental studies of bR unfolding, based on steered MD simulations and SMFS measurements, respectively. Specifically, prior experimental studies supplied evidence for one or two intermediate unfolding states per helix pair (~50 amino acids), whereas some MD simulations predicted a far denser series of unfolding intermediates, occurring once every approximately two to eight
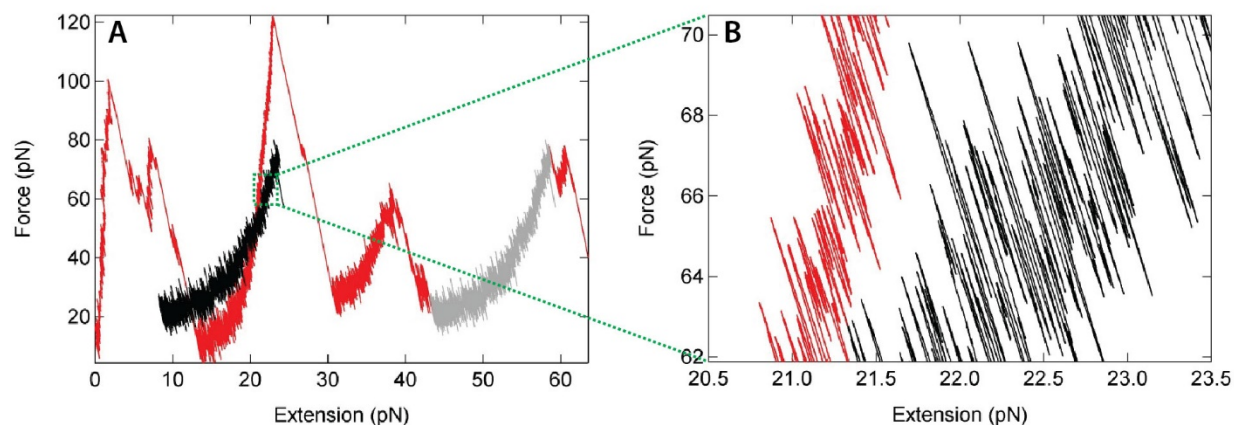
Figure 3-6: Comparison of noise levels for ED major state and A major state. (A) The force-extension curve of a single molecule, a section from the A major state (fourth peak, black) has been shifted left for the purposes of comparing its noise level to the ED major state's (second peak) noise level. **(B)** A zoom of (A), showing the relative noise levels of the two major states. As can clearly be observed, the A major state, even at similar forces, are significantly noisier than ED major state.

amino acids (*79*), while others predicted densities similar to the Müller intermediates (*76*). It should be noted this discrepancy is due in part to the different modeling techniques, where the former is an all-atom simulation, and the latter is coarse grained.

3.4.1) Comparison to all-atom simulation (*79*)

The all-atom simulation published by Kappel and Grubmüller agrees with the closely spaced array of intermediates we detect. In fact, ~60% of the unfolding intermediate states predicted with MD simulations were observed in these experiments. Expressed the other way around, ~55% of the intermediate states observed in our experiments were predicted with MD simulations (Fig 3-7, purple circles). For helix E, in which we managed the highest resolution, a supermajority (80%) of the intermediates predicted with MD could be identified in our experimental data. Given the detail and density of unfolding intermediates, it is equally notable that SMFS detected intermediates that were otherwise absent from MD simulations, over several comparatively large regions (more than five amino acids) (Fig 3-7). We attribute this difference to the seven orders-of-magnitude greater pulling rates used in MD simulations, as compared with those of actual experiments (v = 1 m/s, versus 300 nm/s, respectively). Under the circumstances, the excellent correspondence between MD and SMFS provides increased
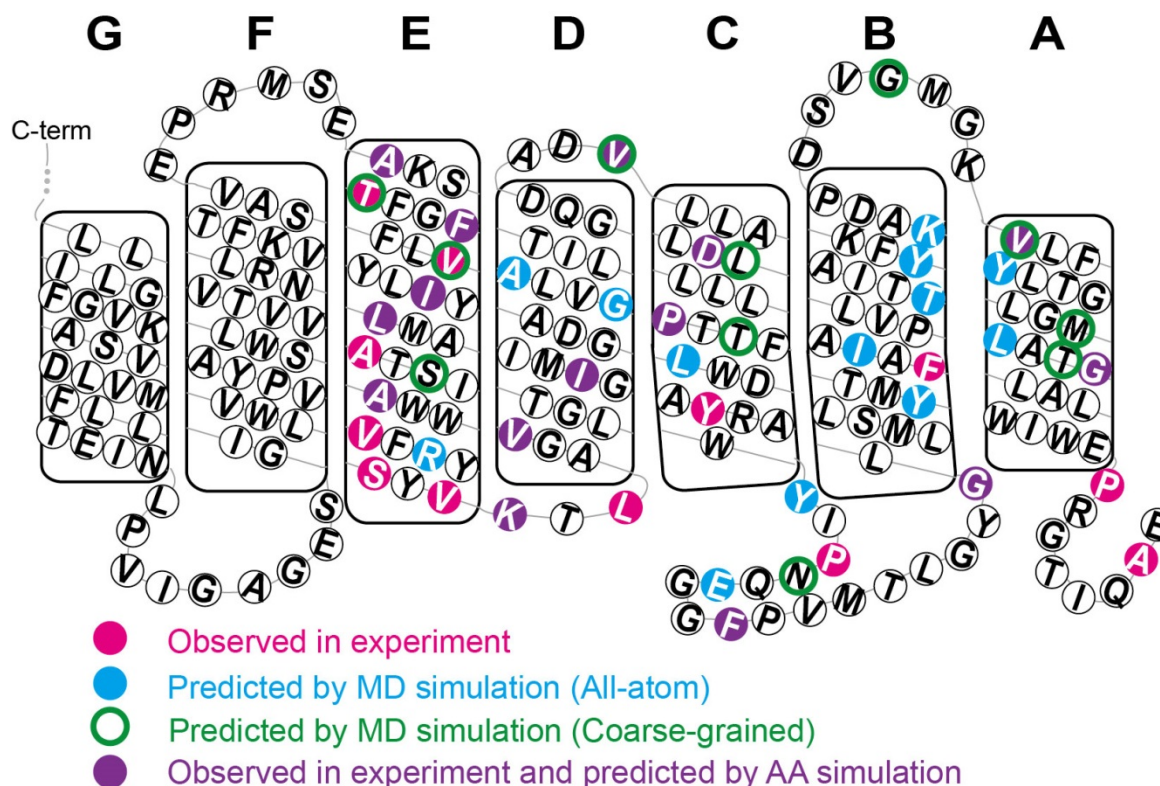
Figure 3-7: Intermediate assignment from experiment vs. MD simulations. A cartoon depicting the primary (labeled residues) and secondary structure (helices) of bR. Circles filled in with pink correspond to the residues mapped as intermediates from experiments, circles filled in with blue correspond to predicted transition residues by the all-atom simulations (*79*), and circles filled in with purple correspond to transition residues predicted by the simulations, and mapped from experiment. The green rings around residues correspond to transition points predicted by the coarse-grained simulation (*76*). Adapted from Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.

confidence for using theoretical simulations to explore the unfolding and energetics of membrane

proteins.

3.4.2) Comparison to coarse-grained simulation (*76*)

Yamada and co-authors published a coarse-grained model excluding interhelical bond

interactions in 2016, which only predicts 2 unfolding intermediates in helices ED, 3 intermediates for

helices CB, and 3 in helix A (Fig 3-7, green circles). In terms of location, the 2 of the intermediates in

helices ED agree well with our mapping, while one falls outside of our uncertainty. It should be noted that

the intermediate that does not fall within the agreement range arises due to breaking of the FRINL

assumptions, as it shows that the folded section of helix E and helix D moves significantly from its native

location (Fig 3-8). For helices CB, the intermediates are proportionately less in agreement, with only 5 of
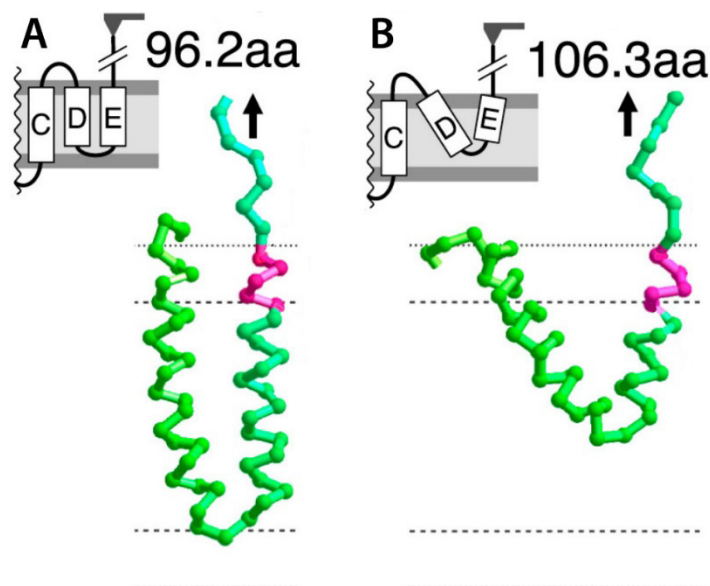
Figure 3-8: **Non-FRINL unfolding behavior**, as simulated in (*76*). Notice that relative location of helix D at intermediate 97.2 aa (left), which is roughly perpendicular to the membrane plane, the position adopted at intermediate 107.3 aa (mapped to an intermediate about at half unfolded helix E), shows helix D significantly shifted. This simulation did not account for the interhelical bonds, which would serve to strengthen FRINL behavior, but this serves as a good demonstration of how the helices could remain in their secondary structure, and remain within the membrane, but significantly diverge from FRINL behavior. The 1 aa difference between numbers and figure arises from fact Yamada *et al.* used a 247 aa model of bR for simulations, as opposed to the 248 aa in native bR. Reprinted from Biophysical Journal, **111**, T. Yamada, T. Yamato, S. Mitaku, "Forced Unfolding Mechanism of Bacteriorhodopsin as Revealed by Coarse-Grained Molecular Dynamics", 2086-2098, Copyright 2016, with permission from Elsevier

the 9 falling within our experimental measurements uncertainty. Lastly, for helix A, of the 4 detected

intermediates, only one falls within the uncertainty of our measured states. Of the 4 states predicted by

Yamada in helices CB, 3 fall within the experimental uncertainty of our measurements. In this

circumstance, the disagreeing intermediate conforms rather well to FRINL model. Lastly, helix A has the

weakest agreement with our experimental measurements: only 2 of the 4 predicted states fall within our

experimental uncertainty limits. This is understandable from the perspective of the breakdown of FRINL

with the loss of interhelical bonds, and the lack of steric restrictions from other folded helices. It should

be noted, half of the intermediates that were not matched within our uncertainty range occurred at

significantly lower forces, and more infrequently (<30%), than the other intermediates observed,

explaining some of the disagreement.

3.5) Continuous transitions vs. discrete transitions

One possible indication of helix reorientation is a phenomenon we are naming continuous transitions. Occurring in ~13% of our unfolding records, it corresponds to divergence from stepwise WLC behavior, with neither the force increasing in accordance with a single WLC curve, nor decreasing in discrete transitions. Rather the unfolding force increases at a less steep rate, which when plotted against WLC curves makes it appear that it is unfolding. Non-discrete folding behavior has been observed before in a globular protein (*94*), with a similar variability in the force, arising from hydrophobically driven aggregation of the newly unfolded α-helices. This explanation is unlikely here, as if the cantilever is exerting a force sufficient to unfold the native (and highly stable) α-helix (~100 pN in Fig 3-9A), it seems unlikely that a hydrophobic aggregation would persist for any significant length of time. Another explanation is the observance of non-FRINL behavior, but the locations where this continuous transition behavior are observed (near the top of helices), are the regions where the contour length would be minimally impacted by helical rearrangement within the membrane: at or above the membrane interface. More study is necessary to explain these events.
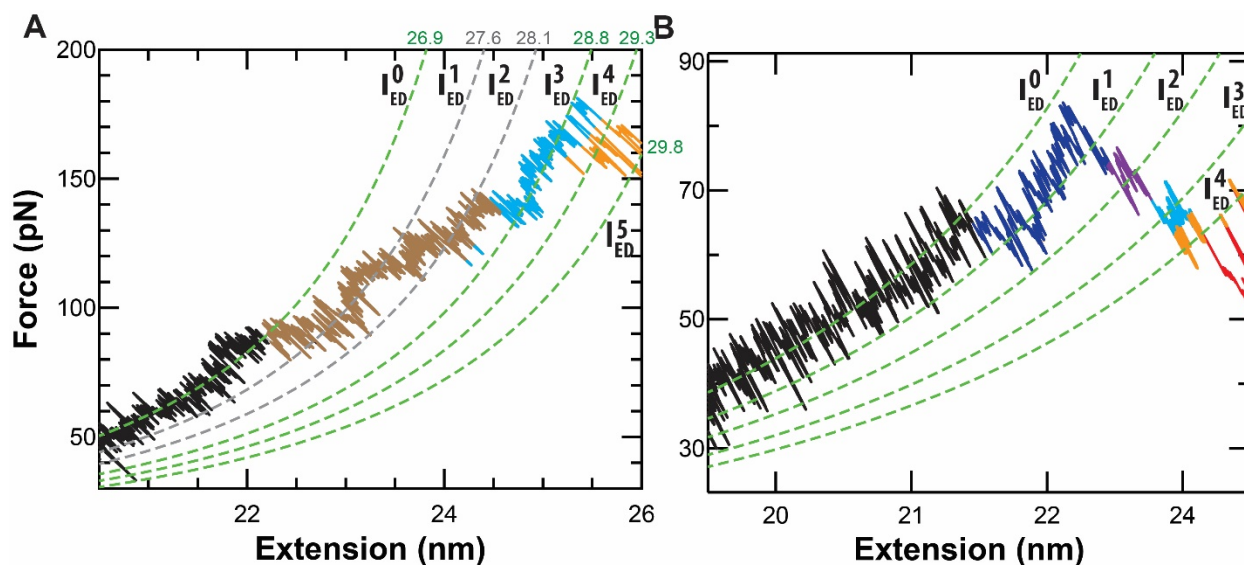
Figure 3-9: Continuous transitions. (A) A FEC shows an unfolding trajectory with a continuous transition where the unfolding trajectory (brown) does not strictly follow a set of WLC curves (dashed grey) between transitions. State occupancy (*i.e.*, dwells) in I_{ED}^0, I_{ED}^3, and I_{ED}^4 are indicated by color (black, cyan, and orange, respectively). At least one instance of a continuous transition occurred in ~13% of molecules within the ED helix pair. **(B)** A typical FEC curve exhibits a staircase-like set of non-continuous transitions where the molecule unfolds through I_{ED}^0 (black), I_{ED}^1 (blue), I_{ED}^2 (purple), I_{ED}^3 (cyan), and I_{ED}^4 (orange). From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.

3.6) Modeling hydrophobic contributions to the energy landscape

One concept explored by Yamada *et al.* is the hydrophobic contribution to the formation/stabilization of unfolding intermediates. Within the paper, they proposed a model to account for the interfacial hydrophobic contributions to the free energy of the molecule. We replicate part of their free energy model with our data, substituting their atomic knowledge of the location of the residues for an approximation based on the relative extension of the unfolding curves and the FRINL model (Fig 3-10). Using their hydrophobicity model for the lipid interface, as well as the augmented Wimley-White hydrophobicity scale (*95*), Fig 3-11 was created to depict the change in the free energy for unfolding each individual residue. The model does answer some open questions in our detected unfolding intermediates: specifically, the mechanism behind intermediates assigned to residues within the
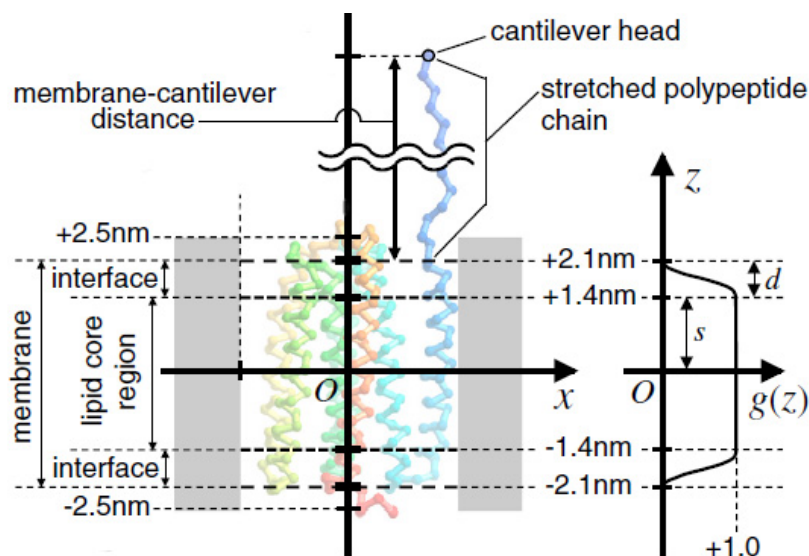
Figure 3-10: Membrane hydrophobicity model from Yamada *et al.*. (Left) A model of the membrane with bR embedded. In this model, the membrane is modeled as being 4.2 nm in height, with interfaces of 0.7 nm on both the cytoplasmic and extracellular sides. **(Right)** A model of the strength of hydrophobic interactions, g(z), as a function of height (z), with maximum strength between – 1.4 nm and +1.4 nm, and diminishing strength within the membrane interfaces (-2.1 nm to -1.4 nm, and 1.4 nm to 2.1 nm). Beyond -2.1 nm and 2.1 nm, the model indicates that there is no hydrophobic interaction with the membrane. We utilized this model of the hydrophobicity, along with the Wimley-White hydrophobicity scale for side chains (*21)*, to estimate the change in the interfacial hydrophobic contribution to the free energy associated with unfolding and extracting every residue. Reprinted from Biophysical Journal, **111**, T. Yamada, T. Yamato, S. Mitaku, "Forced Unfolding Mechanism of Bacteriorhodopsin as Revealed by Coarse-Grained Molecular Dynamics", 2086-2098, Copyright 2016, with permission from Elsevier
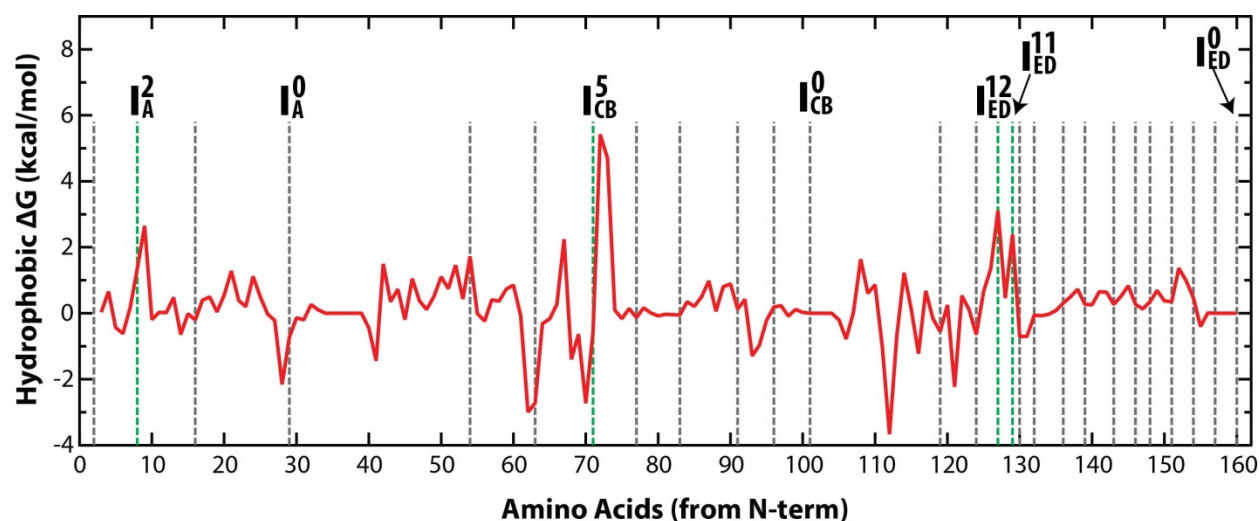
Figure 3-11: Hydrophobic energy change to unfold each residue. This plot depicts the change in the hydrophobic contributions to the free energy associated with unfolding single amino acids in bR. The dashed lines are the locations of the measured transition residues from the experiments, with major states and discussed intermediates labeled. Note the proximity of large positive peaks in the hydrophobic ΔG (blue dashed lines), especially outside of the ED helical region.

extrahelical loops. Loops are regions in interhelical bonds are unlikely to form. For example, the model predicts a large hydrophobic energy costs (>2 kcal/mol/aa) associated with pulling I140, W138, and W137 through the cytoplasmic membrane interface. Using the FRINL model, this would map in our assignment to the intermediates I_{ED}^{11} and I_{ED}^{12}, which are located in the ED loop. Similarly, the model predicts a large hydrophobic energy costs to bringing F88, L87 and W87 through the cytoplasmic membrane interface, which in our FRINL model maps to the intermediate I_{CB}^{5}, located within the CB loop. Lastly, it helps explain the intermediates that we mapped to the unstructured N-terminal tail (I_{A}^{2}), as this intermediate exhibits high hydrophilic energetic costs of E8 resisting transitioning across the extracellular interface.

3.7) Unfolding pathway analysis

3.7.1) Complex unfolding pathways observed

One of the most surprising results observed in our experiments was the high frequency of refolding (especially near the top of some of our helices, see section 3.7.4. Refolding has only been observed in globular proteins at a constant velocity (on only two occasions) while pulling at 30 nm/s or slower (*96*, *97*). This was the first time that it has been observed for a membrane protein at all, and for any protein while being unfolded at a pulling speed of 300 nm/s.

Widespread refolding, together with the large number of resolved intermediates, led to highly varied trajectories along a multistate unfolding pathway (Fig 3-3). Displayed in Fig 3-12 is the unfolding (Fig 3-12, orange lines) and refolding (Fig 3-12, purple line) transitions observed during the unfolding of the ED helix pair. A similar figure, for the unfolding and refolding transitions of helices CB and helix A are displayed in Fig 3-13. A heat map, combined with the actual transition count for all Helices ED, CD and A is displayed in Fig 3-14.
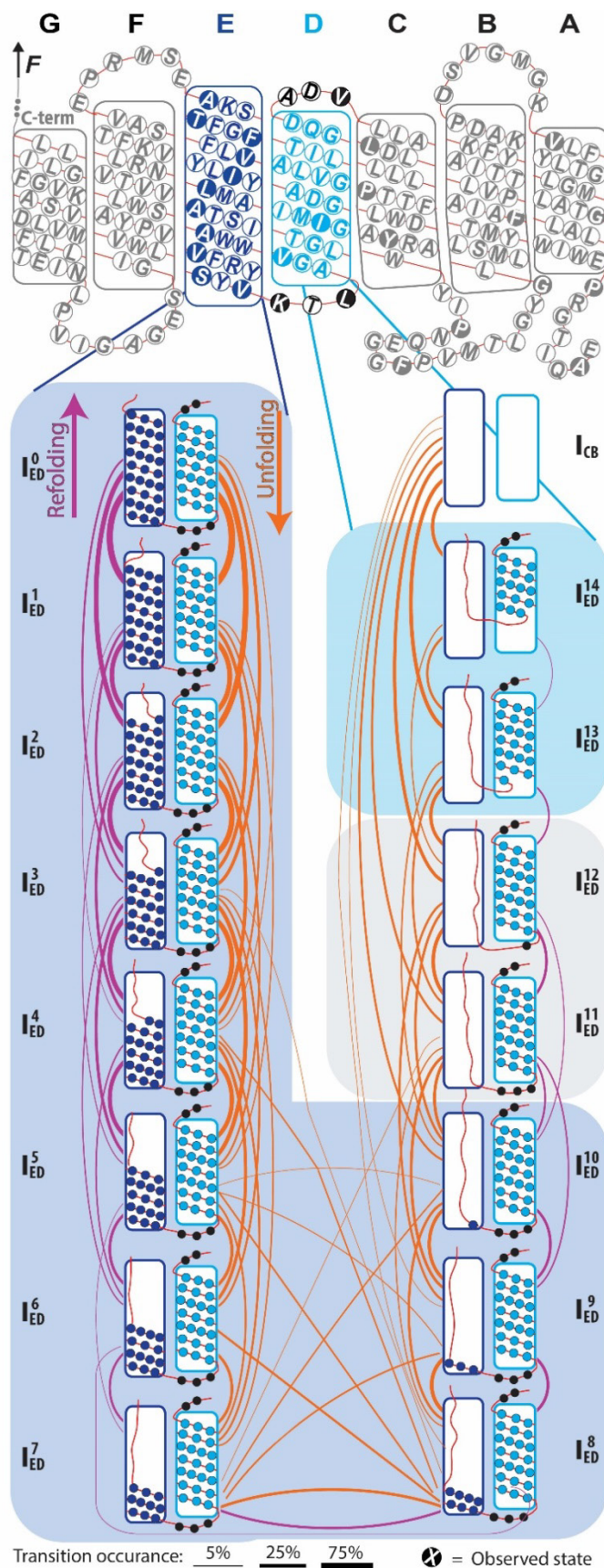
Figure 3-12: Unfolding pathway for the ED helix pair. (Top) Cartoon of the primary and secondary structure of bR. Locations of observed folding intermediates are shown by residues with filled-in circles. **(Bottom)** Each helix pair diagram depicts an observed intermediate state, with connecting lines representing transitions observed in at least 4 (of 98) different molecules containing a total of 1,399 transitions. Orange lines represent unfolding transitions, while purple lines show refolding transitions. Line-widths represent the frequency of observing a particular transition. An alternative matrix representation shown in Fig 3-14. From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.
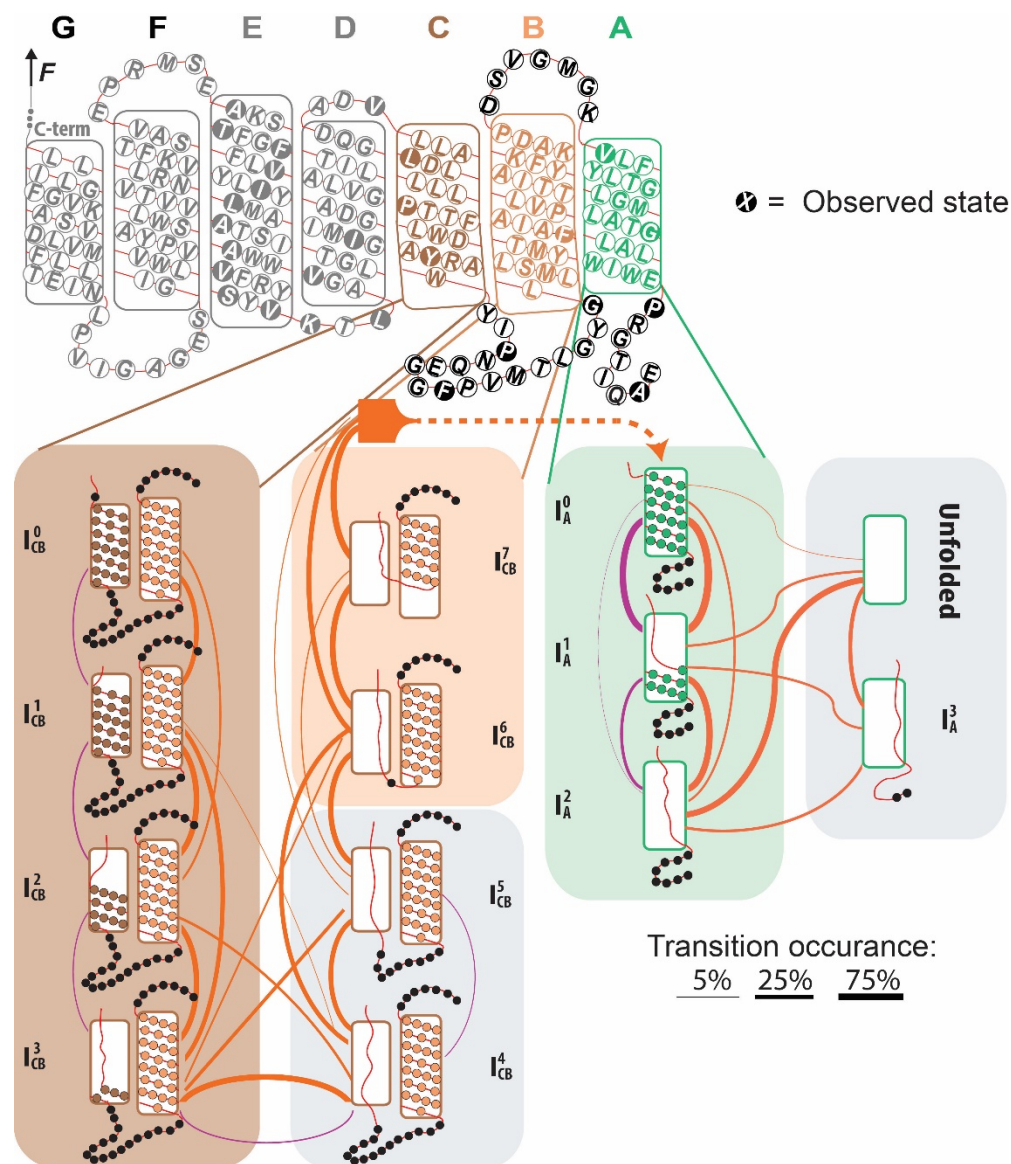
Figure 3-13: The unfolding pathway for the CB helix pair and the A helix. (Top) Cartoon of the primary and secondary structure of bR. The 7 helices (A–G) are shown with their constituent residues, including locations of detected unfolding intermediates (residues with a filled-in circle). Helix C, B, and A are highlighted in brown, light orange, and green, respectively. **(Left)** Unfolding pathway diagram for the CB helix pair deduced from 743 transitions in 129 molecules. Each helix bundle diagram represents an observed intermediate, with connecting lines representing transitions (observed in more than 3 separate unfolding traces). We note that with our improved resolution, I_{CB}^0 is no longer observed to be an obligate intermediate (Table 3-1); rather, I_{CB}^1 is occupied at a higher percentage at $v > 300$ nm/s. **(Right)** Unfolding pathway for the A helix deduced from 599 transitions in 159 molecules. The last state labeled "Unfolded" represents the removal of the bR molecule from the membrane. The orange interior lines represent unfolding transitions, while the purple lines represent refolding traces. Line thickness represents the transition occurrence out of all molecules studied, so the percentage out of any given state is not 100%. For simplicity, repeat transitions are not included in this representation. See Fig 3-13 for the exact frequencies of these transitions. From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.
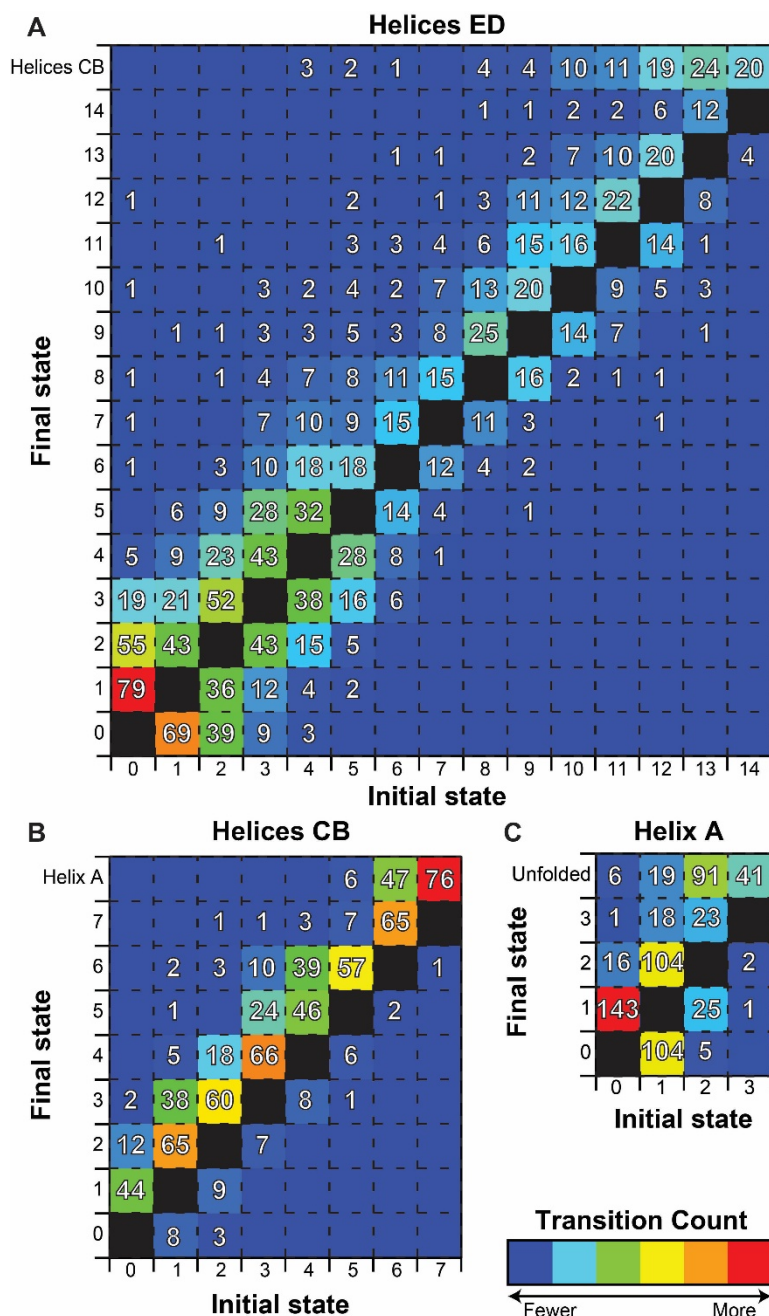
Figure 3-14: Transition heat maps. Each table has columns denoted by the initial state and rows by the final state. Within each column, the number of molecules observed to end up in a state from the specified starting state (black colored cell) is listed. Note, since molecules can refold, the number of transitions in each vertical column do not add up to the number of molecules studied. Multiple back-and-forth transitions between the same two states within a single record (*e.g.*, Fig 3-18) were only counted once in this analysis. (**A**) Transition heat map for the ED helix pair unfolding determined from 98 molecules with a total of 1,399 transitions. (**B**) Transition heat map for the CB helix pair determined from 129 molecules with a total of 743 transitions. (**C**) Transition heat map for helix A determined from 159 molecules with a total of 599 transitions. From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.

The sheer complexity of the data set obtained here prevents broad generalizations, but a few trends are noticeable. Refolding transitions tend be relatively common in the top half of helix E and helix A, with refolding transitions (especially larger than 1 residue) rarely observed elsewhere (though still occurring through the E, D, C, and A helices). Proximal and two stage (e.g. $I^3_{CB} \rightarrow I^5_{CB}$) unfolding transitions dominate the unfolding of helices CB, and helix A, whereas long range (> 2 intermediate) unfolding steps are quite common in helix D, especially to the completely unfolded step. A quantitative comparison would require a significantly larger number of unfolding records, which is beyond the scope of this thesis.

3.7.2) Missed intermediates versus parallel pathways

With the discovery of a diverse array of different pathways the protein can take between the folded and the unfolded states, questions arise as to their cause. Two possibilities seem likely: either the protein actually proceeds through all the intermediates every time, but some are not detected, either because they are too short lived to be detected, or the activation barrier does not exist at the force that that section of the energy landscape was explored (known as the missed intermediates theory), or the intermediates differ between the unfolding records because different areas of the energy landscape are being explored (known as the parallel pathway theory). These are mutually exclusive, and deciding which is correct can speak to the ability of these experiments map out the multidimensional energy landscape.

Unfortunately, to definitively determine the answer to this question, the hundreds of unfolding (and if possible, refolding) transitions need to be observed for each unfolding pathway (so different records for $I^0_{ED} \rightarrow I^2_{ED}$ and $I^0_{ED} \rightarrow I^1_{ED} \rightarrow I^2_{ED}$), and used to the reconstruct the energy landscape underlying each transition. This will take orders of magnitude more data that was is reasonable on a thesis time scale. Other less data intensive means of analysis are possible to suggest which model is correct.

The first means is to test for correlations between rupture force and the next intermediate occupied, for transitions from the ED major state. If strong correlations do exist (especially longer transitions at higher forces, and shorter transitions at lower forces), that is suggestive that differences in unfolding pathways can be explained by the stochastic variation in the major state rupture force as

opposed to by parallel pathways. A histogram of the bottom and top quintiles by force (< 75.8 pN and $>$ 107.9 pN, respectively) was constructed base on the next observed intermediate (Fig 3-15). High-force transitions tend to unfold farther (I_{ED}^2 on average) than low-force transitions (I_{ED}^1, on average). Additionally, the top quintile had rare long force transitions (to I_{ED}^8 and I_{ED}^{12}) that were completely absent from all other quintiles. This is weakly suggestive of the missed intermediate model.

A second means of investigating the parallel pathway versus missed intermediates question is to look for any correlation between rupture forces, and the number of transitions observed during the subsequent unfolding of the ED helices. A strong anti-correlation would imply that certain transitions only occur in a specific force range, and thus certain intermediates are unlikely to be observed at high force, evidence that would be support the missed intermediates theory. Here, the two relate via an exponential decay function, which is (to first-order) the theoretical relationship between state lifetimes (and the detectability of the intermediates) and function of force (Fig 3-16). This relationship is a moderate fit, with r^2 of approximately 0.3. This is not a strong enough correlation to be definitive evidence against a parallel pathway conception of the unfolding, but is mildly suggestive of the missed intermediates model.

The last means of investigating the missed intermediate versus parallel pathways question is to look if any correlations exist within the unfolding pathways. For example if I_{ED}^3 tends occur in the same records as I_{ED}^7 significantly higher than the population averages (while controlling for force), that would imply that they exist along a pathway that contains both those intermediates. We did such analysis (albeit without the force controls, but the force ranges were relatively small for most transitions), were we looked at the occurrence rate of a second state in the set of unfolding records where a first state appears, and subtracted the population occurrence rate (Table 3-2), and no significant correlations existed, yielding a null result.
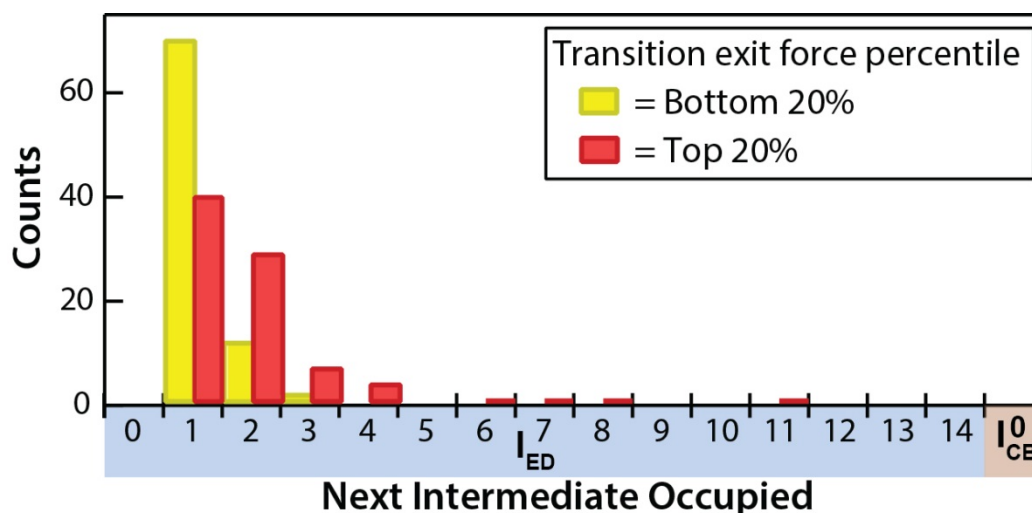
Figure 3-15: Unfolding pathway dependence on the rupture force. A histogram depicts the frequency of the first state occupied after leaving the major state I_{ED}^0, which corresponds to pulling on the top of the E helix. As expected, rupture force are broadly distributed. In this analysis, the transitions in the top quintile (red) and bottom quintile (yellow) of rupture forces observed in transitions out of I_{ED}^0 are histogrammed. On average, high-force transitions tend to unfold farther than low-force transitions.
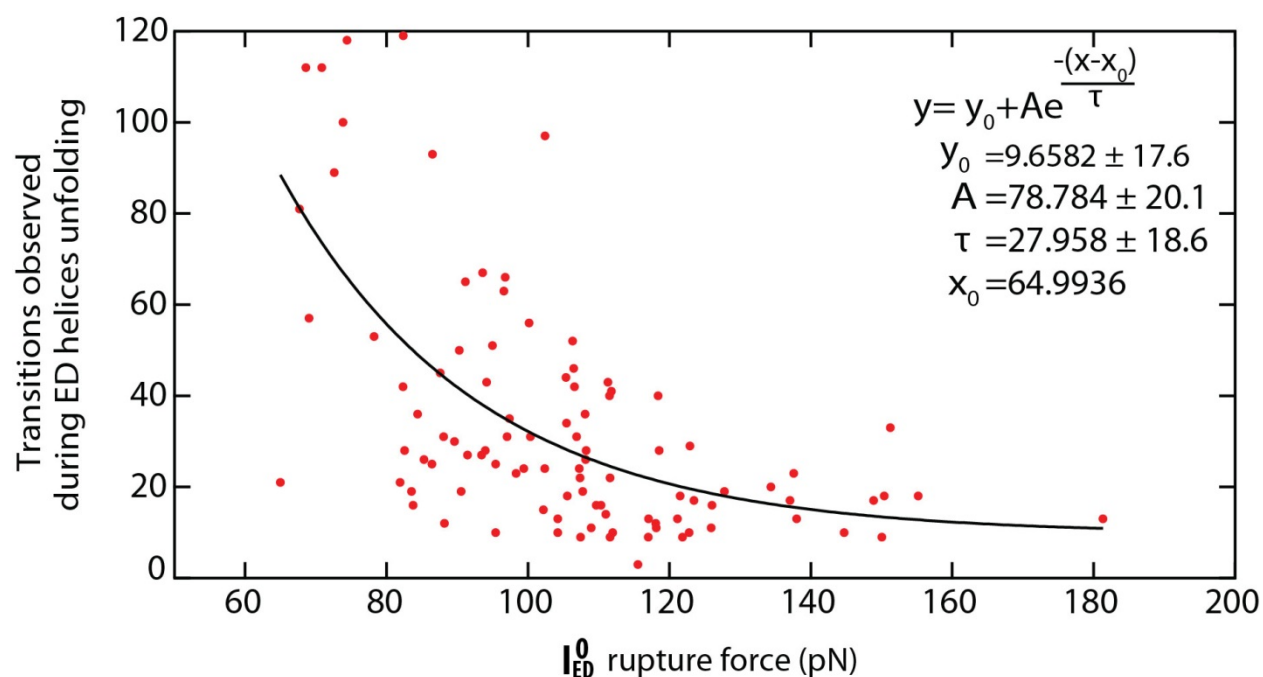
Figure 3-16: Correlations between transitions in an unfolding pathway and the major state rupture force. Plot comparing the number of transitions observed in the unfolding pathway compared to the maximum rupture force observed for the major state (I_{ED}^0). An exponential decay (with offsets) is fit to the equation, similar to an inverse Arrhenius equation. The relationship is moderately strong ($r^2 \sim 0.3$), suggesting that the number of intermediates is dictated by the rupture force of the first intermediate.

Table 3-2: Correlation table between of occurrence rates for ED unfolding intermediates compared to population occurrence rate. The table should be read as following: For every unfolding record that occupies intermediate I_{ED}^X (where X is the row number) during unfolding, it also occupies intermediate I_{ED}^Y (where Y is the column number) z% different than the population occurrence rate (where z is the number listed in the table). For example, "For every unfolding record that occupies intermediate I_{ED}^1 during unfolding, it also occupies intermediate I_{ED}^6 8.4% less than the population occurrence rate". If the intermediate of the row and column were randomly occurring, then this number should be close to 0. Note that the larger divergences from population average near the end of the unfolding I_{ED}^{13} and I_{ED}^{14} are partially a reflection of the low statistics for dwells in those intermediates relative to the earlier intermediates. No strong correlations exist in this table.

	…it also occupies I_{ED}^Y different the population occurrence rate													
X\Y	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		1.6	2.6	-1.4	2.8	-8.4	2.1	-2.5	1.8	-4.7	-4.7	2.8	1.4	6.6
2	1.5		5.7	1.7	2.3	7.8	7.8	5.0	4.9	4.4	-1.6	5.9	-3.9	5.3
3	2.5	5.6		5.8	6.9	1.0	1.0	8.3	4.0	3.4	-2.8	-3.5	-2.8	-2.4
4	-1.1	1.4	4.8		7.7	16.9	1.1	8.2	-3.9	-5.3	-2.9	-4.8	2.1	-2.1
5	1.9	1.6	4.9	6.5		13.9	8.6	-8.3	-5.2	0.9	0.6	-5.1	1.2	8.1
6	-3.8	3.7	0.5	9.7	9.4		8.6	10.1	1.6	5.2	12.3	5.0	11.2	4.7
7	1.0	3.7	0.5	0.6	5.9	8.6		-1.0	1.6	5.2	14.6	7.1	5.7	-4.0
8	-1.3	2.8	4.7	5.6	-6.6	12.0	-1.2		11.8	4.1	3.0	2.0	6.9	-2.4
9	1.2	3.2	2.6	-3.1	-4.8	2.2	2.2	13.6		-3.1	0.3	5.3	8.0	7.8
10	-2.8	2.7	2.1	-4.1	0.8	6.9	6.9	4.5	-2.9		-2.1	7.3	1.8	-7.5
11	-2.5	-0.9	-1.6	-2.0	0.5	14.6	17.2	3.0	0.2	-1.9		4.1	12.4	15.0
12	1.6	3.5	-2.1	-3.5	-4.3	6.3	8.9	2.1	4.9	7.0	4.4		17.7	11.9
13	0.6	-1.7	-1.3	1.1	0.8	10.6	5.4	5.5	5.6	1.3	9.9	13.3		19.8
14	1.8	1.5	-0.7	-0.7	3.3	2.8	-2.4	-1.2	3.5	-3.5	7.6	5.7	12.6	

When an unfolding molecule occupies I_{ED}^x …

3.7.3) I_{CB}^0 occurrence frequency

One interesting divergence in pathway observed from the previous literature concerns the occurrence frequency of the CB major state (I_{CB}^0). Since 2002, this state was understood to be present in all unfolding trajectories (with an exception in Oesterhelt *et al.* 2000, where it was mentioned "…the positions and shapes of the first and third peak varied considerably" (*17*) but this variability is not addressed elsewhere in the literature).

We found this not to be the case, as only one of two states (I_{CB}^0 and I_{CB}^1) were mandatory in a given unfolding trajectory (though many traces exist at v=300 nm/s that occupy both I_{CB}^0 and I_{CB}^1). Further proof that this behavior was real was provided a small set of unfolding records we performed at 30 nm/s (Fig 3-17B) and 3000 nm/s (Fig 3-17C),. The proportion of trajectories that goes directly into I_{CB}^1 without a detectable occupancy in I_{CB}^0 increases as pulling velocity is increased from v = 30 to 300 nm/s, and (to a lesser extent) from 300 nm/s to 3000 nm/s, indicating that this behavior is not merely due to dwells within I_{CB}^0 that transition to I_{CB}^1 below our detection threshold (~25 pN at this extension). If that was the case, one would expect dwells within I_{CB}^0 to become more detectable at higher velocities, due to the positive relationship between rupture force and loading rate.

Structurally, I_{CB}^0 is located above the membrane interfacial region, and I_{CB}^1 is located near the bottom of the membrane interfacial region. Also, as many as 4 residues within helix C (D96, T89, D85 and R82) are shown the undergo conformational motion during the photocycle (*58*), so perhaps the lack of a single obligate state at the top of helix C reflects the well-established functional configurations of the protein.
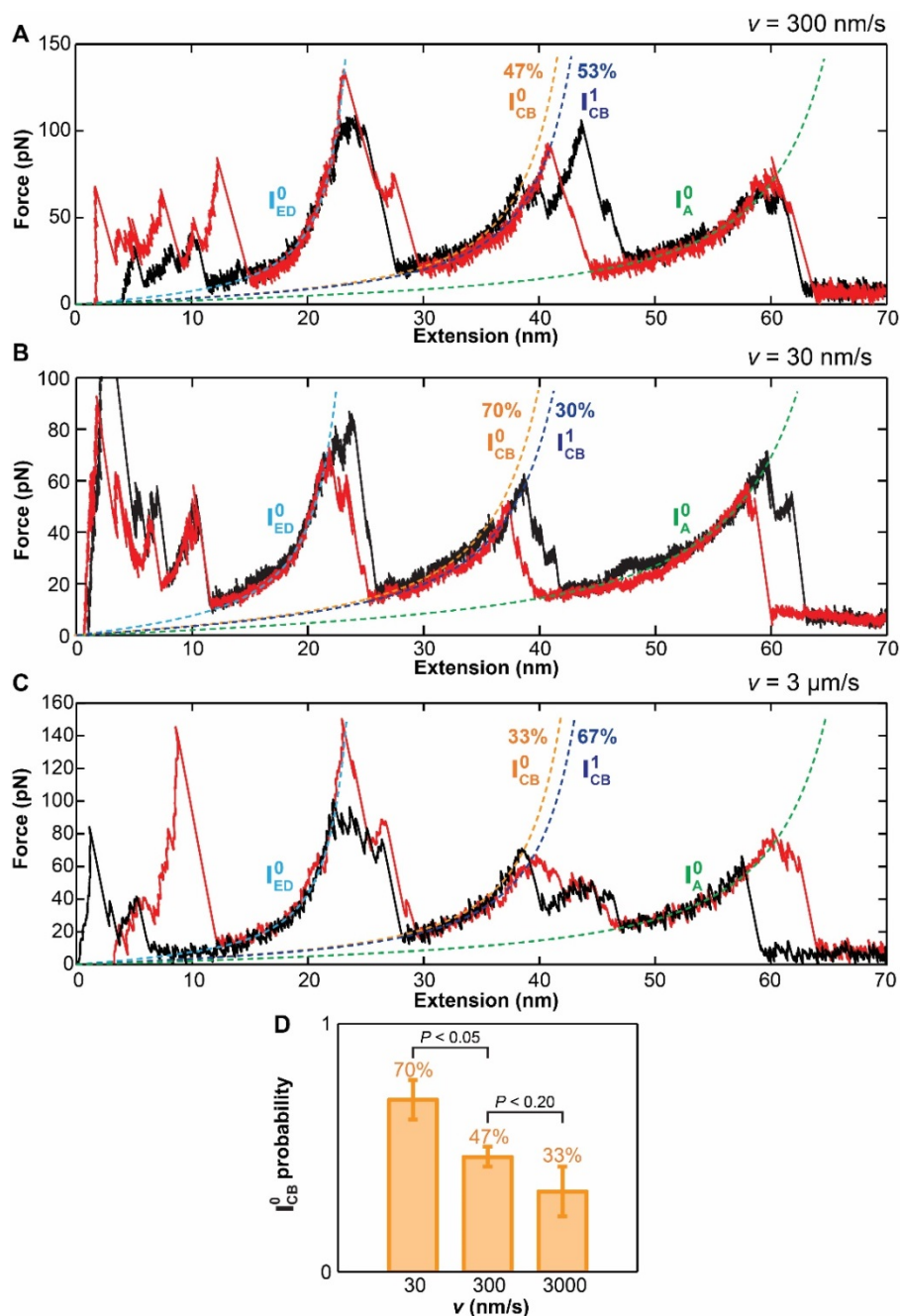
Figure 3-17: Non-obligate behavior in I_{CB}^0 influenced by pulling speed. A previously identified obligate major intermediate of the CB helix is composed of two non-obligatory states separated by 5 amino acids. Representative FECs show two unfolding trajectories for the CB helix pair where the first observed transition in the CB helix was from I_{CB}^0 (black) and I_{CB}^1 (red), respectively. Measurements were performed at three pulling velocities: **(A)** 300 nm/s, **(B)** 30 nm/s and **(C)** 3 μm/s. Data was taken at 5 MHz and smoothed at (A) 10 kHz, (B) 1 kHz, (C) 10 kHz respectively. **(D)** The proportion of trajectories showing I_{CB}^0 decreases as pulling velocity is increased from $v = 30$ to 3,000 nm/s. Error bars represent the standard error of the proportions. *P* values represent the level of significance of the z-test. From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.

3.7.4) Near-equilibrium fluctuations on top of helix E and A

Rapid, back-and-forth transitions have been called a hallmark of equilibrium between states (*96*), but technically, these states are only *near* equilibrium, because the force acting on each state varies slightly in time. Previously, refolding of individual bR molecules was observed by bringing an AFM tip closer to the surface ($v < 0$ nm/s) (*71*), which lowers the force on the unfolded molecule and thereby promotes folding. In contrast, bR refolding had not been detected while retracting the cantilever ($v > 0$ nm/s), implying that the standard rapid stretching assay was far from equilibrium. With our improved spatiotemporal resolution, we now routinely detect reversible transitions between two (Fig 3-3A, lower inset) and even three states (Fig 3-2) while stretching. One surprising feature of this data was the frequency of refolding events that occurred near the top of the helical pairs, even at high pulling speeds. Indeed, back-and-forth transitions in bR persisted up to the highest velocities examined (5,000 nm/s) (Fig 3-18) and were observed in all the helix pairs (ED, CB, and A). These states exhibited brief lifetimes (<1 ms) and could refold against comparatively high loads ($F \approx 30$–160 pN). Refolding was most frequently observed between states located near the top of helix E and helix A (70% of molecules refolded from I_{ED}^1 to I_{ED}^0; 90% of molecules refolded from I_A^1 to I_A^0). Moreover, refolding at $v = 300$ nm/s was ubiquitous: all records exhibited at least one refolding event in the ED helix pair.
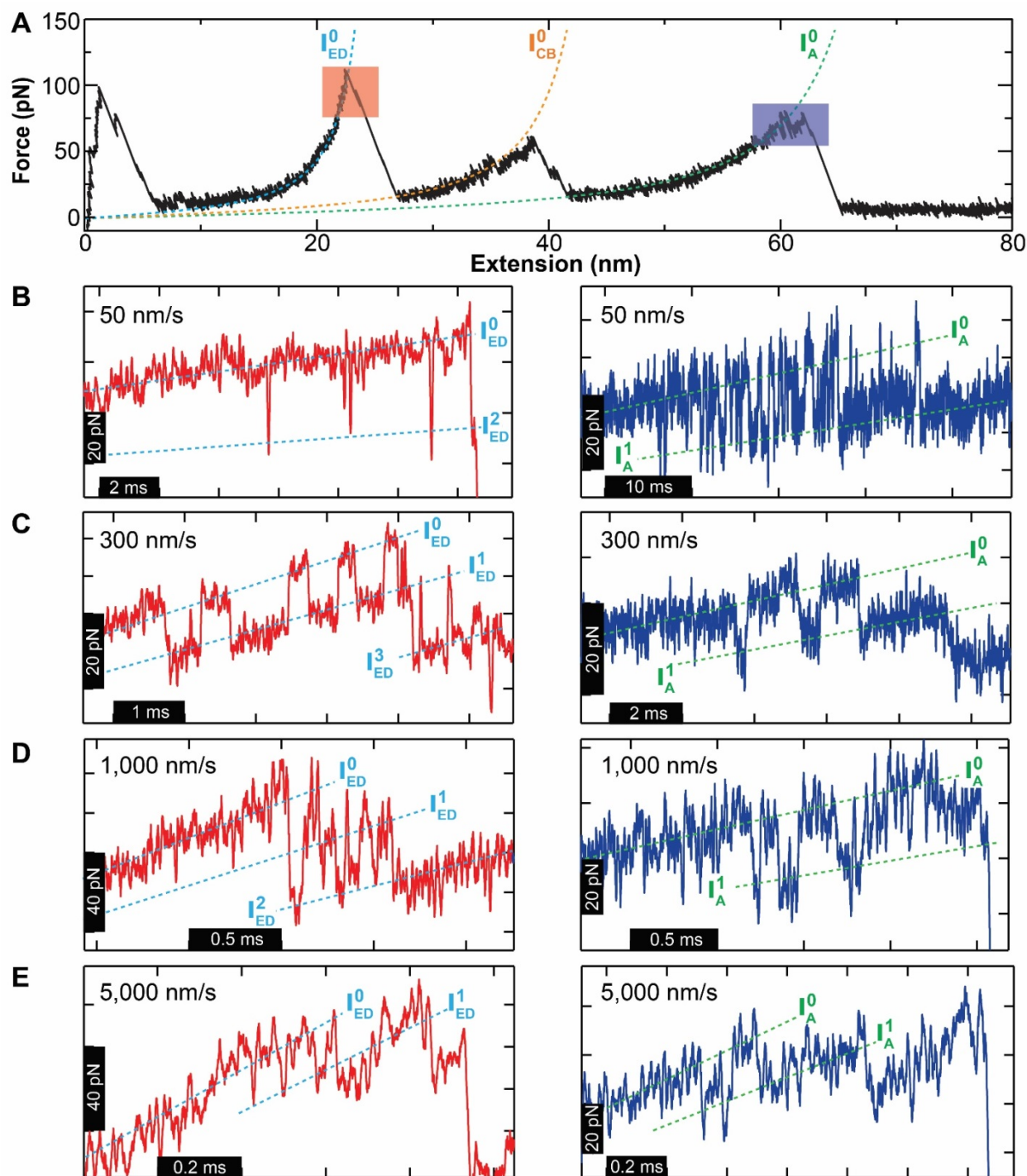
Figure 3-18: Near-equilibrium folding at different pulling velocities. (**A**) A FEC for bR taken at 300 nm/s. Colored boxes highlight two locations near the top of helix E and A where near-equilibrium dynamics were frequently observed. Such back-and-forth transitions between states were observed over a wide range of pulling velocities: (**B**) 50 nm/s, (**C**) 300 nm/s, (**D**) 1,000 nm/s, and (**E**) 5,000 nm/s. Transitions at the top of helix E are shown in red and transitions at the top of helix A are shown in blue. Data was taken at 5 MHz and smoothed at (**B**) 10 kHz, (**C–E**) 50 kHz, respectively. Dashed lines in **B–E** added to guide the eye. From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.

3.8) Equilibrium energy landscape reconstruction

One useful feature of this rapid unfolding and refolding is that each transition can be used as an independent data point in the process of reconstructing the energy landscape. In order to maximize the number of transitions we observed from an individual molecule, we changed the experimental protocol from a constant velocity pull (300 nm/s) to a stepping protocol (Fig 3-19A), which had a significantly lower data success rate, due to the instability of non-specific bonds at forces > 90 pN. In work led by my co-first author Hao Yu, but aided by myself, we collected 11 molecules, from which we extracted ~100 transitions per molecule. On this dataset, Hao performed p-fold (*98*), as well as Inverse-Boltzmann analysis (*99*), and Bell-Evans analysis (*100*), to reconstruct the local energy landscape between I_{ED}^1 and I_{ED}^2 at $F_{1/2}$, the force where both states are equally occupied. For the sake of completeness, I will list the results here, as I was part of the data collection and processing, but acknowledge that the analysis was entirely done by Hao Yu. Extrapolating the energy difference between states to zero applied force yielded $\Delta G_0 = 8.0 \pm 0.4$ kcal/mol. Because this transition involves unfolding 3 amino acids (T-F-G) of the transmembrane helix, the average ΔG_0 per aa was 2.7 kcal/mol. This result is higher than previous single-molecule (0.5–1.5 kcal/mol) measurements, which were averaged over entire transmembrane helices (*71*, *101*). An insertion energy for a full transmembrane helix (20–30 aa) of ~12 kcal/mol is deduced from traditional ensemble measurements, yielding ~0.5 kcal/mol per aa (*48*) .
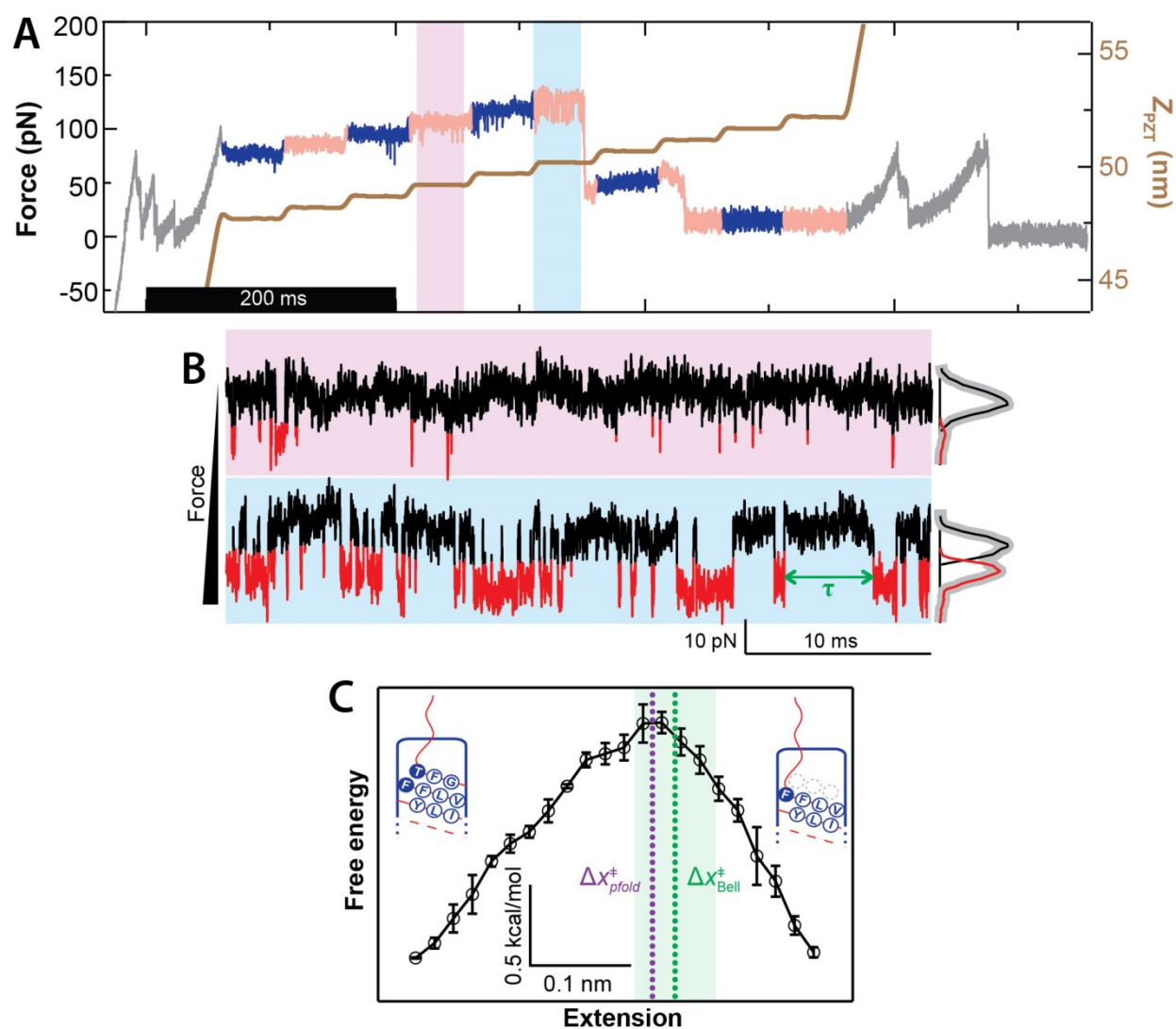
Figure 3-19: Equilibrium protocol and energy landscape reconstruction (A) Force-vs.-time for unfolding bR using the equilibrium protocol, highlighting the ten sequential 0.25-nm steps (blue and pink, alternatively) taken during unfolding of bR. 2 unfolding steps are highlighted (in pink and blue), which are also highlighted in (B) On the right, motion of the base of the cantilever controlled by vertical motion of the PZT stage (Z_{PZT}) as a function of time. (**B**) Two force-vs.-time traces (same as in A), with pink and blue highlights matched) show reversible transitions between two previously unresolved intermediates (I_{ED}^1, black; I_{ED}^2, red) at $v = 0$ nm/s. The cantilever was retracted from the surface between these two traces, increasing the force applied to the bR and thereby shifting the equilibrium towards I_{ED}^2. Data were smoothed at 25 kHz. (**C**) A reconstructed 1D free-energy landscape at $F_{1/2}$ based on the equilibrium data shown in panel A constructed using f_{old} analysis (*98*). The barrier position determined by p_{fold} (purple line) agrees with the result of an independent analysis based on the Bell model (green line). Error bars represent the SEM and the light green shading represents the uncertainly in determination of $\Delta x^{\ddagger}_{Bell}$. From Yu, Siewny *et al.* Science 2017. Reprinted with permission from AAAS.

3.9) Conclusions

By using ultrashort cantilevers optimized for improved spatiotemporal resolution, we have added a significant amount of detail to the bR unfolding process. Force spectroscopy has revealed a multiplicity of closely spaced, transiently occupied intermediate states, representing small changes in the molecular conformation. The widely held notion that the mechanical unfolding of bR at standard stretching rates occurred far from equilibrium is likely to be incorrect; refolding is, in fact, widespread but masked by experimental limitations when using standard cantilevers. In retrospect, elements of bR secondary structure likely unfolded and refolded during past SMFS experiments but did so faster than the force probe could respond. The next stages of this work will look to replicate previous studies on the N-terminal side of the protein, and see if an increase in the number of states is also present there, and possibly learn about the underlying causes of the stabilizing forces to the structure of bR by comparing the two.

3.10) Methods

3.10.1) Alignment, scaling and force offset

To account for small differences in the location of the attachment point between the tip and the bR, we aligned all of the FECs to the major folding intermediate that occurred at the top of the E helix, using a reference-free alignment algorithm (*102*). Briefly, FECs were converted to contour length L_c histograms. A small constant force offset (7.1 pN) was used when fitting the FECs in order to obtain better fits, similar to previous studies (*64*, *75*) (see Fig 3-20). A cross correlation algorithm (*102*) was applied to the L_c histograms that correspond to the ED helix unfolding to align the ED major unfolding peak. One modification to the original Bosshart *et al*. method (*102*) was that the position of the top of E helix was fixed to its averaged L_c before the alignment (26.9 nm), as opposed to letting it vary. The unaligned and aligned data sets are shown in Fig 3-21. Small errors in cantilever deflection sensitivities, which are particularly difficult to determine for these modified ultrashort cantilevers, may result in inaccurately measured extensions. To minimize this problem, we scaled FECs based on the difference in

L_c between the major state of the E helix and the major state of the A helix within each FEC, assuming the positions of these two major states are preserved between different molecules, as evidenced by previous work (*64, 67, 74, 75, 88*). In essence, we are using an intrinsic property of the molecule, the number of unfolded amino acids between these two major states, as a length scale. An initial average of this distance (47.6 nm) was used to scale all FECs. FECs were aligned and scaled iteratively. The final scaling was less than 10%, an excellent result given the difficulties in calibrating ultrashort cantilevers (*53, 91*).
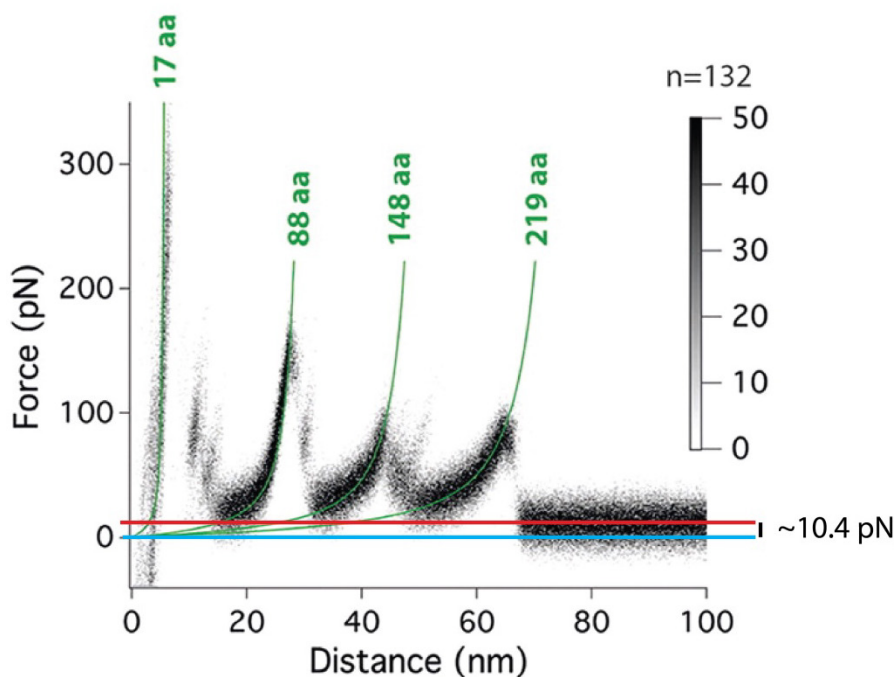
Figure 3-20: Demonstration of force offset from literature. A plot from Petrosyan *et al.* 2015 shows a force-extension plot taken on mica, teal line indicates the zero-force line, while the red line labels the average post rupture force, and which should be 0 pN (*11*). As measured using pixel measurements, these two lines are different by approximately 10 pN, indicating that the data has been shifted up to make the fitting more accurate. This is a technique we implemented to improve the quality of our fits. Adapted with permission from R. Petrosyan *et al*. *Nano Lett.* **15**, 3624–3633. Copyright 2015 American Chemical Society.
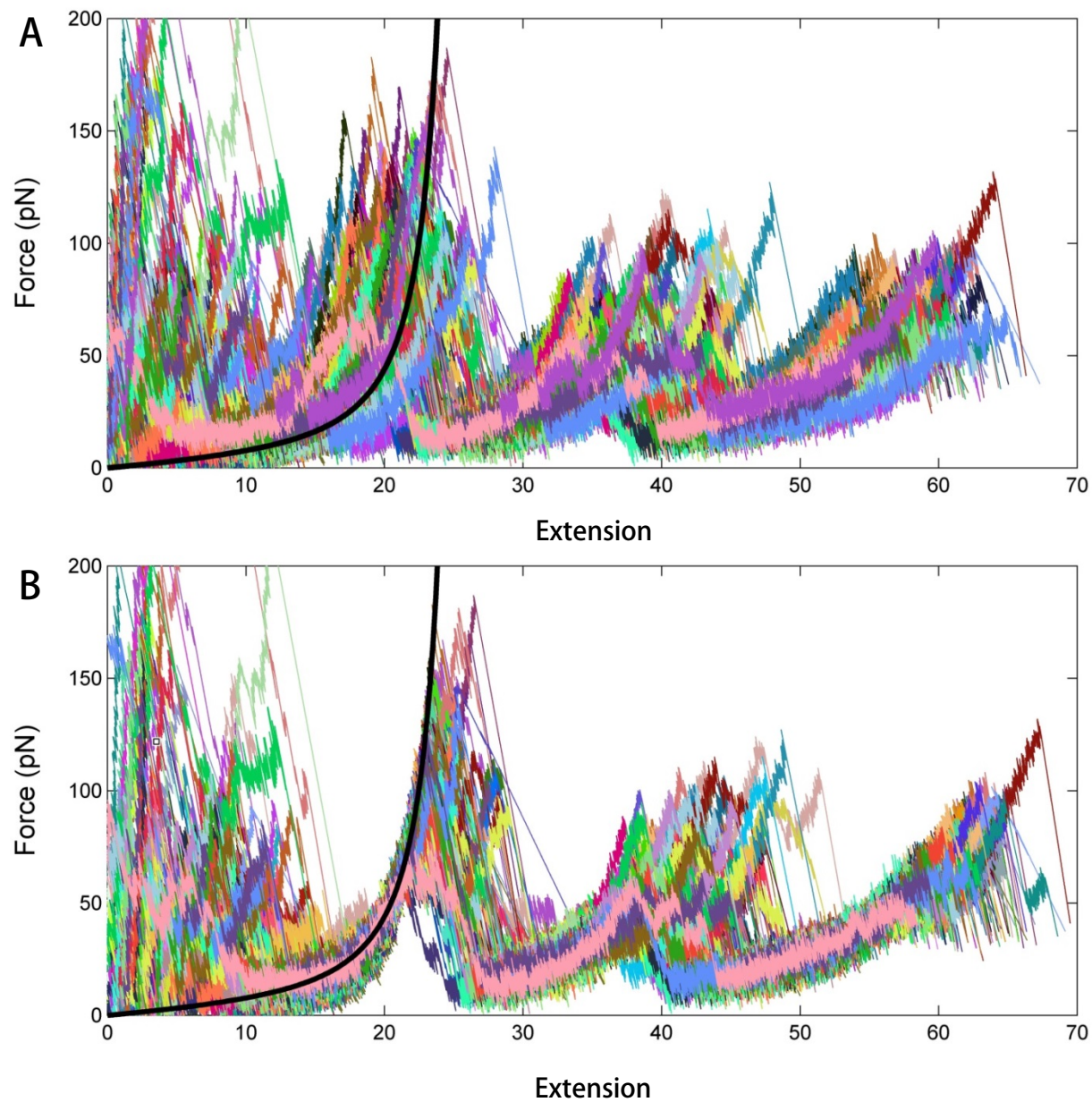
Figure 3-21: Force-extension alignment. This shows all 98 unfolding molecules before and after (A and B, respectively) contour-length based alignment, which is assigned to align to the black WLC curve (L_c=26.9 nm). Note that these curves are presented prior to scaling.

3.10.2) Segmentation

Having aligned the FECs above, we next needed to determine the total number of intermediates and assign each intermediate to a structural state. We initially segmented a full FEC into sections that aligned with worm-like chain (WLC) curves at constant L_c, which correspond to structural states with a constant number of unfolded amino acids. We refer to the sections as dwells for succinctness. The L_c of these dwells was compared to the L_c of dwells from other molecules. When a common L_c was identified in at least 5 molecules, it was defined as a state. The initial values of the total number of states and the L_c of each state were therefore defined.

An example of the segmentation process is demonstrated below in Fig 3-22 and 3-23. Fig 3-22A-B depict straightforward segmentation, with large spacing force and extension between the different segments while Fig 3-22C-D and Fig 3-23C-D examples of segmentation on less clear examples. Ultimately, the segments were delineated based on either large changes in L_c relative to the force noise or small changes in L_c with moderately longer dwells, indicating that the change in contour length is not arising from a transient noise spike.
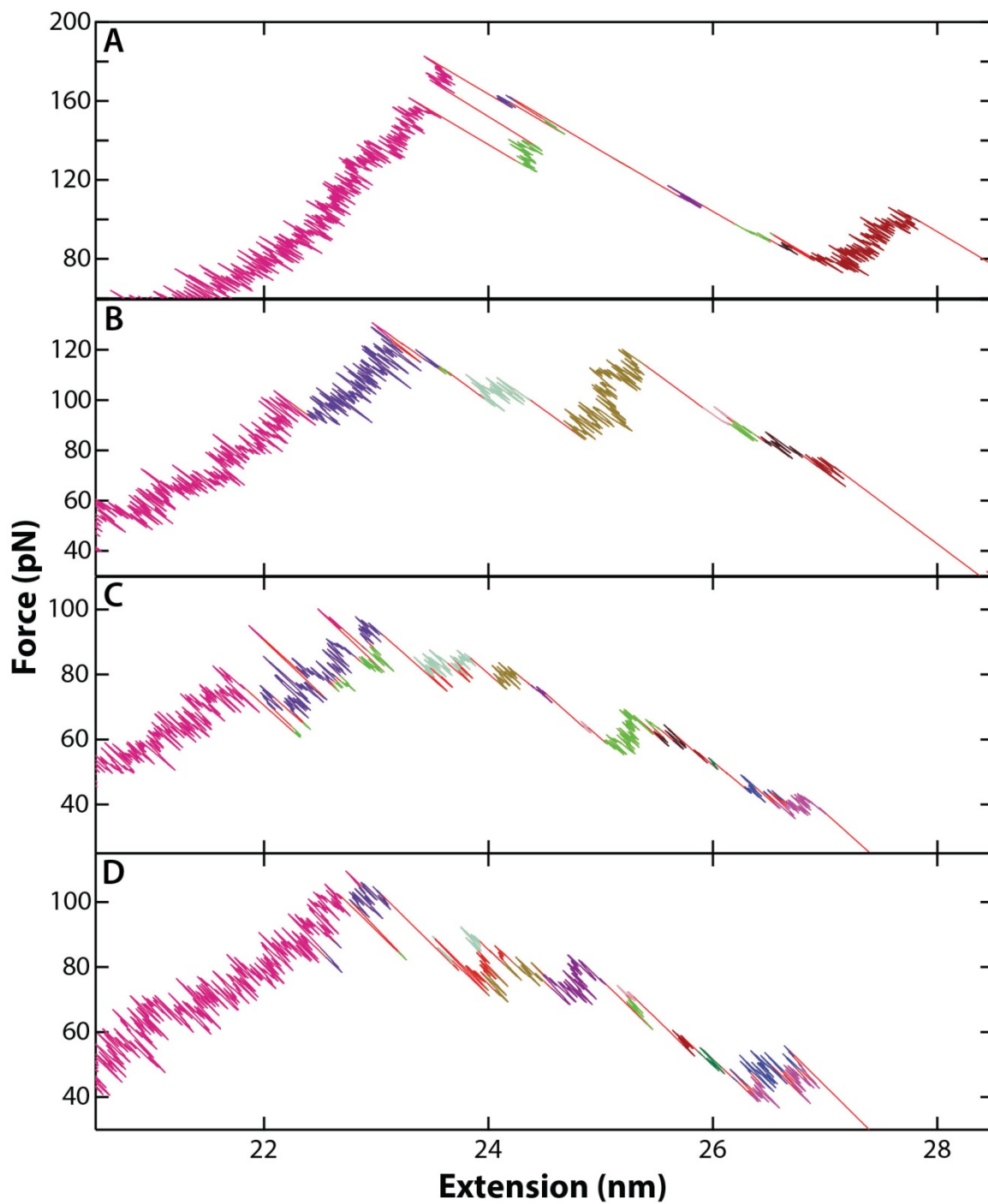
Figure 3-22: Force extension plots demonstrating segmentation of 4 records. (A) and (B) are relatively straightforward, while (C) and (D) have more closely spaced segments, which require careful work segmenting B, C and D records are converted to contour length space in Fig 3-22. Segments colored.

Figure 3-23: Contour length vs. time segmentation plots, for 3 bR molecules unfolding for ED helices. Same lettering used as in Fig 3-22. **(B)** Depicts a relatively straightforward segmentation. **(C)** A more complex segmentation, with tight transition, especially between purple and green (~ 27.5 nm). Even though these transitions are closely spaced, but there are a few dwells that persisted for long enough times (~300 μs, black arrows) that gives confidence that we can record the transitions. A few very short lived "transitions" that are not labeled in the green segment, due to the shortness of their lifetime (~ 30 μs). There is a reasonable argument to be made that these two transitions should have been grouped into the green segment, but we erred on the conservative side for short lived transitions not in immediate proximity to other intermediates, as the transition might have been larger before the smoothing. The asterisk labeled green segment (~32 nm)

Figure 3-23 (cont.): spreads out over a larger range of contour lengths, but is lumped into one segment as there is no clear transitions occurring (the contour length drifts over many milliseconds.
(D) Another more complex segmentation, this one specifically one informative region of segmentation Blue and magenta (~35.25 nm, dagger). Here, two closely spaced segments are separated, aided by the longevity of a portion of the blue segment (~800 μs), that allows shorter lived transitions in proximity to be segmented.

3.10.3) State determination

Based on an initial value from a histogram of contour lengths (that can be constructed from transforms such as Fig 3-23) we assigned all the dwells to the intermediates and update the L_c^k of each state k, through an iterative algorithm. First, each dwell was assigned to the closest state based on the current L_c values of the states. Here, each dwell, with the index j, was assigned to the state that gives the minimal value of the objective function:

$$S_j = \sum_{i=1}^{n}\left(F_{i,j} - f_{WLC}(x_{i,j}, L_c^k)\right)^2,$$
<div align="right">(Eqn. 3-1)</div>

where i is the point index within each dwell, n is the total number of points, $F_{i,j}$ is the force value of each point, $x_{i,j}$ is the extension value of each point, L_c^k is the contour length of each state k, and f_{WLC} is the worm-like chain function (Eqn 2-1) (*103*). Second, L_c values of the states were updated based on the new assignment of the dwells. Here, the L_c^k of each state was varied until the minimization was reached for the objective function:

$$S_{state} = \sum_{j=1}^{N}S_j = \sum_{j=1}^{N}\frac{1}{n}\sum_{i=1}^{n}\left(F_{i,j} - f_{WLC}(x_{i,j}, L_c^k)\right)^2,$$
<div align="right">(Eqn. 3-2)</div>

where N is the total number of dwells assigned to the state. These two steps were repeated iteratively until variations of contour length of the assigned intermediate by 0.01 nm yielded no improvement in state assignment. The L_c of each intermediate was therefore obtained. The error in L_c (Table 3-1) was reported as the standard deviation (s.d.) of the final fitting process:

$$\sigma_{L_c} = \left(\frac{S_{state}}{N-1}C_{LL}^{-1}\right)^{1/2},$$
<div align="right">(Eqn. 3-3)</div>

where
$$C_{LL} = \sum_{j=1}^{N}\left(\frac{\partial f_{WLC}}{\partial L_c}\bigg|_{x_{i,j}}\right)^2$$

After applying this assignment algorithm, two plots of the dwells associated with each intermediate were made. The first was simply all the dwells assigned to a particular intermediate (Fig 3-23A). If the dwells look bifurcated around the contour length, a new intermediate was introduced (assuming a minimum of 5 dwells in each of the new states) (Fig 3-24B). The second graph was used to determine if there were too many states. This analysis was accomplished by plotting the dwells associated with three sequential intermediates and looking for gaps in the distribution of dwells assigned to each intermediate (Fig 3-25). Conceptually, this is equivalent to plotting histograms of the contour length of each dwell and looking for a dip in the histogram between states (Fig 3-25C). However, limited data due to short dwell times in each state prevents routine application of this more sophisticated analysis. This state assignment procedure was run multiple times to ensure consistent convergence. Ongoing efforts in the lab seek to further improve state identification.

Using this algorithm, we identified 14 intermediates within the unfolding of the ED helix pair (Fig 3-3A), 7 intermediates within the CB helix pair (Fig 3-3B), and 3 intermediates within the A helix (Fig 3-3C). To demonstrate the existence of a given state within the ED helix pair, we constructed a set of FEC plots that show individual trajectories that link adjacent states (Fig 3-5). Such transitions within individual unfolding trajectories help validate the existence of closely spaced states. While many of the trajectories could be well described by these states, it is expected that some intermediates did not map perfectly into these identified states. The most likely confounding factors are technical issues such as variations in the measured extension that arise from variability in where the tip attaches to the protein, differences in height of the attachment point of the protein to the AFM tip from the true apex of the AFM tip, force drift, and difficulties in calibrating the stiffness and sensitivity of ultrashort cantilevers. Additional technical advances are needed in the spatiotemporal precision of AFM, calibrations of ultrashort cantilevers, and site-specific coupling of membrane proteins to AFM tips to reliably detect transitions of < 2 amino acids. The contour length change (ΔL_c) between each state is obtained from the procedure described above (and listed in Table 3-1). The consensus positions of previously observed intermediates (*17*, *64*, *67*, *68*, *70*, *72–75*) were also listed in the table for comparison. In this work, we

compared our results directly with these intermediates that have been repeatedly reported over the past 18 years (*17, 64, 67, 68, 70, 72–75*). However, for completeness, we do note that Kessler & Gaub reported one more intermediate within the unfolding of the ED helix pair, CB helix pair, and A helix, respectively (*49*).
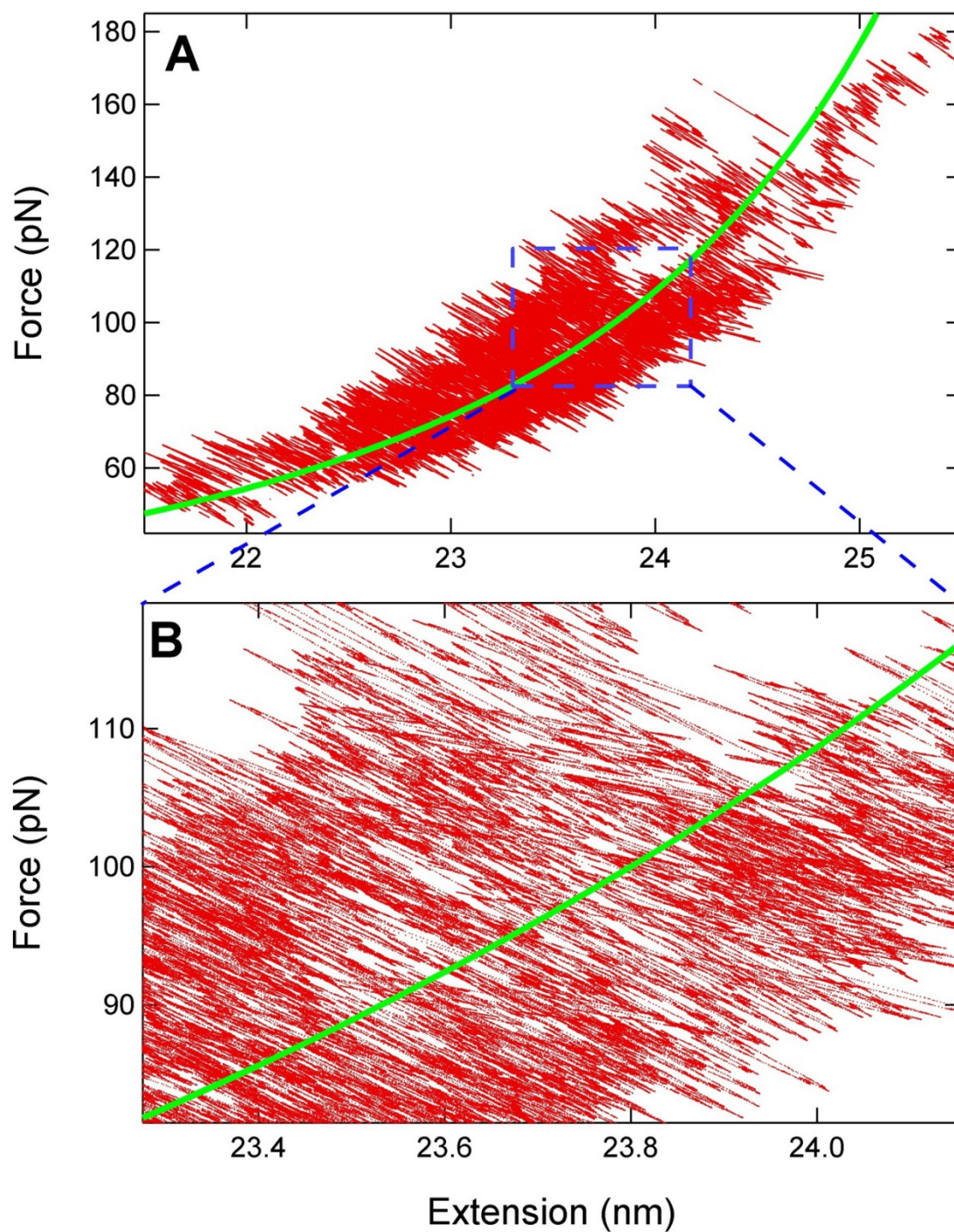
Figure 3-24: Example of test to determine dwells underfit (not enough states). (A) All the dwells (red) assigned to a given intermediate (in this case, C_L=28.48 nm, green). (B) Zoom on blue region of (A), demonstrates a bifurcation of the data above and below the WLC curve, indicating that these dwells likely represent two (or more) molecular intermediates.
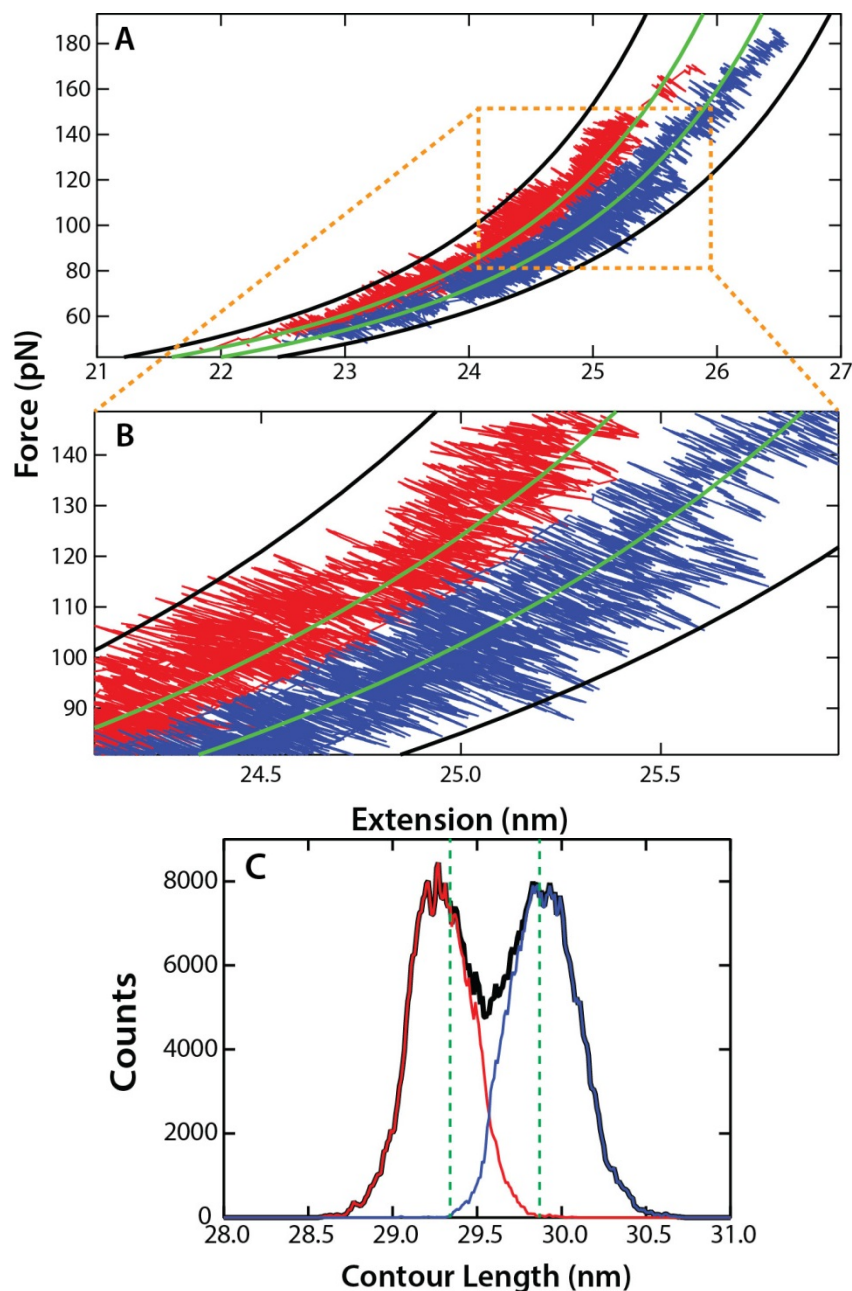
Figure 3-25: Example of curve plotting to determine dwells overfit (too many states). (A) Dwells assigned to two proximal intermediates (I_{ED}^4 and I_{ED}^5, red and blue, respectively). WLC curves representing I_{ED}^3 (black), I_{ED}^4 (green), I_{ED}^5 (green) and I_{ED}^6 (black) are overlaid on top. A zoom in region of this plot is depicted in (B). (B) Demonstrates a rarefaction of dwells in the region between the states. While not completely void of records, certainly they are less dense than around the center of I_{ED}^4 and I_{ED}^5. (C) A contour length histogram plot of the above data, for dwells assigned to I_{ED}^4 (red) and I_{ED}^5 (blue), and their superposition (black). The contour lengths of these intermediates are marked (dashed green lines) The superposition of the histograms clearly demonstrates a rarefaction of dwells between the two states. It should be noted that these histograms will not always demonstrate this rarefaction, as short dwell time are not weighted appropriately.

CHAPTER IV: Extracellular forced unfolding of Bacteriorhodopsin

4.0) Introduction

One of the longstanding goals of SMFS studies is to determine the specific interactions and forces that give rise to the stability of a folded protein. The sources of these forces are well understood: the strongest three are intrahelical bonds, interhelical bonds (hydrogen bonds or salt bridges), and the hydrophobic interactions (*17*, *104*). What is less well understood is the relative significance of these interactions in the stabilization of each intermediate (i.e. is I_{EF}^3 mostly stabilized by an interhelical hydrogen bond, or is it stabilized largely by hydrophobic interactions with the membrane?) Understanding the strength of these interactions is crucial for accurate simulations of protein folding, as modeling these interactions is the functional core of a molecular dynamics simulation (*76*, *79*). More experimental data differentiating these forces strengthen the predictive strength of current simulations.

By leveraging the improved spatiotemporal resolution of our system, we can increasingly localize these forces, such as relating a given intermediate to a specific interhelical interaction, but only if the unfolding is observed from multiple conditions. This is necessary to separate the relative contributions from the different sources. For example, by unfolding at different pulling speeds, one can change the force at which a given intermediate is likely to unfold, which will change the extension of the unfolded peptide chain. This changes which residues are located at the membrane interface, thus altering the hydrophobic forces acting on the unfolded peptide chain. Another, more powerful means of delineating these forces is to pull from a different geometry. Often accomplished in globular proteins by mutating non-native disulfide bonds (*96*), a similar effect can be accomplished on integral membrane proteins by pulling on the opposite terminus, reversing the unfolding direction. Due to the sequential unfolding of the protein, by unfolding from the N-terminus, a given helix will unfold with the other helices in the opposite state (folded/unfolded) as they were during a C-terminal pulling experiment. For example, the B helix unfolds in a C-terminal pulling experiment in the absence of the folded G, F, E, D, and C helices, but in the presence of the folded A helix, while in an N-terminal pulling experiment, the B helix will unfold in

the opposite conditions (in the presence of the folded G, F, E, D, and C helices, but in the absence of the folded A helix). This reversal allows one to disrupt interhelical bonds from opposite ends in the N and C-terminal pulling experiments. By combining the intermediate locations from these experiments, one can provide strong evidence to the presence of an interhelical bond and, with the aid of a crystal structure, assign the given interhelical bond to a specific pair of residues. Another possibility is that if a given intermediate does not correspond to a known hydrogen bond pair, but occurs when a hydrophobic residue is located at the membrane interface, this is strong evidence that the bond is being formed by the residues resistance to solvation. Thus, by comparing the location, strength and frequency of the unfolding intermediates that occur when pulling on the N-terminus (extracellular side), with the intermediates observed in the C-terminal pulling experiment (chapter 3), we can provide information to help differentiate between the known causes for each intermediates.

This experiment is not the first time this analysis was performed on bR. Indeed, the work in this chapter is largely inspired by Kessler & Gaub Structure 2006 (*49*), (see section 1.2.4). Representing the only previous N-terminal SMFS of bR[2], they correlated the position of C-terminal and N-terminal intermediates to a few hydrogen bonds. While groundbreaking, that work ultimately is lacks in intermediate precision resolution (with uncertainties as large as 4 aa) and in statistics (only 12 unfolding records from the N-terminal pulling experiments). With 100x improvement to temporal resolution, and 10x improvement in force precision offered by the FIB-modified ultrashort cantilevers, as well as a change in the attachment chemistry, we achieve significant improvements over this previous study

4.1) Chemistry Changes

4.1.1) Motivation

The low number of N-terminal unfolding records present in Kessler-Gaub arises from the limitations of non-specific attachment chemistry. Non-specific attachment chemistry is the method of achieving connection between the tip and bR by simply pressing the cantilever into the protein at high

---

[2] With comparable buffer conditions (>100 mM monovalent ions (i.e. not Voïtchiovsky *et al Biophys. J.* 2007)

force, a strategy used in the data presented in Chapter 3, and essentially all other AFM studies of

membrane proteins. While the exact mechanism of the attachment is poorly understood, it is established

that the attachment rate on membrane proteins is positively correlated to the amount of protein outside the

membrane (*77*). The significant percentage of the total peptide that resides beyond the cytoplasmic

interface (Fig 4-1, blue spheres), not to mention the significantly longer the C-terminal tail (25 aa),

together enabled non-specific attachment rates of ~2%. In contrast, the amount of peptide that resides

beyond the extracellular interface (Fig 4-1, red spheres) is significantly less, and the N-terminal tail is

significantly shorter (8 aa). From correspondence with the Gaub lab (and verified by our own

experiments), we learned that attachment of any kind to the N-terminal tail of bR occurs over 100-fold

less frequently than on the C-terminus (0.02% vs. 2%, respectively). Furthermore, even if a successful

attachment is made to the N-terminus, the length of the tail is short enough that the tip will probably still

be in contact with the surface during the unfolding of the AB helices pair. This means that it is

overwhelmingly likely that the AB helices unfolding record will be distorted, if not completely

suppressed. Kessler-Gaub notes this when they report that only ~7% of adhesion events will show the

characteristic four unfolding peaks, relative to 30% observed for the C-terminal data (*49*). Even if one is

willing to abandon analysis of the AB helices pairs due to this effect, their presence is important for

interpretation of the remaining unfolding events. If no semblance of AB helices unfolding occurs in a

given record then the assignment of the 3 remaining unfolding peaks becomes ambiguous, as the peaks

may correspond to the unfolding of helices CD, EF and G, or to the unfolding of the AB, CD and EF

helices (with an early rupture). Combining these probabilities, it would likely take somewhere between

5,000 and 20,000 unfolding attempts to yield a single interpretable N-terminal record. This is an

insufficient throughput to collect the enough data for a useful, statistical study. Thus, a different

attachment chemistry needed to be adapted.

4.1.2) New attachment chemistries

The chemistry that we adopted for the protein is a covalent attachment between a DBCO and an

Azide molecule, commonly referred to as "Copper-free click chemistry". It was initially adapted in the

Cytoplasmic/C-terminal side

Extracellular/N-terminal side
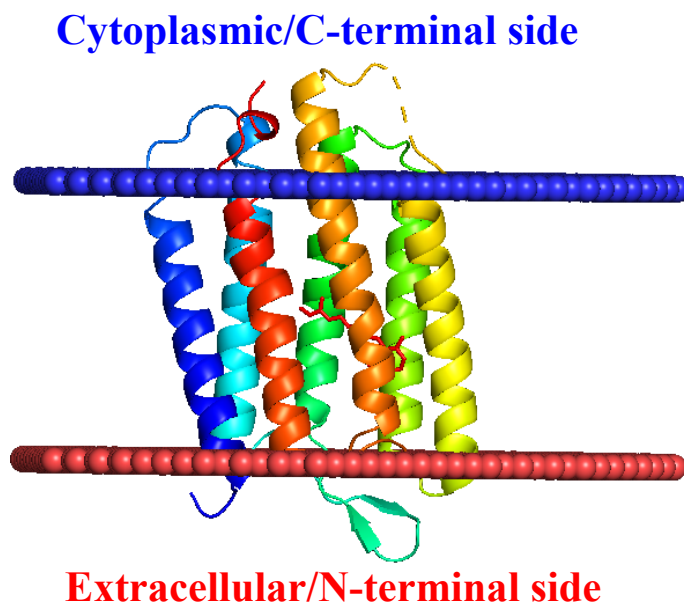
Figure 4-1: 3D model of bacteriorhodopsin in membrane, with approximations of the extracellular (red spheres) and cytoplasmic (blue spheres) membrane interfaces, constructed from the Orientation of Membrane Proteins database (*105*). While some residues are absent from the figure (especially on the C-terminus, where it has 17 of 25 residues are absent), it conveys the relative amount of peptide outside the membrane on the cytoplasmic side relative to the extracellular side. Model constructed with data from (*106*).

lab by Rob Walder for SMFS studies on globular protein (*107*). This chemistry is achieved by changes to both the N-terminal and cantilever functionalities. Azide-PEG molecules are attached to the tip using silane chemistry, while the N-terminus is mutated to a cysteine and a cystine-thiol attachment is used to attach melimide-dibenzocyclooctyl (DBCO) molecules onto the bR N-termini (See sections 2.3.4 and 2.4, and Fig 2-1 for the procedure.)

By pressing the functionalized tip into the functionalized bR at forces below 250 pN for 3-5s, we achieved ~2% attachment rate over the first 1,000 attempts on a membrane patch. After these 1,000 attempts the attachment rate drops to 0%, as it is presumed the active sites become occupied by unfolded bR molecules. Other than the improvement in attachment rate, this method offers several other advantages. For one, the PEG linking the Azide to the cantilever helps passivate the tip to non-specific adhesions, which would likely complicate analysis. Reducing the force that the cantilever exerts on the patch to below 250 pN further makes nonspecific attachments increasingly unlikely. In contrast, we had

to press at ~900 pN to achieve non-specific attachment, as discussed in Section 2.8.1. Moreover, the bond is covalent so it will survive for the entire unfolding, evidenced by the absence of partial bR unfolding curves in our records (which were observed in chapter 3 experiments).

4.1.3) New surface chemistry to change deposition sidedness

Another challenge that needed to be addressed to perform these N-terminal unfolding experiments was to change the sidedness by which PM patches deposited onto the substrate (which is largely dictated by electrostatics). To achieve this, the substrate was changed from mica, which has a slight negative charge when cleaved (*108*), to glass with a surface amino-silane functionalization (See section 2.1.2 for further details). This functionalization will leave the surface with a positive charge, and attract the negatively charged amino acids located on the cytoplasmic side of the membrane to adhere to the substrate. This charged residue attraction to the surface may cause some additional intermediates (see section 4.2.3). The sidedness of the surface was verified in earlier experiments where the membrane patches were imaged: the ratio of cytoplasmic sided patches to extracellular side patches went from ~9:1 for mica, to ~1:1 for APDMES functionalized glass.

4.2) Intermediate identification

Due to the presence of the retinal during the unfolding of every single helix, the intermediates observed in the forced unfolding from the N-terminal were predicted to be fewer and have higher rupture forces compared to those observed on the C-terminal (*79*). These predictions were made despite slightly contrary evidence presented in Kessler-Gaub (with slightly more intermediates observed with similar average rupture forces as they observed on the C-terminus). The present work will provide a higher resolution test of this prediction. The forces will not be an identical comparison to the C-terminal unfolding experiment for a few reasons. First the cantilevers were slightly softer, and include a PEG linker in the attachment, causing the loading rates to be significantly smaller here (< 5% of Kessler-Gaub), and consequentially, the rupture forces will be smaller.

Utilizing the same state determination algorithms established in the C-terminal experiments (section 3.10.3), we identified a significantly larger amount of unfolding intermediates, especially within

the CD helices. Overall, the assignment strategy remained unchanged, except for a few minor details: for alignment, still accomplished by the algorithm listed in section 3.10.1, but the vertical force offset now is 12.5 pN, and the second peak major state is set to be at 28.2 nm, with the fourth peak major state set to 77.8 nm (based on contour length difference listed in Kessler-Gaub 2006 between the C helix major state and the G helix major state).

4.2.1) CD helices intermediates

A four-fold increase (compared to previous literature (*49*)) in the number of intermediates in the CD helices was observed. In the past, they observed 5 intermediates, while we observe 20 intermediates (Fig 4-2A). Additionally, many of these intermediates are very closely spaced, with I_{CD}^5/I_{CD}^6, I_{CD}^{14}/I_{CD}^{15} and I_{CD}^{16}/I_{CD}^{17} spaced by less than 2 aa, and the majority of the remaining intermediates spaced by less than 3 aa (Table 4-1). One interesting behavior observed is that the average rupture force is approximately constant for the major state and the first 6 intermediates, then slowly diminishing. This is similar to the behavior seen in the Kessler-Gaub, where the force is roughly constant from the intermediate labeled 83 aa to the intermediate labeled 98 aa.
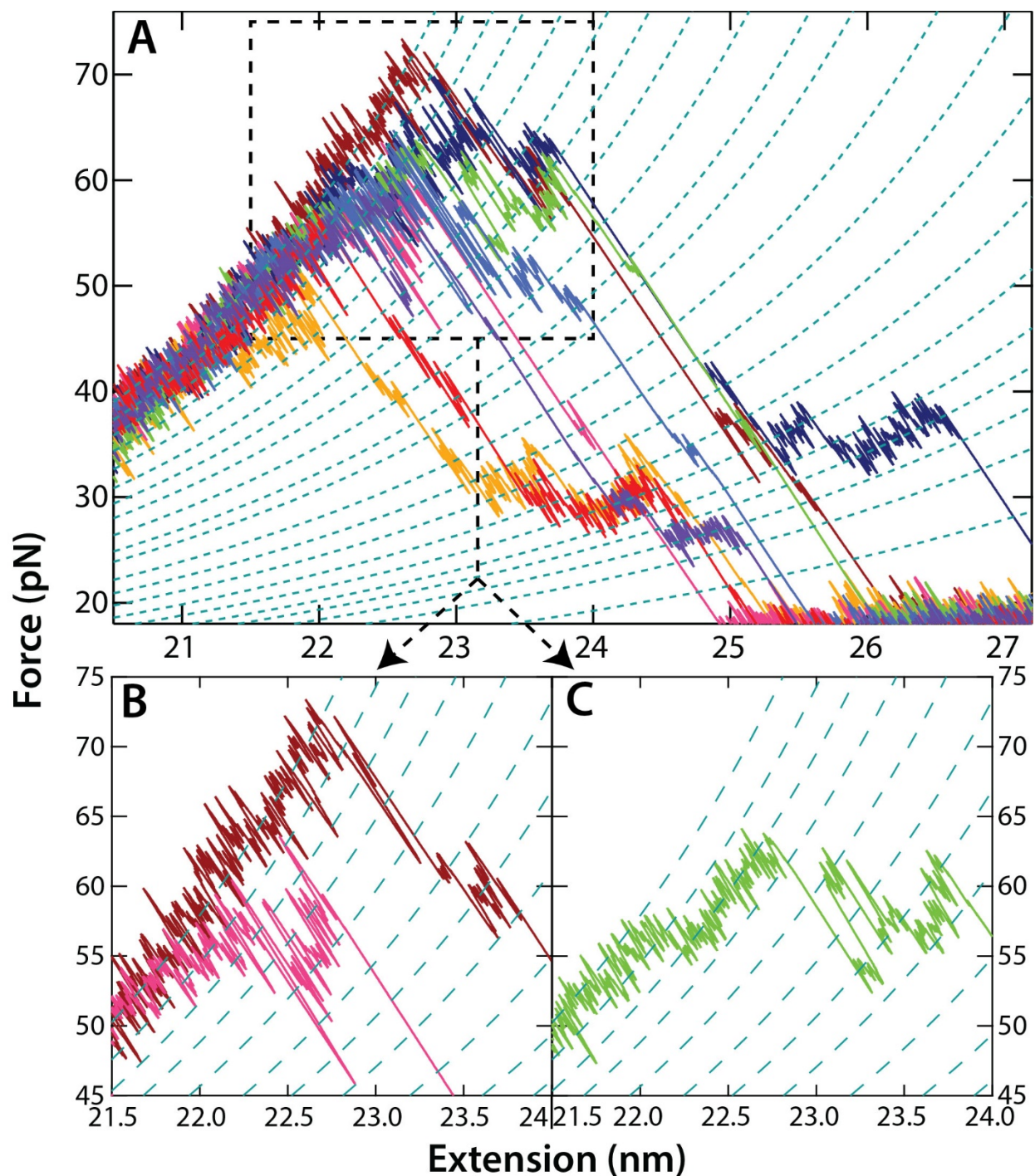
Figure 4-2: CD helices unfolding intermediates. Force-extension plots of individual molecular unfolding records of the CD helices pair. **(A)** An overview, showing 8 unfolding traces. The major state and 20 intermediates are identified (blue dashed lines). **(B)** A zoom on the black dotted box in (A), showing 2 unfolding traces in the transitioning between intermediates located in the top half of helix C. **(C)** The same region as shown in (B), but now with a different unfolding trace displayed, which shows near equilibrium fluctuations further from the helices start than was observed for C-terminal experiments.

4.2.2) EF helices intermediates

For helices EF, there was an increase of observed intermediates from 5 in Kessler-Gaub to 11 with no distinct major state. Compared to the CD intermediates, the EF intermediates spacing was significantly further apart, with an average spacing of 5 aa. In addition to the change in the number of intermediates, intermediates located in similar locations differed in their dynamics. Whereas Kessler-Gaub observed a major state at the same location as I_{EF}^0, we observe only 70% occupancy in I_{EF}^0, Many curves showed non-WLC behavior at low forces (< 20 pN) around I_{EF}^0, and thus points at these low force are excluded from analysis, which may be suppressing the I_{EF}^0 occupancy rate. I_{EF}^0 tends to rupture at very low forces (32 pN). In contrast, I_{EF}^2 and I_{EF}^3 have near obligatory occupancy (94% and 97%, respectively), and rupture at higher forces (46 pN and 43 pN, respectively).Therefore, they exhibit mechanical properties that are closer to the major state behavior than I_{EF}^0 or I_{EF}^1. We note one weakness of assignment in this region is the increased uncertainty of each assignment, which is as high as 2 amino acids within the E helix, due to the lower rupture forces within this region.
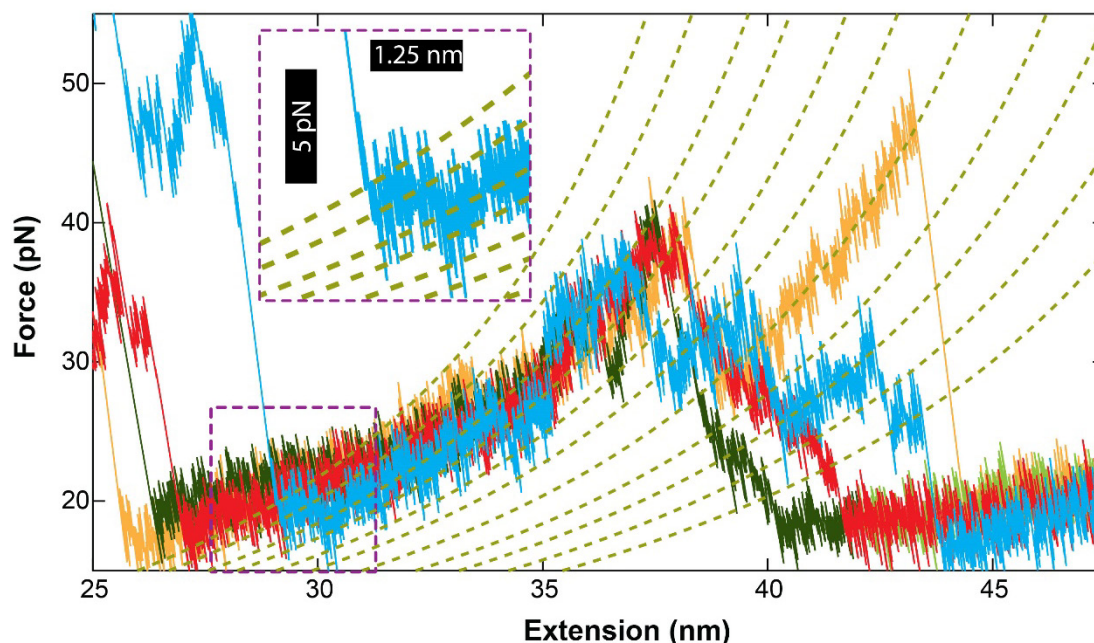
Figure 4-3: EF helices unfolding intermediates. Four force-extension plots of the unfolding of the EF helices pair. The 11 intermediates identified are shown overlaid on top (brown dashed lines). Note that no singular obligate state was observed, although I_{EF}^2 and I_{EF}^3 are highly occupied (> 94%). Further, the intermediate with the highest average rupture force (I_{EF}^2) is not the earliest intermediate observed. **(Inset)** This is an example of a molecule (blue) that demonstrates a short occupancy in I_{EF}^0 (~ 180 μs), and a longer occupancy in I_{EF}^1, but one that still ruptures at ~20 pN. This inset highlights behavior of low rupture forces common to these intermediates.

4.2.3) G helix and C-terminal tail intermediates

For helix G, there was a significant increase in number of unfolding intermediates: a major state and 5 intermediates (Fig 4-4), compared to only a major state observed in previous literature. Major state behavior is observed for I_G^0. It was obligate, as well as having the highest average rupture force in this region. The remaining intermediates, however, are difficult to analyze because their measured contour lengths place them past than the structured section of the G helix. Contour length analysis would place these intermediates in the unstructured C-terminal tail of the protein. The C-terminal tail, being unstructured, is mostly absent in crystallographic structures. While I_G^1 does fall within the part of the structure mapped via crystallography and thus can be assigned (assuming FRINL), I_G^2, I_G^3, I_G^4 and I_G^5 are not assignable and therefore do not have locations for their likely transition residues. The appearance of
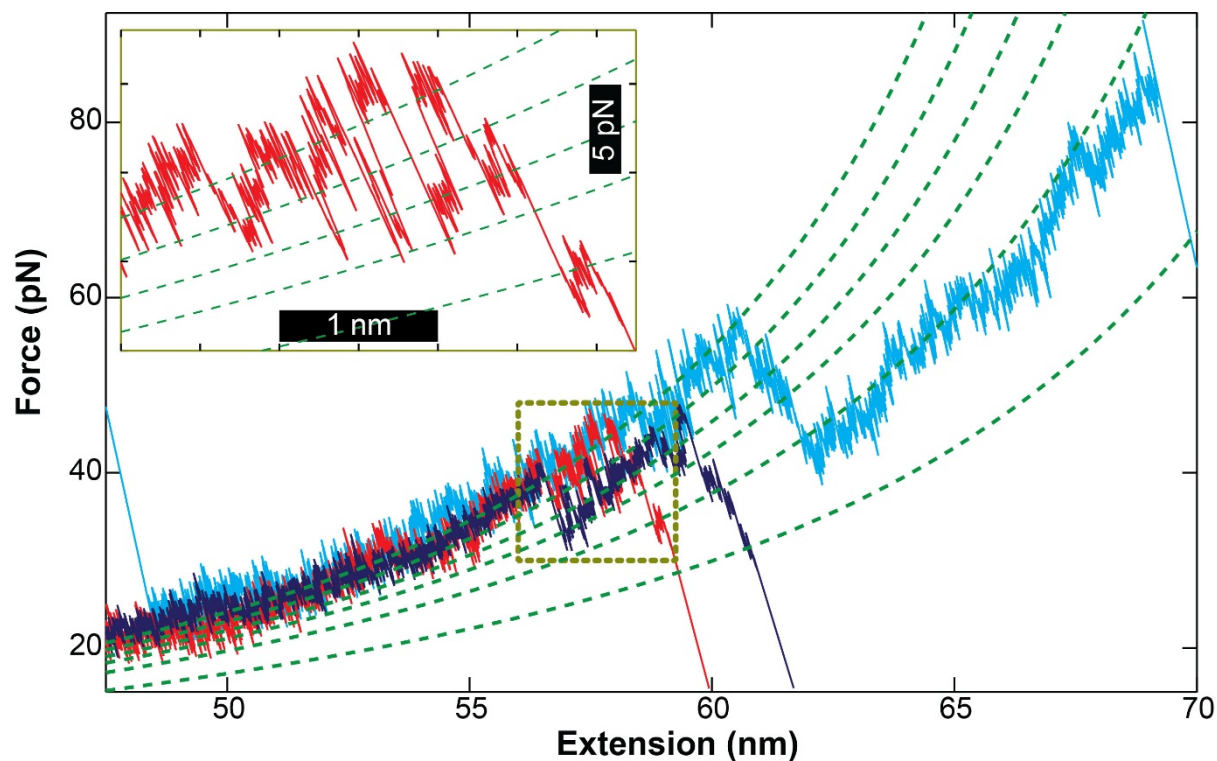
Figure 4-4: G helices and C-terminal tail intermediates. 3 unfolding traces demonstrating unfolding of the G-helix. The major state and 5 identified intermediates are depicted as WLC curves (dashed green lines). Interestingly, I_G^2, I_G^3, I_G^4 and I_G^5 map to the region after the end of the G-helix and present additional difficulties in assignment, as they have no experimentally known structure. **(Inset)** A zoomed in view of the dashed dark yellow region in the larger plot, showing near equilibrium fluctuations occurring at the top of the G helix.

intermediates at contour lengths where there should be no secondary structure raises interesting questions about the causes of these intermediates. Two possible theories are resistance to solvating hydrophilic residues (see section 4.4.2) and adhesion of negatively charged residues (E232, E234, E237, D242, and D249) to the positively charged surface.

4.2.4) AB helices intermediate determination

In the bR literature, as well as in in chapter 3, the analysis of the first peak (for C-terminal pulling: helices GF) was largely futile due to the interactions between the tip apex and the substrate interfering with the unfolding deflection signal. By changing to a specific attachment chemistry with a long polymer linker, a sufficient distance between the tip and the surface should have been achieved before the unfolding of the A helix occurred. Unfortunately, other issues have arisen that have made the

analysis of the AB helices pair intermediates difficult. Oddly, the surface interaction region occurred nearly coincidental with the A major state rupture, distorting all of the unfolding curves collected. This is possibly due to single attachment only occurring after the attachment sites near the apex are depleted (See section 4.7.3 for further discussion). Partial analysis of the AB helices intermediates can be achieved as long as the optical interference artifact is accurately modeled and removed near the surface. Unfortunately, for about half of the unfolding records, this was not the case, as demonstrated in Fig 4-5 (red traces). Because the interference artifact diverges from its sinusoidal behavior during the cantilever extension prior to the surface, part of the retraction is unable to have the interference artifact removed, as an accurate spline was unable to be constructed from the extension data. This divergence distorted the entire AB helices for nearly half of these records and the A helix major state in all records.

The half of records that had a non-distorted region of the AB helices unfolding (e.g. Fig 4-5, black and grey trace) were analyzed (only portions of these unfolded traces greater than 5 nm from the surface were analyzed). Further, the standard state determination algorithm (described in Section 3.10.3) was not used here, as it can be extremely sensitive to errors, especially at low extensions. Instead, state determination was accomplished using peak analysis of contour length histograms (*102*). This method is less sensitive to noise in individual unfolding traces but is poorer at detecting to short lived and/or rarely occupied intermediates.

Four intermediates were identified (Fig 4-6). They all assigned to residues located on half of the B helix located closer to the cytoplasmic side. None of them were obligate, and all exhibit modest mean unfolding forces of 45 pN or less. These four intermediates should not be interpreted as a comprehensive listing of intermediates in the AB helix unfolding. Even in Fig 4-6, there are visible dwells located at other contour lengths. Rather, the identified intermediates are the only ones that met the threshold for statistical significance using the peak analysis method. Additional records with less artefactual distortion will be necessary to identify them.
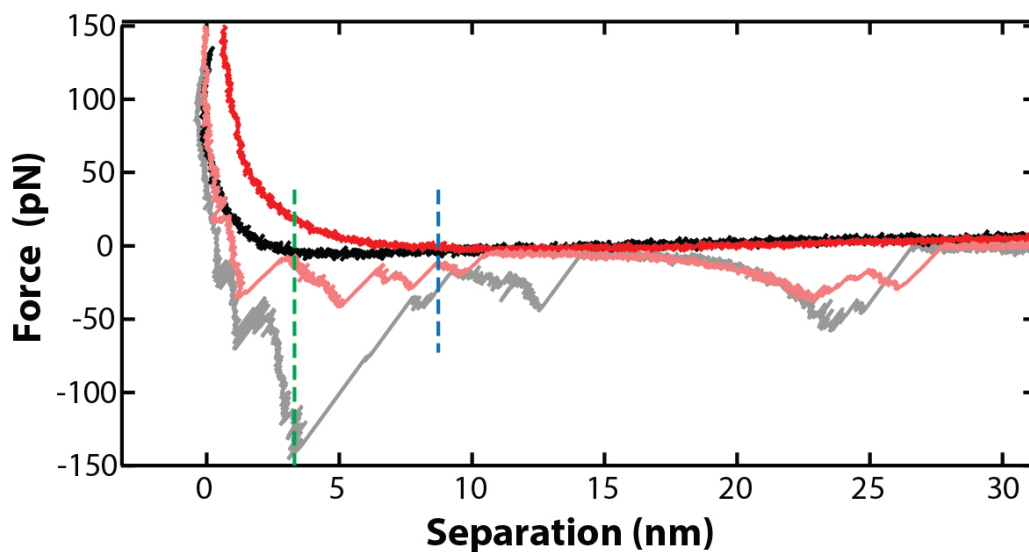
Figure 4-5: Interference artifact subtraction issues for AB helices. Two different unfolding records with the extension (tip motion towards the surface, darker traces) and retraction (tip motion away from the surface, lighter traces) are displayed, both prior to their interference artifact subtraction. The red curve has an interference artifact that diverges from its predictable sinusoidal behavior (blue dotted line) at an extension of ~8 nm, 5 nm before the black curve that diverges from the predictable sinusoidal behavior (~ 3 nm, green dotted line). This means that the interference artifact subtraction will yield a largely distorted AB helices unfolding for the red trace, but not for the black trace.
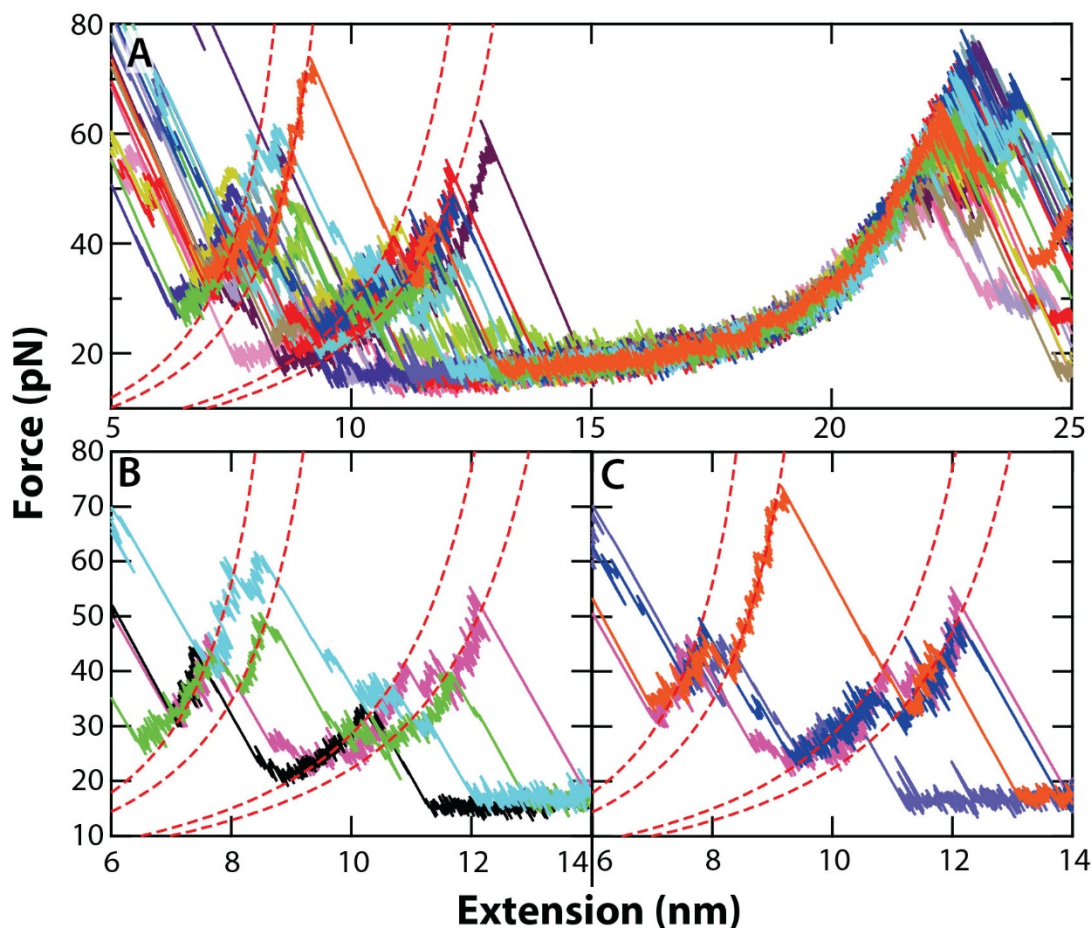
Figure 4-6: Partial determination of AB helices unfolding intermediates. Likely intermediate locations identified for the AB helices for the N-terminal protein are indicated by the dashed red lines. **(A)** All unfolding records used in our overall state determination were collected on a single day, and are shown in the interpretable region of AB helices unfolding (extensions > 5 nm). **(B)** Four unfolding records that feature dwells in multiple intermediates identified. **(C)** Four additional unfolding traces that feature dwells in multiple identified intermediates.

4.2.5) Improves state resolution compared to C-terminal experiment

As mentioned, two intermediates on downwards winding helices mapped to separations of less than 2 amino acids: I_{CD}^5 and I_{CD}^6. Examples of proximal transitions between these two intermediates are depicted (Fig 4-7) The contour length change between these two intermediates is 0.37 nm, which is (within uncertainty) the accepted value of the contour length increase by adding a single amino acid to an unstructured polypeptide chain. As established in section 3.3.1, the calculation by which one converts a contour length change to a number of unfolded residues incorporates the change in the contour length and
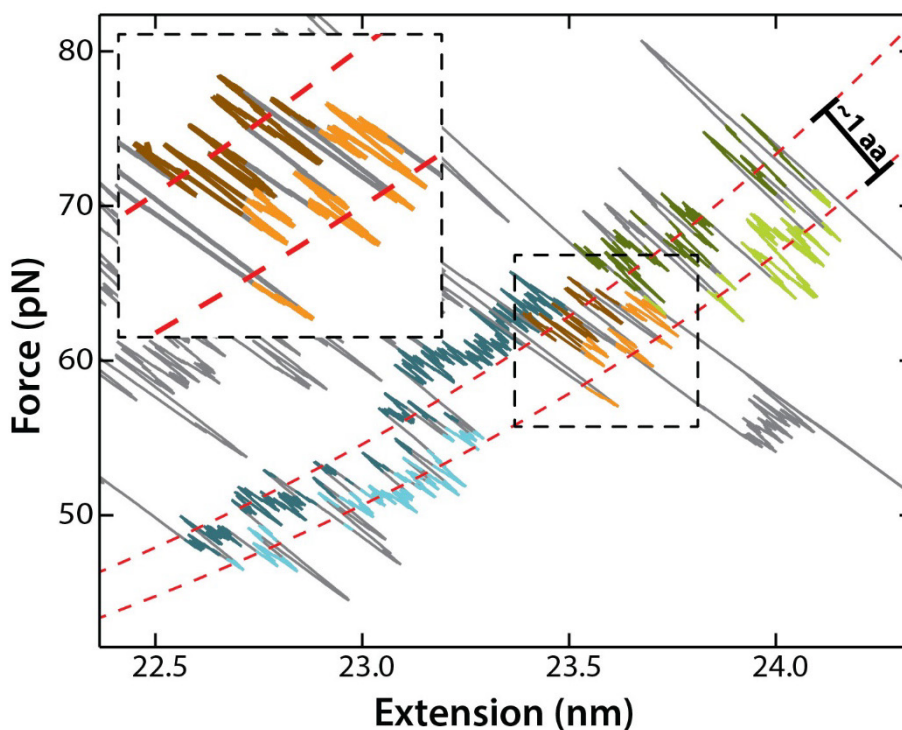
Figure 4-7: **Small contour length change observed.** Three unfolding traces (blue, orange, and green), demonstrating transitions between I_{CD}^5 and I_{CD}^6. The segments assigned to I_{CD}^5 are depicted with darker shades and the segments assigned to I_{CD}^6 are depicted with lighter shades. **(Inset)** A zoom in on the black dotted box, depicting orange trace without the other traces. These traces depict transitions between states separated by a contour length of 0.37 nm, which, for a globular protein, corresponds to the contour length change that a single amino acid unfolding demonstrates.

the change in the transition residue Z position (*49*), which in this region is ~0.15 nm/residue.

Incorporating this latter term, the change in contour length equates to the addition of ~1.4 ± 0.1 aa. While

a single amino acid unfolding is not within uncertainty of this calculation, the contour length change of

0.37 nm does correspond to a single amino acid in proteins that do not have changing transition residues

height (i.e. globular and non-integral membrane), indicating that use of these cantilevers in future

experiments can detect truly one amino acid unfolding steps (assuming the protein unfolds in intervals

that small).

4.2.6) CD helices intermediates proximal transitions

By reporting the densest array of states to date, reasonable questions of the certainty we have of

the distinctness of the states. Once again, there are frequent enough proximal transitions that we can

construct a "ladder figure" of proximal transitions, showing the distinctness of every proximal

intermediate (Fig 4-8). Ladder figures support the existence of every state by showing a transition

surrounded by sufficiently long dwells on each side, showing that the change was not the result of

transient noise. This lends credence to the existence of (at least) 20 intermediates. Note that for states with

small lifetimes, and lower occupancies (e.g. I_{CD}^{20}), that the lifetime of the depicted states is necessarily

shorter.
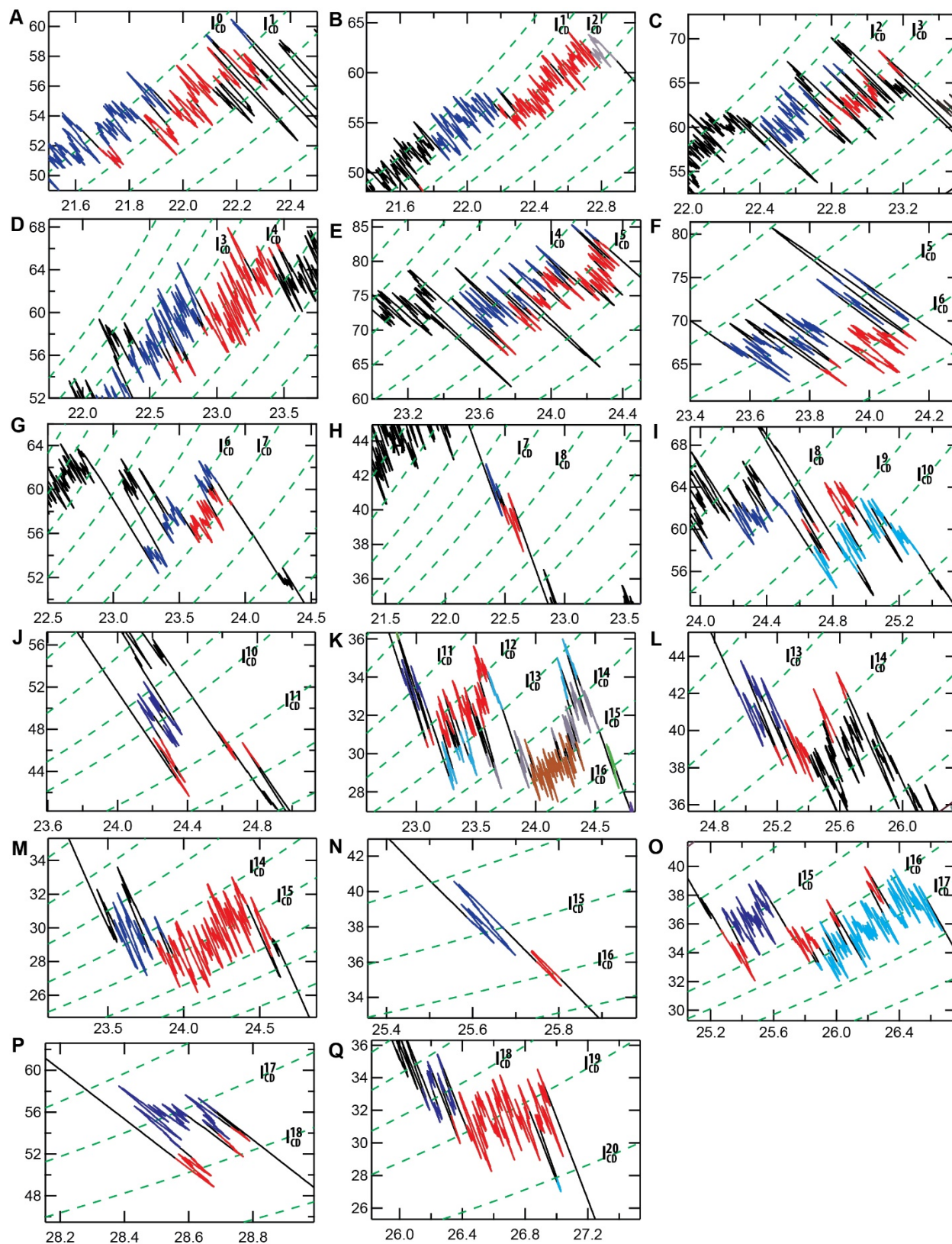
Figure 4-8: Demonstration of intermediates via proximal transitions (previous page). The FECs shown in the subsequent panels illustrate resolution of closely spaced states. Green lines indicate WLC models of the indicated contour length (in nm) for each intermediate. (Top and right). These FECs link adjacent states within the unfolding pathway of the CD helix pair, with **(A)** I_{CD}^0 and I_{CD}^1; **(B)** I_{CD}^1 and I_{CD}^2; **(C)** I_{CD}^2 and I_{CD}^3; **(D)** I_{CD}^3 and I_{CD}^4; **(E)** I_{CD}^4 and I_{CD}^5; **(F)** I_{CD}^5 and I_{CD}^6; **(G)** I_{CD}^6 and I_{CD}^7; **(H)** I_{CD}^7 and I_{CD}^8; **(I)** I_{CD}^8, I_{CD}^9 and I_{CD}^{10}; **(J)** I_{CD}^{10} and I_{CD}^{11}; **(K)** I_{CD}^{11}, I_{CD}^{12}, I_{CD}^{13}, I_{CD}^{14} I_{CD}^{15}, and I_{CD}^{16}; **(L)** I_{CD}^{13} and I_{CD}^{14}; **(M)** I_{CD}^{14} and I_{CD}^{15}; **(N)** I_{CD}^{15} and I_{CD}^{16}; **(O)** I_{CD}^{15}, I_{CD}^{16} and I_{CD}^{17}; **(P)** I_{CD}^{17} and I_{CD}^{18}; **(Q)** I_{CD}^{18}, I_{CD}^{19} and I_{CD}^{20}. The portion of the trace shaded blue in each panel represents the more folded state, and the red trace represents the more unfolded state (except in **I, O** and **Q**, where cyan represents the most unfolded state, and **K** where cyan represents the second unfolded state, grey represents the third unfolded state, brown represents the fourth unfolded state and green represents the fifth unfolded state), with black indicating regions of the trace not assigned to the states of interest. Records smoothed to 10 kHz using a Savitzky-Golay filter. While these segmentats were originally derived by previously discussed methods (see Section 3.10.2), a hidden Markov model was able to approximately recreate these segments.

4.3) Residue assignment & comparison to C-terminal intermediates

4.3.1) Residue assignment of N-terminal intermediates

Using the same assignment strategy as previously established (Section 3.10.3), we assigned each of the listed intermediates to residues within the structure of bR. The resulting intermediates are identified within Fig 4-9 and are listed in Table 4-1.

There is a dense network of intermediates located throughout the second and third pair of helices (C & D and E & F, respectively) while an absence of intermediates located within the G helix, in agreement with Kessler-Gaub (*49*). The quantity of intermediates within helix D and helix F are somewhat surprising, as these are helices that unwind upward. One would expect that the relative entropic favorability (Section 3.3.4) of unwinding helices in this configuration would make their state lifetimes below detectability thresholds, as we presumed the case in the C-terminal unfolding experiments (with the assumption of frequent, difficult to detect refolding occurring). We will go into the possible physical causes of these intermediates later in the chapter (Section 4.2.3 & Section 4.4.2). Also assigned are 4 intermediates within the B helix, all within the portion of the helix closer to the cytoplasmic interface. It is somewhat interesting that no other intermediates were detected within the extracellular half of the B helix. One possible explanation is offered by Yamada et al. (*76*), which predicts a significant deviation from FRINL behavior, making intermediates less distinct, and therefore difficult to detect.
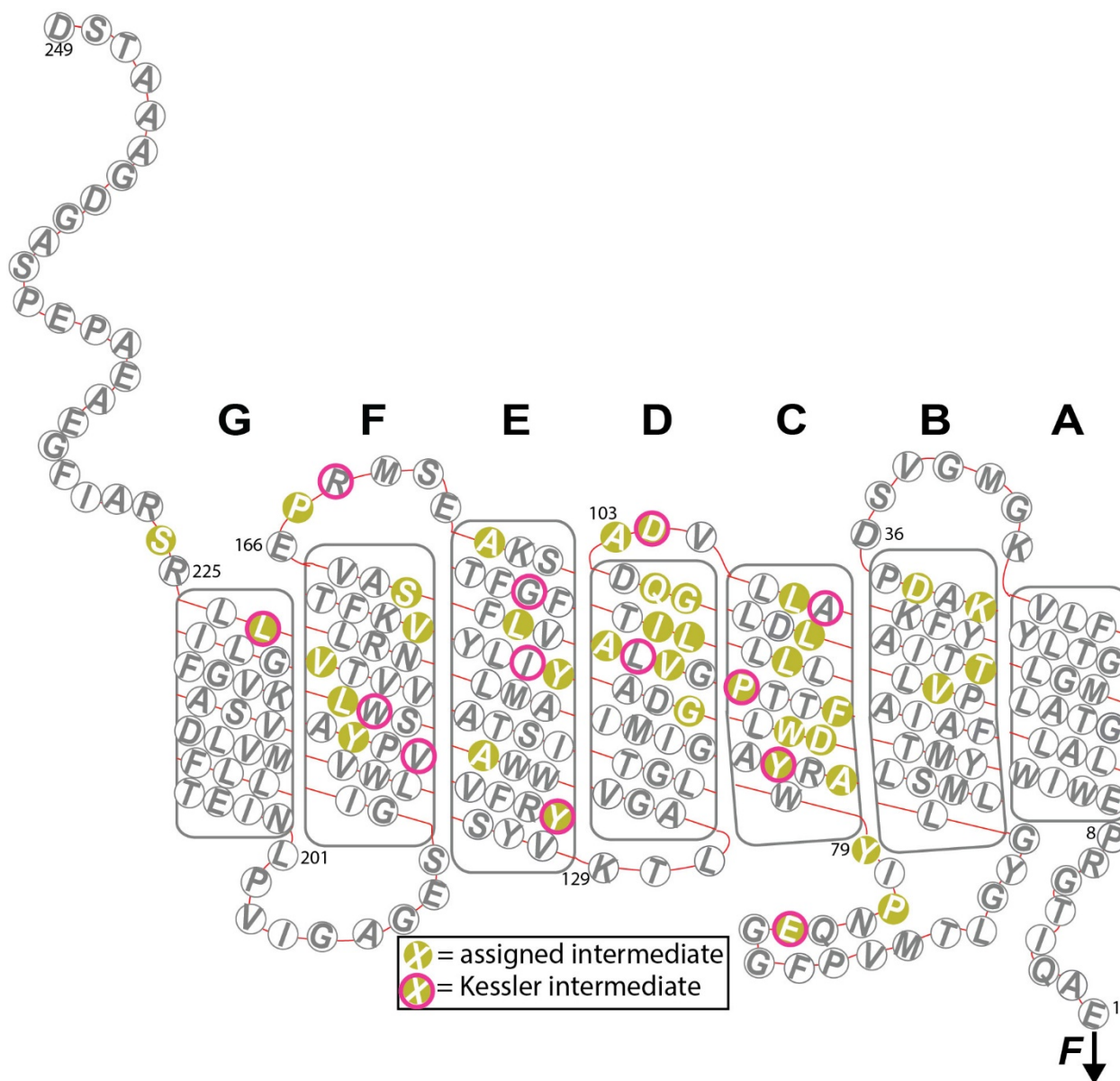
Figure 4-9: **Intermediate assignment from N-terminal experiment.** N-terminal unfolding intermediates mapped onto the primary and secondary structure of bR. Pink circles correspond to locations where unfolding intermediates were located in Kessler-Gaub 2006 (*49*). Labels of residue number are placed at the start of each helix. The application of force to induce unfolding is labeled on the N-terminus.

Table 4-1: Structural parameters of N-terminal intermediates (Next page). The contour length, structural position, probability, average unfolding force of each intermediate, and median dwell time detected in each state are listed. The contour lengths are based on assignments with assigned contour lengths of 28.25 nm for I_{CD}^0 and 77.8 nm for I_G^0. The presented data were derived from 33 molecules for the CD, EF and G helices and 19 molecules for the AB helices. The median dwell time detected in each state only include at forces above 20 pN, due to the prevalence of non-WLC behavior below 20 pN.

	State	Contour length ± s.d. (nm).	Nearest AA	Occupancy rate (%)	Mean unfolding force ± S.E.M (pN)	Median dwell (μs)*	Description (Exact residue)
B Helix	0	10.3 ± 0.3	38	63	45 ± 2	1900	B helix start (36)
	1	11.3 ± 0.1	40	53	45 ± 3	3741	
	2	14.8 ± 0.2	47	79	37 ± 2	1500	
	3	15.9 ± 0.7	49	68	43 ± 3	1307	
	0	28.2 ± 0	74	100	60 ± 3	24947	Beta sheet start
	1	28.5 ±0.2	77	91	63 ± 2	2078	Beta sheet end
C Helix	2	28.8 ± 0.1	79	100	63 ± 2	877	C helix start
	3	29.0 ± 0.1	81	94	64 ± 2	598	
	4	29.4 ± 0.2	83	94	61 ± 2	595	
	5	29.8 ± 0.1	85	88	63 ± 2	315	
	6	30.2 ± 0.1	86	85	60 ± 2	344	
	7	30.6 ± 0.2	88	79	58 ± 2	266	
	8	31.1 ± 0.2	91	67	56 ± 2	253	
	9	31.6 ± 0.2	93	48	56 ± 2	477	
	10	32.0 ± 0.2	95	27	50 ± 4	416	
	11	32.8 ± 0.2	99	27	50 ± 3	231	C helix end (101)
	12	33.4 ± 0.2	102	33	47 ± 2	106	
	13	34.1± 0.3	103	45	44 ± 3	333	
D Helix	14	34.9± 0.3	105	73	45 ±2	606	D helix start (104)
	15	35.6 ± 0.3	106	76	40 ± 2	272	
	16	36.3 ± 0.3	108	51	37 ± 1	80	
	17	36.9 ± 0.3	109	54	39 ± 2	165	
	18	37.6 ± 0.3	110	30	34 ± 3	245	
	19	38.5 ± 0.4	112	18	33 ± 3	167	
	20	40.3 ± 0.4	116	18	26 ± 2	54	
E Helix	0	48.3 ± 0.6	133	70	32 ± 1	14546	E helix start (130)
	1	49.6 ± 0.6	139	70	40 ± 2	12070	
	2	51.2 ± 0.5	147	94	46 ± 3	7768	
	3	52.4 ± 0.5	152	97	43 ± 2	3968	
	4	54.0 ± 0.5	159	79	41 ± 1	661	E helix end
	5	55.5 ± 0.6	165	76	39 ± 2	851	
F Helix	6	57.6 ± 0.7	169	97	37.4 ± 0.9	698	F helix start
	7	59.8 ± 0.7	173	64	36.7 ± 2	1578	
	8	61.6 ± 0.9	177	73	37 ± 1	4054	
	9	63.6 ± 0.5	181	42	26 ± 2	979	
	10	65.9 ± 0.5	185	51	26 ± 2	213	
G	0	77.8 ± 0	223	100	46.1 ± 0.2	40116	
C-term tail	1	79.0 ± 0.6	226	94	44.1 ± 0.4	2962	G helix end (225)
	2	80.1 ± 0.5		94	43 ± 2	1628	
	3	81.3 ± 0.7		85	42 ± 2	1397	
	4	83.3 ± 0.9		79	42 ± 2	1591	
	5	88 ± 2		21	51 ± 2	461	

*Above 20 pN

4.3.2) Residue comparison to C-terminal intermediates

Intrahelical bonds can be detected by identifying intermediates that map to the same (or proximal) residues when pulling in both C and N terminal unfolding experiments. To facilitate this, we combine the intermediates identified from both pulling directions (Fig 4-9 and Fig 3-4), into a single figure (Fig 4-10). Note that we add a few more intermediates mapped to the C-terminal data in helix G from additional unpublished work done pulling on the C-terminus with a specific attachment by Hao Yu (*109*).

A quick overview of the residues identified in both pulling directions reveals a few expected amino acids, such as proline (P77 and P91), due to the conformational rigidity that makes it a favorable location for force-induced breaks in the secondary structure. The rest of the co-incident intermediates cannot be easily explained by residues with high energetic costs for form a helix, as many of them are alanine or leucine (L95, A139, L146, L152, A160) which are favorable to the formation of α-helices (*110*) and thus would be expected to be unfavorable location for unfolding intermediates. This lack of consistency may indicate that the phenomenon causing intermediates may not be as simplistic as being caused by individual residues (except in the case of proline). This hypothesis can be verified by doing force spectroscopy on point mutants for comparison.

In terms of the location of the intermediates solely in the C-terminal unfolding experiments or solely in the N-terminal unfolding experiments, a few interesting trends do emerge. Every interhelical loop opposite the pulling side (not within the first peak) has at least one intermediate mapped (cytoplasmic for N-terminal pulling experiments, extracellular for C-terminal pulling experiments). As discussed in section 3.6, this can plausibly be explained by hydrophobic interactions between hydrophobic residues that transition out of the nearside membrane interface, or hydrophilic residues that transition across the far-side membrane interface. A discussion of this takes place in section 4.4.2.
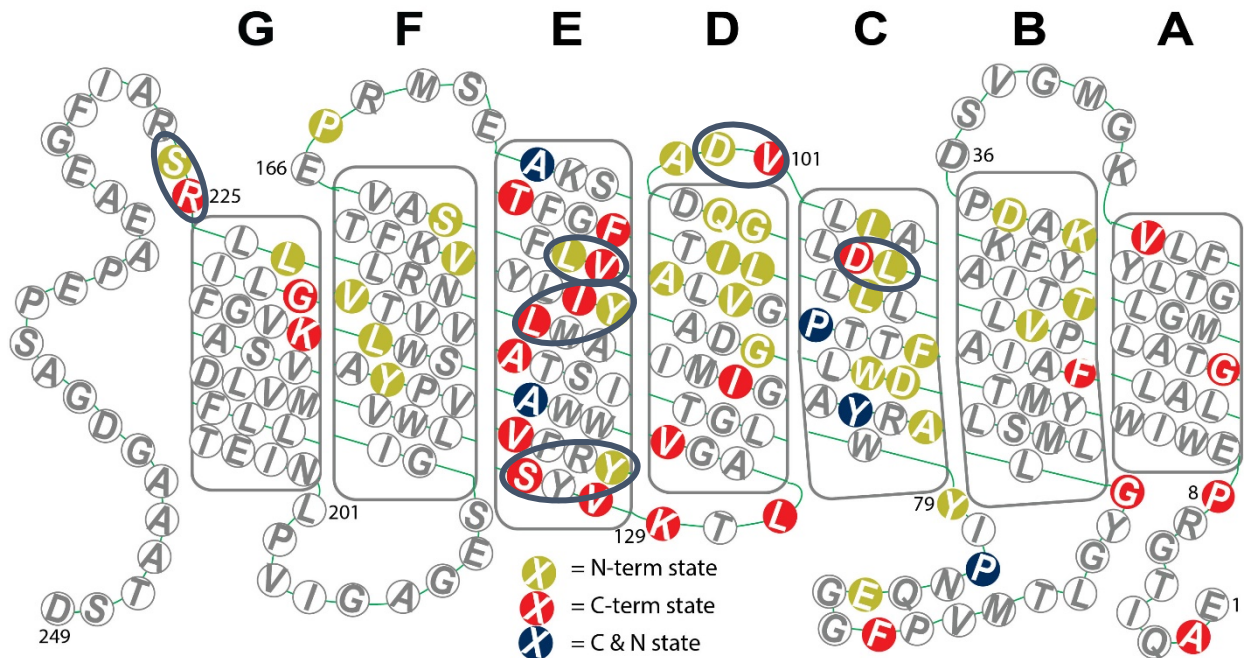
Figure 4-10: Comparison of intermediate locations for C and N-terminal unfolding experiments. N-terminal unfolding intermediates (yellow) and C-terminal unfolding intermediates (red) mapped onto a cartoon representation of the primary and secondary structure of bR. Residues that map to the location of intermediates from both directions are shown in dark blue. Intermediates that do not directly agree, but agree within uncertainty are represented by dark blue circles. Note that the C-terminal tail is oriented down here for compactness, and not to indicate that it is embedded in the membrane.

One of the most surprising results from either experiment is the density of intermediates located on the cytoplasmic side of the E helix for the C-terminal experiments and the cytoplasmic side of the D helix from the N-terminal unfolding experiments. The physical proximity of these two regions suggests that there might be some interhelical interactions between them (Fig 4-11).
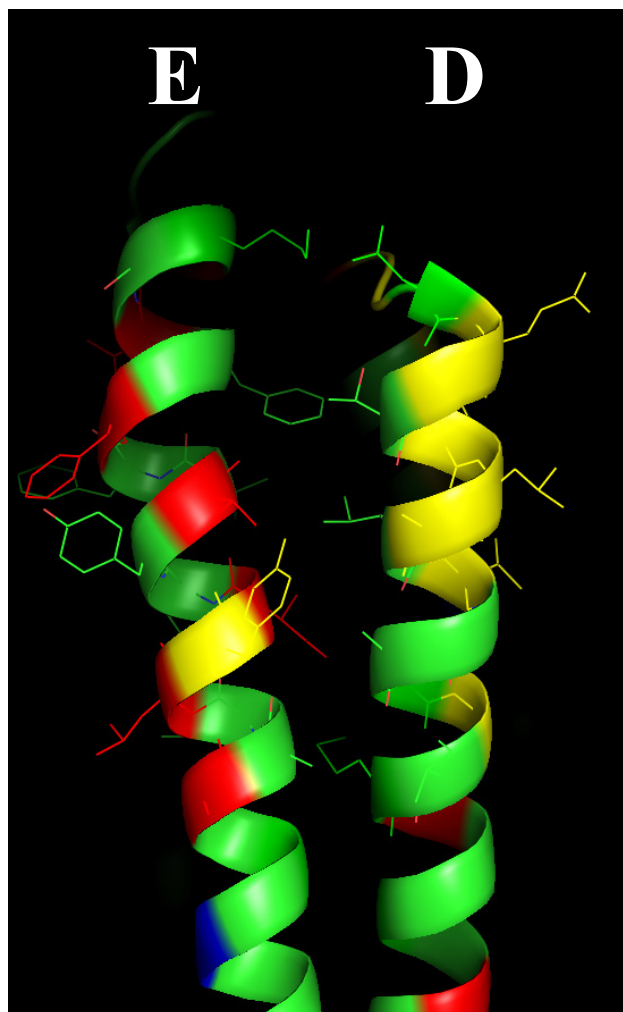
Figure 4-11: **Proximity of dense regions of intermediates suggests interaction.** Cartoon depiction of intermediates as observed on the cytoplasmic half of the E and D helices. Side chains are depicted in wire mode. The density of C-terminal intermediates in the E helix (highlighted in red) and the density of N-terminal intermediates in the D helix (highlighted in yellow) are shown here. The exact residues assignment of these intermediates, which includes residues both facing towards and away the ED helices interface, does not strongly suggest interactions between these two regions. But considering the uncertainty in the assignment (~1 aa), these intermediates could correspond to residues facing each other.

4.4) Exploration of causes of intermediates

4.4.1) Comparison to known or predicted interhelical bonds

A main goal of this study was to combine the set of intermediates from both pulling directions in order to shed light on the causes of the individual intermediates. Specifically, by comparing the intermediates detected against known and predicted interhelical hydrogen, intermediates largely stabilized by interhelical bonds can be identified. Note that for an interhelical bond explanation be consistent, the N-

terminal intermediate must occur closer to the N-terminus than the C-terminal intermediate, and vice versa.

Comparing these two sets, there are four pairs, each comprised of one C-terminal and one N-terminal intermediate (or with a residue within uncertainty) that corresponded to a interhelical hydrogen bonds that has been experimentally verified via an ensemble mutation study (*46*), or predicted from structural determinations (*47*, *111*). They include T46/D96, D85/K216, D104/K159, and T170/S226 (Table 4-2, "Matched pairs"). Additionally, there are two additional pairs of intermediates comprised of intermediates 2 aa from assignment (1 residue beyond uncertainty) Y83/G122, and M118/S141 (Table 4-2, "Near pairs"). These intermediates contain one intermediate that resides at the very end measurable range of the upward winding helices, where assignment (due to non-FRINL) is assumed to be most uncertain, and thus may well be identified. All six of these pairs are depicted within the 3D bR structure in Fig 4-12.

Structurally, these interhelical bonds are located throughout the helices, except in helix A (due to the lack of intermediates detected in the A helix in this experiment). It is noteworthy that there is a lack of detected interhelical bonds near the extracellular interface, as they all either are located near the cytoplasmic interface, or the membrane core (center). More generally, the number of interhelical bonds detected seems low, as more bonds have been proven via double-mutant cycle analysis (*46*). A few possible explanations exist. First, it is unclear if the predicted hydrogen bonds actually stabilize the structure. For example, one bond, which was predicted by crystal structure, was measured to be destabilizing to the protein (the mutated protein actually became harder to unfold) (*46*). Our technique identifies interhelical bonds that strongly stabilize structure. A second possibility is that many hydrogen bonds were missed due to the regions of the protein where the assay is more sensitive intermediates, namely the cytoplasmic side of the A, C and E helices in C-terminal unfolding experiments, and the extracellular side of the C, E and G helices for N-terminal unfolding experiments. Unfortunately, due to the arrangement of the helices, these two sides tend to be far apart, thus detection of half of the pair should be easy, and detection of the other half difficult. The lack of state resolution in the C-terminal

experiment within the F helix, where most of the predicted extracellular-proximal intermediates are located, fits with this explanation. A future study based on a cleaved bR, where an even number of helices are pulled on would rectify this lack of state resolution with the B, D and F helices, similar to work with a cleaved bR construct in the original bR unfolding study (*17*) . While this correlation exists between the previously identified hydrogen bonds and our identified intermediates, further SMFS experiments on point mutants are necessary for stronger proof of the interhelical bonds.

Table 4-2: Predicted interhelical bonds predicted and observed in experiment. The N-terminal and C-terminal intermediates are listed for each interhelical pair that we observed that corresponded one predicted by MD structure (*47*, *111*), or observed in a double-cycle mutation study (*46*). The last column, labeled photocycle, notes the inclusion of at least one intermediate known to participate in the photocycle of bR.

	N-terminal residue	C-terminal residue	N/C helix location	Position	Photocycle
Matched pairs	T46	D96	B/C	Near CP interface	x
	D85	K216	C/G	Core	x
	D104	K159	D/E	Within CP interface	
	T170	S226	F/G	Within CP interface	
Near pairs	Y83	G122	C/D	Core	x
	M118	S141	D/E	Core	
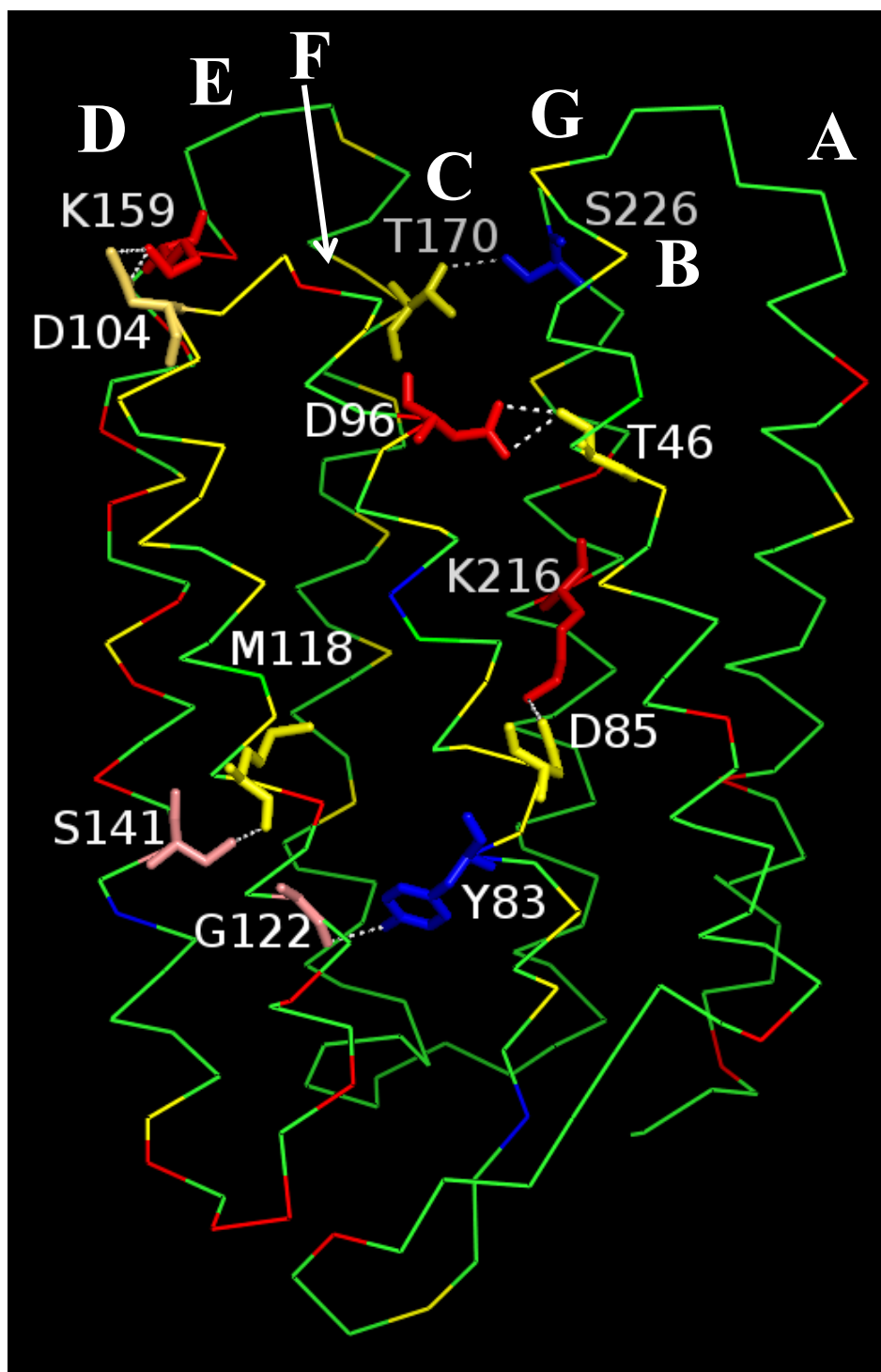
Figure 4-12: Interhelical bonds detected. Ribbon depiction of bR, with the identified intermediates from both the C and N terminal experiments. The same color coding used in Fig 4-10 (red for C-terminal intermediates, yellow for N-terminal intermediates, and blue for both) was used here. Within these intermediates, we identified six pairs (labeled in grey) that correspond to previously observed or predicted hydrogen bonds (shown in dotted white lines). Three of these highlighted intermediates (G122, M118, and S141) are near the identified intermediates, and not exactly assigned by the analysis.

4.4.2) Hydrophobic contributions

Besides hydrogen bonds, another major cause of intermediates is hydrophobic interactions between the membrane and the residues. Using the same method of analysis adapted from Yamada et al. (*76*) as outlined in Section 3.6 we calculated the approximate change in the free energy due to hydrophobic interactions for each residue to transition from folded to unfolded (Fig 4-13).
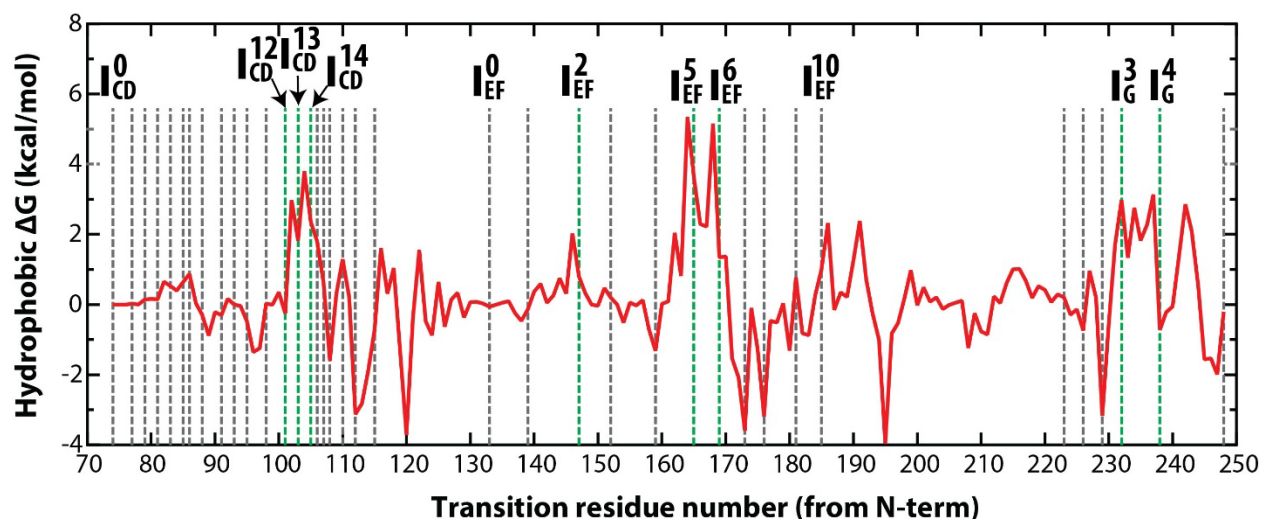
Figure 4-13: Hydrophobic energy changes for each residue unfolding, N-terminus. The above plot depicts the change in the hydrophobic contributions to the free energy associated with unfolding single amino acids in bR. The dotted lines are the locations of the measured transition residues, with major states labeled and intermediates that are coincidental with large positive changes in free energy both labeled and green.

Within Fig 4-13, there are three regions where there are large changes in free energy (aa102-105, aa164-168, and aa232-237), which correspond to regions where strongly hydrophobic residues are extracted from the extracellular membrane interface (W86, L87, F88, L91, L92 [for aa102-105], L146, F153 [for aa164-168], and F218 [for aa232-237]) and strongly hydrophilic helices are being pulled into the cytoplasmic interface (D102, D104 [for aa102-105], K159, E161, E166 [for aa164-168] and E232, E234, E237 [for aa232-237]). This offers a possible explanation for the presence of intermediates within the cytoplasmic loops, as I_G^3 and I_G^4 coincide to regions that have large energetic costs associated with desolvating hydrophilic residues.

There has been a long lasting debate in the field about the contacts exhibited by the C-terminal tail (*112–114*). Suggestions that the C-terminal tail plays a role in the structure of the protein are provided by assays which have modified the C-terminal tail. For example, one study deleted the C-terminal tail and observed changes in the packing of the outer helices but not in the retinal core via electron microscopy (*114*). Another NMR study observed conformational changes in the CD loop position in the point mutant A228G (*115*). These studies suggest that these the C-terminal is interacting with many parts of the protein. The results here suggest that the G-helix exhibits some positional stability arising from interactions with the membrane alone, as all other possible contact have been unfolded by this point.

4.5) Pathways

The pathways that the protein traversed as it unfolded the C and D helices (the second peak) shared many of the same qualitative features demonstrated in the C-terminal unfolding experiment's second peak: both unfolding and refolding transitions are most common near the major state yet refolding persists to the end of the helix pair, albeit less frequently. A subtle difference between the CD unfolding pathway compared to the C-terminal ED unfolding pathway is that all unfolding trajectories exhibit at least one Helix D (I_{CD}^{14}-I_{CD}^{20}) unfolding intermediate.

The same analysis was repeated on the EF and G helices (Fig 4-15). A significant difference compared to the third peak (CB helices) pathway (Fig 3-12, left) from the C-terminal experiments is the relative lack of transitions from I_{EF}^{0} and I_{EF}^{1}, with very little refolding transitions back to I_{EF}^{0} (9%) in particular. In addition to being less frequently occupied (70%) than in Kessler-Gaub (100%, obligate), we find that it is transiently occupied when it is occupied and, additionally, is rarely transitioned back to despite the low average rupture force (~34 pN). Interestingly, the relatively lower occupancy rate from I_{EF}^{0} and I_{EF}^{1}, is shared by I_{ED}^{7}, I_{ED}^{8} and I_{ED}^{9} from the C-terminal unfolding experiments, the intermediates with the common transition residues. This shared lower transition and occupancy rates implies that the energy barriers that generate these intermediates are weaker than those that generate the other intermediates in helix E.
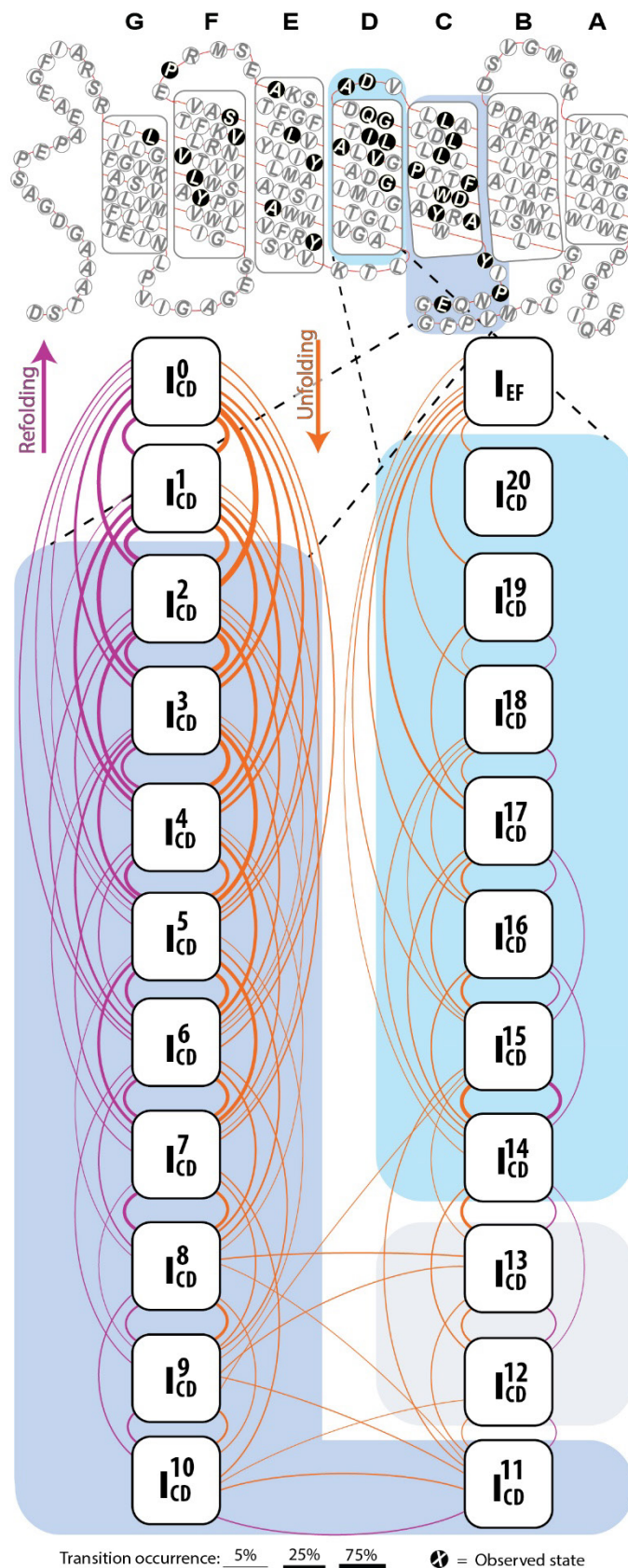
Figure 4-14: CD helices unfolding pathways. Cartoon of the primary and secondary structure of bR. Locations of observed folding intermediates are shown by residues with filled in circle. **(Bottom)** Each helix pair diagram depicts an observed intermediate state within helices CD, with connecting lines representing transitions observed in at least 4 (of 33) different molecules containing a total of 899 transitions. Orange lines represent unfolding transitions, and purple lines represent refolding transitions. Line-widths represent the frequency of observing a particular transition with a legend at the bottom of the figure. For simplicity, repeat transitions are not included in this representation
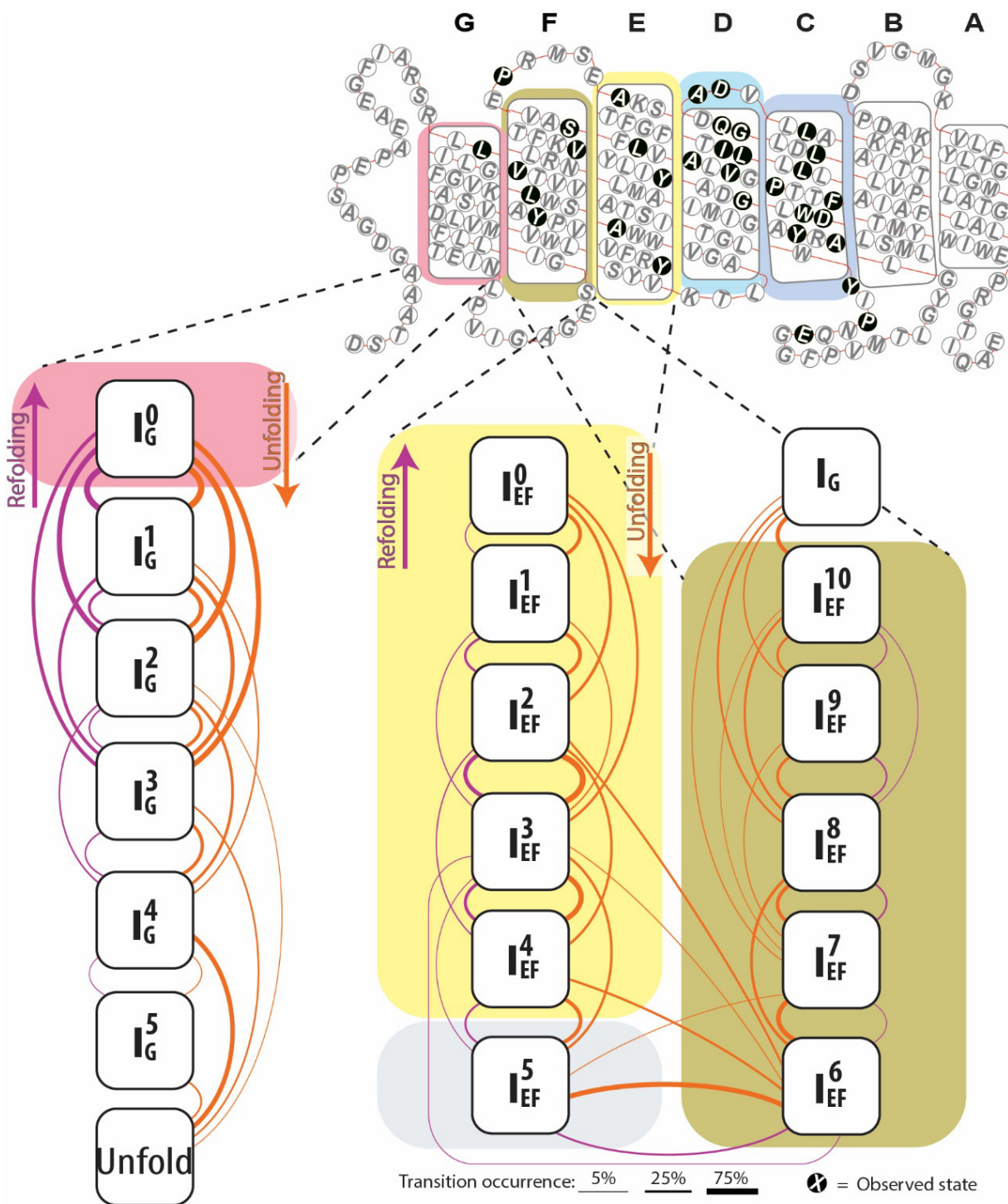
Figure 4-15: The unfolding pathway for the EF helices pair and the G helix. (Top) Cartoon of the primary and secondary structure of bR. The seven helices (A–G) are shown with their constituent residues, including locations of detected unfolding intermediates (residues with a filled-in circle). Helix G,F,E,D and C are highlighted in red, dark green, yellow, light blue, and dark blue, respectively. **(Right)** Unfolding pathway diagram for the EF helix pair deduced from 368 transitions in 33 molecules. Each helix bundle diagram represents an observed intermediate, with connecting lines representing transitions (observed in more than 3 separate unfolding traces). With our improved resolution, I_{EF}^0 is no longer

Figure 4-15 (cont.) observed as an obligate intermediate (Table 4-1); rather, I_{EF}^2 is occupied at a higher percentage. (**Left**) Unfolding pathway for the G helix deduced from 288 transitions in 39 molecules. The last state labeled "Unfolded" represents the removal of the bR molecule from the membrane. The orange interior lines represent unfolding transitions, while the purple lines represent refolding traces. Line thickness represents the transition occurrence out of all molecules studied, so the percentage out of any given state is not 100%. For simplicity, repeat transitions are not included in this representation.

Lastly, one bit of analysis that we performed on the pathways was to examine whether the smoothing that was applied to the data is obscuring a near-equilibrium fluctuation behavior near the top of helix CD. Instead of performing Savitsky-Golay with a window size of 501 points (100 µs), we used a window size of 51 (10 µs). While a few more short-lived back and forth transitions were observed in traces that demonstrated near-equilibrium fluctuations when normally filtered, no molecules that showed an absence of near-equilibrium fluctuations at normally filtered conditions showed near equilibrium fluctuations with less filtered records. This suggests that near-equilibrium fluctuations are not a ubiquitous behavior in the unfolding of bR, but rather a discrete condition that is occasionally present. This could be possibly be the result of the force range entered into by the transitions: if the major state stochastically unfolds at a high force, then the forces that are exerted on the molecule when in the same extension region where near-equilibrium fluctuations tend to occur are too high to have refolding. Follow up analysis of curves that exhibit near-equilibrium fluctuations compared to major state rupture force will be accomplished.

4.6) Conclusions

The N-terminal pulling assay was predicted to contain to a smaller set (compared to the C-terminal unfolding experiment) of very stable intermediates (*79, 116*). We observed the opposite: a slightly larger number of intermediates, with lower rupture forces than were observed in the C-terminal unfolding experiments. We note that the comparison of the unfolding forces is not perfect, as C-terminal pulling assay used softer cantilevers and a linker. This increase in the number of intermediates relative to the C-terminal unfolding assay suggests that the impact of retinal on the protein structure is different than previously predicted, not necessarily weakening, but rather exchanging the rigidity of a few strong bonds for a network of weaker bonds. Considering the range of motion the protein is known to undergo during

the photocycle (*58*), perhaps this array of weaker bonds aids the function of the protein. Additionally, the combined data has indicated that hydrophobic forces cause the intermediates localized interhelical loops on the opposite side of the membrane, indicating that this assay is able to measure intermediates caused by multiple different causes. One distinct difference is that the experiment conducted in this chapter is not a perfect analogue to the C-terminal unfolding experiments, due to the presence of the retinal for the majority of the unfolding process. When forcibly unfolded from the C-terminus, the retinal is removed very early on in the unfolding progress, to the point where the unfolding of helices F, E, D, C, B and A helices are more akin to bacterioopsin (retinal removed bR). In the next chapter, we will repeat this N-terminal unfolding experiment on bacterioopsin, where the retinal is absent for the entirety of the unfolding.

4.7) Methods & techniques

4.7.1) Experimental results

Other than the changes to the attachment protocol, the pulling experiments proceeded in the same manner as performed in Chapter 3. 88 molecules were unfolded over five experimental runs. Unfortunately, due to inconsistencies in the FIB modification process (see section 1.1), the cantilevers used during the first three runs had poor sensitivity (invOLS, see section 5.2.1 for more explanation) and thus high force noise, making their ability to be utilized to state determination limited. Of the 44 remaining unfolding traces, we were able to remove the interference artifact from 33. Albeit a lower number of records utilized than in the C-terminal work (98), the records they exhibit better quality, specifically force noise. One cantilever had a stiffness of 25 pN/nm, and the other had a stiffness of 19 pN/nm, ~ 45% less stiff than the average cantilever stiffness from the C-terminal experiments.

4.7.2) DBCO:bR labeling

Utilizing the DBCO-Azide chemistry listed in Section 2.3.4, a remaining question was the ideal labeling ratio of DBCO to bR to achieve frequent single attachments, and infrequent double attachments. There are 3 competing parameters to optimize: the first is the relative number of bR's that will be within proximity of the cantilever during our attachment protocol: with a nominal tip radius of 20 nm, and approximately 1 bR per every 5.71 nm^2, that results in ~220 bR molecules that will be directly under the tip apex. This does not count the radius of gyration from the tip functionalization of the protein. The second competing force is the efficiency of formation of the maleimide-thiol bond used to affix the DBCO to the protein, which is not known (due to variability in the bulk from PM aggregation), nor was attempted to be measured directly. The last consideration is the kinetics of the formation of the DBCO-Azide bond, which is relatively slow (3-5 s at ~150 pN). We used an empirical approach to determine determined the optical value. In the end we settled upon two different relative concentrations were utilized 1 DBCO : 10 bR, and 1 DBCO : 1 bR. Occasional double attachments were observed, more so for the 1:1 concentration than for the 1:10, but at a sufficiently low rate compared to single attachment.

4.7.3) PEG linker

Somewhat unexpectantly, we observed lower attachment frequency to the N-terminus that the C-terminal when using the same attachment scheme. To overcome this problem we used that emerged from this difficulty was the use of longer PEG linker in the Silane-PEG-Azide (SPA) construct used to functionalize the tip. In the work by Hao Yu, the attachment rate using PEG_{3400} was 3-4x greater than PEG_{600}.

Another change in the chemistry that significantly improved the attachment rate was switching from a maleimide-PEG1-DBCO construct, to a maleimide-PEG4-DBCO construct in the functionalization of the bR. Our attachment rate went up significantly to 2% rate over ~500 attempts on a viable patch. This suggests that part of our difficulty in achieving attachments was from the lack of conformational freedom of the DBCO, as a change from PEG1 to PEG4 will not significantly increase the radius of gyration.

4.7.4) Lack of PEG linker in measurements, and certainty of N-terminal attachments.

We note that despite the use of a longer PEG, the resulting Force-extension curves do not exhibit the expected shift of unfolding intermediates to longer extensions. Specifically, unfolding records made with PEG_{3400} linkers (nominal Lc~ 21.6 nm), yet the average contour lengths identified (I_{CD}^0 L_c~ 28.25 nm) are just slightly larger than the predicted contour length of the CD major state by itself (26.64 nm). Ensuring that the attachment is to the proper terminus of the protein is paramount, as it would complicate analysis if each curve had its individual contour length. A series of tests were performed to ensure that non-specific attachment was not occurring. The first one was to obtain long lived attachments for single bR attachments (Fig 4-16). Here, an attachment was achieved, and unfolded to 80 nm above the surface, making sure no evidence of double attachment was present. Then at 80 nm from the surface, the cantilever stopped. Ideally, an infinitely long attachment would be observed, but realistically at any sufficiently high force (>30 pN), the bR will extract of the membrane eventually, so one that persists for more than a second at forces >40 pN is the strongest proof reasonably attainable of a covalent bond. Indeed that is what is presented in Fig 4-16, which persists at 45 pN for 4 seconds. Thus, a non-covalent

bond is unlikely to be the cause. Other evidence that suggests the attachment that was observed was the long period of non-attachment that occurred prior to having successful attachment.

On the N-terminus side, as stated in Chapter 2, the lab's specific attachment protocol was a thiol-maleimide attached to a single PEG linker further attached to the DBCO group. The same protocol was successfully implemented on the C-terminus. On the N-terminal experiments, success was fleeting (attachment rate <0.1%). This rate was raised to (~2%) by lengthening the PEG linker to four-mer. This suggests that the DBCO is part of the bond that is being formed. More evidence to the same effect is the length of the PEG linker on the cantilever side. Two lengths of PEG linker were used in the Silane-PEG-Azide functionalization: 600 and 3400. For both the C-terminal and N-terminal experiments, PEG_{3400} achieved a higher attachment rate than PEG_{600}. For PEG_{600} the C-terminal had a low but meaningful attachment rate but there was no such success for the N-terminal pulling assay. Switching to a PEG_{3400} yielded a 2% or greater attachment rate was achievable using PEG_{3400} however. If a different attachment chemistry that excluded the PEG linker was occurring for PEG_{3400}, it stands to reasons that the different attachment chemistry would succeed regardless of PEG length.

The last test we conducted to ensure that the DBCO-azide attachment was occurring and was not interfered with by the APDMES functionalization. We functionalized two tips in the same batch of Silane-PEG-Azide, while also preparing a clean glass slide with APDMES, but without bR deposited on it. We then thoroughly imaged and indented one tip into the APDMES surface at high force, exposing the tip to more contact with APDMES than would occur on a typical experimental run. We then took that tip, as well as the functionalized, but never-exposed-to-APDMES cantilever, and used both on C-terminal functionalized bR on mica. Both showed an attachment positions varying by approximately 20 nm. Furthermore, both showed lower forces than anticipated. This suggests that the APDMES is not interfering with the attachment, but that something odd is occurring with the Silane-PEG-Azide construct, possible aggregation of the PEG.

With the variability in the extensions explored, the question of variability in the attachment point is addressed. The difference is the extension longer extension observed for Biolever minis compared to

the Biolever fast, and the difference in the attachment rate for PEG_{3400} and PEG_{600} suggests an answer: too many attachments. The Biolever Mini, with a smaller nominal tip radius than the Biolever Fast, has less surface area, thus will have fewer PEG-azides within bonding distance to the DBCO-functionalized bR. Both cantilevers have multiple attachments initially, evidenced by huge rupture forces. The rate of multi-attachment decreases as the azide sites become occupied, which happens quicker for a tip with a smaller radius (i.e. the Biolever mini). Successful single attachments occur as the tip becomes partially occupied. A sharper tip will have a geometry that will suppress multi-attachment quicker, but still have enough unoccupied sites within range of the surface that the singe attachment will still occur. The use of longer PEG linkers, with boost the ratio of single attachments by reducing the overall concentration of linkers, as well as having a larger portion of the tip surface area able to reach the bR. This is speculation, but does fit with all the evidence.
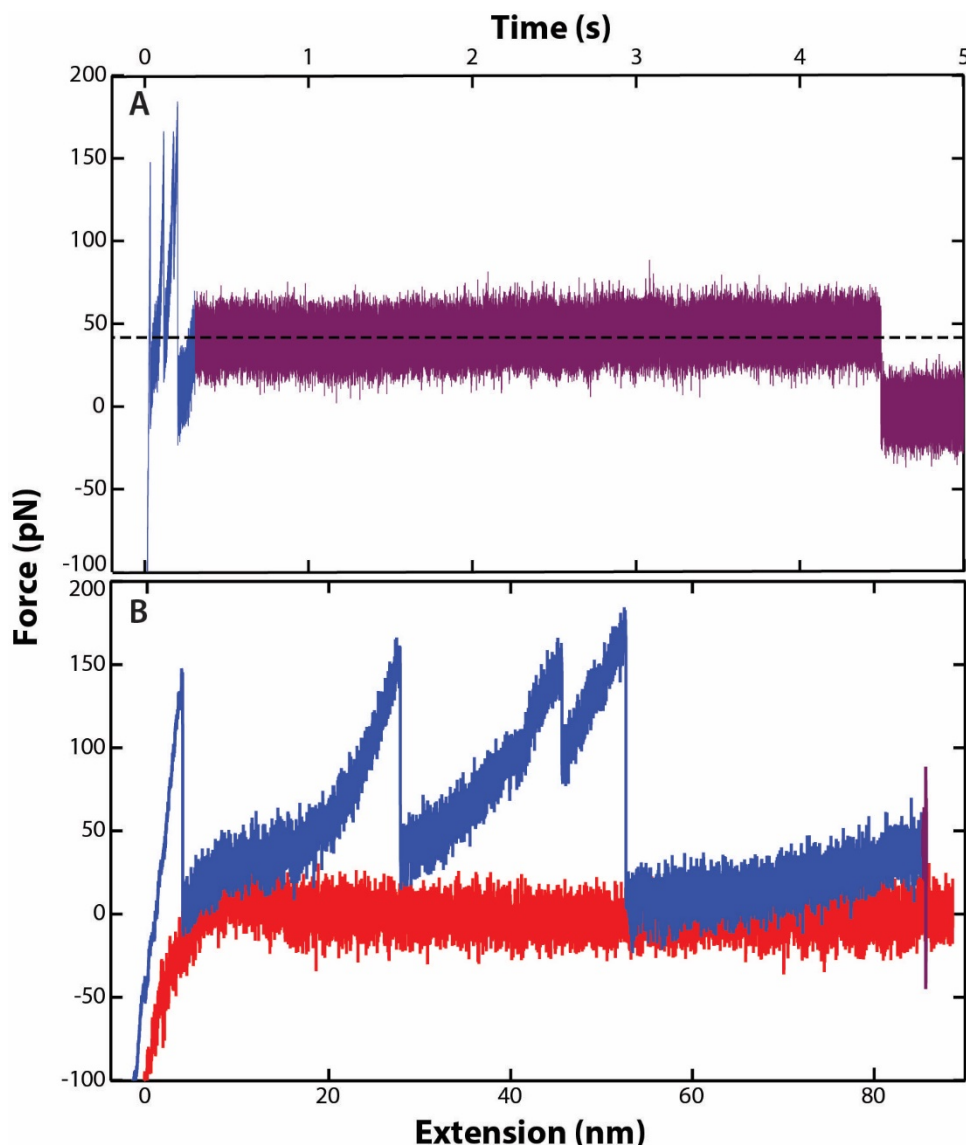
Figure 4-16: Specific attachment verification. Two plots of a single bR unfolding record that demonstrate long-lived covalent bonds. The extension (red), retraction (blue) and constant extension (purple) are depicted. **(A)** Force-Time plot and **(B)** Force-Extension plot. The protocol was to pull at a constant velocity until a certain extension is reached, then to hold the cantilever at that extension (80 nm). This resulted in a protein held at ~47 pN in the fourth unfolding peak that persisted for ~4 seconds. This is an unlikely result for a non-specific attachment chemistry, as they tend to not hold at any significant force (>30 pN) for greater than a second. This is one of 3 records with similar behavior. Note that the force and extension are greater here, as the cantilever used for this experiment was a Biolever Mini, a larger, stiffer cantilever (thus a higher loading rate, leading to higher rupture force) with a shaper apex (indicating that the attachment is likely to take place to a SPA molecule attached closer to the tip).

CHAPTER V: Extracellular forced unfolding of Bacterioopsin

5.0) Introduction

5.0.1) Ongoing work

The work presented in this chapter represents an ongoing study, and as such, the information presented can potentially change when published by a journal due to new results.

5.0.2) Introduction and motivation

After demonstrating a successful N-terminal bR forced unfolding assay, we investigated an open question in the field: the effect of retinal on the unfolding of bR. Retinal, a vitamin A aldehyde, is covalently linked to bR via a Schiff base at Lysine-216, located within the G helix. It is a vital molecule in the photocycle of the protein, as removal halts the protein's functionality. Visually, this is detectable by a loss of its distinctive opaque purple coloring, and for this reason is referred to by a different name: bacterioopsin (bO).

The forced unfolding of bO is more physiologically relevant than the forced unfolding of bR to the native folding of bR. During bR's native folding, the protein forms its secondary structure, inserts and orients in the membrane second, and partially forms its tertiary structure before the retinal is inserted into the protein (*117*). As most of this folding process takes place in the absence of retinal, bO unfolding mirrors this native folding energetics better than bR unfolding. Secondly, due to the sequential unfolding behavior of bR and the covalent attachment of the retinal to the G-helix, all C-terminal unfolding studies of bR will extract the retinal during the first helix extraction – prior to the extraction of the final 6 helices. Comparing the unfolding intermediates observed in these 6 helices (F, E, D, C, B, and A) in the C-terminal unfolding experiments to the N-terminal bO intermediates is more valid than comparing them to the N-terminal bR intermediates, where the retinal remains in the membrane until the extraction of the final helix. There is an open question of the influence that the retinal has on the structural stability of bR. It is firmly established that photoisomerization of the retinal drives the photocycle of bR, during which conformational changes occur (*58*). It has also been shown that removing the retinal from bR disrupts the

trigonal lattice in purple membrane (*118*). A computational simulation of bR predicted that the retinal interactions with the helices dictate the unfolding intermediates, leading to a few stronger intermediates, while unfolding in the absence of the retinal will have a larger amount of weaker unfolding intermediates (*79*). The results presented in chapter 4, a dense array of intermediates, already challenge this hypothesis. But a comparison of N-terminal bO to bR will be definitive. This chapters observes that the removal of the retinal largely does not affect the bR unfolding. This result is a significant departure from expectation and will require follow up work to confirm this null result.

5.0.2) Experimental background

The results from previous studies indicate that the retinal stabilizes the folded bR(*116*). A brief overview of these studies helps to establish the strength of this expectation. Previously, two SMFS experiments has been conducted on bO, one using standard cantilevers (*67*), and the other utilizing ultrashort cantilevers (*109*), both on the C-terminus. The latter study was conducted by the Perkins lab, and is currently in preparation. These two studies corroborated the stabilization provided by the retinal, with the standard cantilever study observing an increases in the occurrence rate of intermediates located within the G helix (*67*), and the ultrashort cantilever study observing a reduction in the rupture forces (*109*). It should be noted that there is no new intermediates relative to those observed for bR were identified in either study, and none of the existing intermediates changed positions. The lack of change in the unfolding in the other helices reflects the extraction of the retinal early on in normal C-terminal bR unfolding experiments (i.e. Chapter 3). Unfolding studies have been conducted on bO via non-SMFS assays. In a FRET study on unfolding via SDS, bR was found to have a stability nearly 4-fold higher than bO (23.4 kcal/mol vs. 6.3 kcal/mol, respectively) (*116*). Chemical denaturant studies have often differed from SMFS studies. This is especially true for membrane proteins, as chemical denaturants disrupt the lipid bilayer and weaken the hydrophobic interactions in order to induce unfolding, which is arguably less physiologically relevant than forced unfolding. NMR studies on bO yielded a similar spectrums as the bR spectrum, but with very broad peaks. This indicates that the protein still maintained its secondary structure, but is more flexible, with variability in its tertiary structure (*119*). Thus, although the prediction

of more, weaker intermediates is a logical extension of the above results, it has not been proven that this behavior will be evident in a SMFS study.

5.1) Preparation

5.1.1) Preparing bO

The process of removing the retinal from bR was first established 40 years ago (*120*), and depends on the presence of the water-soluble reagent hydroxylamine. The hydroxylamine can outcompete the Schiff base by which the retinal is attached to the protein, and form retinal oxmine. The hydroxylamine only has access to the Schiff base in configurations that the protein adopts during the photocycle (*121*). Thus, driving the photocycle greatly accelerates the process.

We replicated this process by mixing in hydroxylamine at 200 mM concentration with stock (N-terminal cysteine mutated, but not DBCO labeled) bR and then exposing it to a mercury arc lamp, with a green filter, for 12 hr. (to drive the process to completion), while the solution was placed in a water bath for temperature control. Before and during exposure to the lamp, we sonicated the solution to reduce aggregation that would likely reduce the hydroxylamine's ability to access the bR.

The completion of the process was monitored by UV-visible spectroscopy on a NanoDrop spectrophotometer (Thermo Inc.). The relative absorbance at 280 nm and 560 nm were recorded. The 280 nm peak arises from tryptophan absorption and is not impacted by retinal removal, thus is useful in determining the overall concentration of protein within the solution. The 560 nm peak represents the bR photocycle absorption and reduces as retinal removal progresses. We saw a reduction of this peak from 0.019 to 0.003 (an 85% reduction). We were unable to drive this peak lower, even by significantly increasing the concentration of hydroxylamine, or leaving the solution under the intense light for longer than prescribed, or by additional sonication. Unfortunately, this leaves the possibility that some of the measured unfolding will still have the retinal present, albeit infrequently (15%).

5.1.2) Functionalization

The functionalization was performed specified in section 4.1.2, but after the retinal removal process. Attempts to functionalize before removing the retinal failed to yield attachments. 1:1 bO:DBCO concentration was again applied.

5.1.3) Surface adhesion and density

The protein was mixed into buffers of 300 mM KCl 10 mM Tris-HCl buffered to pH 7.8, and deposited on a APDMES surface, as was observed in chapter 4. Noticeably, the patch size was smaller for the retinal removed PM than it was for the retinal-containing bR, presumably due to the additional sonication that was performed during the retinal removal process.

5.1.4) Experimental statistics

Force spectroscopy of bO was performed on three different days, and sufficiently high attachment rates were achieved on all 3 days. Due improvements in the consistency of FIB modification (compared to section 4.7.1), all resulting unfolding records were sufficiently sensitive enough to yield useable data A total of 68 unfolding molecules were recorded, but 37 records had to be excluded from final analysis, either due to the presence of artifacts during the tip approach which prevented the proper interference artifact removal (e.g. see Fig 5-1) or double attachments. This rejection rate (54%) is more than double the rejection rate than were in chapter 4 (25%), using similar chemistry. We hypothesize that these extension artifacts are from interaction between unfolded bO, attached to another binding site on the tip, interacting with the membrane patch, due to the extent of the artifact in the extension. Similar artifacts were seen in the N-terminal bR unfolding experiment, but with less frequency.
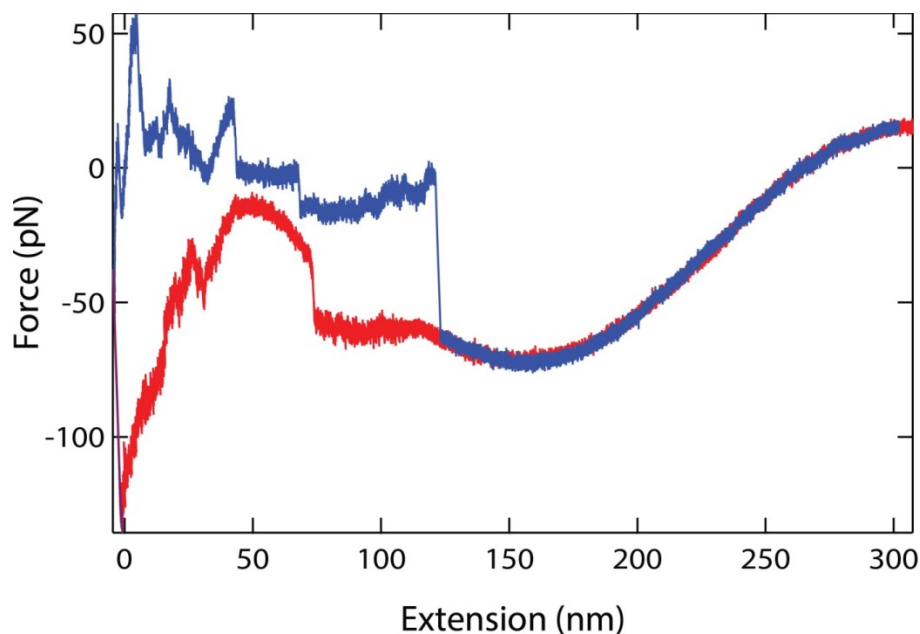
Figure 5-1: Artifacts in extension prevent interference artifact removal. The above graph presents an example the tip approach (red), surface dwell (purple, on left), and retraction (blue) of a forced unfolding attempt. There is a significant divergence in the tip approach from the sinusoidal behavior (at extension 0-100 nm). This prevents accurate determination of the interference artifact, and thus prevent its removal from the unfolding signal.

5.2) Analysis difficulties and abnormalities

Data was collected on three different days (Oct 4th, Oct 20th, and Oct 24th, all 2017). Three abnormal features were observed from these data sets. First, each day had self-consistent rupture forces, but they are inconsistent between different days. Second, the fit contour length of the second peak varied more than for the prior N-terminal unfolding experiments (chapter 4). Last, there are many odd unfolding behaviors (high forces, additional peaks). The first two problems motivated a change in the analysis techniques (section 5.2.3). These abnormalities will be detailed in this section.

5.2.1) Inconsistencies in rupture force

The force curves collected on each day of experimentation have consistent rupture forces compared to those collected on the same day, but inconsistent from day to day. Day 1 (Oct 4th, green Fig 5-2, medium rupture forces) has rupture forces of ~ 75 pN, day 2 (Oct 20th, pink Fig 5-2, low rupture forces), has a rupture force of ~ 55 pN, and day 3 (Oct 24th, blue Fig 5-2, high rupture forces) has rupture forces of ~ 95 pN. This looks like the hallmark of calibration errors with the cantilevers, either in the
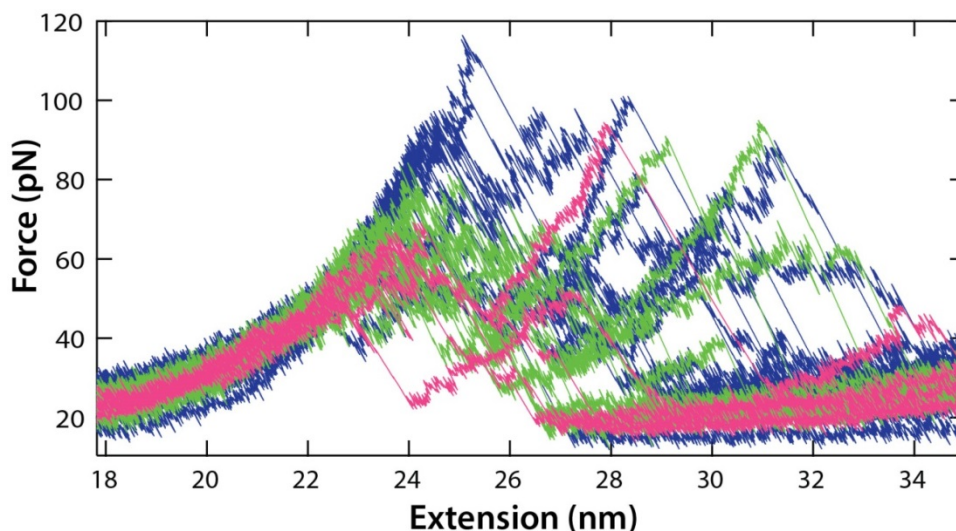
Figure 5-2: Inconsistent rupture forces. The 2nd rupture peak for most of the unfolding curves (excluding some outliners with odd and extra transitions), from each of the 3 experimental days: October 4th (green), October 20th (pink) and October 24th (blue). What is apparent is the average rupture force from each day is different (Oct 4th ~75 pN, Oct 20th ~ 55 pN, Oct 24th ~95 pN).

sensitivity (the change in the deflection signal for a given cantilever deflection) or the spring constant (the change in force associated with a change in deflection).

To determine if the rupture force discrepancy is occurring due to spring constant miscalibration, each cantilever was recalibrated multiple times, in both air and imaging buffer. While small errors were discovered, they were on the order of 10%, which is typical of the thermal calibration method (*122*). Oct 20th, the 'low' rupture force day did have a slightly smaller spring constant (~21.4 pN/nm), than Oct 24th, the 'high' rupture force day (26.1 pN/nm), but these spring constant differences (~20%) are not large enough to explain the disparity (~65%), excluding the possibility of a shortcoming in the thermal calibration method for modified cantilevers.

The other possible miscalibration to explain the differences in rupture force include errors in the cantilever sensitivity, a quantity known as invOLS (INVerse Optical Lever Sensitivity). The value is listed as nanometers of tip motion per volt in the deflection signal (as measured on a quadrant photodiode abbreviated to nm/V). Lower values indicate a more sensitive (and preferable) cantilever. The invOLS is calibrated by utilizing the precision of the piezoelectric motor and pressing the cantilever into the surface a fixed distance, and observing the resulting change in deflection. The invOLS measured (18.8, 26.6 and

26.1 nm/V, for Oct 4th, 20th and 24th, respectively) are all within normal operating range, and do not systematically correlate with the rupture forces (i.e. the lowest invOLS day does not correspond to the lowest rupture force day). Checking the listed invOLS against force curves taken in proximity to the relevant data did not reveal any significant variations. The invOLS vary depending on the position and focus of the laser on the cantilever and as a result, a given cantilever will have different invOLS depending on those two settings (though generally, no large change in invOLS occurs once the position and focus are set). Thus, one cannot verify the invOLS values after those settings have been changed. The best that we can achieve is to vary the focus and position to determine the range of possible values for a given cantilever. This was done for all three cantilevers, all of which yielded ranges that corresponded with the listed invOLS. Another check can be performed by determine how significant of an error in invOLS would be needed to explain the rupture force discrepancy. This analysis suggests it would take a ~50% and ~100% increase in the invOLS to bring the Oct 4th and Oct 20th data into agreement with the Oct 24th data, respectively. Even the most modest scenario: bringing the Oct 24th ("high force") and Oct 20th ("low force") data into agreement with the Oct 4th ("medium force") data, would take changes on the order of a 50% (increase and decrease, respectively) to explain the discrepancy, which are larger than reasonable errors in these values. Thus, despite being the most logical explanations of the disparity, neither possible source of miscalibration sufficiently explained the rupture force discrepancy.

5.2.2) PEG linker variation

Another oddity from the collected data is significant differences in the curvature of the 2nd peak major state from record to record, indicating differences in the contour length. An example of this is presented in Fig 5-3. Most likely, these differences arise from variations in the length of the PEG linker in the silane-PEG-azide (SPA) construct (PG2-AZSL-3k, from Nanocs), possibly due to aggregation. This is odd, as the construct is the same one used in chapter 4 (nominally 3.4 kDa PEG), where contour length variations were not as severe. The SPA used for all experiments also came from the same manufacturer
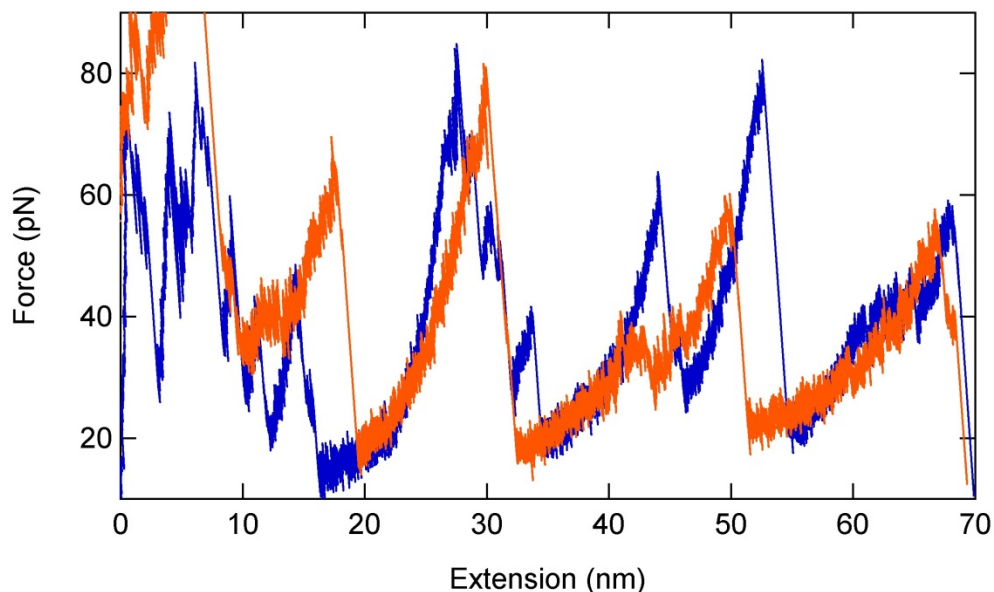
Figure 5-3: Example of contour length differences. Two unfolding curves (orange and blue), both from Oct 4th, are displayed to demonstrate the differences in the curvature (i.e. the contour length) even within a single day dataset. This is especially noticeable at shorter contour length events, (i.e. the 2nd peak) compared to the longer contour length events (i.e. the 4th peak).

lot. This variation in contour length can be quantified by fitting the 2nd peak major state with variable horizontal offset. In chapter 4, the standard deviation was 5.8 nm (and a lot of that variation is from differences between the two datasets collected), compared to 8.3 nm in this chapter. In theory, a 3.4 kDa PEG should have a contour length of 21.6 nm (with variation of ± 10% typical). If one cannot assume a singular contour length for a given structural unfolding event, a new alignment strategy must be adapted that does not rely on that assumption.

To validate this proposed variable contour length approach, we examined the variation in the contour lengths for the 2nd peak major states and the contour length differences between the 2nd peak and 4th peak major states. Assuming the variation in contour length arises from aggregation of the PEG linker, the resulting measured contour lengths of the 2nd peak major state should fall between the contour length of the unfolded section of the protein (26.64 nm, assuming the 2nd peak major state remains at residue 74, as it was in the bR unfolding experiments (*49*)), and the combined length of the unfolded section of the protein and the full PEG linker (21.6 nm ± 10%), which sums to ~50 nm. Combined with the (small)

uncertainty associated with identifying a major states themselves, the contour lengths determined from these fits are likely to include a few slightly above or below this range (25 - 51 nm). Fig 5-4 depicts a histogram of the L_c fits for the second peak major states. As expected, there is a significant amount of variation of this value. While the L_c are not limited to the range limited above (some fitting below 25 nm), they fall only slightly outside the specified range, and a significant majority of fits still fall within this range. We also fit the fourth peak major state (with the same horizontal offset as the 2nd peak fit), in order to determine the contour length difference between those peaks. These fits do roughly conform to the theoretical 2nd to 4th peak ΔL_c of 49.59 nm ($\overline{\Delta L_C}$= 50.35 nm, less than 2% difference), without obvious groupings for the different days (Fig 5-5). Some scaling is still be necessary due to errors in invOLS and spring constant calibration (average of -2 % with a standard deviation ± 4%), but it is within normal calibration errors range. The general agreement in spacing of these intermediates is much more consistent than would be ascertained by aligning the 2nd peaks to a single contour length.
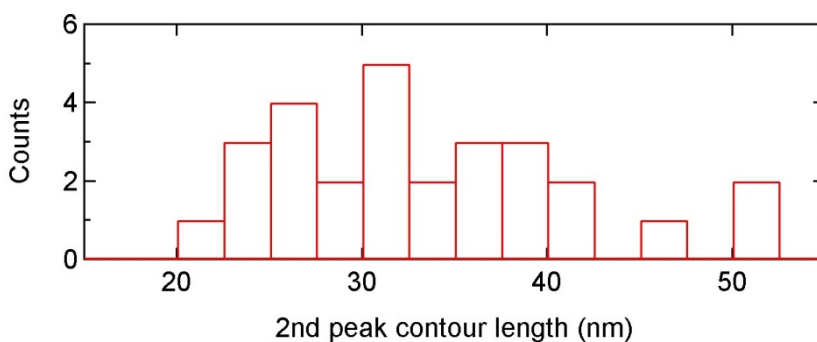
Figure 5-4: 2nd peak major state contour length histogram. Created by fitting each of the 2nd peak major states with a worm-like chains, with fixed persistence lengths, variable contour length and variable horizontal offset, and then histogrammed the resulting contour length fit. In theory, each of the major states should be between ~25 nm and 51 nm, due to possible variability in the length of the PEG linker used in the tip functionalization. Most of them fall within this range, but a few fall just outside. N=31.
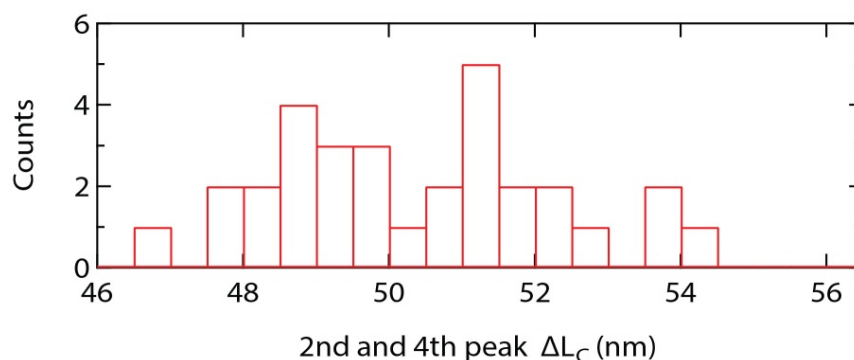
Figure 5-5: ΔL_C between 2nd peak major state, and the 4th peak major state. A histogram created by subtracting the fitted contour length for the 4th peak major state from contour length of the 2nd peak major state, both fitted with the WLC model with variable contour lengths and identical horizontal offsets. Nominally, the contour length change between the 2nd and 4th peak major states should be 49.59 nm, where the average is actually 50.35 nm but a < 2% variation may arise due to spring constant calibration error, invOLS error, or misidentification of the major states. *N=31*.

5.2.3) New analysis method: variable contour length

The variability of the contour lengths of the 2nd peak major state (suspected to arise from variations in the PEG linker length) requires that a new analysis method being implemented. This method will be called the "variable contour length" method. The idea is to fit a fiducial state (chosen as the 2nd peak major state, as it is the shortest and achieves the highest forces, without surface distortion) with a worm-like chain (WLC) model (see section 2.5). This allows the contour length (L_C) as well as the horizontal offset (x_0) to vary (Fig 5-6B) and to account for variations in the linker length, and variations in the attachment point on the tip relative to the apex, respectively. Next, the force-extension record is horizontally shifted by x_0 (Fig 5-6C1) and then each unfolding record is converted from force and extension into contour length. By aligning this data (in contour length space) to a single value, a comparison of relative contour length differences is possible (Fig 5-7). Minor scaling is then performed on the extension data based on the contour length difference between the 2nd and 4th peak major states to account for slight miscalibrations of the spring constant and/or the invOLS as mentioned in Section 5.2.1. This is typically on the order of ± 5% (Fig 5-A1). After this, the intermediates are determined in much the same way as section 3.10.3, except now force is replaced with contour length in Eqn. 3-1 and Eqn. 3-2.
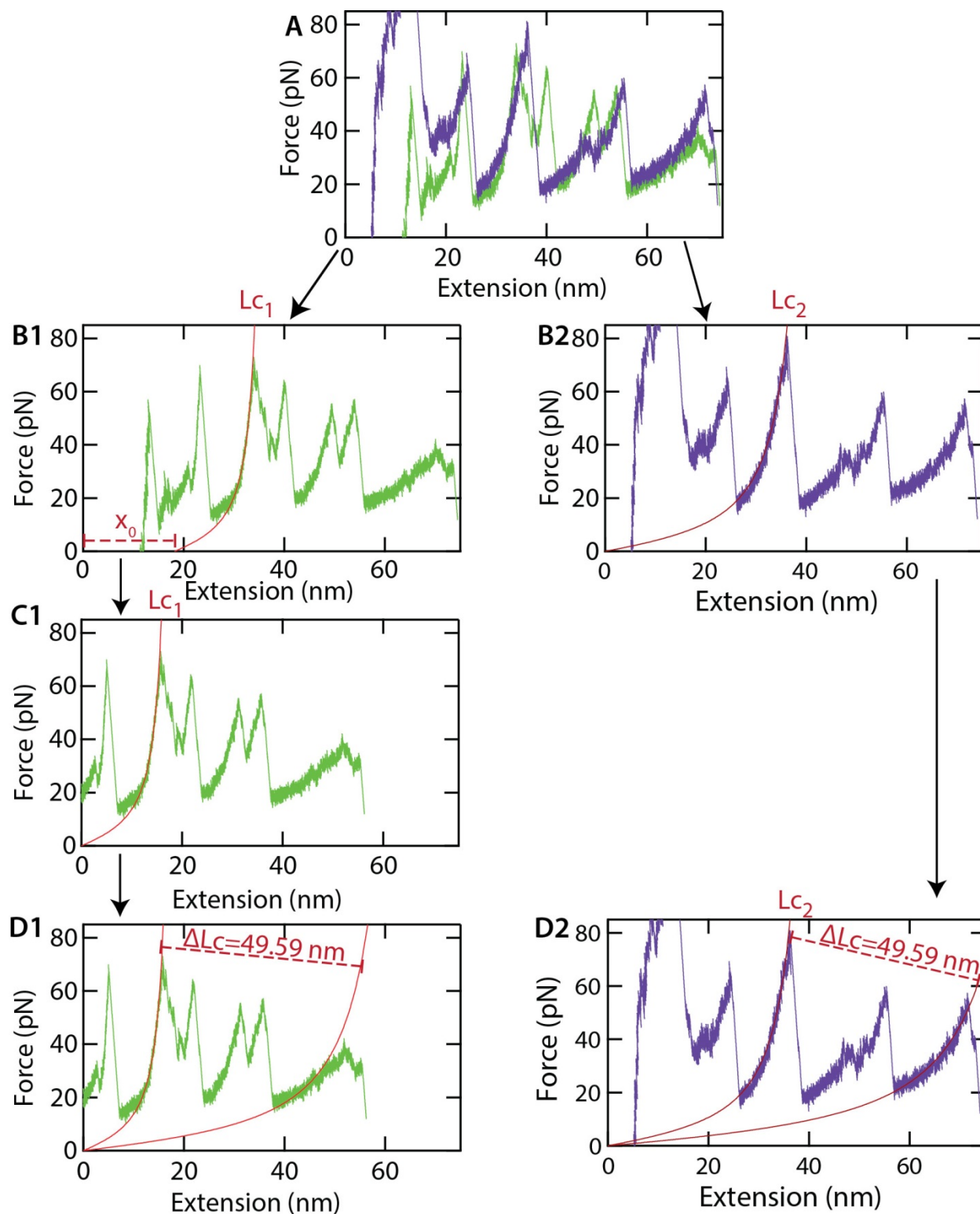
Figure 5-6: Variable contour length method. (A) Two unfolding records (green and purple), approximately aligned by the 2nd peak major state. The curvature of the 2nd peak major state for green and purple records are different, and cannot be brought in agreement by merely scaling the extensions, as any scaling of will cause disagreement in the position of the 4th peak. This motivates the

Figure 5-6 (cont.): move to a new alignment strategy **(B1)** Demonstrates the alternate alignment strategy, based on the idea that there is a variable length of the PEG-linker causing the different absolute contour lengths for each state (Lc_1 & Lc_2) for each unfolding record, but the contour length differences (ΔLc) between states within a given record will still largely agree. To accomplish this alignment, the 2nd peak is fit, with variable contour length (Lc_1) and horizontal offset (x_0) arising from different tip-surface interactions relative to the adhesion point. **(C1)** Then this horizontal offset (x_o) is subtracted from the extension, while keeping the 2nd peak major state constant (Lc_1). **(D1)** The 4th peak major state is fit with a WLC to a contour lengths 49.59 nm beyond Lc_1 the second peak major states. **(B2)**, The purple unfolding record is fit by the 2nd peak major state with a different contour length (Lc_2), and doesn't require a horizontal offset. **(D2)** The 4th peak major state is also well fit to a WLC with a contour length 49.59 nm beyond Lc_2, demonstrating agreement in the relative contour length differences (ΔLc). Note that the WLC model does not fit well to unfolding data at low forces (<20 pN).
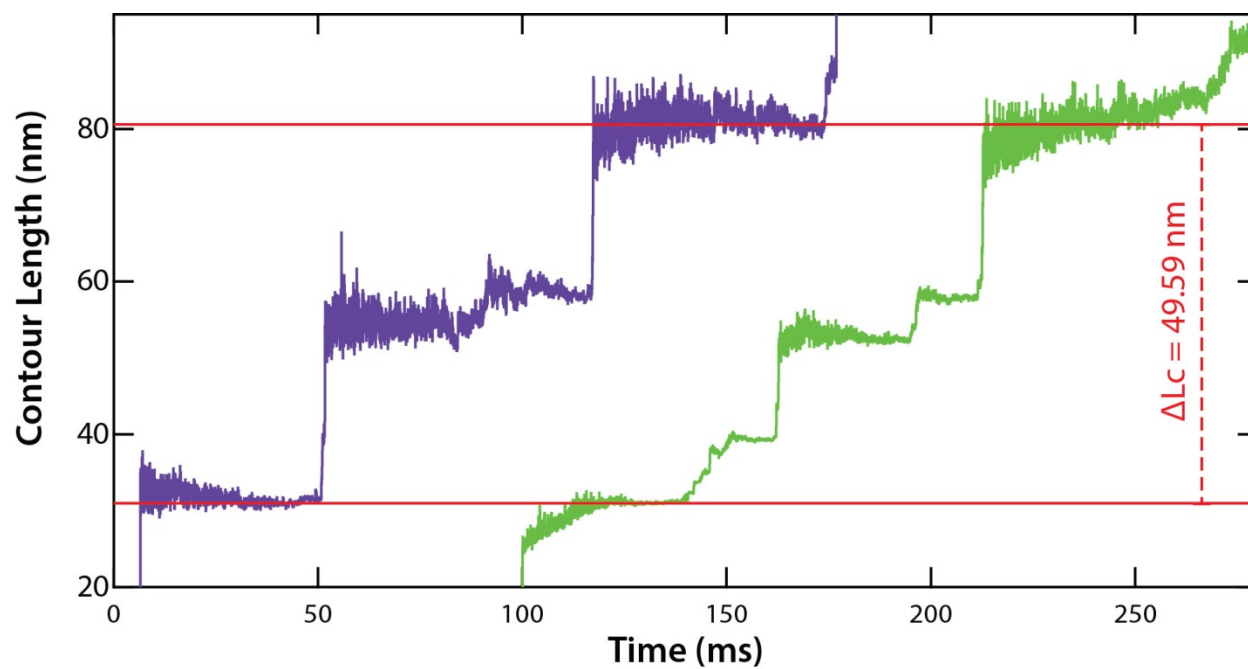
Figure 5-7: Aligned contour length space depiction of records from Fig 5-6. The purple and green unfolding records are converted to contour length space and shifted vertically to align at L_c= 31 nm. They are plotted against time. The red lines are the fitted location of the second peak and fourth peak major states (31 nm and 80.59 nm, respectively).

The quality of the fit is demonstrated in the agreement in ΔL_c between the 2nd and 4th peak and the 'flatness' of the segments in general (as depicted by the major states in Fig 5-7). Another example of the improvement in alignment is presented in Fig 5-A3 and 5-A4. We will utilize this method to align and determine the intermediates for the bO. The variable contour length method is more general than the absolute contour length method, as it assumes less about the polymer lengths. It is susceptible to more noise, due to the reliance on multiple curve fits, as well as properly recognizing two major states (rather

than just one in the constant contour length assignment), and the addition of extension noise to the force

noise inherent in the conversion to contour length space. Thus, the absolute alignment method is

preferable when contour length variation is not observed.

5.2.4) Unusual unfolding patterns

After observing the unfolding of thousands of bR and bO records, an ability to recognize unusual

unfolding patterns that diverge from the four peak characteristic unfolding curve develops. The most

common cause of a divergence in the unfolding pattern is from multiple proteins attaching to the tip (e.g.

Fig 5-A2). Depending on the attachment parameters (concentration of DBCO, pushing force during

attachment, dwell time), one can alter the rate of multiple attachment. This is important, as a multiple

attachment renders an unfolding record uninterpretable. Normally, multi-attachments are excluded from

analysis without further explanation, but here they will be examined here for three main reason. First,

they do not look like typical multi-attachment (which exhibit extraneous high force peaks as demonstrated

in Fig 5-A2). Second, they are relatively rare and will not be noticeable in the overall analysis of the

unfolding (which is one of the key reasons to perform single molecule assays over ensemble). Lastly, we

lack a characteristic N-terminal bO unfolding study to determine they are non-native (i.e., they may

reflect the variability in the structure of the bO).

The first abnormal unfolding behavior is extremely high rupture peaks for the 4th peak (Fig 5-8A,

~300 pN)). High rupture peaks are often a hallmark of multiple attachments, but usually earlier in the

unfolding record (the 2nd or 3rd rupture peak) and often with multiple high-force ruptures per pair of

helices. In these unfolding curves there are no abnormally large ruptures prior to the 4th peak. The second

unusual unfolding behavior is permanent transitions out of major states at low forces (Fig 5-8B, grey and

brown stars, for black and orange curves, respectively). This suggests that the traditional major states are

occasionally destabilized. In its most extreme form, the second rupture peak is almost entirely absent (Fig

5-8C). This is interesting, because this occurs simultaneous with a lack of other unfolding intermediates

in the 2nd and 4th peak (both here and in the one other peak where it occurred). Potentially, this represent a

significant disruption of the ED helices structure. As these features all occur in less than 10% of recorded
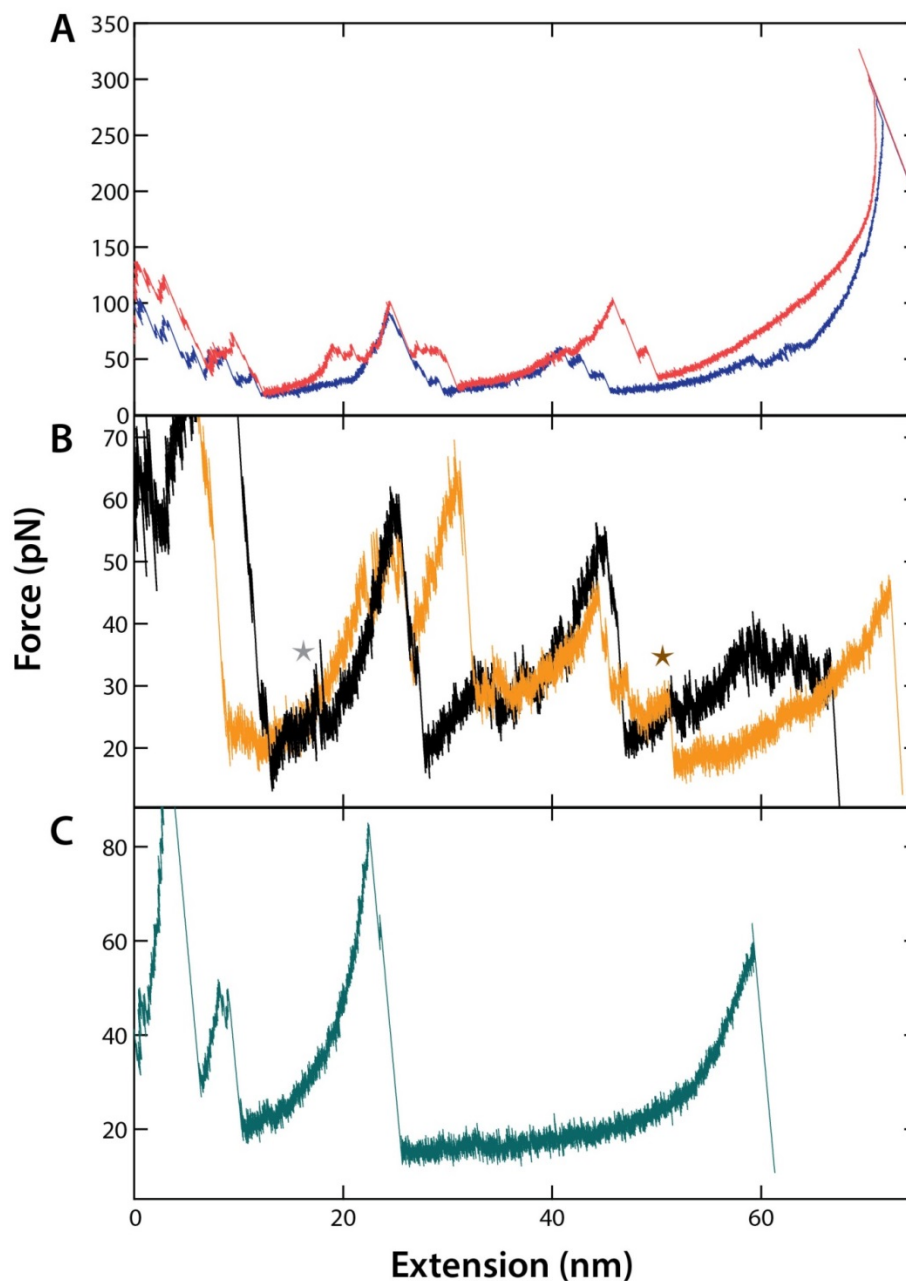
Figure 5-8: Abnormal unfolding behaviors. Three different abnormal unfolding behaviors were observed over the course of the bO unfolding experiment, which cannot be easily explained by double attachment. **(A)** High rupture forces of the fourth peak (~300 pN), after typical unfolding behaviors for the first three peaks. This is approximately 250 pN greater than the average rupture forces for remaining the 4th peak major states. This occurred in 2 out of 38 unfolding records (~5%). **(B)** Low rupture forces for the major states. Two different unfolding traces (black and yellow) and depicted here, with a low force (<30 pN) final transitions out of the major states marked by stars. The grey star depicts a low force transition by the black record out of the CD major state, while the brown star represents a low force transition by the orange record out if the G major state. This occurred in 3 out of 38 unfolding records (~8%) **(C)** An extreme example of low force transitions out of the major state, as the EF major state is completely absent. What's extremely unusual is the lack of subsequent intermediates reaching measurable (>20 pN) forces, as is observed in (B). This occurred in 2 out of 38 unfolding records (~5%).

traces, they will be mostly excluded from the analysis in this chapter (Fig 5-8A will be included in analysis except for the fourth peak), but they may represent something about heterogeneity in the structure of bO, due to the destabilization caused by the retinal removal, or by the additional methods used to remove the retinal removal (i.e. additional sonication).

5.2.5) Exclusion of additional records

In undertaking this new analysis technique, additional care must be taken to exclude potential double attachments from the analysis dataset, as once conversion to contour length space is accomplished, these peaks will not appear as the distinct outliers that they are in force-extension space. To that end, seven records were removed for being likely multiple attachments, based on their presentation of high force peaks at locations incongruous with all other traces (e.g. Fig 5-A2). After their exclusion, we were left with 31 records.

5.3) Bacterioopsin unfolding intermediates

Within this section, we will present two types of figures to demonstrate the intermediates. The first type is contour length vs. time plots, such as Fig 5-7. Dwells within a single intermediate will be displayed as horizontal segments, with transitions corresponding to motion along the vertical axis. This provides the best means of comparing multiple unfolding records. The second figure type is force vs. extension plots. Unlike in chapters 3 and 4, these force vs. extension plots with intermediate WLC curves will be plotted in separate graphs the WLC curves will have different absolute contour lengths, but the relative contour lengths will be the same as demonstrated on Fig 5-6.

Also, to reduce confusion with the identified intermediates in chapter 4 (retinal-containing bR), we will indicate intermediates identified in the retinal removed experiments with an asterisk (*, i.e., I_{CD}^{*0} will be the zeroth CD state for bO). The retinal-containing bR intermediates established in chapter 4 will be referenced without an asterisk (I_{CD}^{0}).

5.3.1) CD intermediates locations

After utilizing the variable contour length alignment and state determination (section 3.10.3, with forces replaced by L_c), we are left with a major state and 12 intermediates. Fig 5-9 depicts seven different unfolding records in a contour length vs. time plot and Fig 5-10 depicts six of those seven unfolding records depicted as force-extension plots (with sublabels A-F maintained). This is a slight decrease in the number of detected intermediates from the N-terminal bR unfolding experiment. It is not clear is if this decrease in the number of intermediates is significant, as switching from an absolute contour length alignment method to the relative contour length alignment method may have diminished the state resolution due to inherent uncertainty in the WLC fitting (Fig 5-6B1). These intermediates assign to a similar range of residues (aa 74 – aa 117) as the intermediates identified for bR, with intermediates absences relative to bR being evenly distributed in the region (with the exception of the CD loop, where all intermediates are preserved).
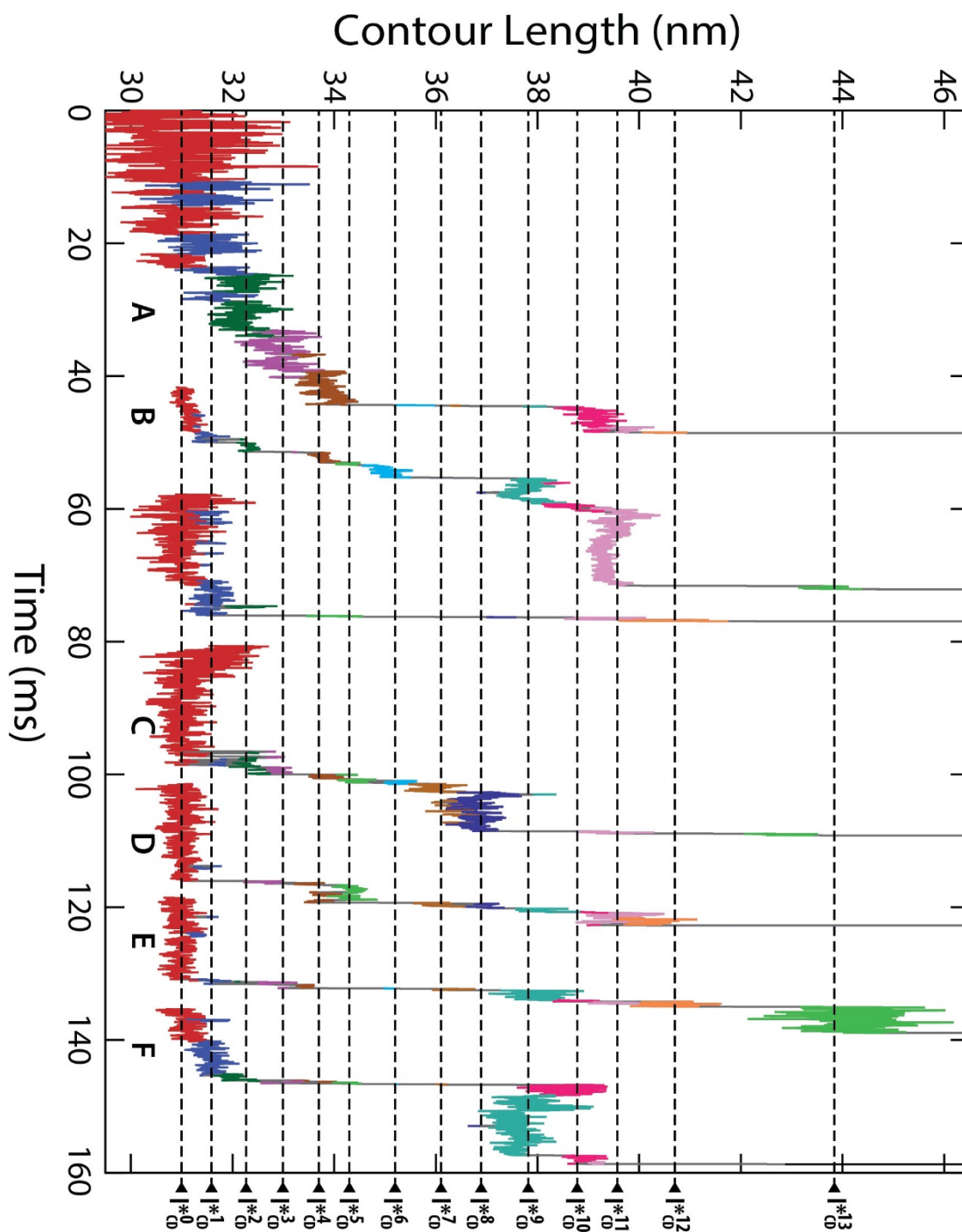
Figure 5-9: Contour length-time records of bO CD helices intermediates. The CD unfolding intermediates are marked by horizontal black dotted lines, and labeled (right axis). 7 unfolding records, with highlights indicating which segments are assigned to each intermediate depicted. I_{CD}^{*0}(red), I_{CD}^{*1}(blue), I_{CD}^{*3}(dark green), I_{CD}^{*4}(purple), I_{CD}^{*5}(brown), I_{CD}^{*6}(light green), I_{CD}^{*7}(light brown), I_{CD}^{*8}(dark blue), I_{CD}^{*9}(teal), I_{CD}^{*10}(magenta), I_{CD}^{*11}(pink), I_{CD}^{*11}(orange) and I_{CD}^{*12}(green). Each record is aligned such that the CD major state lies at 31 nm. 6 of the unfolding records are labeled with (A-F), which are depicted in force-extension space in Fig 5-10. Here that only the parts of the records above 20 pN are depicted here, removing regions of extremely high noise in contour length space.
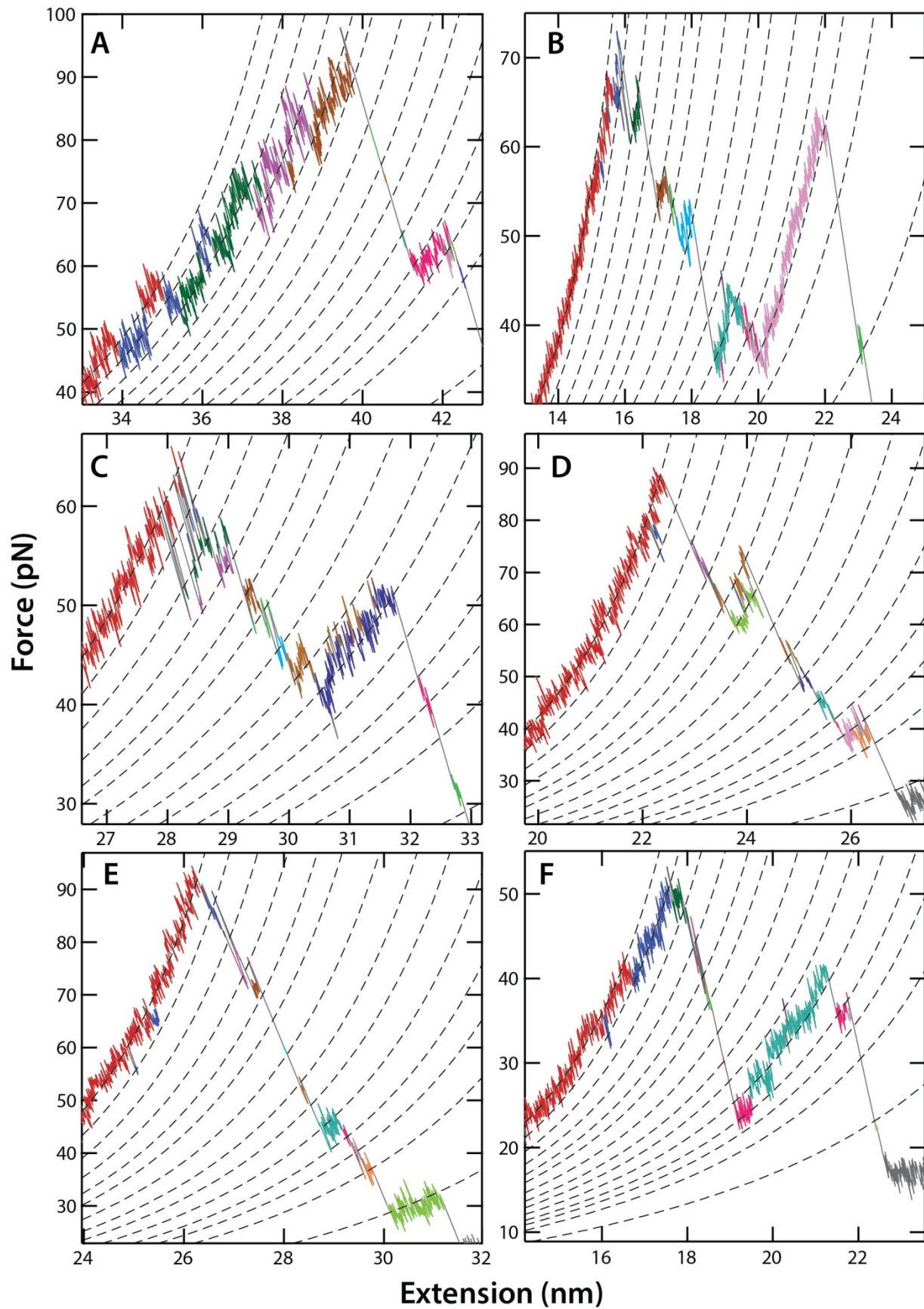
Figure 5-10 (last page): Force-extension plots of bO CD helices intermediates. Each force-extension plot depicts the CD helices unfolding intermediates, as indicated by black dotted lines. These are the same unfolding records as depicted in Fig 5-9, with the same labeling **(A-F)**. Each of these states is at a different contour length for each panel, but the relative separation for the intermediates is the same. The state assignment is the same as it was in Fig 5-9, with the same colors used for each intermediate.

5.3.2) EF intermediates locations

The E and F helices changed from two near obligate states and eight intermediates observed for bR to four near obligate states and three intermediates for bO (Fig 5-11 and 5-12), the largest change in behavior for the three helix pairs analyzed. Once again, the intermediates missing relative to bR are once again are spread throughout the range displayed by bR, which suggests that the reduction in states is likely from the loss in state resolution from the use of the relative alignment method, rather than from any underlying structural change. Further supporting this theory is the presence of dwells in the spaces between these states and more established states (e.g. Fig 5-11A I_{EF}^{*2} and 5-11B I_{EF}^{*7} have dwells slightly both slightly above and below the assigned contour length, hinting that there might be multiple states located here, but we lack the state resolution to statistically differentiate them). Interestingly, we observe an increase in the occurrence rates of all intermediates, similar to the observed in previous C-terminal bO studies (*67*). Another new feature is the presence of a low force non-obligate intermediate (I_{EF}^{*0}) prior to the near obligate state (I_{EF}^{*1} and I_{EF}^{*2}). This maps to either the extracellular side of the D-helix, or the DE loop. Again, it is unclear if this is detectable because of the relative contour length alignment strategy, or this reflects structural changes due to the absence of the retinal. A planned analysis of the N-terminal unfolding date presented in chapter 4 utilizing the relative contour length alignment method will help resolve this question.
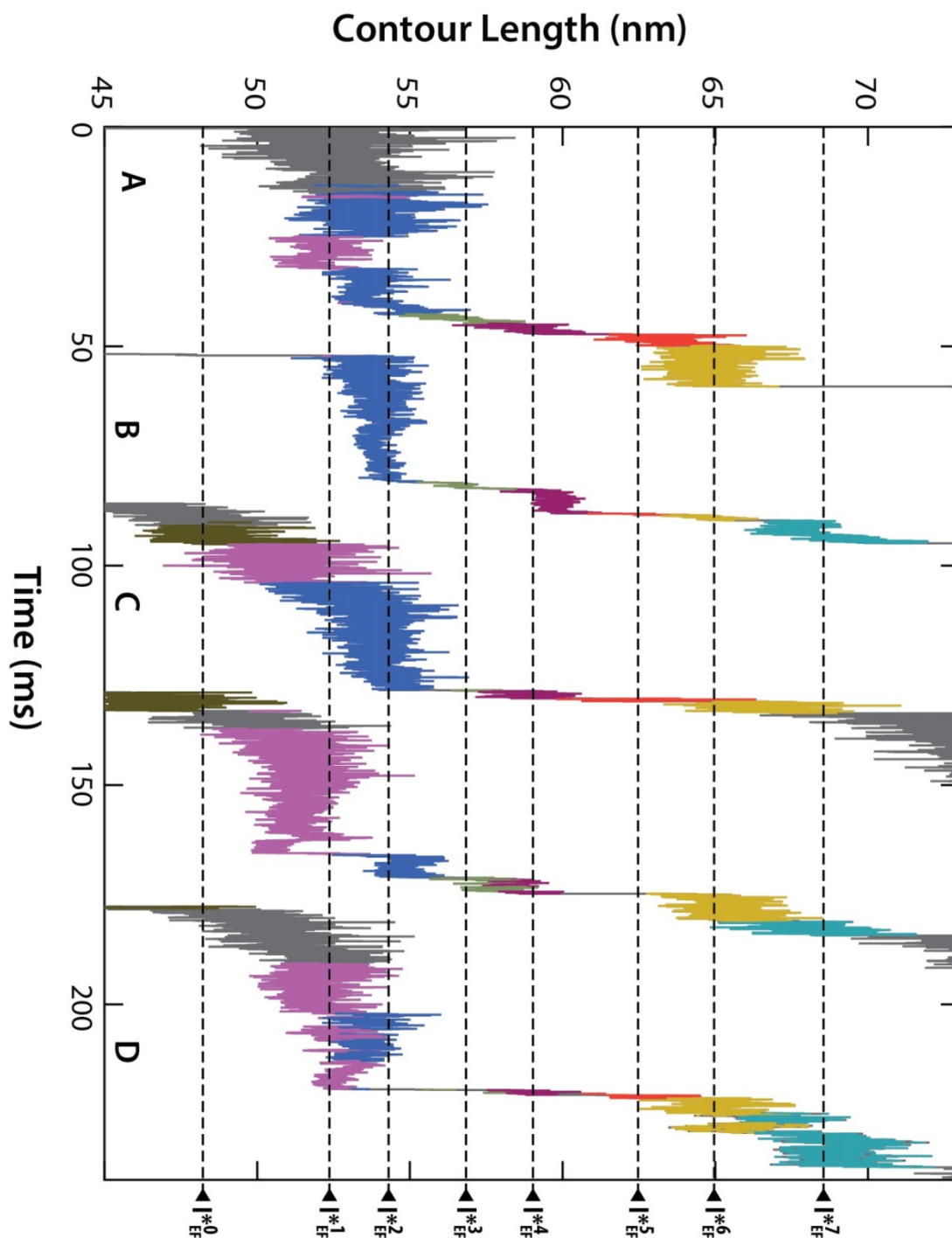
Figure 5-11: Contour length vs. time records of bO EF helices intermediates. The EF unfolding intermediates contour length are marked by horizontal black dotted lines, and labeled (right axis). 5 unfolding records, with highlights indicating which segments are assigned to each intermediate are depicted: I_{EF}^{*0}(brown), I_{EF}^{*1}(pink), I_{EF}^{*3}(blue), I_{EF}^{*4}(green), I_{EF}^{*5}(marooon), I_{EF}^{*6}(orange), I_{EF}^{*7}(yellow), and I_{EF}^{*8}(teal). Unassigned sections (either within transition periods or due to insufficiently high forces to assign state) are shown in grey. 4 of the unfolding records are labeled with (A-D), which are depicted as force-extension curves in Fig 5-12.
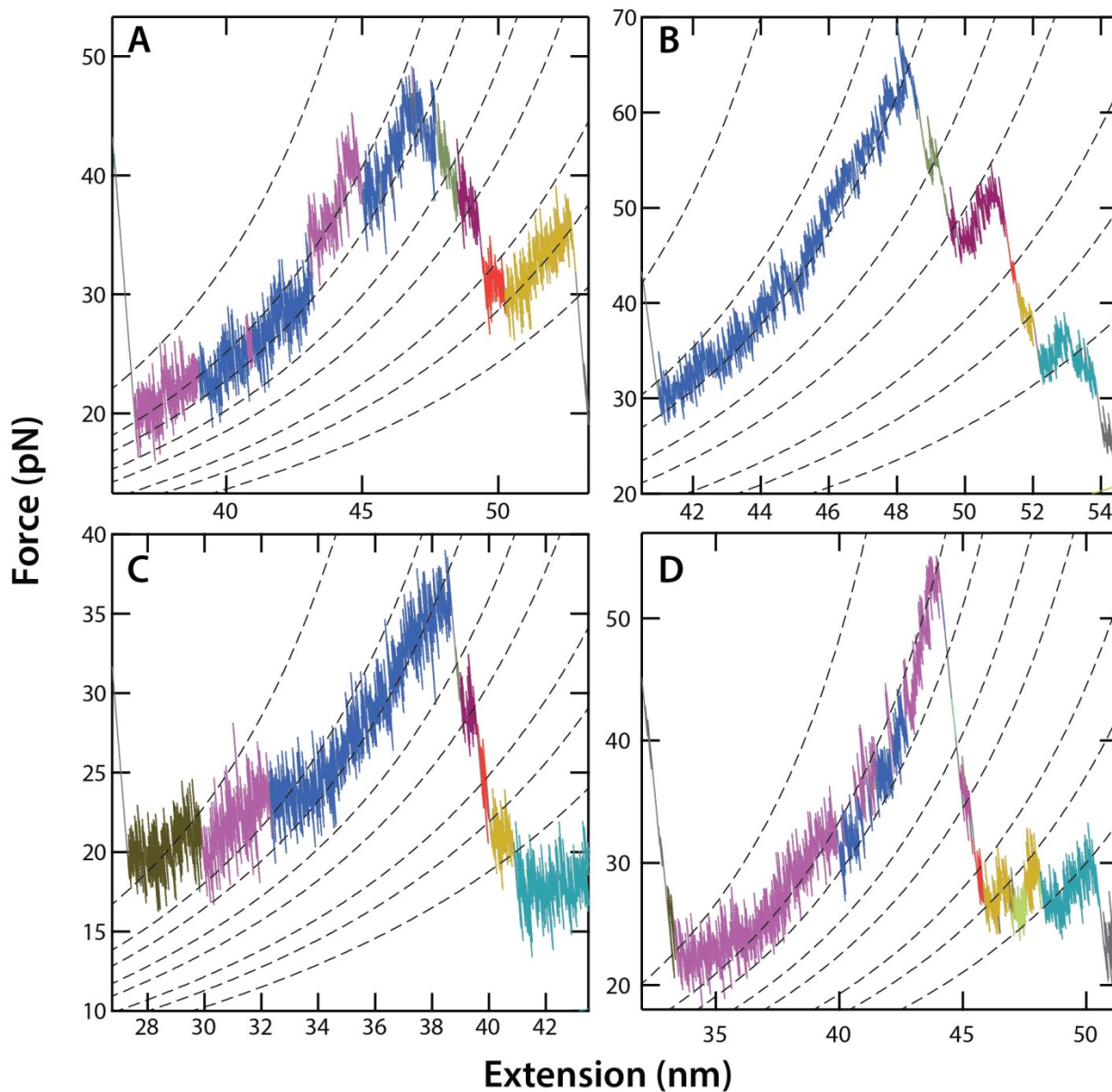
Figure 5-12: Force-extension plots of bO EF helices intermediates. Each force-extension plot depicts the EF helices unfolding intermediates, as indicated by black dotted lines. These are the same unfolding records as depicted in Fig 5-10, with the same labeling **(A-D)**. Each of these states is at a different contour length for each panel, but the relative separation for the intermediates is the same. The state assignment is the same as it was in Fig 5-10, with the same colors used for each intermediate.

5.3.3) G Intermediates locations

The G helix is the site of the covalent attachment to the retinal, but as all of the other helices were extracted previously, thus the retinal has no other helices with which to interact. In fact, within the N-terminal unfolding experiment, no intermediate was even observed at the residue that forms the Schiff base with the retinal. As a result, it was unlikely that the removal of retinal would have a significant impact upon the unfolding. The analysis of the G helix intermediates finds that this expectation was correct, as we find 1 major state, 2 near obligate intermediates and 3 intermediates (the same number of intermediates as in bR unfolding), in essentially the same locations (Fig 5-13 and 5-14). Similar to the EF helices, the presence of a non-obligate, low force intermediate (I_G^{*0}) prior to the major state (I_G^{*1}) is observed. This maps to the loop prior to the start of the G helix, similar to I_{EF}^{*0}. Most intriguing is the addition of a very late intermediate I_G^{*6}, which is located nearly 20 nm after the G helix major state. Similar to the discussion section 4.2.3, this places the intermediate well beyond the structured part of the protein, and even beyond reasonable C-terminal tail locations. As this is ongoing work, this will be investigated further on non-functionalized surfaces to determine the exact origins of the intermediates. Also the same as section 4.2.3, we theorize this is either caused by adhesion of negatively charged intermediates onto the positive charged APDMES surface, or hydrophobic interactions with the membrane by those same charged (hydrophilic) residues.
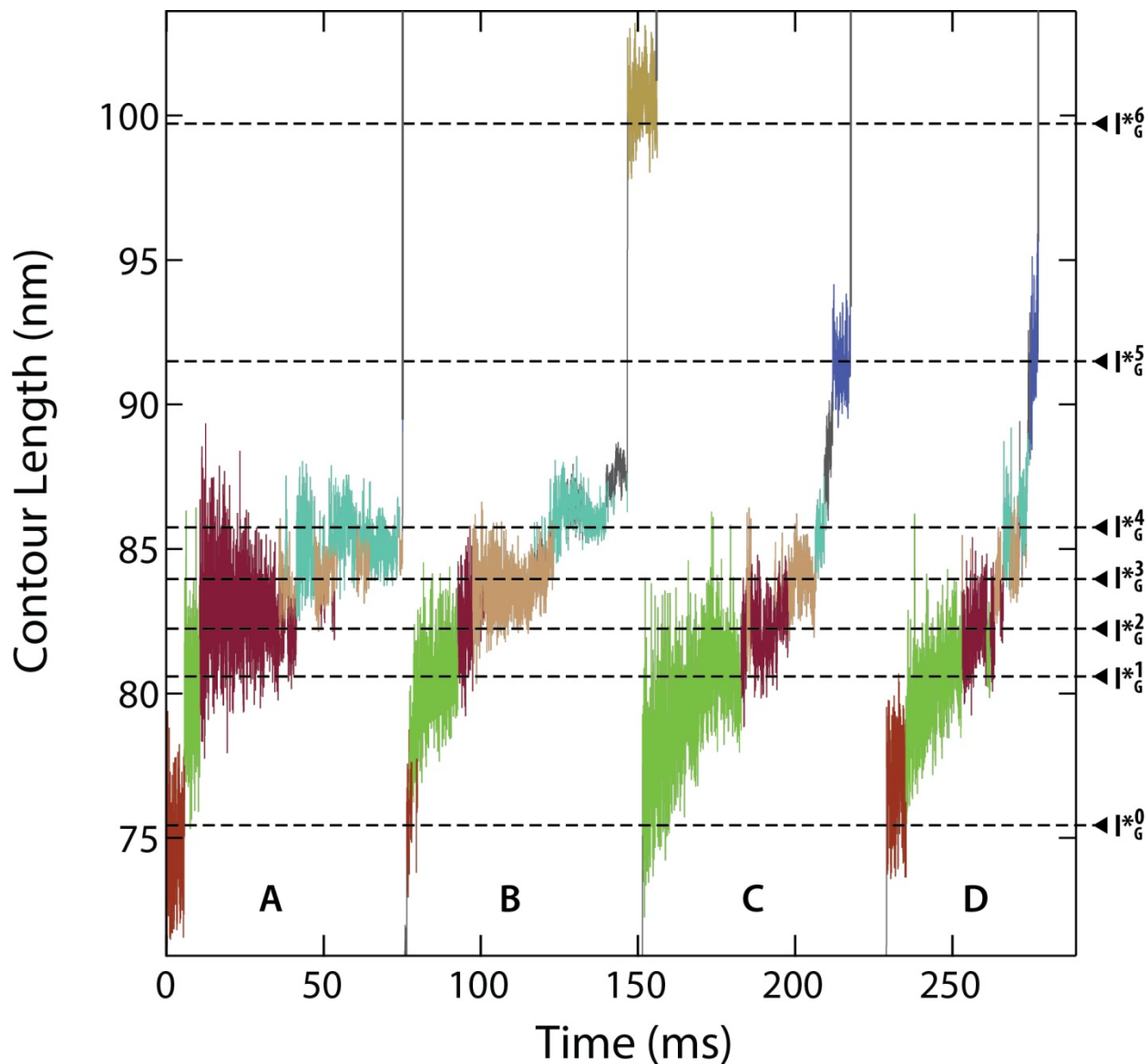
Figure 5-13: Contour length vs. time records of bO G helix intermediates. The G unfolding intermediates contour length are marked by horizontal black dotted lines, and labeled (right axis). 5 unfolding records, with highlights indicating which segments are assigned to each intermediate are depicted: I_G^{*0}(brown), I_G^{*1}(green), I_G^{*2}(maroon), I_G^{*3}(tan), I_G^{*4}(teal), I_G^{*5}(blue), and I_G^{*6}(yellow). Unassigned sections (either within transition periods, or due to insufficiently high forces to assign state) are shown in grey. Each unfolding records is labeled with (A-D), which are depicted in force-extension space in Fig 5-14. Note that sections at low force (e.g. maroon part of A), occur due to low forces.

Figure 5-14: Force-extension plots of bO G helix intermediates. Each force-extension plot depicts the G helices unfolding intermediates, as indicated by black dotted lines of Worm-like chains. These are the same unfolding records as depicted in Fig 5-13, with the same labeling **(A-D)**. Each of these states is at a different contour length for each panel, but the relative separation for the intermediates is the same. The state assignment is the same as it was in Fig 5-13, with the same color assignments used for each intermediate.
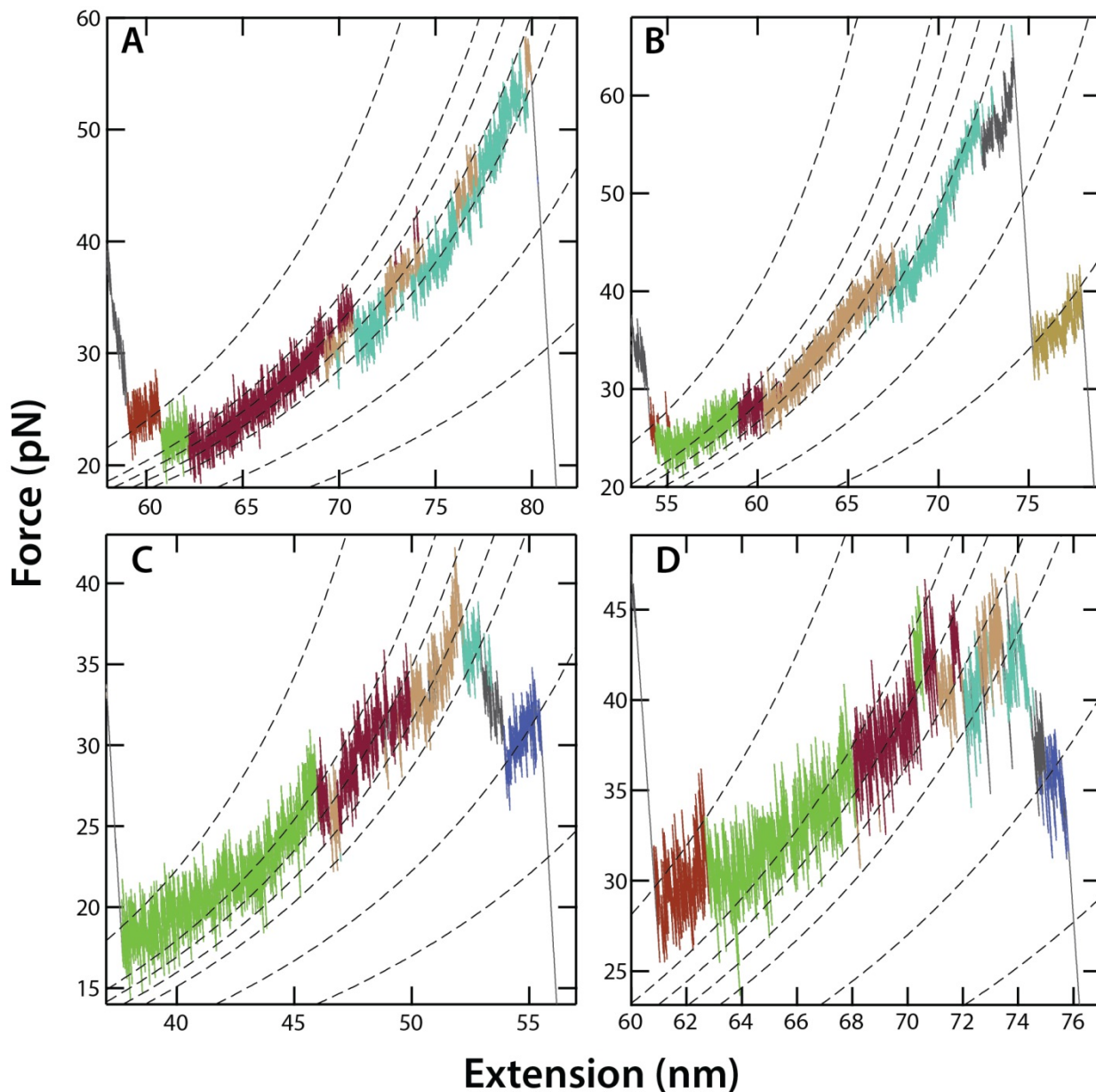
5.3.4) Intermediate properties

Table 5-1 lists the occupancy rates, rupture forces and dwell times for each intermediate.

Table 5-1: Structural parameters of N-terminal bO intermediates. The contour length, structural position, probability, average unfolding force of each intermediate, and median dwell time detected in each state are listed. The contour lengths are based on assignments with assumed contour lengths of 31 nm for I_{CD}^{*0} and 80.6 nm for I_{G}^{*1} (the values for the C and G major states listed in (*49*)). The presented data were derived from 31 molecules for the CD and EF helix pairs and 29 molecules for the G helix, as two records were excluded due to extremely high G rupture forces (i.e. Fig 5-7A). The median dwell time detected in each state are only looking at dwells above 20 pN, due to the prevalence of non-WLC behavior below 20 pN.

	State	Contour length ± s.d. (nm).	Nearest AA	Occupancy rate (%)	Mean unfolding force ± S.E.M (pN)	Median dwell (μs)*	Description (Exact residue)
	0	31 ± 0	74	100	68 ± 3	29317	Beta sheet start
C Helix	1	31.6 ± 0.3	80	90	69 ± 3	2756	C helix start (79)
	2	32.3 ± 0.3	83	87	67 ± 3	952	
	3	33.0 ± 0.3	87	81	68 ± 3	841	
	4	33.7 ± 0.2	90	77	66 ± 3	1531	
	5	34.3 ± 0.3	93	61	64 ± 4	280	
	6	35.2 ± 0.3	97	52	57 ± 3	341	C helix end (101)
	7	36.1 ± 0.3	101	68	55 ± 3	279	
	8	36.9 ± 0.3	103	74	54 ± 3	895	
D Helix	9	37.8 ± 0.3	105	77	49 ± 2	770	D helix start (104)
	10	38.8 ± 0.3	107	68	52 ± 4	519	
	11	39.6 ± 0.3	109	55	53 ± 5	1234	
	12	40.7 ± 0.6	111	61	47 ± 4	255	
	13	44 ± 2	117	39	42 ± 3	251	D helix end (126)
	0	48 ± 3	126	52	35 ± 2	3644	
E	1	52.4 ± 0.8	139	97	45 ± 2	14182	E helix start (130)
	2	54.3 ± 0.7	148	94	52 ± 2	13082	
	3	56.8 ± 0.8	159	97	51 ± 3	3149	E Helix end (160)
	4	59.0 ± 0.8	166	97	50 ± 2	3831	
F	5	62.5 ± 0.8	173	81	53 ± 4	1690	F helix start (167)
	6	65.0 ± 0.8	178	87	46 ± 3	4335	
	7	68.5 ± 1.2	185	61	39 ± 3	991	F helix end (193)
	0	75.4 ± 1.2	200	23	34 ± 2	5359	
G	1	80.6 ± 0	223	97	49 ± 3	29897	G helix start (202)
C-term tail	2	82.2 ± 0.7	226	93	51 ± 3	6534	G helix end (225)
	3	84.0 ± 0.7		97	53 ± 3	3882	
	4	85.8 ± 1		83	52 ± 3	2994	
	5	92 ± 2		30	41 ± 2	3216	
	6	100 ± 2		17	46 ± 2	19090	

*Above 20 pN

5.4) Comparison to bR experiments

The unfolding behavior of N-terminal unfolding bO is relatively straightforward, with little compelling results in and of itself. Rather, the interesting result comes from the contrast to the N-terminal bR unfolding results. To aid the comparison, we list the pertinent data from Table 4-1 (N-terminal bR unfolding data) and Table 5-1 in Table 5-2, with similar intermediates aligned by row. The contour length changes from the major states and the rupture forces are plotted for both the bR and bO unfolding intermediates in Fig 5-15.

Overall, the lack of significant changes in the unfolding forces and locations suggest that the removal of the retinal does not impact the stability of the C, E and G helices significantly. The modest increase (~10-20%) in the occupancy rate for the intermediates D, E, F, and G helices, as well as the bottom of the E helix, suggests that either the sensitivity of the assay increased, or a small destabilization the protein. Overall, the rupture forces are slightly higher ($\Delta F \approx$ 5-10 pN for E, C and A helices intermediates compared to bR unfolding), especially in the upwards winding D and F helices (specifically for I_{CD}^{*9}-I_{CD}^{*13} $\Delta F \approx$ 15 pN and for I_{EF}^{*5}-I_{EF}^{*7} ~ $\Delta F \approx$ 15 pN, compared to bR unfolding). This suggests that the upwards winding helices are slightly stabilized by removing the retinal, a result at odds with previous predictions (*67*, *79*, *109*, *116*). Focusing on the detected intermediates (in both bR and bO experiments) that are known to contact the retinal or to participate to the photocycle: D85 (~I_{CD}^{*3}, I_{CD}^{6}), L93 (I_{CD}^{*5}, I_{CD}^{9}), D96 (I_{CD}^{*6}, I_{CD}^{10}), S141 (I_{EF}^{*1}, I_{EF}^{1}) and M145 (I_{EF}^{*2}, I_{EF}^{2}), reveals no changes, or small increases in force and occupancy rate as the retinal is removed, none of which are significant. These changes are minor at most. In summary, the retinal removal does not significantly impact the stability of the intermediates in the N-terminal forced unfolding assay.

Table 5-2: Comparison between intermediates in N-terminal bR and bO unfolding experiments

	State	ΔL_C from $I_{CD}^{*0} \pm$ s.d. (nm)	AA Assign	Occu. rate (%)	Mean unfolding force ± S.E.M (pN)	Mean unfolding force ± S.E.M (pN)	Occu. rate (%)	AA Assign	ΔL_C from $I_{CD}^{0} \pm$ s.d. (nm)	State
		N-terminal bO unfolding intermediates				N-terminal bR unfolding intermediates				
	*0	0 ± 0	74	100	68 ± 3	60 ± 3	100	74	0 ± 0	0
						63 ± 2	91	77	0.3 ±0.2	1
Helix C	*1	0.6 ± 0.3	80	90	69 ± 3	63 ± 2	100	79	0.5 ± 0.1	2
Helix C						64 ± 2	94	81	0.8 ± 0.1	3
Helix C	*2	1.3 ± 0.3	83	87	67 ± 3	61 ± 2	94	83	1.2 ± 0.2	4
Helix C						63 ± 2	88	85	1.6 ± 0.1	5
Helix C	*3	2.0 ± 0.3	87	81	68 ± 3	60 ± 2	85	86	1.9 ± 0.1	6
Helix C						58 ± 2	79	88	2.3 ± 0.2	7
Helix C	*4	2.7 ± 0.2	90	77	66 ± 3	56 ± 2	67	91	2.8 ± 0.2	8
Helix C	*5	3.3 ± 0.3	93	61	64 ± 4	56 ± 2	48	93	3.4 ± 0.2	9
Helix C	*6	4.2 ± 0.3	97	52	57 ± 3	50 ± 4	27	95	3.8 ± 0.2	10
						50 ± 3	27	99	4.5 ± 0.2	11
	*7	5.1 ± 0.3	101	68	55 ± 3	47 ± 2	33	102	5.2 ± 0.2	12
	*8	5.9 ± 0.3	103	74	54 ± 3	44 ± 3	45	103	5.9± 0.3	13
Helix D	*9	6.8 ± 0.3	105	77	49 ± 2	45 ±2	73	105	6.6 ± 0.3	14
Helix D	*10	7.8 ± 0.3	107	68	52 ± 4	40 ± 2	76	106	7.3 ± 0.3	15
Helix D						37 ± 1	51	108	8.1 ± 0.3	16
Helix D	*11	8.6 ± 0.3	109	55	53 ± 5	39 ± 2	54	109	8.6 ± 0.3	17
Helix D	*12	9.7 ± 0.6	111	61	47 ± 4	34 ± 3	30	110	9.3 ± 0.3	18
Helix D						32 ± 3	18	112	10.2 ± 0.4	19
Helix D	*13	13 ± 2	117	39	42 ± 3	26 ± 2	18	116	12.1 ± 0.4	20
	*0	17 ± 3	126	52	35 ± 2					
Helix E						32 ± 1	70	133	20.1 ± 0.6	0
Helix E	*1	21.4 ± 0.8	139	97	45 ± 2	40 ± 2	70	139	21.4 ± 0.6	1
Helix E	*2	23.3 ± 0.7	148	94	52 ± 2	46 ± 3	94	147	23.0 ± 0.5	2
Helix E						43 ± 2	97	152	24.2 ± 0.5	3
Helix E	*3	25.8 ± 0.8	159	97	51 ± 3	41 ± 1	79	159	25.8 ± 0.5	4
Helix E	*4	28.0 ± 0.8	166	97	50 ± 2	39 ± 2	76	165	27.3 ± 0.6	5
Helix F						37.4 ± 0.9	97	169	29.3 ± 0.7	6
Helix F	*5	31.5 ± 0.8	173	81	53 ± 4	36.7 ± 2	64	173	31.6 ± 0.7	7
Helix F	*6	34.0 ± 0.8	178	87	46 ± 3	37 ± 1	73	177	33.4 ± 0.9	8
Helix F						26 ± 2	42	180	35.4 ± 0.5	9
Helix F	*7	37.5 ± 1.2	185	61	39 ± 3	26 ± 2	51	185	37.6 ± 0.5	10
	*0	44.4 ± 1.2	200	23	34 ± 2					
G	*1	49.6 ± 0	223	100	49 ± 3	46.1 ± 0.2	100	223	49.6 ± 0	0
C-terminal tail	*2	51.2 ± 0.7	226	93	51 ± 3	44.1 ± 0.4	94	226	50.7 ± 0.6	1
C-terminal tail						43 ± 2	94		51.9 ± 0.5	2
C-terminal tail	*3	53.0 ± 0.7		97	53 ± 3	42 ± 2	85		53.0 ± 0.7	3
C-terminal tail	*4	54.8 ± 1		83	52 ± 3	42 ± 2	79		55.1 ± 0.9	4
C-terminal tail	*5	61 ± 2		30	41 ± 2	51 ± 2	21		60 ± 2	5
C-terminal tail	*6	69 ± 2		17	46 ± 2					

Figure 5-15: Comparison between N-terminal bR and bO unfolding intermediates location and rupture force. Each detected intermediate for the bR (blue) and bO (red) unfolding experiments are plotted as mean rupture force vs. contour length change (ΔL_C) from the CD major state. The intermediates within the CD helices are depicted in **(A)**, while the intermediates within EF and G helices are depicted in **(B)**. The standard error of the mean of the contour length change and the mean rupture force are plotted as horizontal and vertical (respectively) error bars around each point.

5.5) Discussion and conclusion

The overall lack of significant change in the unfolding behavior (in terms of intermediate location, occupancy and rupture forces) of bO relative to bR is surprising, as it goes against the predictions that the retinal would significantly stabilize rupture force (*79*). Indeed, this also goes counter the decrease in protein stability observed in the G helix in bO unfolding experiments in C-terminal unfolding experiments. This lack of increased unfolding force and intermediate count can be interpreted in a few different ways:

1) The retinal, while significantly stabilizing to the unfolding of the G helix in the C-terminal unfolding experiments (due to the covalent attachment to Lys-216), has a small stabilizing force on the rest of the protein that is so diffuse that it is difficult to measure in any individual intermediate.

2) The techniques and chemistries applied in the preparation of the bO or the extracellular unfolding of bR destabilize the structure enough that the unfolding is not additionally weakened by the removal of the retinal. The most likely cause would be the charged APDMES surface, which may displace the helices enough due to attraction to positively charged residues.

Until bO unfolding is performed on a different surface chemistry, the APDMES disruption of the membrane cannot be excluded. An exploratory study could be conducted to confirm this on bR, accepting that the throughput will be lower.

One objection that can be raised to the above interpretation is the slight *increase* in the rupture forces (<20% for most intermediates, Table 5-2), which suggest increase stabilization by retinal removal. This is most likely representative of a modest spring constant miscalibration, present either in the bO or the bR experiment. One cannot exclude the possibility that these increased rupture forces are representative of increased stability due to non-native bonds formed after the retinal has been removed. Given the 4-fold change in the folded state energy by removing the retinal (*116*), as well as the variability of the tertiary structure (*119*), this seems unlikely.

The surprising nature of the lack of change seen in the unfolding of bO compared to bR, in addition to the significant aberrations in the behavior (Fig 5-8) means that any conclusions of the data is slightly premature, and a larger data set is required, with a few variations in conditions to ensure that the results are being interpreted correctly (such as other surface chemistries, and changing the PEG linker to reduce aggregation). Thus this work can best be understood as ongoing, and that more definitive results can be made with follow-up studies.

5.6) Supplemental figures for bO
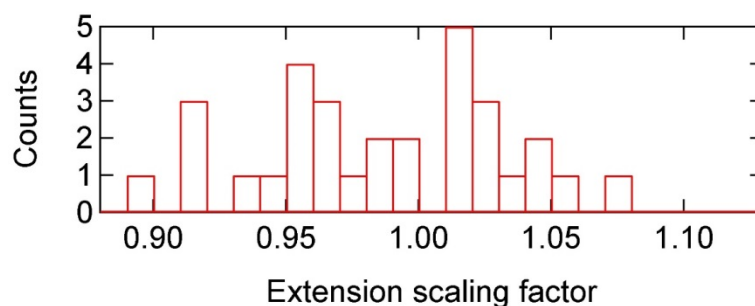
Figure 5-A1: Histogram of scaling factors applied to unfolding records. Each unfolding record's extension is scaled to cause the 2nd and 4th peak major states to align, and a histogram is made of those scaling factors. Most waves are scaled between 95% and 105%. This record bears a resemblance to Fig 5-5, as the scaling factor is determined to set the contour length differences to 49.59 nm, as the N = 31.
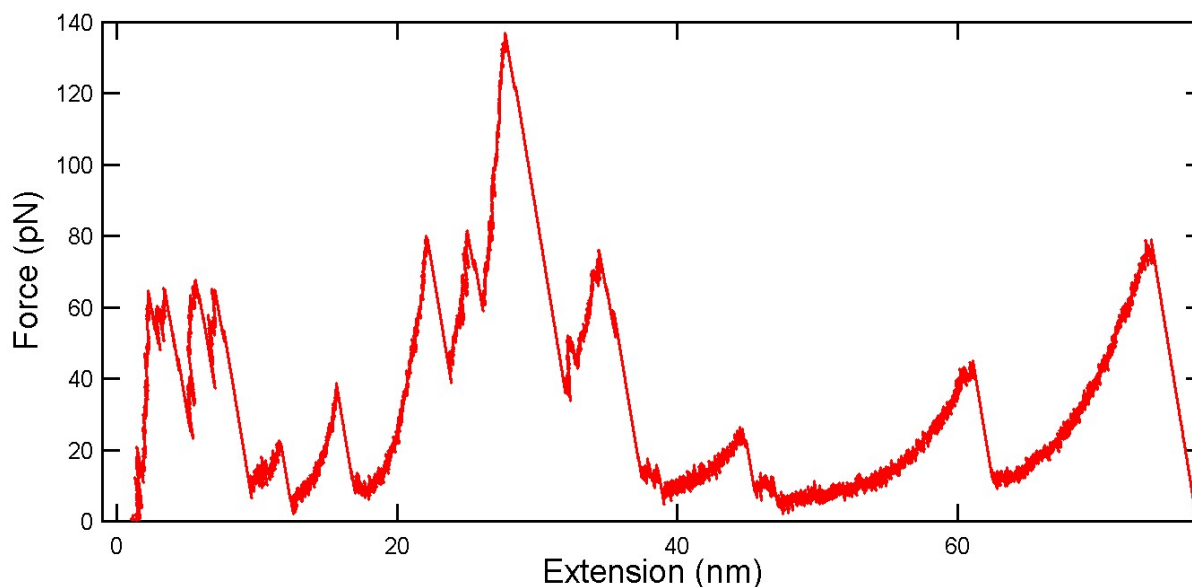
Figure 5-A2: Example of double attachment. Double attachment is evident by the appearance of multiple high force rupture peaks, per helix pair, for multiple helix pairs. A good example of this is presented here, just after the CD major state (around 30 nm), with a peak that goes to 140 pN.
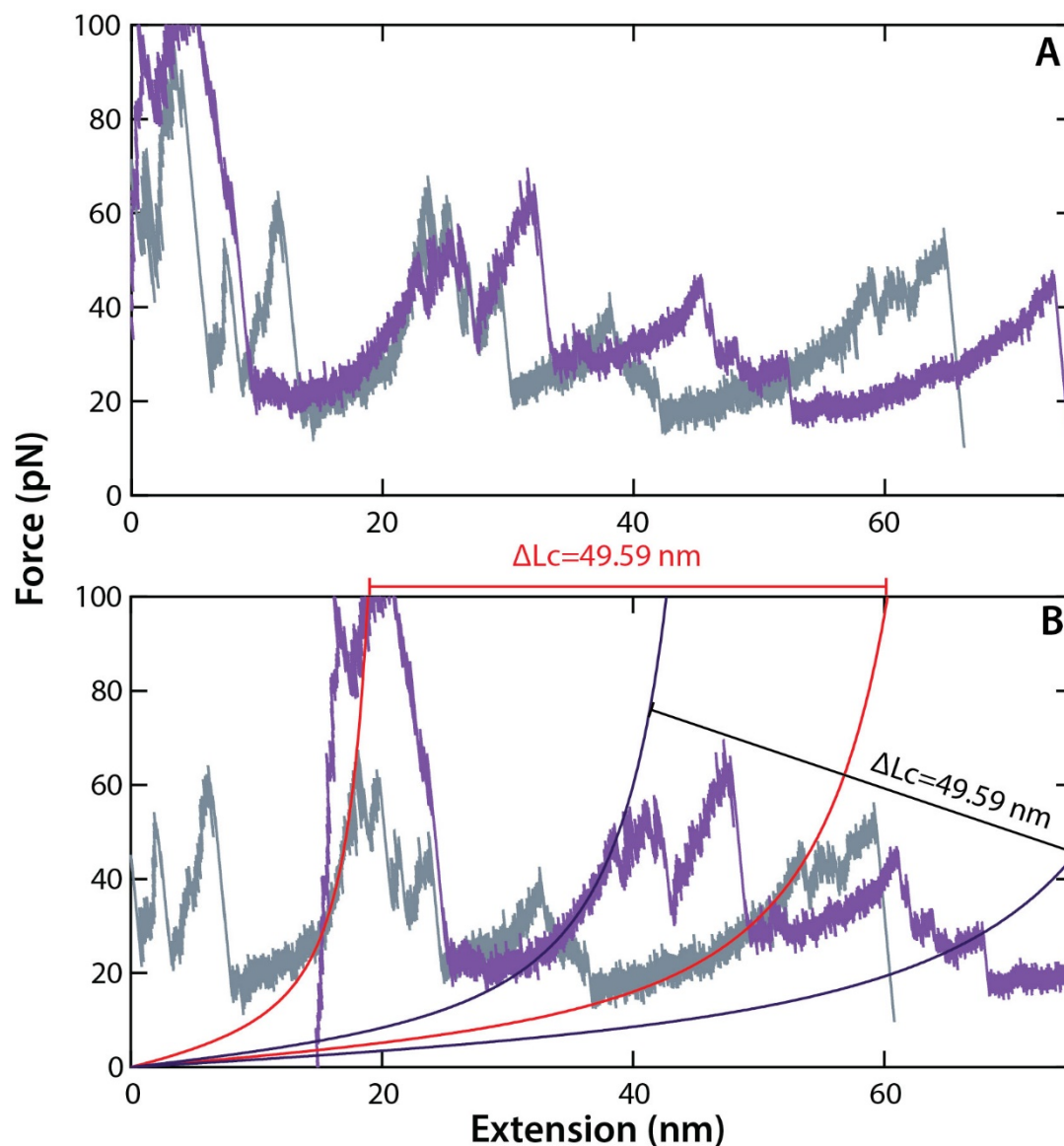
Figure 5-A3: Additional variable contour length demonstrtation. (A) Two unfolding records from October 4th, approximately aligned to the 2nd peak major state. It's apparent that the curvature of the 2nd peak major state for the grey and purple curves are distinctivley different, and with the position of the 4th peak major state, they will not rectify by scaling the extensions, as that will cause the 4th peak to not agree. **(B)** Demonstrates an alternate alignment strategy, based on the variable length of the PEG-linker causing the different curvitures, but the contour length differences (ΔL_c) will still correlate well. To this end, the 2nd peak is fit, with variable contour length and horizontal offset (arising from different tip-surface interactions relative to the adhesion point). Then this horizontal offset is held constant, and the 4th peak major state is fit. As can be seen, both 4th peak major states (albeit short-lived in purple curve) fit to contour lengths 49.59 nm beyond the fit the first peak major states. Note that the WLC model does not fit well to unfolding data at short contour lengths (<30 nm) at low forces (<25 pN).
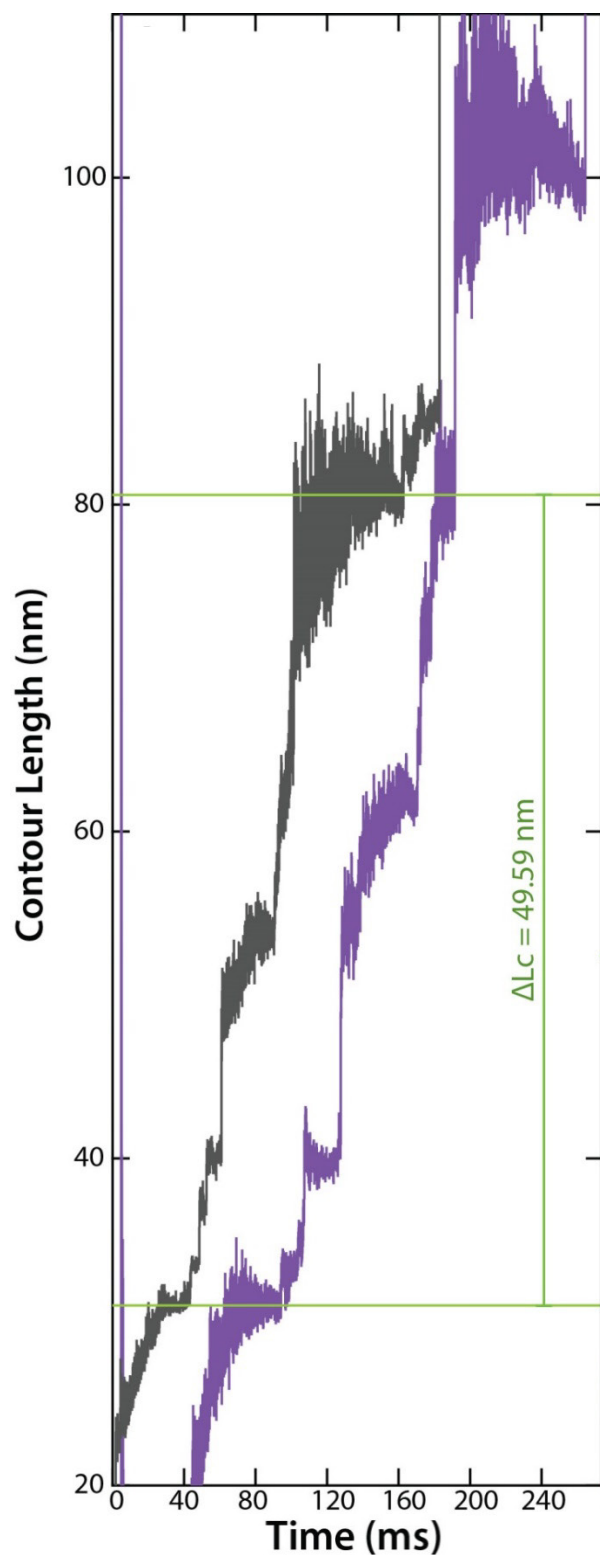
Figure 5-A4: Contour length space depiction of curves from Fig 5-A3. Two unfolding records are presented as contour length vs. time plot, with the colors being the same as they were in Fig 5-A3. The 2nd peak major states are aligned to L_C= 31 nm. The agreement in the spacing of the 2nd peaks major states for the two records (both approximately 49.59 nm) underscores the success of the variable contour length alignment.

CHAPTER VI: Conclusions and Perspectives

The utilization of ultrashort cantilevers (modified to improve spatiotemporal resolution) on the model membrane protein bacteriorhodopsin was expected to reveal short lived states, which it successfully accomplished (resolving dwells as short as 8 µs). The discovery of a significant number of previously hidden intermediates, both from the extensively studied cytoplasmic (C-terminal) side, and the little studied extracellular (N-terminal) side, was unexpected. While the increase in the number of intermediates observed varies significantly by helix pair (from as high as a 7-fold in the C-terminal ED helix pair, to as low as 2 fold increase for the N-terminal EF helix pair), an increase in the number of intermediates is observed for every helix pair studied. It remains to be seen if such a dramatic increase in the number of observed intermediates will be observed in other systems. A study conducted on the globular protein Calmodulin utilizing a high spatiotemporal resolution cantilever observed one additional intermediate previously obscured by standard SMFS cantilevers (*123*).

Coupled with this increase in intermediates is a vast increase in the complexity of observed unfolding pathways. Among these pathways, we observed rapid unfolding and refolding transitions, indicating that we are observing a near equilibrium regime even at the (relatively fast) pulling speed of 300 nm/s, which were presumed to be far from equilibrium. We were even able to utilize this behavior to reconstruct the portion of the energy landscape that underlies those transitions. Widespread refolding might be present in systems beyond bR but has been obscured by the spatiotemporal resolutions of standard cantilevers. It remains to be determined if the different unfolding pathways reflects the exploration of different regions in the energy landscape or stochastic variation within the same region of the energy landscape. The initial analysis of transitions by force suggests the latter theory is at least partially correct. Follow up dynamic force spectroscopy studies on bR can help elucidate this difference.

By comparing the location of the detected intermediates from the N-terminal and C-terminal unfolding experiments, we were able to find evidence for the causes of forces that stabilize a few of the intermediates observed. Four interhelical bonds previously observed in SDS unfolding assays and

predicted by crystallographic determination were observed, making a rare connection between force and chemical denaturation experiments. Also, by comparing the location of intermediates from both sets of unfolding experiments, we were able to observe a strong correlation between intermediates mapped to interhelical loops and the residues solvating during their detection. This supports the theory that hydrophobic interactions between specific residues and the membrane cause those intermediates. The conclusions from the C-terminal and N-terminal comparisons are more speculative than the intermediate identification and would be strengthened with SFMS studies on bR point mutants or repeating the experiments at higher temperatures (weakening the hydrophobic contribution).

Lastly, an initial comparison of the N-terminal forced unfolding behavior of bacteriorhodopsin and its apo-protein, bacterioopsin revealed that retinal does not stabilize bR in any detectable means in an N-terminal unfolding experiment. This is a surprising result and one that we find premature to defend before follow-up studies to verify that the protein is in its native configuration on the APDMES surface.

We are prevented in making stronger conclusion from the comparisons of the results by the inconsistencies in the methodologies: i.e. the use of non-specific attachment chemistry in chapter 3, and specific attachment chemistry with an added polymer linker in chapters 4, and the change from an absolute contour length method (in chapters 3 & 4) to a variable contour length method (chapter 5). The latter discrepancy can be remedied by reanalyzing the data from chapter 4 using a variable contour length alignment method. Currently, my co-first author on the *Science* paper, Hao Yu, has a paper in submission that details an unfolding experiment performed on the C-terminus of both bR and bO with the same attachment chemistry utilized in chapters 4 and 5. This will provide a better C-terminal bR comparison for the N-terminal bR unfolding documented in chapter 4.

Looking forward, the use of high-spatiotemporal resolution cantilevers can readily be applied to other membrane proteins systems. OmpG, a model β-barrel membrane protein, seems like an ideal target. In additional to the existence of previous AFM SMFS studies on OmpG which can serve as a comparison to this hypothetical study (*124*), β-sheet rich proteins are believed to fold significantly slower than α-helical proteins (*125*) with β-hairpin formation being measured to take place on the order of microseconds

(compared to hundreds of nanoseconds for α-helix formation). Intriguingly, this opens the potential for the 1-microsecond resolution cantilevers to observe all unfolding intermediates. It is my hope that the newly revealed intermediates will provide more precise information into structural changes arising from ligand binding and the impact of mutations on the folding process. Beyond membrane proteins, it will be exciting to see what new dynamics are uncovered by the use of high-spatiotemporal resolution cantilevers on globular proteins, nucleic acid structures, and mechanosensitive enzymes.

References

1. J. Liang, H. Naveed, D. Jimenez-Morales, L. Adamian, M. Lin, Computational studies of membrane proteins: Models and predictions for biological understanding. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1818**, 927–941 (2012).

2. J. P. Overington, B. Al-Lazikani, A. L. Hopkins, How many drug targets are there? *Nature Reviews Drug Discovery*. **5**, 993–996 (2006).

3. G. L. Lukacs, A. S. Verkman, CFTR: folding, misfolding and correcting the ΔF508 conformational defect. *Trends in Molecular Medicine*. **18**, 81–91 (2012).

4. S. White, Membrane Proteins of Known Structure, (available at http://blanco.biomol.uci.edu/mpstruc/).

5. H.-X. Zhou, T. A. Cross, Influences of Membrane Mimetic Environments on Membrane Protein Structures. *Annual Review of Biophysics*. **42**, 361–392 (2013).

6. E. A. Roman, F. L. González Flecha, Kinetics and Thermodynamics of Membrane Protein Folding. *Biomolecules*. **4**, 354–373 (2014).

7. J.-L. Popot, S.-E. Gerchman, D. M. Engelman, Refolding of bacteriorhodopsin in lipid bilayers: A thermodynamically controlled two-stage process. *Journal of Molecular Biology*. **198**, 655–676 (1987).

8. W. Tian, M. Lin, K. Tang, J. Liang, H. Naveed, High-resolution structure prediction of β-barrel membrane proteins. *PNAS*, 201716817 (2018).

9. G. A. Khoury, J. Smadbeck, C. A. Kieslich, C. A. Floudas, Protein folding and de novo protein design for biotechnological applications. *Trends in Biotechnology*. **32**, 99–109 (2014).

10. K. A. Dill, J. L. MacCallum, The Protein-Folding Problem, 50 Years On. *Science*. **338**, 1042–1046 (2012).

11. B. L. Kim, N. P. Schafer, P. G. Wolynes, Predictive energy landscapes for folding α-helical transmembrane proteins. *PNAS*. **111**, 11031–11036 (2014).

12. A. Kishino, T. Yanagida, Force measurements by micromanipulation of a single actin filament by glass needles. *Nature*. **334**, 74–76 (1988).

13. G. Lee, Sensing Discrete Stepaviding-Biotin Interaction with Atomic Force Microscopy. *Langmuir* (1994).

14. E. L. Florin, V. T. Moy, H. E. Gaub, Adhesion forces between individual ligand-receptor pairs. *Science*. **264**, 415–417 (1994).

15. S. B. Smith, L. Finzi, C. Bustamante, Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*. **258**, 1122–1126 (1992).

16. M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, H. E. Gaub, Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science*. **276**, 1109–1112 (1997).

17. F. Oesterhelt *et al.*, Unfolding Pathways of Individual Bacteriorhodopsins. *Science*. **288**, 143–146 (2000).

18. A. Engel, H. E. Gaub, D. J. Müller, Atomic force microscopy: A forceful way with single molecules. *Current Biology*. **9**, R133–R136 (1999).

19. S. Kawamura, A. T. Colozo, D. J. Mueller, P. S.-H. Park, Conservation of Molecular Interactions Stabilizing Bovine and Mouse Rhodopsin. *Biochemistry*. **49**, 10412–10420 (2010).

20. K. T. Sapra, P. S.-H. Park, K. Palezewski, D. J. Muller, Mechanical properties of bovine rhodopsin and bacteriorhodopsin: Possible roles in folding and function. *Langmuir*. **24**, 1330–1337 (2008).

21. S. Kawamura *et al.*, Kinetic, Energetic, and Mechanical Differences between Dark-State Rhodopsin and Opsin. *Structure*. **21**, 426–437 (2013).

22. D. A. Cisneros, D. Oesterhelt, D. J. Muller, Probing origins of molecular interactions stabilizing the membrane proteins halorhodopsin and bacteriorhodopsin. *Structure*. **13**, 235–242 (2005).

23. A. L. Klyszejko *et al.*, Folding and assembly of proteorhodopsin. *Journal of Molecular Biology*. **376**, 35–41 (2008).

24. A. Kedrov, M. Krieg, C. Ziegler, W. Kuhlbrandt, D. J. Muller, Locating ligand binding and activation of a single antiporter. *Embo Reports*. **6**, 668–674 (2005).

25. C. A. Bippes *et al.*, Substrate Binding Tunes Conformational Flexibility and Kinetic Stability of an Amino Acid Antiporter. *J. Biol. Chem.* **284**, 18651–18663 (2009).

26. A. Kedrov *et al.*, Probing the Interactions of Carboxy-atractyloside and Atractyloside with the Yeast Mitochondrial ADP/ATP Carrier. *Structure*. **18**, 39–46 (2010).

27. C. Moller *et al.*, Determining molecular forces that stabilize human aquaporin-1. *Journal of Structural Biology*. **142**, 369–378 (2003).

28. P. D. Bosshart *et al.*, The Transmembrane Protein KpOmpA Anchoring the Outer Membrane of Klebsiella pneumoniae Unfolds and Refolds in Response to Tensile Load. *Structure*. **20**, 121–127 (2012).

29. L. Ge, C. Perez, I. Waclawska, C. Ziegler, D. J. Muller, Locating an extracellular K+-dependent interaction site that modulates betaine-binding of the Na+-coupled betaine symporter BetP. *PNAS*. **108**, E890–E898 (2011).

30. A. Kedrov *et al.*, Detecting molecular interactions that stabilize, activate and guide ligand-binding of the sodium/proton antiporter MjNhaP1 from Methanococcus jannaschii. *Journal of Structural Biology*. **159**, 290–301 (2007).

31. T. Serdiuk *et al.*, YidC assists the stepwise and stochastic folding of membrane proteins. *Nature Chemical Biology*. **12**, 911–917 (2016).

32.   T. Serdiuk, S. A. Mari, D. J. Müller, Pull-and-Paste of Single Transmembrane Proteins. *Nano Lett*. **17**, 4478–4488 (2017).

33.   D. Min, R. E. Jefferson, J. U. Bowie, T.-Y. Yoon, Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat Chem Biol*. **11**, 981–987 (2015).

34.   M. Zocher, J. J. Fung, B. K. Kobilka, D. J. Mueller, Ligand-Specific Interactions Modulate Kinetic, Energetic, and Mechanical Properties of the Human beta(2) Adrenergic Receptor. *Structure*. **20**, 1391–1402 (2012).

35.   M. Damaghi *et al.*, pH-Dependent Interactions Guide the Folding and Gate the Transmembrane Pore of the beta-Barrel Membrane Protein OmpG. *Journal of Molecular Biology*. **397**, 878–882 (2010).

36.   J. Thoma, B. M. Burmann, S. Hiller, D. J. Müller, Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. *Nature Structural & Molecular Biology*. **22**, 795–802 (2015).

37.   J. Thoma *et al.*, Maltoporin LamB Unfolds β Hairpins along Mechanical Stress-Dependent Unfolding Pathways. *Structure*. **25**, 1139-1144.e2 (2017).

38.   L.-N. Liu, Forces guiding assembly of light-harvesting complex 2 in native membranes. *PNAS* (2011).

39.   D. Seiwert, H. Witt, A. Janshoff, H. Paulsen, The non-bilayer lipid MGDG stabilizes the major light-harvesting complex (LHCII) against unfolding. *Scientific Reports*. **7**, 5158 (2017).

40.   N. L. Stephenson, J. M. Avis, Direct observation of proteolytic cleavage at the S2 site upon forced unfolding of the Notch negative regulatory region. *PNAS*. **109**, E2757–E2765 (2012).

41.   D. N. Ganchev, D. T. S. Rijkers, M. M. E. Snel, J. A. Killian, B. de Kruijff, Strength of Integration of Transmembrane α-Helical Peptides in Lipid Bilayers As Determined by Atomic Force Spectroscopy. *Biochemistry*. **43**, 14987–14993 (2004).

42.   S. A. Contera, V. Lemaître, M. R. R. de Planque, A. Watts, J. F. Ryan, Unfolding and Extraction of a Transmembrane α-Helical Peptide: Dynamic Force Spectroscopy and Molecular Dynamics Simulations. *Biophysical Journal*. **89**, 3129–3140 (2005).

43.   G. Binnig, C. F. Quate, C. Gerber, Atomic Force Microscope. *Phys. Rev. Lett*. **56**, 930–933 (1986).

44.   K. C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Meth*. **5**, 491–505 (2008).

45.   G. Stirnemann, S. Kang, R. Zhou, B. J. Berne, How force unfolding differs from chemical denaturation. *PNAS*. **111**, 3413–3418 (2014).

46.   N. H. Joh *et al.*, Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature*. **453**, 1266 (2008).

47.   H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailler, J. K. Lanyi, Structure of bacteriorhodopsin at 1.55 Å resolution 1. *Journal of Molecular Biology*. **291**, 899–911 (1999).

48.   F. Cymer, G. von Heijne, S. H. White, Mechanisms of Integral Membrane Protein Insertion and Folding. *Journal of Molecular Biology*. **427**, 999–1022 (2015).

49.   M. Kessler, H. E. Gaub, Unfolding Barriers in Bacteriorhodopsin Probed from the Cytoplasmic and the Extracellular Side by AFM. *Structure*. **14**, 521–527 (2006).

50.   M. B. Viani *et al.*, Small cantilevers for force spectroscopy of single molecules. *Journal of Applied Physics*. **86**, 2258–2262 (1999).

51.   M. Gruebele, The Fast Protein Folding Problem. *Annual Review of Physical Chemistry*. **50**, 485–516 (1999).

52.   M. S. Bull, R. M. A. Sullan, H. Li, T. T. Perkins, Improved Single Molecule Force Spectroscopy Using Micromachined Cantilevers. *ACS Nano*. **8**, 4984–4995 (2014).

53.   D. T. Edwards *et al.*, Optimizing 1-μs-Resolution Single-Molecule Force Spectroscopy on a Commercial Atomic Force Microscope. *Nano Lett.* **15**, 7091–7098 (2015).

54.   S. Subramaniam, R. Henderson, Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature*. **406**, 653–657 (2000).

55.   R. Henderson, P. N. T. Unwin, Three-dimensional model of purple membrane obtained by electron microscopy. *Nature*. **257**, 28–32 (1975).

56.   K. J. Rothschild, FTIR difference spectroscopy of bacteriorhodopsin: Toward a molecular model. *J Bioenerg Biomembr*. **24**, 147–167 (1992).

57.   M. Etzkorn *et al.*, Cell-free Expressed Bacteriorhodopsin in Different Soluble Membrane Mimetics: Biophysical Properties and NMR Accessibility. *Structure*. **21**, 394–401 (2013).

58.   E. Nango *et al.*, A three-dimensional movie of structural changes in bacteriorhodopsin. *Science*. **354**, 1552–1557 (2016).

59.   D. Voet, J. Voet, *Biochemistry* (John Wiley & Sons, Hoboken, NJ, Fourth., 2011).

60.   S. H. White, W. C. Wimley, Membrane protein folding and stability: Physical principles. *Annual Review of Biophysics and Biomolecular Structure; Palo Alto*. **28**, 319 (1999).

61.   L. Maglova, B. Atanasov, L. Keszthelyi, Unfolding of monomeric bacteriorhodopsin in water-urea solution. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. **975**, 271–276 (1989).

62.   V. Krishnamani, B. G. Hegde, R. Langen, J. K. Lanyi, Secondary and Tertiary Structure of Bacteriorhodopsin in the SDS Denatured State. *Biochemistry*. **51**, 1051–1060 (2012).

63.   L. S. Brown, O. P. Ernst, Recent advances in biophysical studies of rhodopsins – Oligomerization, folding, and structure. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. **1865**, 1512–1521 (2017).

64.   K. T. Sapra, S. Besir, D. Oesterhelt, D. J. Muller, Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy. *J. Mol. Biol.* **355**, 640–650 (2006).

65. Intermediates in the Assembly of Bacteriorhodopsin Investigated by Time-Resolved Absorption Spectroscopy - Booth - 1997 - European Journal of Biochemistry - Wiley Online Library, (available at https://onlinelibrary.wiley.com/doi/full/10.1111/j.1432-1033.1997.00674.x).

66. D. J. Müller, H.-J. Sass, S. A. Müller, G. Büldt, A. Engel, Surface structures of native bacteriorhodopsin depend on the molecular packing arrangement in the membrane. *Journal of Molecular Biology*. **285**, 1903–1909 (1999).

67. D. J. Müller *et al.*, Stability of Bacteriorhodopsin α-Helices and Loops Analyzed by Single-Molecule Force Spectroscopy. *Biophysical Journal*. **83**, 3578–3588 (2002).

68. H. Janovjak, M. Kessler, D. Oesterhelt, H. Gaub, D. J. Muller, Unfolding pathways of native bacteriorhodopsin depend on temperature. *Embo Journal*. **22**, 5220–5229 (2003).

69. C. Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* (Wiley, New York, 2nd edition., 1980).

70. H. Janovjak *et al.*, Probing the Energy Landscape of the Membrane Protein Bacteriorhodopsin. *Structure*. **12**, 871–879 (2004).

71. M. Kessler, K. E. Gottschalk, H. Janovjak, D. J. Muller, H. E. Gaub, Bacteriorhodopsin folds into the membrane against an external force. *Journal of Molecular Biology*. **357**, 644–654 (2006).

72. K. T. Sapra, G. P. Balasubramanian, D. Labudde, J. U. Bowie, D. J. Muller, Point mutations in membrane proteins reshape energy landscape and populate different unfolding pathways. *Journal of Molecular Biology*. **376**, 1076–1090 (2008).

73. K. T. Sapra, J. Doehner, V. Renugopalakrishnan, E. Padros, D. J. Muller, Role of extracellular glutamic acids in the stability and energy landscape of bacteriorhodopsin. *Biophysical Journal*. **95**, 3407–3418 (2008).

74. M. Zocher *et al.*, Single-Molecule Force Spectroscopy from Nanodiscs: An Assay to Quantify Folding, Stability, and Interactions of Native Membrane Proteins. *ACS Nano*. **6**, 961–971 (2012).

75. R. Petrosyan *et al.*, Single-Molecule Force Spectroscopy of Membrane Proteins from Membranes Freely Spanning Across Nanoscopic Pores. *Nano Lett*. **15**, 3624–3633 (2015).

76. T. Yamada, T. Yamato, S. Mitaku, Forced Unfolding Mechanism of Bacteriorhodopsin as Revealed by Coarse-Grained Molecular Dynamics. *Biophysical Journal*. **111**, 2086–2098 (2016).

77. K. Voïtchovsky, S. A. Contera, J. F. Ryan, Electrostatic and Steric Interactions Determine Bacteriorhodopsin Single-Molecule Biomechanics. *Biophysical Journal*. **93**, 2024–2037 (2007).

78. M. Cieplak, S. Filipek, H. Janovjak, K. A. Krzyśko, Pulling single bacteriorhodopsin out of a membrane: Comparison of simulation and experiment. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1758**, 537–544 (2006).

79. C. Kappel, H. Grubmüller, Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network. *Biophysical Journal*. **100**, 1109–1119 (2011).

80. M. A. Roseman, Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *Journal of Molecular Biology*. **200**, 513–522 (1988).

81. H. Yu, M. G. W. Siewny, D. T. Edwards, A. W. Sanders, T. T. Perkins, Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science*. **355**, 945–950 (2017).

82. A. Carpenter, Human TATA-Binding Protein Interactions with DNA Characterized by Stabilized, Axial Optical Trapping. *Physics Graduate Theses & Dissertations* (2011) (available at https://scholar.colorado.edu/phys_gradetds/10).

83. D. Oesterhelt, W. Stoeckenius, in *Methods in Enzymology* (Academic Press, 1974; http://www.sciencedirect.com/science/article/pii/0076687974310725), vol. 31 of *Biomembranes Part A*, pp. 667–678.

84. Röntgenuntersuchung gelöster Fadenmoleküle - Kratky - 1949 - Recueil des Travaux Chimiques des Pays-Bas - Wiley Online Library, (available at https://onlinelibrary.wiley.com/doi/abs/10.1002/recl.19490681203).

85. J. F. Marko, E. D. Siggia, Stretching DNA. *Macromolecules*. **28**, 8759–8770 (1995).

86. SPIETV, *Thomas Perkins: Atomic force microscopy measures properties of proteins and protein folding* (https://www.youtube.com/watch?v=xATMav9-cf0).

87. R. M. A. Sullan, J. K. Li, S. Zou, Direct Correlation of Structures and Nanomechanical Properties of Multicomponent Lipid Bilayers. *Langmuir*. **25**, 7471–7477 (2009).

88. C. A. Bippes, D. J. Muller, High-resolution atomic force microscopy and spectroscopy of native membrane proteins. *Rep. Prog. Phys.* **74**, 086601 (2011).

89. S. W. Englander, L. Mayne, The case for defined protein folding pathways. *PNAS*. **114**, 8253–8258 (2017).

90. J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gebhardt, M. Rief, The Complex Folding Network of Single Calmodulin Molecules. *Science*. **334**, 512–516 (2011).

91. F. Rico, L. Gonzalez, I. Casuso, M. Puig-Vidal, S. Scheuring, High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations. *Science*. **342**, 741–743 (2013).

92. K. Neupane *et al.*, Direct observation of transition paths during the folding of proteins and nucleic acids. *Science*. **352**, 239–242 (2016).

93. H. S. Chung, K. McHale, J. M. Louis, W. A. Eaton, Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times. *Science*. **335**, 981–984 (2012).

94. J. M. Fernandez, H. Li, Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein. *Science*. **303**, 1674–1678 (2004).

95. S. Jayasinghe, K. Hristova, S. H. White, Energetics, stability, and prediction of transmembrane helices. *Journal of Molecular Biology*. **312**, 927–934 (2001).

96. J. P. Junker, F. Ziegler, M. Rief, Ligand-Dependent Equilibrium Fluctuations of Single Calmodulin Molecules. *Science*. **323**, 633–637 (2009).

97.   M. Kim *et al.*, Fast and Forceful Refolding of Stretched α-Helical Solenoid Proteins. *Biophysical Journal*. **98**, 3086–3092 (2010).

98.   Y. M. Rhee, V. S. Pande, One-Dimensional Reaction Coordinate and the Corresponding Potential of Mean Force from Commitment Probability Distribution. *J. Phys. Chem. B*. **109**, 6780–6786 (2005).

99.   M. T. Woodside *et al.*, Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science*. **314**, 1001–1004 (2006).

100.   G. I. Bell, Models for the specific adhesion of cells to cells. *Science*. **200**, 618–627 (1978).

101.   J. Preiner *et al.*, Free Energy of Membrane Protein Unfolding Derived from Single-Molecule Force Measurements. *Biophysical Journal*. **93**, 930–937 (2007).

102.   P. D. Bosshart, P. L. T. M. Frederix, A. Engel, Reference-Free Alignment and Sorting of Single-Molecule Force Spectroscopy Data. *Biophysical Journal*. **102**, 2202–2211 (2012).

103.   C. Bustamante, J. F. Marko, E. D. Siggia, S. Smith, Entropic elasticity of lambda-phage DNA. *Science*. **265**, 1599–1600 (1994).

104.   N. J. Gibson, J. Y. Cassim, Nature of forces stabilizing the transmembrane protein bacteriorhodopsin in purple membrane. *Biophysical Journal*. **56**, 769–780 (1989).

105.   Orientations of Proteins in Membranes (OPM) database, (available at http://opm.phar.umich.edu/).

106.   B. Schobert, J. Cupp-Vickery, V. Hornak, S. O. Smith, J. K. Lanyi, Crystallographic Structure of the K Intermediate of Bacteriorhodopsin: Conservation of Free Energy after Photoisomerization of the Retinal. *Journal of Molecular Biology*. **321**, 715–726 (2002).

107.   R. Walder *et al.*, Rapid Characterization of a Mechanically Labile α-Helical Protein Enabled by Efficient Site-Specific Bioconjugation. *J. Am. Chem. Soc.* **139**, 9867–9875 (2017).

108.   M. V. Maslova, L. G. Gerasimova, W. Forsling, Surface Properties of Cleaved Mica. *Colloid Journal*. **66**, 322–328 (2004).

109.   H. Yu, D. T. Edwards, L. Uyetake, T. T. Perkins, P. R. Heenan, Characterizing the Initial Unfolding of Bacteriorhodopsin Reveals a Retinal-Stabilize Obligate Intermediate. *In Preparation*.

110.   C. Nick Pace, J. Martin Scholtz, A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal*. **75**, 422–427 (1998).

111.   L.-O. Essen, R. Siegert, W. D. Lehmann, D. Oesterhelt, Lipid patches in membrane protein oligomers: Crystal structure of the bacteriorhodopsin-lipid complex. *PNAS*. **95**, 11673–11678 (1998).

112.   R. Renthal, G. J. Harris, R. Parrish, Reaction of the purple membrane with a carbodimide. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. **547**, 258–269 (1979).

113.   O. Edholm, O. Berger, F. Jähnig, Structure and Fluctuations of Bacteriorhodopsin in the Purple Membrane: A Molecular Dynamics Study. *Journal of Molecular Biology*. **250**, 94–111 (1995).

114. G. J. Turner *et al.*, The Bacteriorhodopsin Carboxyl-Terminus Contributes to Proton Recruitment and Protein Stability. *Biochemistry*. **48**, 1112–1122 (2009).

115. S. Yamaguchi *et al.*, Cytoplasmic surface structure of bacteriorhodopsin consisting of interhelical loops and C-terminal α helix, modified by a variety of environmental factors as studied by 13C-NMR. *European Journal of Biochemistry*. **268**, 2218–2228 (2001).

116. P. Curnow, P. J. Booth, The Contribution of a Covalently Bound Cofactor to the Folding and Thermodynamic Stability of an Integral Membrane Protein. *Journal of Molecular Biology*. **403**, 630–642 (2010).

117. D. M. Engelman *et al.*, Membrane protein folding: beyond the two stage model. *FEBS Letters*. **555**, 122–125 (2003).

118. C. Moller, G. Buldt, N. A. Dencher, A. Engel, D. J. Muller, Reversible loss of crystallinity on photobleaching purple membrane in the presence of hydroxylamine. *Journal of Molecular Biology*. **301**, 869–879 (2000).

119. S. Yamaguchi *et al.*, Irreversible Conformational Change of Bacterio-opsin Induced by Binding of Retinal during Its Reconstitution to Bacteriorhodopsin, as Studied by 13C NMR. *J Biochem (Tokyo)*. **127**, 861–869 (2000).

120. D. Oesterhelt, L. Schuhmann, H. Gruber, Light-dependent reaction of bacteriorhodopsin with hydroxylamine in cell suspensions of Halobacterium halobium: Demonstration of an APO-membrane. *FEBS Letters*. **44**, 257–261 (1974).

121. M. Iwamoto, Y. Sudo, K. Shimono, N. Kamo, Selective reaction of hydroxylamine with chromophore during the photocycle of pharaonis phoborhodopsin. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1514**, 152–158 (2001).

122. J. L. Hutter, J. Bechhoefer, Calibration of atomic-force microscope tips. *Review of Scientific Instruments*. **64**, 1868–1873 (1993).

123. D. T. Edwards, J. K. Faulk, M.-A. LeBlanc, T. T. Perkins, Force Spectroscopy with 9-μs Resolution and Sub-pN Stability by Tailoring AFM Cantilever Geometry. *Biophysical Journal*. **113**, 2595–2600 (2017).

124. M. Damaghi, S. Koester, C. A. Bippes, O. Yildiz, D. J. Mueller, One beta Hairpin Follows the Other: Exploring Refolding Pathways and Kinetics of the Transmembrane beta-Barrel Protein OmpG. *Angewandte Chemie-International Edition*. **50**, 7422–7424 (2011).

125. K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, The Protein Folding Problem. *Annu Rev Biophys*. **37**, 289–316 (2008).