2015

Examination Of The Dynamic Assembly Equilibrium For E. Coli Clpb

Jiabei Lin
University of Alabama at Birmingham

EXAMINATION OF THE DYNAMIC ASSEMBLY EQUILIBRIUM FOR *E. COLI* CLPB

by

JIABEI LIN

AARON L. LUCIUS, COMMITTEE CHAIR
DAVID E. GRAVES
DONALD D. MUCCIO
PETER E. PREVELIGE
BINGDONG SHA

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2015

EXAMINATION OF THE DYNAMIC ASSEMBLY EQUILIBRIUM FOR *E. COLI* CLPB

JIABEI LIN

CHEMISTRY

ABSTRACT

As a member of the Clp/Hsp100 chaperone family, *E. coli* ClpB is able to disaggregate denatured proteins with assistance from DnaKJE co-chaperones in cell. However, the working mechanism of ClpB disaggregation remains unclear. The active structure of ClpB is known to be a hexameric ring. It is known that ATP binding and hydrolysis are required for ClpB to perform its chaperone activity. Therefore, studying the energetics and kinetics of the ATP linked ClpB assembly equilibrium is essential for the quantitative examination of ClpB-protein substrate interaction to fully reveal its disaggregation mechanism. ATPγS (slowly hydrolysable ATP analog) is used as a model for ATP in this study. In order to examine the ligand liked ClpB assembly, ClpB self-assembly in the absence of nucleotide needs to be determined. In the first part of this study, we introduce to you the methods that were applied to perform this study using analytical ultracentrifugation. By performing NLLS analysis on the simulated sedimentation velocity data, we presented that both thermodynamic and kinetic parameters of a complex assembly system can be determined accurately with certain limitations. Further, the linkage of ligand binding can be determined by analyzing the assembly equilibrium constants as a function of [ligand]. In the second part, we applied the methods discussed in the first section to the determination of the assembly energetics and kinetics for ClpB in the absence of nucleotide. Here, we show that ClpB can form

hexamers in the absence of nucleotide through two intermediates, dimers and tetramers. The assembly equilibrium constants and dissociation rate constants were determined for each oligomer in the absence of nucleotide. With these results, we examined the linkage of [ATPγS] binding to ClpB assembly. It has been assumed that ClpB forms only hexamer in the presence of ATP/ATPγS; however, here we show that ClpB exhibits a dynamic equilibrium in the presence of both limiting and excess ATPγS. ClpB monomer, dimer, tetramer, and hexamer were observed and their assembly equilibrium constants were determined. These interaction constants make it possible to predict the concentration of hexamers present and are able to bind to co-chaperones and polypeptide substrates. Such information is essential for the interpretation of many *in vitro* studies. Moreover, the ATPγS binding equilibrium constant and stoichiometry for each oligomer were determined for the first time. All twelve NBDs of the hexameric ring are saturated with ATPγS bound, however, the binding stoichiometry of dimers and tetramers is one fewer than the maximum number of the NBDs, which suggests an open conformation. Our results are consistent with the previously published structure studies. Finally, the strategies presented here are broadly applicable to a large number of AAA+ molecular motors that assemble upon nucleotide binding.

Key word: Analytical ultracentrifugation, Assembly, ATPase, ClpB, Ligand linked, Kinetics, Thermodynamics

DEDICATION

To my family.

ACKNOWLEDGEMENTS

I would like to thank my mentor Dr. Aaron Lucius. Thank you for all your patience and efforts spent on training me to think, write, and communicate as a scientist. Thank you for all your encouragements and challenges that kept me going forward. I can't count how many bets we had on disagreements and how many times you said "Prove I'm wrong.", but I do remember the enthusiasm and pressure that push me to solve the problems. Regardless of all the arguments we had, thank you for allowing me to pursue my ideas but also try to keep me focused at the same time. I would like to express my gratitude to my current and former colleagues for the training and suggestions. These include Elizabeth Duran, Zac Ingram, Nate Scull, Ryan Stafford, Clarissa Weaver, Tao Li, Justin Miller, and Keith Veronese. Special thanks to Tao Li for being a good friend and great company for many years in graduate school. I would like to thank my graduate committee, David Graves, Aaron Lucius, Donald Muccio, Peter Prevelige and Bingdong Sha, for their instructive suggestions. Special thanks to Peter Prevelige and Bingdong Sha for the valuable trainings during my time working in your labs. Finally, I want to thank my beloved husband, Ali Zamani, for your support and encouragement that allowed me to pursue my dream. Thanks to my parents, who always love and trust me. Thanks to my expecting baby, Sara, who has been so great and allowed me to focus on preparing three manuscripts and the final defense.

TABLE OF CONTENT

ABSTRACT iii
DEDICATION v
ACKNOWLEDGEMENTS vi
LIST OF TABLES xi
LIST OF FIGURES xiii
LIST OF ABBREVIATIONS xv
INTRODUCTION 1
The methodology of using analytical ultracentrifugation to study assembly equilibria 4
ClpB forms hexamer in the absence of nucleotide 5
ClpB exhibits a dynamic assembly 7
ATPγS binding affinity and stoichiometry for each ClpB oligomer 10
ANALYSIS OF LINKED EQUILIBRIA 13
Abstract 14
Introduction 15
Determination of $L_{n,0}$ for an assembling system 23
Global Fitting of Sedimentation Velocity Data as a Function of Protein Concentration 24
Kinetic Considerations 24
Global analysis using the 1-2-4-6 model with rapid dissociating oligomers 31
Global analysis using the 1-2-4-6 model with slow dissociation of oligomers 34
Global analysis with rate constants in the detectable range 36
Conclusions 37
References 40
EXAMINATION OF THE DYNAMIC ASSEMBLY EQUILIBRIUM FOR *E.COLI* CLPB 50
Abstract 51
Introduction 52
Materials and Methods 56

Buffers 56
Strains, plasmid and ClpB protein 56
Analytical ultracentrifugation 59
Analysis of Sedimentation velocity data 60
Analysis of Sedimentation Equilibrium Data 61
Sedimentation coefficients (s) estimation from global fitting of the time difference curves 61
Global fitting of sedimentation velocity data using the time difference curve method 62
Results 65
ClpB exhibits dynamic assembly in the absence of nucleotide 65
NaCl concentration has a strong impact on ClpB assembly 66
One ClpB oligomer is predominately populated at 100 mM NaCl 67
Determination of the molecular weight of the largely populated ClpB oligomer in the absence of nucleotide in Buffer H with 100 mM NaCl 67
Determination of the assembly model for ClpB 69
Determination of the assembly path for ClpB 71
Global analysis of Sedimentation Velocity data collected in the presence of 100 and 200 mM NaCl 74
Discussion 76
ClpB forms hexamer in the absence of nucleotide 79
ClpB hexamer exhibits rapid dissociation 80
Acknowledgements: 83
Supporting Information 100
EXAMINATION OF CLPB QUATERNARY STRUCTURE AND LINKAGE TO NUCLEOTIDE BINDING 111
Abstract 112
Introduction 113
Materials and Methods 115
Protein and buffer 115
Sedimentation velocity experiment using interference optics 115
Sedimentation velocity experiment using absorbance optics 116

Analysis of sedimentation velocity data 116
Global analysis of the time difference curves from sedimentation velocity experiments performed at various ClpB concentrations for a fixed ATPγS concentration 118
Analysis of the difference curves from sedimentation velocity data for a single ClpB concentration when the [ATPγS] is not in large excess over the [ClpB] 118
Thermodynamic reversibility tests for ATPγS linked ClpB assembly 119
Results: 119
ClpB exhibits dynamic assembly in the presence of nucleotide 119
ClpB resides in a monomer-dimer-tetramer-hexamer equilibrium in the presence of ATPγS 122
Measurement of the free ATPγS concentration in solution 125
Determination of the binding density of ClpB oligomers 126
Global analysis of $L_{n,app}$ data as a function of $[ATP\gamma S]_f$ 130
ClpB monomer binds ATPγS 135
ATPγS driven ClpB assembly reaction is thermodynamically reversible 136
Discussion 136
ClpB resides in a dynamic equilibrium in the presence of ATPγS 137
ClpB hexamer dissociation rate constant has an ATPγS concentration dependence 139
Determination of the binding stoichiometry and affinity of ATPγS to each ClpB oligomer 140
Author Contributions 143
Acknowledgements 144
References: 145
Supporting information 147
CONCLUSION 153
Analytical ultracentrifugation as a tool to study assembly equilibria 154
ClpB forms hexamer in the absence of nucleotide 155
ClpB exhibits a dynamic assembly 155
ATPγS binding affinity and stoichiometry 156

Summary 157
REFERENCE FOR INTRODUCTION AND CONCLUSIONS 159

LIST OF TABLES

Tables Page

ANALYSIS OF LINKED EQUILIBRIA

1 Parameters used to simulate the sedimentation velocity data for 1 - 15 μM macromolecule concentration 42

2 Examination of data simulated with fast dissociation using the "1-2-4-6" model 43

3 Examination of data simulated with slow dissociation using the "1-2-4-6" model44

4 Parameters used to simulate the sedimentation velocity data for 1 – 15 μM macromolecule with rate constants indicated in the table. 45

5 Analysis of data simulated in Table 4 using "1-2, 1-4, 1-6" model to analyze the data 46

EXAMINATION OF THE DYNAMIC ASSEMBLY EQUILIBRIUM FOR *E.COLI* CLPB

1 Sedimentation coefficients for ClpB oligomers used in global analysis of sedimentation velocity data.....89

2 Model determination for ClpB assembly in buffer H with 300 mM NaCl.....90

3 Global fitting results of sedimentation velocity experimental data at multiple ClpB concentrations in buffer H with the NaCl concentration indicated in the table.....91

4 Kinetic parameters determined from global NLLS analysis that are associated with 1-2-4-6 model presented in table 3-a. 92

5 Comparison of the goodness of fit for three replicates to the 1-2-3-6 vs. the 1-2-4-6 model.....93

EXAMINATION OF NUCLEOTIDE BINDING AND LINKAGE TO CLPB ASSEMBLY

1 The assembly equilibrium constants as a function of $[ATP\gamma S]_f$. 123

2 The dissociation rate constantans as a function of ATPγS. model 124

3 Prediction for the extent of binding of ClpB dimer $\bar{X}_2$, tetramer $\bar{X}_4$ and hexamer $\bar{X}_6$ using the slopes determined from the Wyman Plots.....129

4 ATPγS binding constants and binding stoichiometry determined from global analysis of $logL_{n,app}$ as a function of $[ATP\gamma S]_f$. 134

LIST OF FIGURES

Figure Page

ANALYSIS OF LINKED EQUILIBRIA

1 Predicted binding isotherms from Equation 20 47

2 *c*(*s*) distributions resulting from c(s) analysis of data simulated from 1, 9 and 15 μM protein 48

3 Global analysis of simulated sedimentation velocity data 49

EXAMINATION OF THE DYNAMIC ASSEMBLY EQUILIBRIUM FOR *E.COLI* CLPB

1 ClpB exhibits dynamic assembly in the absence of nucleotide 94

2 Salt effect on ClpB assembly 95

3 One ClpB oligomer is predominately populated in 100 mM NaCl 96

4 Global fitting of sedimentation equilibrium scans 97

5 Time difference curves from Sedimentation velocity experiments performed on multiple ClpB concentrations in the presence of 300 mM NaCl 98

6 Species fractions simulations of ClpB 99

EXAMINATION OF NUCLEOTIDE BINDING AND LINKAGE TO CLPB ASSEMBLY

1 c(s) analysis of the sedimentation velocity data for ClpB assembly in the presence of various concentrations of ATPγS 121

2 Sedimentation velocity experimental measurements of $[ATP\gamma S]_f$ for ClpB assembly in the presence of 20 and 50 μM $[ATP\gamma S]_t$. 127

3 (A) $ln(L_{6,app})$, (B) $ln(L_{4,app})$ and (C) $ln(L_{2,app})$ as a function of $ln[ATP\gamma S]$. 130

4 Global analysis of the (A) $L_{6,app}$, (B) $L_{4,app}$ and (C) $L_{2,app}$ as a function of $[ATP\gamma S]_f$ using n-independent identical model 133

5 *c(s)* distributions for sedimentation velocity experiments on 2 μM ClpB in the absence (red curve) or presence of 10 μM $[ATP\gamma S]_t$ (black curve). 135

LIST OF ABBREVIATIONS

AAA+	ATPase associated with various cellular activities
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
ATPγS	Adenosine 5'-[γ-thio] triphosphate
ClpB:	Caseinolytic peptidase B
Hsp:	Heat shock protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
EDTA	Ethylenediaminetetraacetic acid disodium salt
FRET	Förster resonance energy transfer
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
NBD	Nucleotide binding domains
NLLS	Non-linear least squares
SDS-PAGE	Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis
RMSD:	Root mean squared deviation
RPM:	Revolutions per minute

INTRODUCTION

Escherichia coli ClpB (Caseinolytic peptidase B) belongs to the Clp/Hsp100 protein family.[1] Unlike its name suggests, ClpB cannot interact with a proteolytic subunit, but rather works with DnaKJE co-chaperones to disaggregate large protein aggregates.[2] The disaggregation activity of ClpB (Hsp104 for eukaryotic cells) is essential for cell survival under stress. ClpB homologues have been found in plants and mitochondria, but not in the mammalian cytosol.[3] Understanding the ClpB disaggregation mechanism may aid in developing treatments for a variety of human neurodegenerative diseases that involve protein aggregation.[4, 5] To date, the mechanism of ClpB catalyzed protein disaggregation is not fully understood.

As a member of the AAA+ (ATPases Associated with diverse cellular Activities) superfamily, ClpB is composed by one N-terminal domain, two nucleotide-biding domains (NBDs), one M-domain and one C-terminal domain.[2] It has been reported that ClpB requires nucleotide binding to form a barrel-like hexamer to perform its chaperon activity.[3] However, the crystal structure of the ClpB hexamer has not been resolved. Previous studies on the determination of ClpB quaternary structure reported that ClpB forms hexamer in the presence of a large excess of ATP/ATPγS.[4, 5] Hence, it has been widely accepted that only hexamer reside in solution in the studies performed at a variety of ClpB concentrations to investigate ClpB chaperon activity.[5-8]

With the assumption that ClpB forms hexamer only in the presence of a large excess of ATPγS, our lab performed anisotropy titration experiments to study the binding specificity of ClpB-polypeptide interaction.[9] In this study, we titrated ClpB into the

fluorescein labeled polypeptide in the presence of 1 mM ATPγS and monitored the change of anisotropy signal. Multiple titrations were performed at various polypeptide concentrations. To our surprise, the resulting binding isotherms couldn't be described if assuming all ClpB to be hexameric. The incompetence to describe the binding isotherms led us to ask the question: Dose ClpB only forms hexamer in the presence of ATPγS?

Interestingly, the study performed by Werbeck et al. concluded that ClpB hexamers undergo rapid subunit exchange in the presence of a large excess of ATP/ATPγS.[7] They performed stopped-flow experiments to rapid mix FRET (Förster resonance energy transfer) donor labeled ClpB with FRET acceptor labeled ClpB, where both ClpB samples were assumed to form only hexamers with single type of fluorophore label. Upon mixing, they observed ClpB hexamers that contains both FRET donor and acceptor labeled ClpB protomer. This suggest that ClpB hexamers with one type of fluorophore label must dissociate first to smaller oligomers and then re-associate to form a new hexamer that contains both types of labels. This observation suggests that ClpB can't be uniformly hexamer in the presence of large excess of nucleotide.

Understanding ClpB assembly mechanism therefore becomes to be an urgent need for studying ClpB-polypeptide interactions and further revealing its disaggregation mechanism. Several models for ClpB disaggregation mechanism have been proposed based on the assumption that ClpB forms stable hexamer in the presence of a large excess of ATP/ATPγS. One example is that ClpB hexamer binds and translocates polypeptide substrate throughout the central pore.[8] To be able to do this, it requires that ClpB forms a stable hexamer and this hexamer utilizes the energy generated from ATP hydrolysis to mechanically pull polypeptide through its central cavity. However, according to Werbeck

et al.'s results, ClpB hexamer has gone through several rounds of dissociation and reassociation during one ATP hydrolysis cycle, which suggests that the energy from ATP hydrolysis can't be used to translocate peptide processively.

Moreover, to study of the assembly-linked ClpB disaggregation mechanism requires the knowledge of the linkage of ATP binding to ClpB assembly. One of the import aspects of the ClpB disaggregation mechanism is how ClpB utilizes the energy from ATP hydrolysis to perform its chaperone activity. This requires the quantitative analyses of the binding of nucleotide to the twelve ATP binding sites in the hexamer. Moreover, whether other ClpB oligomers are present in solution and what are their roles in ATP hydrolysis and peptide disaggregation also need to be investigated. The association equilibrium constants will govern the population of hexamers or other oligomers that are present in solution. The dissociation rate constants reveal how long ClpB oligomers can exist in solution.

The objective of this dissertation is to rigorously determine the energetics and kinetics of ClpB assembly. The quaternary structures of ClpB in the absence and presence of nucleotide are examined accordingly. The thermodynamic and kinetic stabilities of ClpB oligomers are determined. The binding stoichiometry of ATPγS (slowly hydrolysable ATP analogue) to each existing oligomer is resolved. On the foundation of this study, using a "step by step" strategy, the mechanism of ClpB disaggregation can be investigated. In the second chapter of this dissertation, we presented a method to study the energetics and kinetic of a complicated dynamic monomer-dimer-tetramer-hexamer assembly using analytical ultracentrifugation. In the

third and fourth chapters, we applied this method to study ClpB assembly in the absence and presence of nucleotide, respectively.

The methodology of using analytical ultracentrifugation to study assembly equilibria

In chapter 2, we presented a method of using analytical ultracentrifugation technique to monitor the assembly and dissociation of multiple oligomers during sedimentation. For the first time, we demonstrated detailed analyses on using sedimentation velocity experiments to quantitatively investigate the energetics and kinetics of a complicated interacting system.

In order to determine the energetics and kinetics of a complicated dynamic assembly system, protein assembly at multiple concentrations needs to be examined. This is because the population of each oligomer can be adjusted by varying the total protein concentrations. In the selected concentration range, all oligomers need to be significantly populated to ensure an accurate determination of the thermodynamic and kinetic parameters.

We introduced three plausible ways to describe the sedimentation velocity data using SedAnal when considering reaction kinetics: 1, if the dissociation rate constants of oligomers are faster than 0.01 s^{-1}, the assembly reactions can be considered at instantaneous equilibrium on the time scale of sedimentation. The data can be then analyzed using a stoichiometry model with the “Newton-Raphson” method to numerically solve the Lamm Equation and the partition functions of all oligomers. 2. If the dissociation rate constant is in the range of 10^{-5} - 0.01 s^{-1}, the dissociation rate constant has a significant impact on the shape of concentration boundaries and can be

quantitatively determined.[10-13] In such a case, a bimolecular reaction model needs to be generated and the "kinetic integrator" needs to be selected to analyze "equilibrium during the run". 3. For oligomers whose dissociation rate constants are slower than 10^{-5} s^{-1}, the system can be considered as non-interacting. The data can then be described using a "non-interacting components" model.

Simulated sedimentation velocity data at multiple protein concentrations were generated based on the experimental conditions that were used to determine ClpB assembly with dissociation rate constants > 0.01 s^{-1}, < 10^{-5} s^{-1}, or in the range of 10^{-5} to 0.01 s^{-1}, respectively. Different analysis strategies were applied to demonstrate and compare the three fitting methods introduced above. Here we show that for a monomer-dimer-tetramer-hexamer assembly system, both energetics and kinetics can be determined with constrains. These fitting results provide us confidence to apply this method to examine ClpB assembly in the absence and presence of nucleotide.

ClpB forms hexamer in the absence of nucleotide

ClpB assembly has been examined using various techniques, such as polyacrylamide gel electrophoresis, size exclusion column, electron microscope, and analytical ultracentrifugation.[3, 4, 14-19] Conflicting results have been presented on ClpB assembly in the absence of nucleotide.[3, 4, 14, 19] Early studies performed by Woo et al. conclude that ClpB appears to be a tetrameric complex. A single symmetric elution peak with an apparent molecular weight with of ~35 kDa was observed upon analysis using a Superose-6 gel filtration column.[19] However, the chromatogram was not provided. Later, Zolkiewski et al. performed similar experiments using gel filtration and observed an elution peak with an apparent molecular weight of ~16 kDa, which is less than the

molecular weight of ClpB dimer. On the other hand, their sedimentation velocity experiment results suggest ClpB association has a protein concentration dependence, which indicates that ClpB oligomers may dissociate upon dilution in the gel filtration column. Moreover, a *c(s)* distribution peak with $s_{20,w}$ = 17.2 S was observed at high ClpB concentration. $s_{20,w}$ = 17.2 S could be interpreted as the sedimentation coefficient for ClpB hexamer or heptamer.[3]

Whether or not ClpB forms hexamers or heptamers in the absence of nucleotide was still unclear due to conflicting results reported by different research groups.[4, 14] Zolkiewski et al. concluded that ClpB resides in a monomer-dimer-heptamer equilibrium based on sedimentation equilibrium experiment results. One angular velocity (8000 rpm) was chosen to examine the energetics of 0.96, 2.0, and 4.0 mg/ml ClpB assembly at 4 °C.[4] In contrast, del Castillo et al. reported that ClpB assembly can be best described by a monomer-hexamer-dodecamer model. The assembly of 1-50 μM ClpB (actual concentrations were not reported) was examined by using sedimentation equilibrium experiments at 4000, 6000, and 9000 rpm.[14]

Notably, both sedimentation equilibrium studies discussed above were performed at ClpB concentrations that were up to 50 μM. At such a high protein concentration, the assembly system may no longer be ideal. Moreover, non-specific aggregation of ClpB may occur and contribute to the species distribution of the assembly. Moreover, for assembly system containing multiple oligomers, NLLS fitting sedimentation equilibrium scans by sum of exponentials can be problematic. Cross correlations between parameters (molecular weight and assembly equilibrium constants) can be introduced. One way to overcome this problem is to involve another type of experiment to determine oligomeric

states of ClpB. As such, the model that can be used to describe the data can be selected with more precision.

In our study, both sedimentation velocity and sedimentation equilibrium experiments are performed to rigorously examine the energetics of ClpB assembly as a function of [*ClpB*] and [*NaCl*]. Thus, the assembly equilibrium was perturbed to populate only one ClpB oligomer. Accordingly, its oligomeric state can be accurately determined. Our results show that ClpB forms hexamer in the absence of nucleotide. Moreover, the intermediates for ClpB monomer assembling into hexamer were also determined. Our results indicate no evidence for the previously reported heptamers and dodecamers.

ClpB exhibits a dynamic assembly

In the absence of nucleotide, ClpB was considered unable to assemble into hexamer when using the gel filtration method to examine the assembly.[3, 19] Studies using analytical ultracentrifugation observed large ClpB oligomers that may be hexamers or heptamers in the absence of nucleotide and the assembly has a ClpB concentration dependence. [4, 14] The discrepancy of the conclusions on ClpB assembly stoichiometry between using gel filtration methods and analytical ultracentrifugation can be a consequence of three possibilities: 1. ClpB can form small oligomers and hexamers in the absence of nucleotide. Their populations in solution have ClpB concentration dependence. 2. ClpB hexamers are dissociating on the gel filtration column. 3. A combination of possibilities 1 and 2. In fact, all three possibilities indicate that ClpB may exhibit a dynamic assembly in the absence of nucleotide. Thus, quantitative examinations of the assembly energetics and kinetics as a function of ClpB concentration are required to fully understand the ClpB assembly mechanism.

Our study, for the first time, rigorously determined the assembly pathway for ClpB to form hexamers in the absence of nucleotide. Our results show that ClpB resides in a monomer-dimer-tetramer-hexamer equilibrium. We report the self-association equilibrium constants for each of these oligomers. Furthermore, the dissociation kinetics for ClpB are incorporated into the data analysis. The dissociation rate constant for each oligomer was found to be around or greater than 0.01 s^{-1} in buffer H supplemented with 200 mM or 300 mM NaCl, indicating that *E. coli* ClpB oligomers dissociate on the time scale of minutes or shorter. In buffer H with 100 mM NaCl, the dissociation rate constant for ClpB hexamer was measured to be $1.3\ (1.0,\ 1.6) \times 10^{-3}\ s^{-1}$, indicating that dissociation is slower but still on the time scale of minutes.

On the other hand, in the presence of nucleoside triphosphate Zolkiewski et al. observed only hexameric ClpB using the sedimentation equilibrium approach at a single ClpB concentration.[3] Using gel filtration chromatography Zolkiewski et al.,[3] Schlee et al.,[15] and Mogk et al.[20] reported chromatograms with broad elution peaks for ClpB in the presence of a large excess of ATP. The retention time of the first elution peak indicates the presence of hexameric ClpB; however, the asymmetric broadening of this peak still suggests that ClpB hexamer is dissociating on the column or that there are also small oligomers in the presence of ATP. This indicates that ClpB may exhibit dynamic equilibrium in the presence of nucleotide.

In support of the interpretation that ClpB resides in a dynamic equilibrium, Werbeck et al. showed that ClpB exhibited fast subunit exchange.[7] They used stopped-flow FRET experiments to show that 0.2 μM *T. thermophilus* ClpB exhibit rapid subunit exchange in the presence 200 μM ATP. The authors conclude that ClpB forms mainly

hexamer at such a condition, however, the data that support this conclusion were not shown.

Interestingly, Aguado et al performed stopped-flow experiments using a similar experimental design to examine 0.4 μM FRET pair labeled *E. coli* ClpB in the presence of 2 mM ATP. They observed a slow subunit exchange for ClpB hexamer (or other oligomers).[21] The observed rate constant reported by Aguado et al. is two orders of magnitude smaller than the slowest rate constant reported by Werbeck et al. Indeed, the two studies were performed using enzymes from different organisms; however, another clear difference was that each study was performed at different protein and nucleotide concentrations. While both studies stated that ClpB forms mainly hexamer, ClpB assembly at their experimental conditions was not quantitatively determined.

Above all, the discrepancies on whether ClpB forms hexamer only in the presence of a large excess of ATP/ATPγS and how stable this hexamer (or other oligomer) is are due to the fact that none of the above studies examined ClpB assembly as a function of both [ClpB] and [nucleotide]. The statement of "large excess of nucleotide" misleads researchers to simply use the ratio of [ClpB] to [nucleotide] to predict the population of oligomers. Instead, the assembly equilibrium constants should be measured and further used to predict the population of oligomers.

In our work, the energetics and kinetics of ClpB assembly as a function of [ATPγS] are quantitatively determined, where ATPγS serves as a model for ATP. Here we show that ClpB resides in a monomer-dimer-tetramer-hexamer equilibrium in the presence of ATPγS. The assembly equilibrium constants as a function of nucleotide concentration were determined, which makes it possible to predict the concentration of

hexamers and other oligomers present at any total ClpB concentration and total nucleotide concentration. Finally, our results show that ClpB is in rapid subunit exchange at nucleotide concentrations equal to or below 100 μM ATPγS. However, at concentrations above 500 μM ATPγS the observed dissociation rate constants are slowed to a range consistent with a non-dissociating system. Thus, nucleotide binding affects both the thermodynamic state of the system and the kinetics.

ATPγS binding affinity and stoichiometry for each ClpB oligomer

An accurate determination of the binding affinity and stoichiometry of nucleotide to ClpB is essential for the study of ClpB ATP hydrolysis and ClpB-peptide substrate interaction. This study is complicated by the fact that there are two nucleotide binding domains (NBDs) per ClpB monomer and thus twelve NBDs per hexamer.[2, 19] In order to simplify the examination of nucleotide binding to each individual NBD, a mutagenesis strategy has been employed. [15-17, 20, 22-27] Much has been learned from those works. However, how the modifications on the primary structure of ClpB interrupt the protein assembly and their subsequent impact on nucleotide binding, ATP hydrolysis, and peptide recognition is not clear.

ClpB has been considered to form hexamer only if the ratio of [*nucleotide*] to [*ClpB*] is large enough. Because of that, the assembly state of ClpB mutants were often examined at a very high ClpB concentration (40 μM as an example) in the presence of large excess of ATP/ATPγS (such as 2 mM) using the gel filtration method. Once the elution peak corresponding to the molecular weight of ClpB hexamer was observed, the mutants were concluded to have the same assembly properties as wild type ClpB. One

clear problem is that the concentration of ClpB and nucleotide used there were too high to observe the differences of the assembly energetics between wild type ClpB and its mutants. More importantly, ClpB was then assumed to form hexamer only at a much lower ClpB and ATP/ATPγS concentration as long as [nucleotide] was in large excess. For example, 0.1 μM ClpB in the presence of 200 μM ATP were used in the experiments determining the nucleotide binding affinity for hexameric ClpB mutants.

As discussed in the last section, the concentration of each ClpB oligomer at any given [ClpB] and [nucleotide] can be calculated using equilibrium constants determined by this study. According to our results, the population of ClpB hexamer has both ClpB and nucleotide concentration dependence. The concentration of ClpB hexamer was not significantly populated at low ClpB and nucleotide concentration regardless of the ratio of [nucleotide] to [ClpB]. Therefore, the determination of nucleotide binding affinity and stoichiometry for ClpB and its linkage to ClpB assembly needs to be examined as a function of both [ClpB] and [nucleotide].

Our study, for the first time, examined the linkage of nucleotide binding to ClpB assembly. With these results, we determined the binding constants and stoichiometry for nucleotide binding to each of the oligomers without performing any modifications to ClpB. Here we show that all twelve sites are bound in the hexameric ring of ClpB and monomer, dimer, and tetramer exhibit stoichiometry of 1, 3, and 7, respectively. The binding stoichiometry of the smaller oligomers is one fewer than the maximum number of binding sites, which suggests an open conformation rather than a ring structure. With results determined by our study, the number of ATPγS bound to each ClpB oligomer can be predicted at any given [ATPγS] and [ClpB]. This work provides an essential

foundation for quantitative studies on ClpB ATP hydrolysis and ClpB-protein interactions. It allows us to better quantify how ClpB utilizes the energy from ATP hydrolysis to perform its disaggregation function.

ANALYSIS OF LINKED EQUILIBRIA

by

JIABEI LIN AND AARON L. LUCIUS

Methods in Enzymology, in Press

Format adapted for dissertation

Abstract

The ATPases associated with diverse cellular activities (AAA+) is a large superfamily of proteins involved in a broad array of biological processes. Many members of this family require nucleotide binding to assemble into their final active hexameric form. We have been studying two example members, *E. coli* ClpA and ClpB. These two enzymes are active as hexameric rings that both require nucleotide binding for assembly. Our studies have shown that they both reside in a monomer, dimer, tetramer, hexamer equilibrium and this equilibrium is thermodynamically linked to nucleotide binding. Moreover, we are finding that the kinetics of the assembly reaction are very different for the two enzymes. Here we present our strategy for determining the self-association constants in the absence of nucleotide to set the stage for the analysis of nucleotide binding from other experimental approaches including analytical ultracentrifugation.

Introduction

Determining a binding constant for a protein ligand interaction is quite possibly one of the most common experiments in biophysics. Techniques such as ITC, fluorescence titrations, equilibrium dialysis, and many others are commonly used. Analysis of these data is fairly straightforward if the protein of interest does not change its oligomeric state as the free concentration of the ligand (chemical potential of the ligand) is increased. However, if the protein does change its oligomeric state as the chemical potential of the ligand is increased then the analysis becomes substantially more complex.

It is important to recall that a binding constant measured using any of the approaches stated above is an apparent binding constant[1]. Although we typically write down a binding reaction as an association between a macromolecule, M, and a ligand, X, as schematized in Equation 3, the binding is actually much more complex.

$$M + X \overset{K_{app}}{\rightleftharpoons} MX \qquad 1$$

In fact, the interaction also involves the removal of water and/or ions from the binding pocket as well as from the ligand that is entering the binding pocket. Although there could be many solution condition components, including protons, involved in this interaction an example scheme that represents only ions and water exchange upon ligand binding is given by Equation 4.

$$M \bullet Na_i^+ \bullet \left(H_2O\right)_j + X \bullet Cl_k^- \bullet \left(H_2O\right)_l \overset{K}{\rightleftharpoons} MX + iNa^+ + kCl^- + l\left(H_2O\right) \qquad 2$$

These exchanges of water and ions with bulk solvent all contribute to the energetics of the binding interaction. Therefore, the binding equilibrium constant, Kapp, given in

Equation 3 will exhibit a dependence on solution condition variables such as ion concentration, pH, water concentration, etc. In other words, the binding constant will be thermodynamically linked to all of the components involved in the binding reaction. Thus, linkage analysis can be used to deconvolute the energetics of a 'second ligand' (ion, water, protons, etc.) binding. This is accomplished by determining the apparent equilibrium constant for X binding M as a function of, for example, salt concentration. Then, this apparent equilibrium constant can be treated as a signal for a 'second ligand' binding event (water, proton, ions, etc.).

With the above in mind, one can recognize that a ligand binding constant for an assembling system will exhibit a protein concentration dependence. Likewise, a protein-protein interaction constant will exhibit a ligand concentration dependence for an assembling system. That is to say, the binding constants are thermodynamically linked to either the protein concentration or ligand concentration.

(This might be a good place to explain the meaning of linkage, and explain qualitatively how the binding constants will depended on [Protein] and how the self-association parameters will depend on [ligand]. A figure would help.

The ATPases associated with diverse cellular activities (AAA+) is a large superfamily of proteins involved in a broad array of biological processes [2]. Examples include: microtubule severing catalyzed by katanin [3]; membrane fusion involving N-Ethylmaleimide-sensitive fusion proteins (NSF) [4]; morphogenesis and trafficking of endosomes by VPs4p [5]; protein disaggregation by ClpB/Hsp104; and enzyme catalyzed protein unfolding and translocation by ClpA or ClpX for ATP dependent proteolysis [6, 7].

Another example is human VCP/p97, which has been connected to ubiquitin-dependent reactions, but is implicated in an expanding number of physiological processes [8].

Many of these proteins require nucleoside triphosphate binding to assemble into their final active hexameric form. Thus, ligand linked assembly is an integral component of the mechanism driving assembly. For many of these examples, the oligomers present in solution will change as the chemical potential of the ligand (nucleotide) is increased.

Analytical ultracentrifugation is the technique of choice for examining the energetics of macromolecular assembly. The technique has been used extensively to do so and a large number of computer applications are available to aid in the analysis of the experimental results, and these have been discussed in many places [9-13] (see chapters in this volume; equilibrium by Rowe and sedimentation velocity by Stafford and Correia). Moreover, an enormous body of literature exists on the application of this approach to examine assembling systems [9, 14-16]. However, substantially less has been done on examining ligand linked assembly problems, but some examples can be found [17-20].

Here we outline our strategy for elucidating the thermodynamic mechanism for an assembling system that exists in a complex, dynamic equilibrium of monomers, dimers, tetramers, and hexamers. This is being done in order to set the stage for an examination of the thermodynamic linkage to the nucleotide driven hexamer formation for two example AAA+ macromolecular machines, *E. coli* ClpA and ClpB. We have found that both proteins form hexamers and both exist as mixtures of oligomeric states both in the presence and absence of nucleotide (Lin and Lucius, manuscript submitted), [21-24]. This is a difficult problem to address for a variety of reasons, some of which will be discussed

here. Nevertheless, the number of systems that exhibit ligand linked assembly is large and there is a pressing need for a set of strategies and approaches to solve the problem.

Oftentimes, without being able to predict the concentration of each species in solution, quantitatively interpreting binding and catalytic data is not possible. However, solving these problems is needed for more than just interpretation of in vitro studies. We are finding that the nucleotide affinity of these motor proteins is in the range of 10 – 100 μM, which is at least an order of magnitude below the concentration of nucleotide in the cell (5 – 10 mM). This indicates that the nucleotide binding sites on these macromolecules would likely be saturated in the cell and thus nucleotide is not likely to be a regulatory molecule. On the other hand, the concentration of these proteins in the cell [25-27] have been found to be similar to their ligand linked assembly dissociation equilibrium constant, indicating that the linkage between nucleotide binding and assembly may be an important regulatory component of their function.

To begin to quantitatively address the linkage of ligand binding to macromolecular assembly we can start by writing down a partition function, *Q*, that represents the sum of all of the macromolecular states for an arbitrary solution containing monomers (M_1), dimers (M_2), tetramers (M_4), and hexamers (M_6) as follows.

$$Q = [M_1] + [M_2] + [M_4] + [M_6] \qquad 3$$

We can define self-association equilibrium constants for each thermodynamic state. Here, we will use '*L*' for stoichiometric or overall protein-protein interaction constants and '*K*' for both step-wise protein-protein interaction constants and for ligand binding constants. For the monomer, dimer, tetramer, hexamer system the following three reactions given by Equations 6 - 8 can define the equilibria

$$2M_1 \overset{L_{2,0}}{\rightleftharpoons} M_2 \qquad 4$$

$$4M_1 \overset{L_{4,0}}{\rightleftharpoons} M_4 \qquad 5$$

$$6M_1 \overset{L_{6,0}}{\rightleftharpoons} M_6 \qquad 6$$

The three stoichiometric equilibrium constants that result are given by Equations 9 - 11

$$L_{2,0} = \frac{[M_2]}{[M_1]^2} \qquad 7$$

$$L_{4,0} = \frac{[M_4]}{[M_1]^4} \qquad 8$$

$$L_{6,0} = \frac{[M_6]}{[M_1]^6} \qquad 9$$

Where the first subscript represents the oligomeric state and the second subscript represents the nucleotide ligation state, in this case the zero represents no ligand bound. The partition function given in Equation 5 can be simplified by algebraically solving Equations 9 - 11 for each oligomer and substituting the solutions into Equation 5 to yield Equation 12.

$$Q = [M_1] + L_{2,0}[M_1]^2 + L_{4,0}[M_1]^4 + L_{6,0}[M_1]^6 \qquad 10$$

Where the partition function given by Equation 12 is only a function of the free monomer concentration and the equilibrium constants for each oligomer formed. If each of the oligomers can bind ligand, then each term in the partition function will be multiplied by a partition function for binding of ligand to that particular oligomer to yield Equation 13.

$$Q = [M_1]P_1 + L_{2,0}[M_1]^2 P_2 + L_{4,0}[M_1]^4 P_4 + L_{6,0}[M_1]^6 P_6 \qquad 11$$

where P_1, P_2, P_4, and P_6 represent the partition functions for the binding of ligand to the monomer, dimer, tetramer, and hexamer, respectively. The steps for deriving Equation 13 are straightforward and will not be reproduced here; for a review see Wyman and Gill [28]. For simplicity, if the monomer binds one ligand, X, then the partition function for ligand binding is given by the sum of all of the monomeric states normalized to the unligated state given by Equation 14.

$$P_1 = \frac{[M_1] + [M_1X]}{[M_1]} \qquad 12$$

If we define an equilibrium constant for nucleotide binding to the monomer as:

$$K_{1,1} = \frac{[M_1X]}{[M_1]} \qquad 13$$

Where the first subscript represents the oligomeric state and the second subscript represents the number of ligands bound. We can simplify Equation 14 to be:

$$P_1 = 1 + K_{1,1}[X] \qquad 14$$

If we assume that the monomer has n-independent and identical binding sites then the partition function given by Equation 16 would be expressed as Equation 17.

$$P_1 = \left(1 + K_1[X]\right)^{n_1} \qquad 15$$

where n_1 represents the number of binding sites per monomer. If we assume that each oligomer binds n number of ligands independently and identically than the partition function for the entire system is given by

$$Q = [M_1]\left(1 + K_1[X]\right)^{n_1} + L_{2,0}[M_1]^2\left(1 + K_2[X]\right)^{n_2} + L_{4,0}[M_1]^4\left(1 + K_4[X]\right)^{n_4} + L_{6,0}[M_1]^6\left(1 + K_6[X]\right)^{n_6} \qquad 16$$

Where n_1, n_2, n_4, and n_6 represent the number of binding sites on the monomer, dimer, tetramer, and hexamer, respectively and K_1, K_2, K_4, and K_6 represent the average binding constant for ligand binding to monomers, dimers, tetramers, and hexamers, respectively, where the average binding constant has been corrected with statistical factors [28]. Since, in this model, all of the binding sites are assumed to be the same, there is no second subscript on K_x to denote the number of ligands bound.

For any experiment that would seek to examine ligand binding to such a complex system, whether it be ITC, fluorescence titrations, or some other approach, the signal is typically proportional to ligand bound divided by total macromolecule, which is defined as 'extent of binding' or $\overline{X}$. The importance of expressing the partition function is that the partition function can now be used to derive an equation that represents the extent of binding, ligand bound over total macromolecule, which could be used to analyze ligand binding data. The extent of binding is given by Equation 19 [28]

$$\frac{[X]_{Bound}}{[M]_{Total}} = \overline{X} = \frac{\dfrac{dQ}{d\ln[X]}}{\dfrac{dQ}{d\ln[M]}} = \frac{[X]}{[M]}\frac{\dfrac{dQ}{d[X]}}{\dfrac{dQ}{d[M]}} \qquad 17$$

If the derivatives in Equation 19 are applied to the partition function given by Equation 18 then the extent of binding equation is given by Equation 20

$$\overline{X} = \frac{[M_1]n_1K_1[X](1+K_1[X])^{n_1-1} + L_{2,0}[M_1]^2 n_2K_2[X](1+K_2[X])^{n_2-1} + L_{4,0}[M_1]^4 n_4K_4[X](1+K_4[X])^{n_4-1} + L_{6,0}[M_1]^6 n_6K_6[X](1+K_6[X])^{n_6-1}}{[M_1](1+K_1[X])^{n_1} + 2L_{2,0}[M_1]^2(1+K_2[X])^{n_2} + 4L_{4,0}[M_1]^4(1+K_4[X])^{n_4} + 6L_{6,0}[M_1]^6(1+K_6[X])^{n_6}} \qquad 18$$

Clearly, Equation 20 would have entirely too many parameters to apply to a single binding isotherm collected with any technique. However, several important predictions

can be made from inspection of Equation 20 or from inspection of the partition function given in Equation 18. First, these equations are functions of the self-association constants in the absence of ligands, $L_{n,0}$, the ligand binding constants to each oligomer, K_n, the free monomer concentration $[M_1]$, and the free ligand concentration $[X]$. Second, Equation 20, tells the experimentalist that there is a need to first define the self-association equilibrium constants in the absence of nucleotide, $L_{n,0}$. Third, unlike binding isotherms for simple systems, a binding system that is linked to macromolecular assembly will exhibit a dependence on the free protein concentration. This tells the experimentalist that binding studies will have to be executed over a range of protein concentrations

Figure 1 shows a series of isotherms simulated using Equation 20 with several different total macromolecule concentrations ranging from 1 – 10 μM. In this example $L_{2,0}$ = 1 x 10^4 M^{-1}, $L_{4,0}$ = 1 x 10^{14} M^{-3}, and $L_{6,0}$ = 1 x 10^{24} M^{-5}, each monomer is considered to bind one ligand so that n_1 = 1, n_2 = 2, n_4 = 4, and n_6 = 6, the ligand binding constants for the monomers through tetramers are all identical so that $K_1 = K_2 = K_4$ = 1 x 10^5 M^{-1} and the hexamer binding constant is an order of magnitude tighter, K_6 = 1 x 10^6 M^{-1}.

The most salient feature of the binding isotherms shown in Figure 1 is that there is a shift of the midpoint to lower free ligand concentration as the macromolecule concentration increases. This is the consequence of the fact that as the macromolecule concentration increases there is a corresponding increase in the concentration of hexamers. This observation is general because the thermodynamic driving force for an assembly reaction is the chemical potential of the free monomer. Thus, there will always

be an increase in the population of higher order oligomers with increasing protein concentration. The apparent increase in ligand binding affinity as the macromolecule concentration is increased is because the ligand binds to the hexamer with an affinity constant one order of magnitude tighter than the other oligomers.

When examining ligand binding to a macromolecule it is always advisable to perform multiple titrations at several total macromolecule concentrations [29]. If such a strategy is invoked and an apparent change in the affinity constant is observed as a function of macromolecule concentration then this would be the first indicator that a ligand linked assembly process is occurring.

Determination of $L_{n,0}$ for an assembling system

If assembly is suspected from an experiment such as that illustrated by Figure 1 then Equation 20 suggests that the first objective would be to determine the self-association equilibrium constants in the absence of any ligand, $L_{n,0}$, so that this parameter could be constrained in the examination of the titration curves. To determine $L_{n,0}$ we perform sedimentation velocity experiments over a range of protein concentrations.

Sedimentation velocity has been our experiment of choice over sedimentation equilibrium. This is because we have found that the nucleotide bound hexamers are not stable over the timescale required for sedimentation equilibrium, which is often several days. In contrast to a sedimentation equilibrium experiment, a sufficient number of sedimentation boundaries can be collected within 2 – 3 hours in a sedimentation velocity experiment. Further, deconvoluting three or more species from exponential fitting performed on sedimentation equilibrium boundaries can be difficult.

Naturally, one would not know the self-association equilibrium constants, so knowing what concentrations to examine will not be initially certain. However, if a series of isotherms were collected like those shown in Figure 1, one could judge that the assembly state is making a transition over the range of 1 – 10 μM and this would be a reasonable starting point.

Global Fitting of Sedimentation Velocity Data as a Function of Protein Concentration

Kinetic Considerations

Sedimentation velocity data is potentially sensitive to the assembly kinetics if the dissociation rate constants for the species are in the range of 10^{-2} – 10^{-5} s^{-1} [11, 12, 30, 31]. Consequently, there are two ways to globally analyze sedimentation velocity data. The first is to assume that the system is always at thermodynamic equilibrium. This assumption holds if the dissociation rate constants are faster than 10^{-2} s^{-1}. The second is that the dissociation rate constant is found to be within or slower than the empirical range 10^{-2} – 10^{-5} s^{-1}. If the dissociation rate constants are within this empirical range, modeling the reaction kinetics is required. Otherwise, for dissociations that are slower than 10^{-5} s^{-1}, the components can be considered as non-interacting discrete species.

To illustrate the impact of the assembly kinetics we simulated sedimentation velocity experiments with reverse rate constants of either 1 s^{-1} or 10^{-6} s^{-1}. Dissociation rate constants of k_r = 1 s^{-1} or 10^{-6} s^{-1} correspond to half-lives of 0.7 s and 192 hours, respectively. It is not difficult to conclude that reaction kinetics occurring with these half-lives would be outside of the detectable range in a sedimentation velocity experiment since the boundaries are typically collected on the minutes time scale. Although one

could enhance the temporal resolution with interference experiments performed on only a single cell since these could be collected every 8 s.

To simulate the sedimentation boundaries using SedAnal the concentration of each species needs to be modeled. Here we will use the language of the Gibbs phase rule. A component is defined as a chemical component and a species is made up of products of reactions between the components. Thus, in an experiment containing a single protein, '*M*', that reacts to form dimers, tetramers, and hexamers we define '*M*' as the component and monomers, dimers, tetramers, and hexamers as species. Therefore, this is a single component - four species system.

The first step in the simulation is to relate the total loading concentration of the protein to the concentration of each species. This is done by writing down the conservation of mass equation given by Equation 21.

$$[M_1]_T = [M_1] + 2[M_2] + 4[M_4] + 6[M_6] \qquad 19$$

where $[M_1]_T$ is the total monomer concentration, $[M_1]$, $[M_2]$, $[M_4]$, and $[M_6]$ are the equilibrium concentrations of monomers, dimers, tetramers, and hexamers. The coefficients of 2, 4, and 6 are present because the total concentration is expressed in monomer units, e.g. 2 monomers in a dimer, etc. Equation 21 can be expressed in terms of the equilibrium constants given by Equations 9 - 11 and the free monomer concentration, $[M_1]$, to yield Equation 22.

$$[M_1]_T = [M_1] + 2L_{2,0}[M_1]^2 + 4L_{4,0}[M_1]^4 + 6L_{6,0}[M_1]^6 \qquad 20$$

In a sedimentation velocity experiment there are two stages of equilibrium that one needs to consider. Those stages are before the force of centrifugation is applied and while the force is present. When the force is present, the equilibrium can be perturbed. However,

before the force is applied, the system should be at equilibrium. This 'pre-equilibrium' is achieved if the experimentalist has given sufficient time for the system to fully relax to equilibrium after any perturbations, i.e. preparation of samples, dilutions, temperature equilibration, etc. Thus, if an experiment is intended to be performed at, for example, $[B_1]_T$ = 1 μM then the experimentalist would make up this solution in some vessel. In that vessel the system will distribute itself into free monomers, dimers, tetramers, and hexamers, where the population of each could be defined by Equation 22 if the equilibrium constants are known. However, an important control would be to allow this sample to incubate for increasing amounts of time before performing the run. Then, the results could be compared after, for example, a 6 hour vs. 12 hour pre-incubation time. If the system is at equilibrium one would expect to see results that are independent of incubation time. If incubation time dependent differences are observed then the incubation time should be extended until differences are no longer observed.

To determine the concentration of each species for this system one needs to solve Equation 22, which is a sixth-order polynomial in the free monomer concentration. However, the free monomer concentration, $[M_1]$, is not known. What is known to the experimentalist is the total monomer concentration, $[M_1]_T$. Thankfully, numerically solving a polynomial is not a difficult task.

In SedAnal the roots of Equation 22 and thus the concentrations of each species are determined by numerical methods, specifically the Newton-Raphson method. In practice, this is achieved in SedAnal by going to 'preferences' choosing 'Control extended', "Kinetics/equilibrium control" and under "Initial equilibration" one chooses "No analytic solution" under the Newton-Raphson column. Again, since this model

requires solving a sixth order polynomial; there is no analytic solution and therefore the equation must be solved numerically.

What was just described represents the determination of the concentration of each species in the reaction vessel before the sample is subjected to the force of sedimentation. Since there is not force there is no perturbation of the equilibrium, and therefore the concentrations of each species are fixed. The next task at hand is to define the concentrations of each species upon application of force. Since, in a sedimentation velocity experiment, the experimentalist is observing the time dependent separation of each species the experiment is potentially sensitive to the reaction kinetics.

To simulate the sedimentation boundaries one needs to again determine the concentration of each species and then model the sedimentation of each species by passing the determined concentration to the Lamm Equation, which defines the movement of the particle under the force of centrifugation and back diffusion. For the monomer, dimer, tetramer, hexamer reaction we assume that all species are being formed through bimolecular interactions defined by the following reactions given by Equations 23 - 25.

$$2M_1 \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} M_2 \qquad 21$$

$$2M_2 \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} M_4 \qquad 22$$

$$M_2 + M_4 \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} M_6 \qquad 23$$

where $k_{f,n}$ is the bimolecular association rate consant with units of M^{-1} s^{-1} and $k_{r,n}$ is the dissociation rate constant with units of s^{-1}. We assume that all species are formed

through bimolecular interactions because tri-molecular reactions and above are highly improbable. The equilibrium constants for the reactions in equation 23 - 25 are given by Equations 26 - 28.

$$K_2 = \frac{k_{f2}}{k_{r2}} = \frac{[M_2]}{[M_1]^2} \qquad 24$$

$$K_4 = \frac{k_{f4}}{k_{r4}} = \frac{[M_4]}{[M_2]^2} \qquad 25$$

$$K_6 = \frac{k_{f6}}{k_{r6}} = \frac{[M_6]}{[M_2][M_4]} \qquad 26$$

It is important to note that if the system is an equilibrium system then the step-wise equilibrium constants given by Equations 26 - 28 can be related to the stoichiometric interaction constants given by Equations 9 - 11 as follows.

$$L_{2,0} = K_2 \qquad 27$$

$$L_{4,0} = K_2^2 \cdot K_4 \qquad 28$$

$$L_{6,0} = K_2^3 \cdot K_4 \cdot K_6 \qquad 29$$

Since the equilibrium is potentially being perturbed, determining the concentration of each species is accomplished by numerically solving the following system of coupled differential equation given by Equation 32 - 35.

$$\frac{d[M_1]}{dt} = [M_2]k_{r2} - [M_1]^2 k_{f2} = [M_2]k_{r2} - [M_1]^2 K_2 k_{r2} \qquad 30$$

$$\frac{d[M_2]}{dt} = [M_1]^2 k_{f2} - [M_2]k_{r2} - [M_2]^2 k_{f4} + [M_4]k_{r4} - [M]_2[M_4]k_{f6} + [M_6]k_{r6}$$
$$= [M_1]^2 K_2 k_{r2} - [M_2]k_{r2} - [M_2]^2 K_4 k_{r4} + [M_4]k_{r4} - [M_2][M_4]K_6 k_{r6} + [M_6]k_{r6} \quad 31$$

$$\frac{d[M_4]}{dt} = [M_2]^2 k_{f4} - [M_4]k_{r4} - [M]_2[M_4]k_{f6} + [M_6]k_{r6}$$
$$= [M_2]^2 K_4 k_{r4} - [M_4]k_{r4} - [M_2][M_4]K_6 k_{r6} + [M_6]k_{r6} \quad 32$$

$$\frac{d[M_6]}{dt} = [M_2][M_4]k_{f6} - [M_6]k_{r6} = [M_2][M_4]K_6 k_{r6} - [M_6]k_{r6} \quad 33$$

Where the right hand side of Equations 32 - 35 have been simplified to include only the equilibrium constant and reverse rate constants for each reaction. To model the sedimentation of each species, the system of coupled differential equations are numerically integrated to determine the concentration of each species as a function of time based on the values of the equilibrium constants and dissociation rate constants for each reaction.

In the extreme of rapid dissociation, $k_r > 0.01\ s^{-1}$ ($t_{1/2} < 1.2$ minutes), the system is considered to be in rapid-equilibrium. That is to say, for each infinitely small radial slice of solution the macromolecules that rapidly dissociate within this slice are considered to rapidly re-associate and therefore, the differential equations given in Equations 32 - 35 are all equal to zero, $d[M_n]/dt = 0$. This indicates that the concentration of each oligomer in each infinitely small radial slice is constant and thus at equilibrium.

In contrast, if the dissociation rate constants are on the other extreme, $k_r < 10^{-5}\ s^{-1}$ ($t_{1/2} > 19$ hours), then the differential equations are again equal to zero and the concentrations of each species are again considered to be fixed. However, in this

scenario there is no re-equilibration at each infinitely small radial slice because no dissociation is occurring during the course of the entire experiment. Thus, the species are being separated when force is applied. In this scenario the system can be modeled as non-interacting discrete species because they will not appear to react on the time scale of sedimentation.

To illustrate these points we simulated a series of sedimentation boundaries from the monomer, dimer, tetramer, hexamer model given by Equations 23 - 25 using SedAnal for two total protein concentrations, $[B_1]_T$ = 1 and 15 μM. We refer to this model as the '1-2-4-6' model. For the first simulation all reverse rate constants were considered to be fast relative to sedimentation, $k_{r2} = k_{r4} = k_{r6} = 1\ s^{-1}$, see Table 1. In a second simulation all reverse rate constants were considered to be slow, $k_{r2} = k_{r4} = k_{r6} = 10^{-6}\ s^{-1}$, see Table 1. Both extremes of the rate constants were considered to be outside of $10^{-2} - 10^{-5}\ s^{-1}$, which has been previously reported to be the range over which one would expect to be able to extract meaningful measures of the rate constants from sedimentation velocity data [11, 12, 31].

Our preferred first level of analysis is to analyze the sedimentation boundaries using SedFit to generate c(s) distributions (Peter Shuck, NIH) [9] (REF). Figure 2A shows the results of a c(s) analysis on the simulated data assuming all dissociation rate constants are 1 s^{-1}. In red is the *c(s)* distribution from the analysis of the simulation with $[M_1]_T$ = 1 μM and a single peak is observed that corresponds to the monomer. In blue is the *c(s)* analysis for $[M_1]_T$ = 9 μM. The peak corresponding to monomer shifts to the right slightly and a broad distribution from ~5 S – 8 S emerges. In green is the *c(s)* distribution from the simulation with $[M_1]_T$ = 15 μM and what is observed is a clear

shifting of the peaks further to the right. The peak shifting can be taken as the first indication that the system is exhibiting fast reaction kinetics [31, 32]. Thus, a preliminary *c(s)* analysis that shows this type of broad *c(s)* distribution may serve to be the first indicator that rapid dissociation may be occurring on the time scale of sedimentation and this hypothesis would warrant further testing.

Figure 2B shows a *c(s)* analysis resulting from analysis of simulations performed with all dissociation rate constants $k_r = 10^{-6}$ s^{-1}. At 1 μM total protein two peaks are observed that correspond to the sedimentation coefficient of monomers and dimers. In contrast, at 15 μM all four peaks appear that correspond to the monomers, dimers, tetramers, and hexamers, respectively. Under these conditions of slow dissociation the oligomers sediment as non-interacting discrete species. Under such conditions the area under each of these peaks would represent the equilibrium concentrations of each species and could be analyzed to yield the equilibrium constants.

Global analysis using the 1-2-4-6 model with rapid dissociating oligomers

If the *c(s)* plot given by Figure 2A was experimentally observed our next step would be to globally fit all protein concentrations by direct boundary analysis in SedAnal. Again, the goal here is to determine the self-association equilibrium constants, $L_{n,0}$. Since the *c(s)* plot suggests there may be rapid dissociation the first strategy in this analysis would be to float each of the reverse rate constants starting with a guess of around 0.01 s^{-1}. Thus, the global floating fitting parameters are the loading concentrations, K_2, K_4, K_6, k_{r2}, k_{r4}, and k_{r6} as given by Equations 26 - 28.

As with the simulations described above to fit the data accounting for the kinetics there are several steps that have to be executed in SedAnal. We always assume that the

system is at equilibrium at the start of the run. In practice, this is achieved in SedAnal by going to 'preferences' choosing 'Control extended', "Kinetics/equilibrium control" and under "Initial equilibration" one chooses "No analytic solution" under the Newton-Raphson column. What this accomplishes is numerically solving Equation 22 to determine the free monomer concentration and thus the concentration of each oligomer is determined based on the initial guesses of the equilibrium constants. Next, under "Equilibration during run" we choose either BulSt or SEulEx under "Kinetic integrator" in the "No analytic solution" row. The choice of kinetic integrator primarily impacts the speed of executing the fit and the differences have been discussed elsewhere [11] including the SedAnal manual. What this accomplishes is numerically solving the system of coupled differential equations given by Equations 32 - 35 for the concentrations of each species as a function of time and passes this to the Lamm equation to define the movement of the oligomer in the field.

The other parameters required for this analysis are the molar mass, the sedimentation coefficient, density increment and mass extinction coefficient. The molar mass would be known from sequence information and thus the molar mass of each oligomer calculated. Estimates of the sedimentation coefficients are needed and solution conditions can usually be modified to acquire reasonable estimates of the monomer and largest oligomers. One can estimate the intermediate sedimentation coefficients based on the $s_i = s_1 (n)^{2/3}$ with the assumption that the frictional ratio of the oligomers are the same as monomer's, where s_i is the sedimentation coefficient for the *i-th* species containing n monomer units and s_1 is the sedimentation coefficient for the monomer [33]. Alternatively, if data on the three dimensional structure is available then one can approximate

hydrodynamic information using applications like Hydropro [34] and Chapter 8 of this volume by Rocco and Byron. The extinction coefficient should be rigorously determined by denaturing the protein in 6 M guanidine [35-37]. The density increments, $\frac{d\rho}{dc}$, can be determined experimentally, but for most cases it is acceptable to substitute $\left(1-\bar{v}\rho\right)$, where $\bar{v}$ is the partial specific volume of the protein and ρ is the density of the buffer. These parameters are typically calculated with SednTerp [38] (David Hayes, Magdalen College, Tom Laue, University of New Hampshire, and John Philo, Alliance Protein Laboratories).

Global NLLS analysis of the data simulated from the 1-2-4-6 model with all the dissociation rate constants $k_{r,n}$ = 1 s^{-1} yield good estimates of the equilibrium constants (compare values used to generate data in Table 1 to fitted values in Table 2, first column). The difference curves and their associated fits are shown in Figure 3. Values of the rate constants float somewhere between ~0.3 – 2 s^{-1}, see Table 3. We have interpreted this to indicate that the kinetic parameters are unconstrained because the value used to simulate the data of 1 s^{-1} results in a half-life of ~0.7 s. Thus, there is little information in the sedimentation boundaries that yield constraints on the values of these rate constants.

The next step in our analysis strategy would be to constrain the dissociation rate constants to a value of 0.01 s^{-1} and repeat the analysis. In doing this on the same simulated data as above, we acquire estimates of the equilibrium constants that have the correct order of magnitude but are as much as 25 % higher than the values used to generate the data, Table 2 second column. Moreover, the RMSD is significantly larger than the RMSD determined when allowing the rate constants to float. This cannot simply be the consequence of having three additional floating parameters because the degrees of

freedom (DOF) between the fits are essentially identical. That is to say, since the DOF is the difference between the number of data points and the number of parameters and here the number of data points is ~80,000 the additional three parameters do not influence this number enough to account for the deviation in the RMSD.

Despite the fact that constraining the rate constants to the empirical upper bound for 'fast dissociation' leads to an ~25 % over estimate of the equilibrium constant, the observation that both fits lead to the correct order of magnitude in the equilibrium constant suggests that the data could be modeled with a purely thermodynamic model. This leads to the suggestion that a path independent thermodynamic model could be used to describe the data. To test this, the data were fit by choosing Newton-Raphson under the Equilibration during run in the No analytic solution row. What this accomplishes is numerically solving the sixth order polynomial in the free monomer concentration to yield the concentration of species, thereby modeling the system under the assumption that each infinitely small slice of radial position is at equilibrium. This analysis yields equilibrium constants that are in good agreement with the values used to generate the data and the RMSD is identical to the value acquired when allowing the rate constants to float as fitting parameters, see Table 2 third column.

Global analysis using the 1-2-4-6 model with slow dissociation of oligomers

Figure 2B shows a c(s) analysis from sedimentation velocity experiments simulated for the 1-2-4-6 model with all dissociation rate constants of 10^{-6} s^{-1}. Again, our first level of global analysis is to allow both the kinetic parameters, $k_{r,n}$, and the equilibrium constants, K_n to float as fitting parameters. As seen in Table 3 the values of the equilibrium constants are in good agreement with those used to simulate the data

(compare Table 1 for simulated values to Table 3 for fitted values). However, the rate constants are as much as three orders of magnitude slower than the values used to simulate the data. This is not surprising because a dissociation rate constant of 10^{-6} s^{-1} yields a half-life of ~192 hours if one assumes a simple first order reaction. Thus, on the time scale of sedimentation there is no appreciable dissociation that occurs for the simulated data with a given 10^{-6} s^{-1} rate constant.

When the data are analyzed with the rate constants constrained to 10^{-5} s^{-1} the values of the equilibrium constants determined are in agreement with the values used to simulate the data within 2 – 9 % errors, see Table 3. However, the RMSD is significantly worse than the value when the rate constants are allowed to float as fitting parameters. This is surprising because with such a slow dissociation rate constant one does not expect the data to contain any information on these parameters. One possible explanation is that the empirical bound of 10^{-5} s^{-1} may be on the edge of values that would define no dissociation for the simulated system.

To probe this further, the data were analyzed by constraining all $k_{r,n} \leq 10^{-6}$ s^{-1}. As shown by the analysis performed when allowing $k_{r,n}$ to float, $k_{r,n} = 10^{-6}$ s^{-1} is slow enough to indicate no dissociation for the simulated system. Therefore, rate constants smaller than 10^{-6} s^{-1} should be able to represent reactions with no dissociation and describe the simulated data adequately well. Our results show that an RMSD = 5.000 x 10^{-3} was determined when constraining all $k_{r,n} = 10^{-6}$ s^{-1} and an RMSD = 5.003 x 10^{-3} was determined when constraining all $k_{r,n} = 10^{-7}$ s^{-1}. Both analyses yield identical equilibrium constants and are identical to the value used to generate the data.

Above all, the analyses suggest that the empirical boundary for defining "no dissociation" is in the range of 10^{-6} s^{-1} -10^{-5} s^{-1}. We suspect that for systems with different levels of complexity and sizes of species, the empirical boundaries for rapid and slow dissociation may be slightly larger than the reported 10^{-2} - 10^{-5} s^{-1} range. Based on these observations, we recommend testing the reaction kinetics by allowing the rate constants to float as fitting parameters in the analysis.

Global analysis with rate constants in the detectable range

The next and most obvious question is; how well can the rate constants be determined if they fall between the empirical bound of 10^{-2} – 10^{-5} s^{-1}? To address this question we simulated data using the 1-2-4-6 model with equilibrium constants and rate constants given in Table 4. As shown in Table 4, the rate constants are in the range of 10^{-3} – 10^{-4} s^{-1}. Table 4 shows the results of globally fitting these simulated data to the 1-2-4-6 model. It is not surprising that the data are well described by the correct model and the predicted parameters are well within range of the values used to simulate the data. Although not surprising, one might expect that the information on the kinetic rate constants could be lost due to the complexity of the model. Thus, what this analysis does show is that the global fitting strategy is able to detect these rate constants for this model.

Thermodynamics is path independent. Sedimentation velocity data where the oligomers are either in rapid dissociation or slow dissociation do not contain information about path. Although the specific strategy for doing so is different due to the process of sedimentation; in the two extremes fast and slow kinetics the data can be described by either stoichiometric or step-wise equilibrium constants when allowing the kinetic parameters to float and the results are the same. This is not the case for a system that

exhibits dissociation with rate constants in the range of $10^{-2} – 10^{-5}$ s^{-1}. This is shown in Table 5 where the simulated data were analyzed assuming that each oligomer forms in a single step from monomers as given by Equations 6 - 11 where monomers form dimers, tetramers, or hexamers in a single kinetic step. The analysis of these data does not accurately predict the equilibrium constants or the rate constants, with the exception of dimer formation which is not different for the two models. What this analysis reveals is that when the rate constants are in the detectible range there is information on the path and the data cannot be modeled by simple path independent thermodynamic models. Thus, under these conditions care needs to be taken to write down appropriate path dependent models. Although it may take more computational power to do so, we assert that this should be done by assuming all reactions occur through bimolecular interactions.

Conclusions

Similar simulations and examination of the kinetics as discussed here have also been discussed elsewhere [16, 31]. The advance presented here is an analysis of a more complex model. Most previous discussions have centered on a monomer - dimer equilibrium. Here, we have focused on the monomer, dimer, tetramer, hexamer equilibrium since this is what we are experimentally observing for two AAA+ molecular motors, *E. coli* ClpA[21-23] and ClpB [24] and (J. Lin manuscript submitted). Most importantly, there are some general principles that have been derived from previous studies on less complex systems that do not seem to generalize to the more complex systems we are studying. That is to say, the empirical bound of 0.01 s^{-1} – 10^{-5} s^{-1} may be a bit wider depending on the size of molecules and the rotor speed.

One observation that holds for all values of the kinetic rate constants is that floating the values always seems to lead to acquisition of the values used to simulate the data. So why not always float the kinetic rate constants? The answer is the lack of computational power. Indeed, twenty years ago, globally analyzing upwards of 80,000 data points by numerically solving the Lamm equation combined with numerically solving a complex system of coupled differential equations was outside of our computational reach. However, we now have desktop computers that can accomplish these tasks. Nevertheless, fitting to such complex models is very slow and it can be accelerated when limiting assumptions can be made.

One limiting factor, which is true for all NLLS minimization routines, is that each computation is dependent on the outcome of the last. Thus, such approaches are not amenable to parallel computing. In contrast, routines such as the Genetic algorithm are gaining interest because of the independence of each calculation and thus the ease with which such routines can be parallelized. Ultrascan is an application for the analysis of Sedimentation Velocity data that takes advantage of the genetic algorithm and thus takes advantage of advances in parallel computing [12].

Sedphat does not allow the user to write their own models for global analysis. Thus the experimentalist is constrained to using pre-written models that assume, for example, that tetramers would form in a single kinetic step. One could make an argument for trimolecular interactions but anything higher than three bodies colliding in a single kinetic step is unrealistic. Although Ultrascan allows the user to write down their own models it does not allow the product of one reaction to be the reactant of another. As shown by the simulations presented here, there is a need to be able to write down

kinetic models that reduce the reactions to bimolecular steps when kinetic information is present.

To our knowledge, the only software application that allows the experimentalist to construct their own models and directly fit boundaries by numerically solving the Lamm equation is SedAnal [11]. More importantly, SedAnal is numerically solving a system of coupled differential equations describing the reaction. Thus, the rate constants that emerge have the potential to contain a great deal more information about the reaction than arbitrary rate constants that are being applied in other applications. We assert that the gravity of what SedAnal is doing behind the scenes is not fully appreciated. Indeed, Sedimentation velocity would not be the experiment of choice if one were trying to thoroughly deconvolute a kinetic mechanism. However, not adequately accounting for the kinetics in the analysis of these data may lead to significant errors in the equilibrium parameters of interest. Moreover, it is clear that Sedimentation velocity experiments can lead the experimentalist to pursue more direct kinetic approaches if something interesting is revealed in the modeling.

References

Alberty, R. A. (2003). Thermodynamics of biochemical reactions. Hoboken, N.J., Wiley-Interscience.

Babst, M., T. K. Sato, L. M. Banta and S. D. Emr (1997). "Endosomal transport function in yeast requires a novel AAA-type ATPase, Vps4p." EMBO J **16**(8): 1820-1831.

Cole, J. L. (1996). "Characterization of human cytomegalovirus protease dimerization by analytical centrifugation." Biochemistry **35**(48): 15601-15610.

Cole, J. L. (2004). "Analysis of heterogeneous interactions." Methods Enzymol **384**: 212-232.

Cole, J. L., J. J. Correia and W. F. Stafford (2011). "The use of analytical sedimentation velocity to extract thermodynamic linkage." Biophys Chem **159**(1): 120-128.

Correia, J. J. (2000). "Analysis of weight average sedimentation velocity data." Methods Enzymol **321**: 81-100.

Correia, J. J., P. H. Alday, P. Sherwood and W. F. Stafford (2009). "Effect of kinetics on sedimentation velocity profiles and the role of intermediates." Methods Enzymol **467**: 135-161.

Correia, J. J. and W. F. Stafford (2009). "Extracting equilibrium constants from kinetically limited reacting systems." Methods Enzymol **455**: 419-446.

Dam, J. and P. Schuck (2005). "Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions c(s) and asymptotic boundary profiles from Gilbert-Jenkins theory." Biophys J **89**(1): 651-666.

Dam, J., C. A. Velikovsky, R. A. Mariuzza, C. Urbanke and P. Schuck (2005). "Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions c(s)." Biophys J **89**(1): 619-634.

Demeler, B., E. Brookes, R. Wang, V. Schirf and C. A. Kim (2010). "Characterization of reversible associations by sedimentation velocity with UltraScan." Macromol Biosci **10**(7): 775-782.

Dougan, D. A., B. G. Reid, A. L. Horwich and B. Bukau (2002). "ClpS, a substrate modulator of the ClpAP machine." Mol Cell **9**(3): 673-683.

Edelhoch, H. (1967). "Spectroscopic determination of tryptophan and tyrosine in proteins." Biochemistry **6**(7): 1948-1954.

Farrell, C. M., A. D. Grossman and R. T. Sauer (2005). "Cytoplasmic degradation of ssrA-tagged proteins." Mol Microbiol **57**(6): 1750-1761.

Gill, S. C. and P. H. von Hippel (1989). "Calculation of protein extinction coefficients from amino acid sequence data." Anal Biochem **182**(2): 319-326.

Laue, T. M., Shah, B.D., Ridgeway, T.M., Pelletier, S.L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. Analytical Ultracentrifugation in Biochemistry and Polymer Science. A. J. R. S.E. Harding, J.C. Horton. Cambridge, Royal Society of Chemistry.

Li, T., J. Lin and A. L. Lucius (2015). "Examination of polypeptide substrate specificity for Escherichia coli ClpB." Proteins **83**(1): 117-134.

Lohman, T. M. and D. P. Mascotti (1992). "Nonspecific ligand-DNA equilibrium binding parameters determined by fluorescence methods." Methods Enzymol **212**: 424-458.

Meyer, H. and C. C. Weihl (2014). "The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis." J Cell Sci **127**(Pt 18): 3877-3883.

Mogk, A., T. Tomoyasu, P. Goloubinoff, S. Rudiger, D. Roder, H. Langen and B. Bukau (1999). "Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB." Embo J **18**(24): 6934-6949.

Na, G. C. and S. N. Timasheff (1985). "Measurement and analysis of ligand-binding isotherms linked to protein self-associations." Methods Enzymol **117**: 496-519.
Na, G. C. and S. N. Timasheff (1985). "Velocity sedimentation study of ligand-induced protein self-association." Methods Enzymol **117**: 459-495.
Neuwald, A. F., L. Aravind, J. L. Spouge and E. V. Koonin (1999). "AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes." Genome Res **9**(1): 27-43.
Ogura, T. and A. J. Wilkinson (2001). "AAA+ superfamily ATPases: common structure--diverse function." Genes Cells **6**(7): 575-597.
Ortega, A., D. Amoros and J. Garcia de la Torre (2011). "Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models." Biophys J **101**(4): 892-898.
Pace, C. N., F. Vajdos, L. Fee, G. Grimsley and T. Gray (1995). "How to measure and predict the molar absorption coefficient of a protein." Protein Sci **4**(11): 2411-2423.
Roll-Mecak, A. and F. J. McNally (2010). "Microtubule-severing enzymes." Current opinion in cell biology **22**(1): 96-103.
Sauer, R. T. and T. A. Baker (2011). "AAA+ Proteases: ATP-Fueled Machines of Protein Destruction." Annual review of biochemistry **80**: 587-612.
Schuck, P. (1998). "Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation." Biophys J **75**(3): 1503-1512.
Schuck, P. (2003). "On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation." Anal Biochem **320**(1): 104-124.
Scott, D. J., S. E. Harding, A. J. Rowe and Royal Society of Chemistry (Great Britain) (2005). Analytical ultracentrifugation : techniques and methods. Cambridge, UK, RSC Pub.
Stafford, W. F. and P. J. Sherwood (2004). "Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants." Biophys Chem **108**(1-3): 231-243.
Streaker, E. D., A. Gupta and D. Beckett (2002). "The biotin repressor: thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific DNA binding." Biochemistry **41**(48): 14263-14271.
Veronese, P. K. and A. L. Lucius (2010). "Effect of Temperature on the Self-Assembly of the Escherichia coli ClpA Molecular Chaperone." Biochemistry **49**(45): 9820-9829.
Veronese, P. K., B. Rajendar and A. L. Lucius (2011). "Activity of Escherichia coli ClpA Bound by Nucleoside Di- and Triphosphates." Journal of molecular biology **409**(3): 333-347.
Veronese, P. K., R. P. Stafford and A. L. Lucius (2009). "The Escherichia coli ClpA Molecular Chaperone Self-Assembles into Tetramers." Biochemistry **48**(39): 9221-9233.
Wyman, J. and S. J. Gill (1990). Binding and linkage : functional chemistry of biological macromolecules. Mill Valley, Ca., University Science Books.
Yu, R. C., R. Jahn and A. T. Brunger (1999). "NSF N-terminal domain crystal structure: models of NSF function." Mol Cell **4**(1): 97-107.

Table 1: Parameters used to simulate the sedimentation velocity data for 1 –15 μM macromolecule concentration.

Model used for simulation	Parameters used for fast dissociation		Parameters used for slow dissociation	
	K_n (M^{-1})	$k_{r,n}$ (s^{-1})	K_n (M^{-1})	$k_{r,n}$ (s^{-1})
$2B \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} B_2$	6×10^4	1	6×10^4	1×10^{-6}
$2B_2 \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} B_4$	1.5×10^5	1	1.5×10^5	1×10^{-6}
$B_2 + B_4 \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} B_6$	1.5×10^6	1	1.5×10^6	1×10^{-6}

The stoichiometric equilibrium constants for dimerization, tetramerization and hexamerization can be calculated using Equations 25 – 27 and are: $L_{2,0} = 6 \times 10^4$ M^{-1}, $L_{4,0} = 5.4 \times 10^{14}$ M^{-3} and $L_{6,0} = 4.86 \times 10^{25}$ M^{-5}. The sedimentation coefficients used for simulations are s = 3.0, 5.6, 8.86, and 11.9 for monomer, dimer, tetramer and hexamer, respectively. The standard deviation of noise added to the data is 0.005 absorbance unit.

Table 2: Examination of data simulated with fast dissociation using the "1-2-4-6" model.

Model used for fitting	Fit with kinetic integrator allowing $k_{r,n}$ to float		Fit with kinetic integrator, constraining $k_{r,n}$ = 0.01 s^{-1}		Fit with Newton-Raphson, no $k_{r,n}$ input	
	RMSD=5.003 × 10^{-3}		RMSD=5.929 × 10^{-3}		RMSD=5.003 × 10^{-3}	
	K_n (M^{-1})	$k_{r,n}$ (s^{-1})	K_n (M^{-1})	$k_{r,n}$ (s^{-1})	K_n (M^{-1})	$k_{r,n}$ (s^{-1})
$2B \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} B_2$	6.00 × 10^4	0.29	5.87 × 10^4	0.01	6.00 × 10^4	NA
$2B_2 \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} B_4$	1.49 × 10^5	1.52	2.02 × 10^5	0.01	1.49 × 10^5	NA
$B_2 + B_4 \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} B_6$	1.51 ×10^6	1.94	1.02 ×10^6	0.01	1.52 ×10^6	NA

Table 3: Examination of data simulated with slow dissociation using the "1-2-4-6" model.

Model used for fitting	Fit with kinetic integrator allowing $k_{r,n}$ to float		Fit with kinetic integrator, constraining $k_{r,n}$ = 1× 10^{-5} s^{-1}		Fit with Newton-Raphson, no $k_{r,n}$ input	
	RMSD=5.000× 10^{-3}		RMSD= 5.188 10^{-3}		RMSD=3.178× 10^{-2}	
	K_n (M^{-1})	$k_{r,n}$ (s^{-1})	K_n (M^{-1})	$k_{r,n}$ (s^{-1})	K_n (M^{-1})	$k_{r,n}$ (s^{-1})
$2B \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} B_2$	6.00 × 10^{4}	1.16× 10^{-6}	5.91 × 10^{4}	1× 10^{-5}	5.89 × 10^{3}	NA
$2B_2 \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} B_4$	1.52 × 10^{5}	4.86× 10^{-9}	1.65 × 10^{5}	1× 10^{-5}	6.45 × 10^{6}	NA
$B_2 + B_4 \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} B_6$	1.49 ×10^{6}	5.13× 10^{-7}	1.43 ×10^{6}	1× 10^{-5}	7.00 ×10^{7}	NA

Table 4: Parameters used to simulate the sedimentation velocity data for 1 – 15 μM macromolecule with rate constants indicated in the table.

Model used for simulation	values used for simulation		Parameters resulting from analysis	
			RMSD = 4.972×10^{-3}	
	K_n (M^{-1})	$k_{r,n}$ (s^{-1})	K_n (M^{-1})	$k_{r,n}$ (s^{-1})
$2B \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} B_2$	6×10^{4}	3×10^{-3}	6.00×10^{4}	3.04×10^{-3}
$2B_2 \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} B_4$	1.5×10^{5}	4×10^{-4}	1.50×10^{5}	3.98×10^{-4}
$B_2 + B_4 \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} B_6$	1.5×10^{6}	7×10^{-4}	1.50×10^{6}	7.01×10^{-4}

The stoichiometric equilibrium constants for dimerization, tetramerization and hexamerization can be calculated using Equations 25 – 27 and are: $L_{2,0} = 6 \times 10^{4}$ M^{-1}, $L_{4,0} = 5.4 \times 10^{14}$ M^{-3} and $L_{6,0} = 4.86 \times 10^{25}$ M^{-5}.

Table 5 : Analysis of data simulated in Table 4 using "1-2, 1-4, 1-6" model to analyze the data:

Model used for fitting	Fit with kinetic integrator allow $k_{r,n}$ to float		
	RMSD=5.539× 10^{-3}		
	$L_{n,0}$	Calculated K_n (M^{-1})	$k_{r,n}$ (s^{-1})
$2B \overset{L_{2,0}}{\rightleftharpoons} B_2$	6.24×10^{4} M^{-1}	6.24×10^{4}	1.88×10^{-3}
$4B \overset{L_{4,0}}{\rightleftharpoons} B_4$	3.91×10^{14} M^{-3}	1.00×10^{5}	1.16×10^{-4}
$6B \overset{L_{6,0}}{\rightleftharpoons} B_6$	5.64×10^{25} M^{-5}	2.31×10^{6}	1.80×10^{-4}

Figures

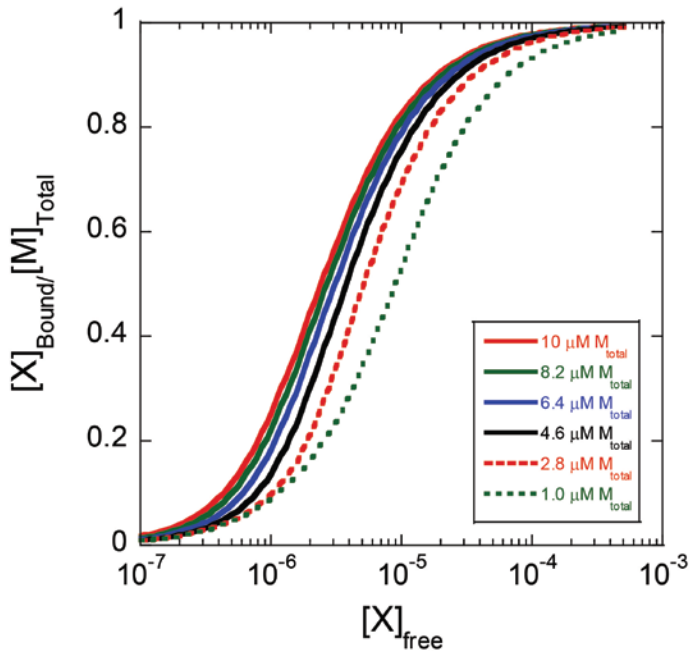

Figure 1: Predicted binding isotherms from Equation 20 with $L_{2,0} = 1 \times 10^4\ M^{-1}$, $L_{4,0} = 1 \times 10^{14}\ M^{-3}$, and $L_{6,0} = 1 \times 10^{24}\ M^{-5}$, each monomer is considered to bind one ligand so that $n_1 = 1$, $n_2 = 2$, $n_4 = 4$, and $n_6 = 6$, the ligand binding constants for the monomers through tetramers are all identical so that $K_1 = K_2 = K_4 = 1 \times 10^5\ M^{-1}$ and the hexamer binding constant is an order of magnitude tighter, $K_6 = 1 \times 10^6\ M^{-1}$.

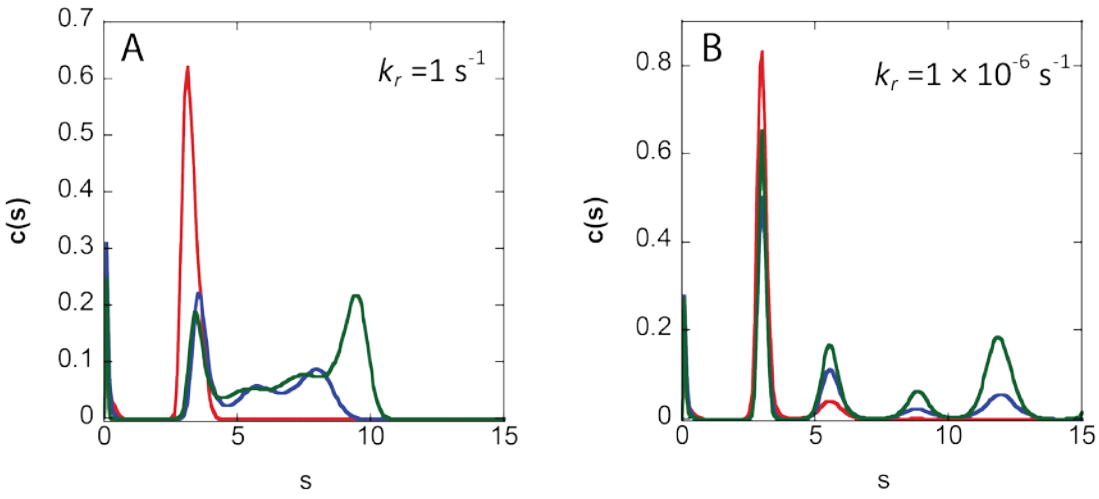

Figure 2: C(s) distributions resulting from c(s) analysis of data simulated from 1, 9 and 15 μM protein. A) parameters are given in Table 1 and B) parameters are given in Table 2. In both cases the sedimentation coefficients used to simulate the data were s = 3.0, 5.6, 8.86, and 11.9 for monomer, dimer, tetramer and hexamer, respectively. The extinction coefficient used in the simulation is 4.2 (mg/ml)$^{-1}$ cm^{-1} at 230 nm and 0.45 (mg/ml)$^{-1}$ cm^{-1} at 280 nm. Time between simulated scans is 4 minutes.

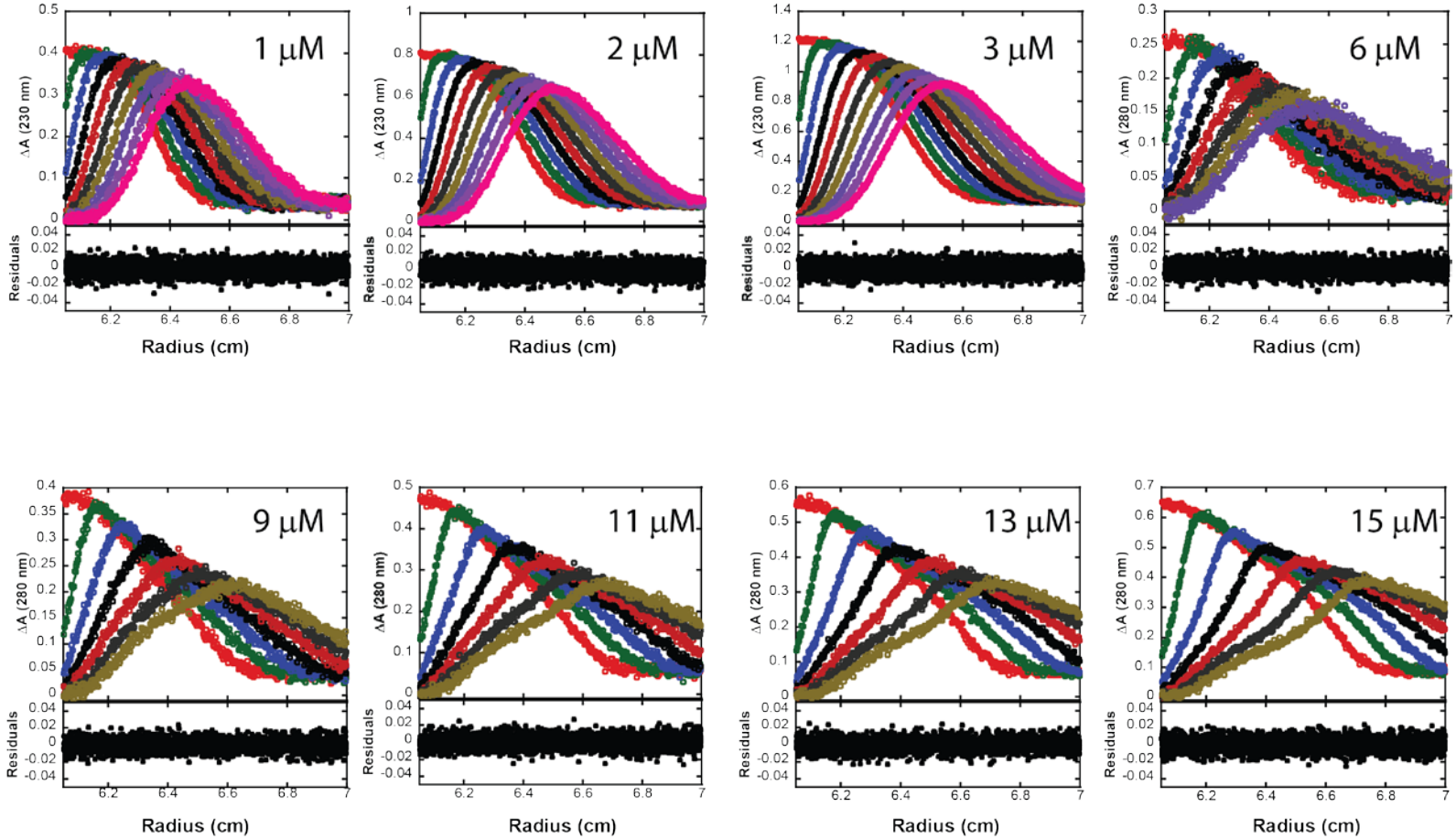

Figure 3: Global analysis of simulated sedimentation velocity data using 1-2-4-6 model by allowing dissociation rate constant, $k_{r,n}$, to float. The data were simulated using parameters presented in Table 1. The fitting results are presented in table 3. The concentrations of protein are indicated on the plots. Every 4th difference curve is presented.

EXAMINATION OF THE DYNAMIC ASSEMBLY EQUILIBRIUM FOR *E.COLI* CLPB

by

JIABEI LIN AND AARON L. LUCIUS

Proteins: Structure, Function, and Bioinformatics, in Press

Format adapted for dissertation

Abstract

Escherichia coli ClpB is a heat shock protein that belongs to the AAA+ protein superfamily. Studies have shown that ClpB and its homologue in yeast, Hsp104, can disrupt protein aggregates *in vivo*. It is thought that ClpB requires binding of nucleoside triphosphate to assemble into hexameric rings with protein binding activity. In addition, it is widely assumed that ClpB is uniformly hexameric in the presence of nucleotides. Here we report, in the absence of nucleotide, that increasing ClpB concentration leads to ClpB hexamer formation, decreasing NaCl concentration stabilizes ClpB hexamers, and the ClpB assembly reaction is best described by a monomer, dimer, tetramer equilibrium under the three salt concentration examined here. Further, we found that ClpB oligomers exhibit relatively fast dissociation on the time scale of sedimentation. We anticipate our studies on ClpB assembly to be a starting point to understand how ClpB assembly is linked to the binding and disaggregation of denatured proteins.

Introduction

The *Escherichia coli* ClpB (Caseinolytic peptidase B) protein belongs to the AAA+ (ATPase Associated with various cellular Activities) superfamily of ATPases.[1, 2] Proteins in this family bind and hydrolyze ATP and utilize the energy to perform their cellular activities. These activities include folding and unfolding of proteins, dissociating protein-protein complexes, unwinding double stranded DNA, cytoskeleton regulation, and associating with proteases to form ATP dependent proteases.[3]

As protein chaperones, ClpB and its eukaryotic ortholog, Hsp104, facilitate the dissociation of large protein aggregates in collaboration with the DnaK/Hsp70 system.[4-6] The disaggregation activity of ClpB and Hsp104 is essential for cell survival under stress. Their homologues have been found in plants and mitochondria, but not in the mammalian cytosol.[7] To date, the mechanism of ClpB catalyzed protein disaggregation is not fully understood. Understanding the ClpB disaggregation mechanism may aid in developing treatments for a variety of human neurodegenerative diseases that involve protein aggregation.[8, 9]

In many cases, proteins in the AAA+ family assemble into hexamers or higher order oligomers when they bind nucleoside triphosphate. Although it is well established that ClpB, with two Nucleotide Binding Domains (NBDs) per monomer, forms hexamers in the presence of nucleoside triphosphate,[10] an accurate model for ClpB assembly in the absence of nucleotide is lacking. This information will be required to quantitatively analyze the binding of nucleotide to the twelve ATP binding sites in the hexamer, address such questions as cooperativity between sites, and examine the subsequent linkage to polypeptide binding. This is because modeling the ligand linked assembly process first

requires an accurate model for the association equilibrium in the absence of the ligand. [11, 12] Consequently, there is a need to determine what oligomers reside in solution and the self-association equilibrium constants that govern the population. Importantly, with the assembly energetics in hand, a precise determination of the concentration of hexamers as a function of both nucleotide and ClpB concentration will be possible, which is essential for the quantitative interpretation of a vast array of in vitro studies.

Multiple reports show that ClpB resides in a dynamic equilibrium of monomers and hexamers or other oligomeric states.[10, 13, 14] Likely due to the complexity of the assembly, conflicting reports have been presented as to whether or not ClpB forms hexamers or heptamers in the absence of nucleotide. [13-16] Zolkiewski et al. concluded that ClpB resides in a monomer-dimer-heptamer equilibrium.[10] Whereas, del Castillo et al. reported that ClpB assembly can best be described by a monomer-hexamer-dodecamer model.[10, 13, 14]

On the other hand, in the presence of nucleoside triphosphate Zolkiewski et al. concluded that only hexamers were observed in their sedimentation equilibrium study which was performed with a single ClpB concentration.[10] Using gel filtration chromatography Zolkiewski et al.,[10] Schlee et al.,[17] and Mogk et al.[18] reported chromatograms with broad elution peaks for ClpB in the presence of a large excess of ATP, indicating that either hexamers are dissociating during the gel filtration run, other oligomers are present in solution, or interactions with the media are occuring.[10]

In support of the interpretation that ClpB resides in a dynamic equilibrium, Werbeck et al. showed that ClpB exhibited what they describe as fast subunit exchange.[19] They used stopped-flow FRET experiments to show that *T. thermophilus* ClpB hexamers

(or other oligomers) exhibit rapid subunit exchange both in the presence and absence of nucleotide. Despite the clear evidence that ClpB does not form stable hexamers, it is widely reported that only hexamers reside in solution in experiments performed at a variety of ClpB concentrations.[19-22]

Rapid subunit exchange is being reported to be an important aspect of the ClpB catalyzed protein disaggregation mechanism.[9, 19, 22] The hypothesis is that the hexameric ring that encounters a stable aggregate will tend to disassemble instead of stalling. In addition, hexamer is reported as the active conformation for its chaperon activity; therefore, it is imperative to fully understand both the association equilibria and kinetics for ClpB in the presence and absence of nucleoside triphosphate.

Here we report a quantitative examination of the self-association equilibrium for *E. coli* ClpB in the absence of nucleotide. Our results show that ClpB resides in a monomer-dimer-tetramer-hexamer equilibrium with no evidence found for the previously reported heptamers. We report the self-association equilibrium constants for each of these oligomers. Furthermore, the dissociation kinetics for ClpB are incorporated into the data analysis. The dissociation rate constant for each oligomer was found to be around or greater than 0.01 s^{-1} in buffer H supplemented with 200 mM or 300 mM NaCl, indicating that *E. coli* ClpB oligomers dissociate on the time scale of minutes or shorter. In buffer H with 100 mM NaCl, the dissociation rate constant for ClpB hexamer was measured to be $1.3\ (1.0,\ 1.6) \times 10^{-3}\ s^{-1}$, indicating dissociation is slower but still on the time scale of minutes. Going forward, these results will make it possible to quantitatively examine the linkage of nucleotide binding to hexamer formation. This will make it possible to predict the concentration of hexamers as a function of both nucleotide and ClpB concentrations.

Moreover, it will allow us to test the hypothesis that disassembly is a component of the ClpB catalyzed protein disaggregation reaction.

Materials and Methods

Buffers

Buffers were prepared with reagent grade chemicals using deionized H_2O purified using the Purelab Ultra Genetic system (Siemens Water Technology). Buffer B is composed of 40 mM Tris, 2 mM 2-mercaptoethanol and 10 % (v/v) glycerol, pH 7.5 at 4 °C. Buffer H contains 25 mM HEPES pH 7.5 at 25 °C, 10 mM $MgCl_2$, 2 mM 2-mercaptoethanol, 10 % (v/v) glycerol. The NaCl concentration is indicated in the text. All experiments were performed in buffer H, which contains 10 % glycerol. The 10 % glycerol is maintained in this work so that identical solution conditions are used as those used to examine polypeptide binding and translocation in our recent reports.[23, 24] All ClpB concentrations referred to in the text are in monomeric units unless otherwise stated.

Strains, plasmid and ClpB protein

The *E. coli* strains used were BL21 (DE3) and DH5α™ (Invitrogen). The gene encoding for the N-terminally His_6-tagged *E. coli* ClpB (95 kDa) with thrombin cleavage site was cloned into the pET28b (+) vector (Novagen) and verified by sequencing. ClpB was over-expressed from the pET28b (+) vector in BL21 (DE3) cells. Six liters of LB media with 30 μg ml^{-1} kanamycin was used for cell culture at 37 °C. The overexpression was induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD_{600} = 0.8. After induction the cells were allowed to grow three more hours at 37 °C to reach OD_{600} = 2. A 23 g cell paste was harvested.

The cell paste was suspended in 100 mL of buffer containing 2 mM 2-mercaptoethanol, 40 mM Tris (pH 7.5), 500 mM NaCl, 20 mM imidazole, 20% (v/v)

glycerol and 10 % (w/v) sucrose at 4 °C. Cells were then lysed with an Ultrasonic liquid processor (Misonix®, USA) and the sample was subjected to centrifugation at 10,500 rpm for two hours in an SLA-3000 rotor to pellet the cell debris.

The supernatant was loaded onto five 5 ml HisTrap FF™ crude (GE Healthcare) columns equilibrated with Buffer B supplemented with 500 mM NaCl and 20 mM imidazole. The column was washed with Buffer B supplemented with 20 mM imidazole for 30 column volumes. The sample was eluted with Buffer B supplemented with 500 mM NaCl and 500 mM imidazole. The fractions from the elution peak were pooled and dialyzed using 15,000 Da molecular weight cut off dialysis tubing against thrombin cleavage buffer (Buffer B supplemented with 300 mM NaCl and 3 mM $CaCl_2$) overnight and then treated by 1 unit thrombin per milligram ClpB for 12 hours at 4 °C in the same dialysis bag against fresh thrombin cleavage buffer. The sample was then switched to a 50,000 Da cut off dialysis bag to remove the thrombin from the reaction and dialyzed against Buffer B supplemented with 500 mM NaCl and 20 mM imidazole for 6 hours at 4 °C.

The dialyzed sample was loaded on the HisTrap FF™ (GE Healthcare) column equilibrated with Buffer B supplemented with 500 mM NaCl. The flow through was then dialyzed against Buffer B supplemented with 80 mM NaCl. The sample was loaded onto a Heparin Sepharose FF (GE Healthcare) column that was equilibrated with Buffer B supplemented with 80 mM NaCl. The column was washed with Buffer B supplemented with 80 mM NaCl for two column volumes and then the protein was eluted with Buffer B supplemented with 1 M NaCl. The fractions containing ClpB were loaded onto Sephacryl S-300 high resolution (GE Healthcare) column that was equilibrated in Buffer B

supplemented with 1 M NaCl. The ClpB fractions were pooled and dialyzed against Buffer B supplemented with 1 M NaCl and 50 % (v/v) glycerol for storage at -80 °C.

The resultant protein was >95 % pure as judged by Coomasie staining and its molecular weight was determined by MALDI-TOF mass spectrometry. One species with a molecular weight of 95,817 Da was observed. This result is consistent with the molecular weight of the monomer of ClpB plus two additional amino acids at the N-terminus after removing the His-tag. From 23 g cell paste 600 mg of >95 % pure ClpB was acquired using this protocol.

The concentration of ClpB was determined spectrophotometrically from the absorption spectra of several aliquots of ClpB in 6 M guanidine-HCl. The extinction coefficient in 6 M guanidine-HCl was calculated from the extinction coefficients of the individual aromatic amino acids in 6 M guanidine-HCl using Sednterp[25, 26] (David Hayes, Magdalen College, Tom Laue, University of New Hampshire, and John Philo, Alliance Protein Laboratories).[27] The extinction coefficient of ClpB in Buffer H was determined by comparing the absorption spectra of three aliquots of ClpB protein in 6 M guanidine-HCl with the absorption spectra of aliquots of native ClpB in Buffer H supplemented with different NaCl concentrations.

The determined extinction coefficients for ClpB at 280 and 230 nm was found to be $\varepsilon_{280} = (3.6 \pm 0.1) \times 10^4$ (M monomer)$^{-1}$ cm^{-1} and $\varepsilon_{230} = (3.5 \pm 0.1) \times 10^5$ (M monomer)$^{-1}$ cm^{-1}. These values represent the average and standard deviation of 10 replicates collected in the presence of 100, 200, and 300 mM NaCl and various protein concentrations. Since ClpB resides in a mixture of oligomers at these three different salt concentrations and protein concentrations (see Results section) and the standard deviation

on the extinction coefficient is less than 3 %, we conclude that there is little detectable difference in the extinction coefficient for each oligomer.

Analytical ultracentrifugation

Analytical ultracentrifugation experiments on ClpB were performed using a Beckman Optima XL-A analytical ultracentrifuge. Sedimentation velocity experiments using absorbance optics were carried out by loading a sample of protein (380 μL) and the protein dialysate (400 μL) into a double sector Epon charcoal-filled centerpiece and subjected to an angular velocity of 40,000 rpm. Absorbance as a function of radial position was collected by scanning the sample cells at a wavelength of 230 nm or 280 nm as indicated in the text with a radial step-size of 0.003 cm. Absorbance scans were collected every 4 minutes.

The sedimentation velocity experiment on 10 % glycerol was performed by using a Beckman ProteomeLab XL-I analytical ultracentrifuge. The sedimentation velocity experiments using interference optics were carried out by loading a sample of 10 % glycerol (425 μL) and water reference (430 μL) into a double sector Epoxy charcoal-filled meniscus matching centerpiece. An angular velocity of 40,000 rpm was used to perform the sedimentation velocity experiments. Interference scans were collected every 30 s at 25 °C.

Sedimentation equilibrium experiments were performed by loading a sample of protein (110 μL) and the protein dialysate (120 μL) into a six-sector Epon charcoal-filled centerpiece. Samples were spun at the velocity indicated in the text until sedimentation equilibrium was achieved as judged by WinMatch (David Yphantis, University of Connecticut, Jeff Lary, National Analytical Ultracentrifugation Center, and Storrs, CT).

Analysis of Sedimentation velocity data

Sedimentation velocity boundaries were analyzed using SedFit[28] version.14 (Peter Schuck, NIH), where the c(s) analysis was applied by modeling the sedimentation boundaries as solutions of the Lamm equation for noninteracting species.[29] The sedimentation velocity data were checked for errors in the timestamp automatically by V.14 SedFit and no timestamp errors were found.[30, 31] The sedimentation coefficient, *s*, is given by Svedberg's equation, as shown in Equation 1:[32]

$$s = \frac{M\left(1-\bar{v}\rho\right)}{Nf} = \frac{MD\left(1-\bar{v}\rho\right)}{RT} \qquad 1$$

where *M* is the molecular weight, $\bar{v}$ is the partial specific volume of the macromolecule, ρ is the density of the buffer, *R* is the ideal gas constant, *T* is absolute temperature in Kelvin, *N* is Avogadro's number, *f* is the frictional coefficient, and *D* is the diffusion coefficient.

The weight average sedimentation coefficient was calculated from the c(s) distribution by integrating over the area of the c(s) distribution. All sedimentation coefficients, *s*, reported in the text if not indicated otherwise are corrected to standard buffer condition, 20 °C in water, i.e. $s_{20,w}$ using Equation 2: [33]

$$s_{20,w} = \frac{\left(1-\rho_{20,w}\bar{v}\right)}{\left(1-\rho\bar{v}\right)} \cdot \frac{\eta}{\eta_{20,w}} \cdot s \qquad 2$$

where $\rho_{20,w}$ and $\eta_{20,w}$ are the density and viscosity of water at 20 °C, respectively. ρ and η are the density and viscosity of buffer at the experimental temperature, respectively.

Analysis of Sedimentation Equilibrium Data

Sedimentation equilibrium boundaries were subjected to global nonlinear-least-squares (NLLS) fitting using HeteroAnalysis (James L. Cole and Jeffrey W. Lary, Storrs CT) and the single ideal species model provided in the software.[34, 35] The partial specific volume for ClpB, $\bar{v}$ = 0.7403 ml g^{-1}, was calculated from the primary sequence. Since these experiments were performed in the presence of 10 % glycerol the partial specific volume was corrected using Equation 3 as reported by Cole[35] and previously applied to the examination of ClpA[36] by us, which accounts for the changes in hydration due to the presence of the glycerol.

$$\frac{\Delta\bar{v}}{\Delta[\text{glycerol}]\%(\text{v}/\text{v})} = (3.33 \pm 0.38) \times 10^{-4} \text{mL g}^{-1} \qquad 3$$

Taking into account the 10 % glycerol used in our experiments results in a correction of $\Delta\bar{v}$ = +0.0033 mL g^{-1}, which gives a partial specific volume for ClpB in 10 % glycerol of $\bar{v}$ = 0.7436 ml g^{-1}. The density of the buffer was calculated from the buffer components using Sednterp[25]. The density of buffer and partial specific volume of ClpB were set as 1.03253 g ml^{-1} and 0.7436 ml g^{-1} in the "options" tab in HeteroAnalysis, respectively.

Sedimentation coefficients (s) estimation from global fitting of the time difference curves

Sedimentation coefficients, *s*, for ClpB monomers and hexamers were determined experimentally. Sedimentation coefficients for other ClpB oligomers were calculated by using WinHydroPRO.[37] This was accomplished using a simulated *E. coli* ClpB hexamer structure based on the monomer crystal structure, which was kindly provided by Dr.

Rebecca Wade.[16, 38] ClpB dimer, trimer, tetramer, and pentamer model structures were generated using VMD[39] by removing four, three, two, or one adjacent protomers, respectively from the hexameric ring model. Thus, the interface between ClpB protomers for these oligomers was preserved to be the same as in the ClpB hexamer model. For the dimer, trimer, tetramer, and pentamer the "shell model from residual-level" model was used for calculation as recommended by A. Ortega et al.[37] The sedimentation coefficients for ClpB oligomers in buffer H supplemented with 100 mM, 200 mM and 300 mM NaCl based on this calculation are constrained in the global fitting of the time difference curves to the values presented in Table 1 for the various models tested.

The sedimentation coefficients were also calculated using Equation 4 with the assumption that ClpB monomer and n-mer have the same frictional ratio.[33]

$$s_n = s_1(n)^{\frac{2}{3}} \qquad 4$$

However, the calculated s_6 is 6 % different from the experimental value when applying the experimentally determined s_1 to Equation 4. On the other hand, s_6 predicted using HydroPro is only 2 % different from the experimental value determined for the hexamer. Moreover, fits by allowing s values for the intermediates to float within in the value range predicted by Equation 4 and HydroPro show that changes in the sedimentation coefficient that fall within the bounds predicted by Equation 4 and HydroPro have no significant impact on the determined equilibrium constants (fitting not shown).

Global fitting of sedimentation velocity data using the time difference curve method

SedAnal[40] was used to globally fit sedimentation velocity data at various ClpB concentrations to obtain thermodynamic and, if present, kinetic information on ClpB

assembly. SedAnal calculates and fits the time difference curves. The data were examined by two different strategies in an effort to rule out models that do not adequately describe the experimental observations. In summary, the first strategy solves the path independent equilibrium equations and no information on kinetic rate constants is obtained. The second strategy solves the system of coupled differential equations describing the kinetics and therefore, under limiting conditions, can yield information on the kinetic rate constants. Specifically, the dissociation rate constants are used as the fitting parameters. It is important to note that since the system of coupled differential equations are being solved for a specific mechanism the rate constants determined are elementary rate constants and not observed rate constants or relaxation times, where both the observed rate constant and the relaxation times are composites of several elementary rate constants. For a detailed description of how this was accomplished see Lin and Lucius, Methods in Enzymology, in press).

Throughout the manuscript the following notation is used to describe reaction schemes. The reactions given by Equations 6 - 8 are noted as "1-2, 1-4, 1-6" to indicate stoichiometric assembly reactions. Whereas, Equations 9 - 11 are denoted as 1-2-4-6 to indicate step-wise assembly reactions.

Sedimentation coefficients, molecular weights, density increment, and extinction coefficient are all constrained in the analysis. The extinction coefficients for ClpB (× 1.2 cm path length) used in the global fitting are (0.45 ± 0.01) ml mg^{-1} at 280 nm and (4.4 ± 0.1) ml mg^{-1} at 230 nm. The protein loading concentrations, dissociation rate constants and association equilibrium constants are floated as indicated in the models.

For each final set of parameters presented here the Nonlinear-Least-Squares fit (NLLS) was performed multiple times by starting with different initial guesses for the parameters. This was done to insure that the examination resulted in the same set of parameters regardless of starting point. Moreover, this strategy provides evidence that the results are not the consequence of local minima.

The F-test was used to compare the goodness of fits between two models applied to the same experimental data set. $F_{calculated}$ is calculated using Equation 5 (derived from Johnson & Straume Equation 45),[41]

$$F_{calculated} = \frac{RMSD_1^2}{RMSD_2^2} \qquad 5$$

where $RMSD_1$ and $RMSD_2$ are the Root Mean Squared Deviation for the two fits being compared. $RMSD_1$ is always chosen to be equal to or greater than $RMSD_2$, so that the $F_{calculated}$ is always larger than unity. $F_{calculated}$ is compared to $F_{critical}$, *where* $F_{critical}$ was determined using the "F-calculator" that is embedded in SedAnal.[41] If $F_{calculated}$ is greater than $F_{critical}$, the two fits are statistically different and the fit with the smaller RMSD can be concluded to be a significantly better fit.

To determine the error space on the resulting fitting parameters, "Johnson & Straume Equation 35"[41] was used to calculate the confidence intervals for each parameter by selecting the "F-statistics" function that is built into SedAnal. That is to say, to determine the errors for each parameter, F-test was also performed by clicking "Calculate F-statistics" in "advance" setting. In addition, the $F_{critical}$ value used to determine the maximum and minimum values for each parameter was calculated using Johnson &

Straume Equation 35.[41] The confidence level for the uncertainty on each parameter was set at 68.3 %.

Results

ClpB exhibits dynamic assembly in the absence of nucleotide

To quantitatively examine the ClpB association mechanism, we performed sedimentation velocity experiments at multiple ClpB concentrations ranging from 1 to 15 μM in buffer H supplemented with 300 mM NaCl. Figure 1a shows a series of absorbance boundaries collected with 6 μM ClpB by monitoring absorbance at 280 nm as a function of radial position and time. The data were subjected to c(s) analysis using SedFit (Peter Shuck, NIH) and the red solid lines in Figure 1a represent the fit. The residuals from the analysis are shown in Figure 1b they are randomly distributed about zero and, on average, less than 0.01 absorbance units, indicating that the fit describes the data well.

Figure 1c shows the c(s) distribution obtained from analyzing the raw data from sedimentation velocity experiments performed with 6 μM (red), 9 μM (blue) and 15 μM (green) ClpB. At 6 μM ClpB, two broad c(s) distributions are observed and a weighted average sedimentation coefficient, $\overline{s}_{20,w}$ = (7.2 ± 0.4) S was determined (see red trace in Figure 1c), where the uncertainty represents the standard deviation determined from three independent measurements. At a ClpB concentration of 9 μM, $s_{20,w}$ increased to (9.0 ± 0.4) S indicating that larger oligomers form at this higher ClpB concentration. The c(s) distribution continues to shift to higher sedimentation coefficient values as the concentration of ClpB is elevated to 15 μM (see green traces in Figure 1c).

In Figure 1 C the distribution around 5 S appears to shift to larger sedimentation coefficient values with increasing protein concentration. This observation suggests that one or more than one ClpB oligomer may exhibit rapid dissociation on the time scale of sedimentation.[42] Hence, the c(s) distributions may represent reaction boundaries of ClpB oligomers and the peaks are not likely representative of discrete species. Consequently, we sought to find conditions where we could perturb the equilibrium in a way that would favor discrete species.

NaCl concentration has a strong impact on ClpB assembly

Schlee et al. reported that decreasing the salt concentration drives ClpB to form large oligomers.[17, 43] Therefore, we performed sedimentation velocity experiments to examine the impact of [NaCl] on the distribution of states. Sedimentation velocity experiments with 2 μM ClpB in buffer H at 300, 200, and 100 mM NaCl were performed as described in materials and methods. The data were subjected to c(s) analysis and the c(s) distribution for ClpB at each [NaCl] is shown in Figure 2 (panels a-c).

For 2 μM ClpB at 300 mM NaCl, the c(s) distribution shows one predominant peak with an $s_{20,w}$ = (5 ± 1) S, where the uncertainty is determined from the analysis performed in SedFit. Figure 2 b shows that decreasing the [NaCl] to 200 mM results in a shift to the right and a broad distribution is observed with sedimentation coefficients ranging from 4.4 S to 18.6 S. Consistently, as the [NaCl] is further decreased to 100 mM the distribution shifts farther to the right (see Figure 2c). Noticeably, under these conditions, one major peak at (17.6 ± 0.6) S begins to emerge. This ~17.6 S distribution peak didn't shift significantly to larger $s_{20,w}$ values when the [NaCl] was decreased further to 50 mM (data not shown).

One ClpB oligomer is predominately populated at 100 mM NaCl

We next examined the ClpB concentration dependence of the assembly reaction at 100 mM NaCl since it appears as though a single ClpB oligomer is emerging at low NaCl concentrations. Sedimentation velocity experiments in buffer H supplemented with 100 mM NaCl were performed at various ClpB concentrations ranging from 4 μM to 18 μM. Strikingly, one predominate c(s) peak was observed at all examined ClpB concentrations (see Figure 3). Increasing the ClpB concentration didn't shift the (17.6 ± 0.6) S c(s) peak indicating that this peak likely represents the largest ClpB oligomer that is significantly populated in the 100 mM NaCl conditions.

In Figure 3, for 15 μM ClpB, the emergence of a minor distribution is observed centered about an $s_{20,w}$ = (23 ± 2) S. This could represent a reaction boundary for a larger oligomer. However, the peak area doesn't appear to have a clear ClpB concentration dependence, which suggests those oligomers, if present, may exhibit weak assembly and their population is not sufficiently large to be accurately determined at these [ClpB]. This is further indicated by the fact that the observed peak area of the 23 S peak is less than 5 % of the entire c(s) distribution.

Determination of the molecular weight of the largely populated ClpB oligomer in the absence of nucleotide in Buffer H with 100 mM NaCl

To determine the molecular weight of the ClpB oligomer that we report here to have $s_{20,w}$ = (17.6 ± 0.6) S, we performed sedimentation equilibrium experiments with 4, 6, 9, 12, 15, and 18 μM ClpB in buffer H with 100 mM NaCl at 25 °C. The sedimentation equilibrium data as a function of radial position and [ClpB] are shown in Figure 4. The data were globally analyzed using the "single ideal species" model in

HeteroAnalysis(James Cole and Jeffery Lary). [34] The molecular weight of the single species was allowed to float as the fitting parameter and the determined molecular weight, M = (578 ± 3) kDa is in good agreement with the value calculated from the primary structure of ClpB as a hexamer of 575 kDa. The fitting Root Mean Squared Deviation (RMSD) is 0.0090 indicating the data are well described by this model. This is consistent with the analysis of the sedimentation velocity data shown in Figure 3.

To examine the precision in the determination of the hexamer, we asked the question; can the sedimentation equilibrium data be equally well described if the predominant oligomer is assumed to be a pentamer or heptamer? Although there has been no report concluding ClpB forms pentamers, the question here is; could our determination of the molecular weight be off by plus or minus one protomer unit? To address this question, we constrained the molecular weight to be either 479 kDa or 671 kDa, the molecular weight of the pentamer or heptamer, respectively. In this analysis the partial specific volume was $\bar{v}$ = 0.7436 ml mg^{-1}, which is corrected for the presence of 10 % glycerol (see Materials and Methods). The analysis resulted in an RMSD = 0.01313 assuming a pentamer and an RMSD = 0.0120 assuming a heptamer. Comparing to the RMSD = 0.0090 where the molecular weight is allowed to float and determined to be M = (578 ± 3) kDa, the analyses assuming pentamer or heptamer are both significantly worse.

An alternate way in which the determined molecular weight could be off by as much as one promoter unit would be if the error in the partial specific volume is sufficiently large. The experiments performed here have been carried out in 10 % glycerol, which can lead to a dynamic gradient of glycerol. This could lead to

uncertainty in the partial specific volume. However, we have performed an extensive analysis of the impact of the inclusion of the 10 % glycerol and conclude that it does not impact the determined parameters, see Supplemental Figure S1, S2 and the corresponding presentation.

Determination of the assembly model for ClpB

From sedimentation equilibrium experiments presented in Figure 4 we conclude the 17.6 S species is hexameric ClpB. In comparison to the c(s) distribution determined in the presence of 100 mM NaCl, the c(s) distribution for ClpB in the presence of 300 mM NaCl did not exhibit a significant distribution at ~17.6 S (compare Figure 2a to 2c). Rather, in the presence of 300 mM NaCl a broad c(s) distribution as a function of ClpB concentration was detected and the apparent c(s) peak positions do not superimpose (see Figure 1c). This indicates that ClpB oligomers may exhibit rapid dissociation on the time scale of sedimentation. Alternatively, oligomers between monomers and hexamers may be present.

To resolve these possibilities we globally analyzed the data using the time difference curve method (see Materials and Methods).[40] Shown in Figure 5 are the difference curves for sedimentation velocity experiments collected with 1, 2, 3, 6, 9, 11, 13, and 15 μM ClpB in Buffer H with 300 mM NaCl. These data were subjected to global NLLS analysis with the simplest monomer to hexamer model. This model assumes that only monomers and hexamers are significantly populated at thermodynamic equilibrium. However, it is important to note that this model does not rule out the possibility that intermediates are present. If the model adequately describes the data it only allows for the conclusion that intermediates are not sufficiently populated to be

detected in the analysis. Nevertheless, the data are not well described by this model with an RMSD = 3.374×10^{-2} (fit not shown). Thus, we conclude that at least one intermediate must be significantly populated.

Since it is clear that monomers and hexamers should be present we globally analyzed the difference curves in Figure 5 with the assumption that at least monomers and hexamers are present and one intermediate is also present. The RMSD values for the analysis including a dimeric, trimeric, or tetrameric intermediate are given in Table 2. All of these "three species" models do not adequately describe the sedimentation boundaries as judged by comparing the RMSDs using F-test (Table 2).

It is possible that the hexamer may not be significantly populated at the elevated salt concentration of 300 mM NaCl. Thus, we analyzed the data with a monomer-dimer, monomer-tetramer model ("1-2, 1-4" model in Table 2). Based on the RMSD this model also did not adequately describe the experimental observations (see Table 2).

Table 2 shows that the data are not adequately described by a two species (1-6) or any of the three species models ("1-2, 1-6", etc). These observations may suggest that there are more intermediates present at thermodynamic equilibrium. To test this possibility, we analyzed the data with the "1-2, 1-3, 1-6" or the "1-2, 1-4, 1-6" models. Both of these models describe the data well. However, the "1-2, 1-4, 1-6" model describes the data significantly better based on the RMSD and the F-tests shown in Table 2.

The data can be well described by including monomers, dimers, tetramers, and hexamers. However, an energetically simpler model is the isodesmic model because it assumes all the equilibrium constants are the same even though it includes every

intermediate from monomers to hexamers. As shown in Table 2, the isodesmic model does not describe the data as well as either the “1-2, 1-3, 1-6” or the “1-2, 1-4, 1-6” model. Moreover, the “1-2, 1-4, 1-6” model yields the lowest RMSD.

Determination of the assembly path for ClpB

To test if the sedimentation velocity experiments reported here are sensitive to the kinetics of dissociation the difference curves in Figure 5 were subjected to global NLLS analysis by modeling the reaction kinetics as described in Materials and Methods. However, analysis of these data by incorporating rate constants for each of the steps results in the model no longer being path-independent as assumed when modeling for only the equilibrium constants. Consequently, the first step in this analysis strategy is to determine the path that best describes the data.

There are many paths for ClpB monomers to form hexamers. However, the probability of three bodies colliding in a single kinetic step is diminishingly small. Thus, we sought to describe the data with the simplest model that requires each oligomer to form through bimolecular interactions. The analysis of the sedimentation velocity data shown in Figure 5 was best described by the “1-2, 1-4, 1-6” path independent stoichiometric model given by the reactions in Equations 6 - 8.

$$2B_1 \underset{}{\overset{L_{2,0}}{\rightleftharpoons}} B_2 \qquad 6$$

$$4B_1 \underset{}{\overset{L_{4,0}}{\rightleftharpoons}} B_4 \qquad 7$$

$$6B_1 \underset{}{\overset{L_{6,0}}{\rightleftharpoons}} B_6 \qquad 8$$

To transition to incorporating the kinetics into this model we subjected the difference curves to analysis using the step-wise 1-2-4-6 model given by the reactions in Equations 9 - 11.

$$2B_1 \overset{K_2}{\rightleftharpoons} B_2 \qquad 9$$

$$2B_2 \overset{K_4}{\rightleftharpoons} B_4 \qquad 10$$

$$B_2 + B_4 \overset{K_6}{\rightleftharpoons} B_6 \qquad 11$$

Figure 5 shows a representative set of difference curves and global NLLS fit using the 1-2-4-6 model . The resultant step-wise equilibrium constants, K_n, are given in Table 3-a and the resultant reverse rate constants, $k_{r,n}$, are given in Table 3-b both under the heading of 300 mM NaCl. Table 3-a also reports the stoichiometric binding constants, $L_{n,0}$ that were calculated using the relationships given by Equations 12- 14.

$$L_{2,0} = K_2 \qquad 12$$

$$L_{4,0} = K_2^2 \cdot K_4 \qquad 13$$

$$L_{6,0} = K_2^3 \cdot K_4 \cdot K_6 \qquad 14$$

The analysis of the difference curves shown in Figure 5 yielded dissociation rate constants that floated to values larger than 0.01 s^{-1} (See Table 3-b for 300 mM NaCl). If all the reactions given by Equations 9 - 11 have reverse rate constants, $k_r > 0.01\ s^{-1}$ then the data should be equally well described by stoichiometric equilibrium constants given by Equation 15 - 17 (Lin and Lucius, Methods in Enzymology manuscript in press).

$$K_2 = \frac{k_{f2}}{k_{r2}} = \frac{[B_2]}{[B_1]^2} \qquad 15$$

$$K_4 = \frac{k_{f4}}{k_{r4}} = \frac{[B_4]}{[B_2]^2} \qquad 16$$

$$K_6 = \frac{k_{f6}}{k_{r6}} = \frac{[B_6]}{[B_2][B_4]} \qquad 17$$

In other words, the difference curves should not contain any information on path. Thus, a model using step-wise equilibrium constants, K_n, (1-2-4-6 model) or a model using stoichiometric binding constants, $L_{n,0}$ (1-2, 1-4, 1-6) as floating parameters should describe the data equally well. However, as seen in Table 3-a, the 1-2-4-6 model, where the reverse rate constants for each step are floated, describes the data significantly better (based on an F-test of the RMSD) than the "1-2, 1-4, 1-6" model, which contains no rate constants.

It is important to note that the 1-2-4-6 model contains three additional floating parameters (three rate constants) compared to the "1-2, 1-4, 1-6" model that does not contain rate constants. However, the three additional parameters are not the reason for the improved RMSD exhibited in Table 3-a for the 300 mM NaCl data. The RMSD for the two fits is the sum of the squared residuals divided by the degrees of freedom. The degrees of freedom are defined as the number of data points minus the number of floating parameters. Since the two fits are performed on ~35,000 data points, the addition of three parameters does not significantly change the degrees of freedom. Thus, the improvement in the fit when the rate constants are allowed to float as fitting parameters cannot be dismissed as the simple consequence of three additional fitting parameters. However, we

acknowledge that the value of the reverse rate constants for this fit (see Table 3-b) are not well constrained because they are outside of the expected measurable range of 10^{-2} – 10^{-5} s^{-1}. Nevertheless, constraining these rate constants to the empirical upper limit of 10^{-2} s^{-1} resulted in a statistically worse fit (data not shown). Consequently, we are forced to conclude that some information on these rate constants must be present in the data.

With respect to the stoichiometric assembly models, Table 2 indicates that the "1-2, 1-3, 1-6" model represents the next best model compared to the "1-2, 1-4, 1-6" model. This conclusion is based on the F-statistic presented in Table 2. Thus, we asked the question; upon incorporation of path information could a step-wise 1-2-3-6 model describe the experimental observations better than the step-wise 1-2-4-6 model? To address this question we subjected the difference curves in Figure 5 to global NLLS analysis using the step-wise 1-2-3-6 model. Further, two additional replicates of the same concentration dependent data were analyzed to the 1-2-3-6 and the 1-2-4-6 models and the RMSD values and F-statistics are reported in Table 4. In all cases the 1-2-4-6 model describes the data significantly better than the 1-2-3-6 model. Thus, we conclude that the step-wise 1-2-4-6 model represents the best description of the experimental observations.

Global analysis of Sedimentation Velocity data collected in the presence of 100 and 200 mM NaCl

Sedimentation velocity data collected in the presence of 200 and 100 mM NaCl were subjected to global NLLS analysis using the "1-2, 1-4, 1-6" model (no kinetic parameters) and the 1-2-4-6 models (including kinetic parameters). The fitting results, together with the results for buffer H supplemented with 300 mM NaCl, are presented in

Table 3-a. The kinetic parameters from the analysis to the 1-2-4-6 model are presented in Table 3-b. The difference curves and the best fits are shown in the supplemental materials as Figure S3 for 200 mM NaCl and Figure S4 for 100 mM NaCl buffer.

From the analysis of sedimentation velocity experiments in the presence of either 200 mM NaCl or 100 mM NaCl with the 1-2-4-6 model and floating the step-wise equilibrium constants, K_n, and reverse rate constants, $k_{r,n}$, we find that most of the rate constants are at or above the empirical boundary for instantaneous dissociation (10^{-2} s^{-1}). However, the dissociation rate constant for hexamer formation, in the presence of 100 mM NaCl, is within the measureable range (see Table 3-b for rate constants). In all cases, based on the fitting RMSD and subsequent F-test, the data are better described by the step-wise 1-2-4-6 model compared to the "1-2, 1-4, 1-6" model.

For comparison purposes, the stoichiometric binding constant was calculated when fitting to the 1-2-4-6 model and the step-wise constant was calculated when fitting to the "1-2,1-4,1-6" model using the relationships in Equations 12 - 14 , see Table 3-a. Although the RMSD is better in all cases when fitting using the 1-2-4-6 model compared to the "1-2, 1-4, 1-6" model the parameters are similar. However, in most cases, the parameters are slightly outside of the uncertainty calculated using the F-statistics (see Table 3-a).

It is important to note that the uncertainties reported on the parameters in Table 3-a are from the F-statistic function built into Sedanal and represent fitting error. The uncertainty that would represent the reproducibility is likely larger. To quantify this assertion we present the average of three experimental replicates of the entire protein concentration range collected in the presence of 300 mM NaCl and subsequent fits.

Table S1 shows the average and standard deviation of $L_{n,0}$ for these three replicates. Although the order of magnitude of these equilibrium constants is highly reproducible, the coefficient has between 29 and 55 % uncertainty, which is likely a better representation of the reproducibility of the parameters than the fitting uncertainty presented in Table 3-a.

Discussion

Our studies, for the first time, reveal the assembly pathway for formation of *E. coli* ClpB hexamers in the absence of nucleotide. Here we show that *E. coli* ClpB resides in a monomer-dimer-tetramer-hexamer equilibrium in the absence of nucleotide at three different salt concentrations. Further, we report that *E. coli* ClpB hexamers, like *T. thermophilus* ClpB and *S. cerevisiae* Hsp104,[9, 19] exhibit a short half-life in solution (several minutes or less). Although we do not have precise measures of the dissociation rate constants, our analysis does indicate that it occurs with a rate constant greater than 0.01 s^{-1}.

ClpB/Hsp104 disrupts disordered protein aggregates in collaboration with the DnaK/Hsp70 system.[6, 44, 45] These enzymes play important roles in cell survival during stress, such as heat shock. To date, the mechanism of ClpB catalyzed protein disaggregation remains unclear.

At the heart of this protein disaggregation function are many protein-protein interactions. These include ClpB self-association, ClpB interactions with protein aggregates, and protein-protein interactions between ClpB hexamers and components of the DnaKJE system of enzymes that are yet to be fully resolved.[6, 44, 45] Consequently, determination of the ClpB assembly mechanism is an imperative first step to

quantitatively understand how ClpB collaborates with DnaKJE to dissociate large protein aggregates.

Here we show that ClpB forms monomers, dimers, tetramers and hexamers in solution in the absence of nucleotide. The concentrations of each of these oligomers can be described as a function of the total ClpB monomer concentration. This can be accomplished using the reported parameters in Table 3-a and the model written in Scientist (Micromath Research, St. Louis MO) provided in the supplemental materials.

The ability to predict the concentration of hexamers in experiments performed at different total ClpB monomer concentrations at a variety of [NaCl] will aid in the design and interpretation of a number of other experiments. For example, as shown in Figure 6A for 300 mM NaCl, ClpB dimers and tetramers are significantly populated while hexamers are less populated. As the [NaCl] is decreased (Figure 6, panel B and C) ClpB hexamers become the most populated oligomer. This knowledge can be used to interpret kinetic data by accounting for the population of hexamers and other oligomers that may also contribute to the chaperone activity, such as ATP binding and hydrolysis or peptide binding and disaggregation.

As seen in Table 3-a there is large uncertainty on the determination of the step-wise dimerization, K_2 and tetramerization, K_4, equilibrium constants in 100 mM NaCl when fitting to the 1-2-4-6 model. Yet the error on the hexamerization equilibrium constant, K_6, is much smaller. As shown in Figure 6C, ClpB monomers and hexamers are significantly populated in 100 mM NaCl but dimers and tetramers are much more sparsely populated. Since monomers and hexamers are more highly populated the constraints on the hexamerization equilibrium constant, K_6, are much better than the

dimerization, K_2, and tetramerization, K_4, equilibrium constants. However, because the stoichiometric assembly constants $L_{4,0}$ and $L_{6,0}$ are calculated using Equation 13 and 14, respectively, which are functions of the measured step-wise constants, the propagated uncertainties on those parameters are large. This observation would seem to imply that the sedimentation velocity data should be able to be described by a simple monomer-hexamer equilibrium. However, attempts to describe these data with the monomer-hexamer model resulted in and RMSD = 0.00989, which, based on F-statistics, is significantly worse than the 0.00965 reported in Table 3-a for the 1-2-4-6 model.

One way to eliminate the dependence of the stoichiometric binding constants on the step-wise binding constants for ClpB assembly in 100 mM NaCl is using the stoichiometric model, "1-2, 1-4, 1-6", to analyze the data. However, the results presented here for 100 mM NaCl indicated that ClpB hexamer doesn't achieve instantaneous equilibrium on the time scale of sedimentation and the "1-2, 1-4, 1-6" model doesn't describe the data as well as the 1-2-4-6 model (see RMSD in Table 3-a). Although the association rate constants for the "1-2, 1-4, 1-6" model make little sense, when the data are examined with this model with $L_{2,0}$, $L_{4,0}$, $L_{6,0}$ k_{r2}, k_{r4} and k_{r6} floating, the fit describes the data equally well with an RMSD = 0.00964 as the 1-2-4-6 model. The dimerization constants floated to values consistent with an insignificant population and rapid dissociation, $L_{2,0}$ = 32.2 M^{-1} and k_{r2} = 258 s^{-1}, consistent with an insignificant population. Thus, the data are really being fit to a "1-4, 1-6" model and the determined parameters are $L_{4,0}$ = 3.4 (2.8, 4.0) × 10^{18} M^{-3}, $L_{6,0}$ = 4.5 (4.0, 5.0) × 10^{32} M^{-5}, k_{r4} = 3 (2, 6) × 10^{-3} S^{-1} and k_{r6} =0.9 (0.8, 1.1) × 10^{-3} S^{-1}. These results agree well with the parameters determined using a 1-2-4-6 model with an error space suggesting better constraints.

ClpB forms hexamer in the absence of nucleotide

There exist conflicting conclusions in the literature regarding whether ClpB forms hexamers or heptamers in the absence of nucleotide. It was reported by many groups that *E. coli* ClpB forms heptamers in the absence of nucleotide.[14, 15, 18] Akoev et al concluded that ClpB resided in a monomer-dimer-heptamer equilibrium in the absence of nucleotide and the binding of nucleotide resulted in a conformational switch of ClpB from heptamer to hexamer.[14] However, del Castillo et al. reported that their experimental data for ClpB assembly can best be described by a monomer-hexamer-dodecamer model in the absence of nucleotide.[13]

Rigorously determining the assembly state for a polydisperse system is challenging. Previously published results on ClpB[17] and our experimental data (see Figure 2) show that [NaCl] perturbs the ClpB assembly equilibrium. Therefore, we varied the [ClpB] and [NaCl] to find a condition that favors monodispersity. In buffer H supplemented with 100 mM NaCl, one predominant c(s) peak was found with an $s_{20,w}$ ~17.6 S at multiple [ClpB] ranging from 4 to 18 μM.

Finding conditions where the system is monodisperse allowed us to determine the molecular weight of this 17.6 S species using sedimentation equilibrium experiments. We showed that the experimentally determined molecular weight of the ~17.6 S oligomer is ~578 kDa (see Figure 3 and Figure 4), which is in good agreement with the molecular weight of hexameric ClpB calculated from its primary structure.

del Castillo et al. observed a population of oligomers with high sedimentation coefficients at high [ClpB] that they hypothesized may be a dodecamer. In our experimental [ClpB] range, no significant population of ClpB dodecamer is observed.

However, the concentrations in our study are lower than those used by del Castillo. We avoided [ClpB] above 18 μM monomeric concentration to limit the impact on the data from effects of non-ideality and nonspecific aggregation. Moreover, the physiological concentration of ClpB in *E. coli* reported by Mogk et al. is in the range of ~9 – 19 μM monomers.[46]

ClpB hexamer exhibits rapid dissociation

Here, in the absence of nucleotide, we report that ClpB oligomers exhibit rapid dissociation. In most cases, from this analysis, we do not have precise determinations of the rate constants. However the analysis does reveal kinetic information and allows us to put a lower limit on the dissociation rate constants. For 300 and 200 mM NaCl all of the reactions occur with a dissociation rate constant of greater than or equal to 0.01 s^{-1}. In the case of 100 mM NaCl the dissociation rate constant for hexamers was found to be within the measurable range and determined to be ~1.3 x 10^{-3} s^{-1}.

The lower limits of the dissociation rate constants, which represent protein concentration independent kinetic parameters, were determined from global fitting of sedimentation velocity experiments at multiple ClpB concentrations. Further, by combining the determined equilibrium constant with the lower limit of the dissociation rate constant we can approximate the lower limits of the association rate constants for each bimolecular step, which represents the protein concentration dependent kinetic parameter (see k_f in Equations 15 - 17). Therefore, a lower limit on the subunit exchange rate could be calculated as the product of the free ClpB concentration and the bimolecular rate constant, k_f.

The kinetics of subunit exchange for the hexamer of *T. thermophilus* ClpB[19] and *S. cerevisiae* Hsp104[9] has been examined by others using a stopped-flow FRET strategy. In those experiments, donor labeled ClpB is rapidly mixed with acceptor labeled ClpB. The donor and acceptor labeled ClpB must dissociate and then reassociate. Upon reassociation a FRET signal will be observed. The kinetic time courses acquired at a single protein concentration were examined by nonlinear-least squares fitting to a sum of three exponentials. The observed rate constants were reported as the rate constants for dissociation of the hexamer. However, it is clear from the work presented here that there are multiple oligomers present in solution that could contribute to this signal. Moreover, the reported apparent rate constant is a convolution of both dissociation and re-association. Thus, to fully understand the mechanism of assembly, an examination of the dependence of this observed rate constant on protein concentration is still needed.

Several groups have employed gel filtration methods to examine ClpB assembly.[10, 17, 18, 47] One study suggested that *E. coli* ClpB forms tetramers in the absence of nucleotide. This conclusion was drawn from the observation of a single elution peak with a retention time corresponding to a molecular weight of ~ 350 kDa, although the chromatogram was not shown.[47] These gel filtration experiments were performed under low salt conditions where our experiments would predict hexamer is the predominantly populated species. Thus, one explanation for the discrepancy between the gel filtration results and our findings would be that ClpB hexamer is rapidly dissociating and re-associating during the run. The free monomers, intermediate oligomers and hexamers are separated by the gel-filtration column. The hexamers are further dissociated because the

smaller oligomers and monomers are more and more separated as the sample moves through the column.

Since the kinetic properties of ClpB assembly were not largely discussed before the studies performed by Werbeck et al.,[19] the dissociation kinetics of ClpB were not included in the interpretation of the gel filtration elution profiles for ClpB. In fact, the broadening of the elution peak can represent ClpB dissociation during the gel filtration run, which is similar to what we observe in c(s) distributions. That is to say, peaks in a c(s) distribution will be broadened if the macromolecule dissociates on the time scale of sedimentation at low ClpB concentrations. Indeed, the c(s) distributions can exhibit broadening for other reasons. Nevertheless, just as our analysis of the difference curves by incorporating kinetics, so too can the gel filtration elution profiles for ClpB be analyzed by incorporating kinetic parameters.[48-51] However, to our knowledge, this analysis has not been applied to the examination of ClpB using gel filtration.

The contribution of our work compared to a similar study performed by del Castillo et al.,[13] is that we incorporated the kinetics of oligomerization into the data analysis and determined a pathway to describe ClpB hexamer assembly. Moreover, our reported association equilibrium constants for ClpB hexamers are several orders of magnitude greater than theirs. For example, they report a $K_6' = 10^{13}$ M^{-5} (in our notation and by our interpretation of their results this is $L_{6,0}$) at 150 mM KCl and as can be seen in Table 3-a for 200 mM NaCl and 100 mM NaCl we determined an $L_{6,0} = \sim 10^{29}$ and 10^{33} M^{-5}, respectively. Although the solution conditions employed by del Castillo et al and us are not identical, with their reported numbers, the concentration of hexamers in solution would be predicted to be 7.8×10^{-23} μM for 10 μM total ClpB monomer concentration in

the presence of 150 mM KCl. This concentration of hexamers is clearly under the detection limits of analytical ultracentrifugation experiments and in direct conflict with the raw data reported in their manuscript.[13]

The standard state Gibbs free energy change, $\Delta G°_6$, for ClpB hexamers formed through monomers can be calculated using Equation 18:

$$\Delta G^{\circ}_{6} = -RT \ln(L_{6,0}) \qquad 18$$

where R is the gas constant and T is the absolute temperature. Using the $L_{6,0}$ presented in Table 3-a, the $\Delta G°_6$ is -190 kJ mol^{-1}, -167 kJ mol^{-1}, and -146 kJ mol^{-1} at 100 mM, 200 mM, and 300 mM NaCl, respectively. Our values are significantly smaller than those reported by del Castillo et al.[13] Nevertheless, our findings show that the formation of hexamers in the absence of nucleotide is energetically favored in solution.[10]

In this paper, we presented a strategy to quantitatively investigate the energetics and kinetics of a large motor protein assembly, a strategy that can be employed to examine other AAA+ protein complexes. These proteins include a diverse array of molecular machines that are engaged in a variety of cellular activities.[2] As pointed out by many researchers, to fully understand the function of a protein machine, we not only need to know the structure at an atomic level, we also need to have knowledge of the energetics and kinetics for each intermediate step in the motor reaction.[52, 53]

Acknowledgements:

We would like to thank Dr. Rebecca Wade and Axel Mogk for providing the ClpB hexamer structure model; thanks to Dr. Peter Prevelige for use of the XL-A analytical ultracentrifuge. Thanks to Dr. Walter Stafford for the XL-I analytical

ultracentrifuge. Thanks to Hongyi Yang for discussion on structural analysis using VMD. Thanks to Elizabeth Duran, Nate Scull, and Ryan Stafford for comments on the manuscript. Special thanks to Francis Appling and Clarissa Weaver for critical reading of the manuscript. This work was supported by NSF grant MCB-1412624 to ALL.

References

[1] Arlt, H., Tauer, R., Feldmann, H., Neupert, W., and Langer, T. (1996) The YTA10-12 complex, an AAA protease with chaperone-like activity in the inner membrane of mitochondria, *Cell 85*, 875-885.
[2] Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes, *Genome research 9*, 27-43.
[3] Vale, R. D. (2000) AAA proteins. Lords of the ring, *J Cell Biol 150*, F13-19.
[4] Motohashi, K., Watanabe, Y., Yohda, M., and Yoshida, M. (1999) Heat-inactivated proteins are rescued by the DnaK.J-GrpE set and ClpB chaperones, *Proc Natl Acad Sci U S A 96*, 7184-7189.
[5] Genevaux, P., Georgopoulos, C., and Kelley, W. L. (2007) The Hsp70 chaperone machines of Escherichia coli: a paradigm for the repartition of chaperone functions, *Molecular microbiology 66*, 840-857.
[6] Glover, J. R., and Lindquist, S. (1998) Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins, *Cell 94*, 73-82.
[7] Schmitt, M., Neupert, W., and Langer, T. (1996) The molecular chaperone Hsp78 confers compartment-specific thermotolerance to mitochondria, *J Cell Biol 134*, 1375-1386.
[8] Shorter, J. (2008) Hsp104: a weapon to combat diverse neurodegenerative disorders, *Neuro-Signals 16*, 63-74.
[9] DeSantis, M. E., Leung, E. H., Sweeny, E. A., Jackrel, M. E., Cushman-Nick, M., Neuhaus-Follini, A., Vashist, S., Sochor, M. A., Knight, M. N., and Shorter, J. (2012) Operational plasticity enables hsp104 to disaggregate diverse amyloid and nonamyloid clients, *Cell 151*, 778-793.
[10] Zolkiewski, M., Kessel, M., Ginsburg, A., and Maurizi, M. R. (1999) Nucleotide-dependent oligomerization of ClpB from Escherichia coli, *Protein Sci 8*, 1899-1903.
[11] Veronese, P. K., Stafford, R. P., and Lucius, A. L. (2009) The Escherichia coli ClpA Molecular Chaperone Self-Assembles into Tetramers, *Biochemistry 48*, 9221-9233.
[12] Li, T., and Lucius, A. L. (2013) Examination of Polypeptide Substrate Specificity for E. coli ClpA, *Biochemistry 52*, 4941-4954.
[13] del Castillo, U., Alfonso, C., Acebron, S. P., Martos, A., Moro, F., Rivas, G., and Muga, A. (2011) A quantitative analysis of the effect of nucleotides and the M domain on the association equilibrium of ClpB, *Biochemistry 50*, 1991-2003.
[14] Akoev, V., Gogol, E. P., Barnett, M. E., and Zolkiewski, M. (2004) Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB, *Protein Sci 13*, 567-574.
[15] Kim, K. I., Cheong, G. W., Park, S. C., Ha, J. S., Woo, K. M., Choi, S. J., and Chung, C. H. (2000) Heptameric ring structure of the heat-shock protein ClpB, a protein-activated ATPase in Escherichia coli, *J Mol Biol 303*, 655-666.
[16] Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M., and Tsai, F. T. (2003) The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state, *Cell 115*, 229-240.

[17] Schlee, S., Groemping, Y., Herde, P., Seidel, R., and Reinstein, J. (2001) The chaperone function of ClpB from Thermus thermophilus depends on allosteric interactions of its two ATP-binding sites, *J Mol Biol 306*, 889-899.
[18] Mogk, A., Schlieker, C., Strub, C., Rist, W., Weibezahn, J., and Bukau, B. (2003) Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity, *J Biol Chem 278*, 17615-17624.
[19] Werbeck, N. D., Schlee, S., and Reinstein, J. (2008) Coupling and dynamics of subunits in the hexameric AAA+ chaperone ClpB, *J Mol Biol 378*, 178-190.
[20] Schlieker, C., Weibezahn, J., Patzelt, H., Tessarz, P., Strub, C., Zeth, K., Erbse, A., Schneider-Mergener, J., Chin, J. W., Schultz, P. G., Bukau, B., and Mogk, A. (2004) Substrate recognition by the AAA+ chaperone ClpB, *Nature structural & molecular biology 11*, 607-615.
[21] Lee, S., Choi, J. M., and Tsai, F. T. (2007) Visualizing the ATPase cycle in a protein disaggregating machine: structural basis for substrate binding by ClpB, *Mol Cell 25*, 261-271.
[22] Haslberger, T., Zdanowicz, A., Brand, I., Kirstein, J., Turgay, K., Mogk, A., and Bukau, B. (2008) Protein disaggregation by the AAA+ chaperone ClpB involves partial threading of looped polypeptide segments, *Nature structural & molecular biology 15*, 641-650.
[23] Li, T., Lin, J., and Lucius, A. L. (2015) Examination of polypeptide substrate specificity for Escherichia coli ClpB, *Proteins 83*, 117-134.
[24] Li, T., Weaver, C. L., Lin, J., Duran, E. C., Miller, J. M., and Lucius, A. L. (2015) E. coli ClpB is a Non-Processive Polypeptide Translocase, *Biochemical Journal*
[25] Laue, T. M., Shah, B.D., Ridgeway, T.M., Pelletier, S.L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins, In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (S.E. Harding, A. J. R., J.C. Horton, Ed.), Royal Society of Chemistry, Cambridge.
[26] T.M. Laue, B. D. S., T.M. Ridgeway and S.L. Pelletier (1992) *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, Royal Society of Chemistry.
[27] Lohman, T. M., Chao, K., Green, J. M., Sage, S., and Runyon, G. T. (1989) Large-scale purification and characterization of the Escherichia coli rep gene product, *J Biol Chem 264*, 10139-10147.
[28] Schuck, P. (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation, *Biophys J 75*, 1503-1512.
[29] Schuck, P., and Rossmanith, P. (2000) Determination of the sedimentation coefficient distribution by least-squares boundary modeling, *Biopolymers 54*, 328-341.
[30] Ghirlando, R., Balbo, A., Piszczek, G., Brown, P. H., Lewis, M. S., Brautigam, C. A., Schuck, P., and Zhao, H. (2013) Improving the thermal, radial, and temporal accuracy of the analytical ultracentrifuge through external references, *Analytical biochemistry 440*, 81-95.
[31] Zhao, H., Ghirlando, R., Piszczek, G., Curth, U., Brautigam, C. A., and Schuck, P. (2013) Recorded scan times can limit the accuracy of sedimentation coefficients in analytical ultracentrifugation, *Analytical biochemistry 437*, 104-108.
[32] Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling, *Biophys J 78*, 1606-1619.

[33] Correia, J. J. (2000) Analysis of weight average sedimentation velocity data, *Methods Enzymol 321*, 81-100.
[34] Cole, J. L. (2004) Analysis of heterogeneous interactions, *Methods Enzymol 384*, 212-232.
[35] Cole, J. L. (1996) Characterization of human cytomegalovirus protease dimerization by analytical centrifugation, *Biochemistry 35*, 15601-15610.
[36] Veronese, P. K., and Lucius, A. L. (2010) Effect of Temperature on the Self-Assembly of the Escherichia coli ClpA Molecular Chaperone, *Biochemistry 49*, 9820-9829.
[37] Ortega, A., Amoros, D., and Garcia de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models, *Biophys J 101*, 892-898.
[38] Oguchi, Y., Kummer, E., Seyffer, F., Berynskyy, M., Anstett, B., Zahn, R., Wade, R. C., Mogk, A., and Bukau, B. (2012) A tightly regulated molecular toggle controls AAA+ disaggregase, *Nature structural & molecular biology 19*, 1338-1346.
[39] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *Journal of molecular graphics 14*, 33-38, 27-38.
[40] Stafford, W. F., and Sherwood, P. J. (2004) Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants, *Biophysical chemistry 108*, 231-243.
[41] Johnson, M. L., and Straume, M. (1994) Comments on the analysis of sedimentation equilibrium experiments, In *Modern Analytical Ultracentrifugation* (Schuster TM and Laue TM, e., Ed.), pp p.37-65, Birkhauser, Boston.
[42] Schuck, P. (2010) Sedimentation patterns of rapidly reversible protein interactions, *Biophysical journal 98*, 2005-2013.
[43] Diamant, S., Rosenthal, D., Azem, A., Eliahu, N., Ben-Zvi, A. P., and Goloubinoff, P. (2003) Dicarboxylic amino acids and glycine-betaine regulate chaperone-mediated protein-disaggregation under stress, *Molecular microbiology 49*, 401-410.
[44] Goloubinoff, P., Mogk, A., Zvi, A. P., Tomoyasu, T., and Bukau, B. (1999) Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network, *Proc Natl Acad Sci U S A 96*, 13732-13737.
[45] Zolkiewski, M. (1999) ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from Escherichia coli, *J Biol Chem 274*, 28083-28086.
[46] Mogk, A., Tomoyasu, T., Goloubinoff, P., Rudiger, S., Roder, D., Langen, H., and Bukau, B. (1999) Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB, *The EMBO journal 18*, 6934-6949.
[47] Woo, K. M., Kim, K. I., Goldberg, A. L., Ha, D. B., and Chung, C. H. (1992) The heat-shock protein ClpB in Escherichia coli is a protein-activated ATPase, *J Biol Chem 267*, 20429-20434.
[48] Ackers, G. K. (1967) Molecular Sieve Studies of Interacting Protein Systems.
[49] Hummel, J. P., and Dreyer, W. J. (1962) Measurement of protein-binding phenomena by gel filtration, *Biochimica et biophysica acta 63*, 530-532.

[50] Wilton, R., Myatt, E. A., and Stevens, F. J. (2004) Analysis of protein-protein interactions by simulation of small-zone gel filtration chromatography, *Methods in molecular biology (Clifton, N.J 261*, 137-154.
[51] Winzor, D. J. (2003) The development of chromatography for the characterization of protein interactions: a personal perspective, *Biochem Soc Trans 31*, 1010-1014.
[52] Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists, *Cell 92*, 291-294.
[53] Bukau, B., and Horwich, A. L. (1998) The Hsp70 and Hsp60 chaperone machines, *Cell 92*, 351-366.
[54] Roark, D. E. (1976) Sedimentation equilibrium techniques: multiple speed analyses and an overspeed procedure, *Biophysical chemistry 5*, 185-196.
[55] Durchschlag, H. (1986) Specific Volumes of Biological Macromolecules and Some Other Molecules of Biological Interest, In *Thermodynamic data for biochemistry and biotechnology* (Hinz, H.-J. r., Ed.), pp 45-128 Springer-Verlag, Berlin ; New York.
[56] Gabrielson, J. P., Arthur, K. K., Kendrick, B. S., Randolph, T. W., and Stoner, M. R. (2009) Common excipients impair detection of protein aggregates during sedimentation velocity analytical ultracentrifugation, *Journal of pharmaceutical sciences 98*, 50-62.

Table 1: Sedimentation coefficients for ClpB oligomers used in global analysis of sedimentation velocity data.

1-2-4-6	Monomer	Dimer	Trimer	Tetramer	Pentamer	Hexamer
M.W. (Da)	**95,817**	**191,634**	**287,451**	**383,268**	**479,085**	**574,902**
100 mM NaCl	3.6	5.88	7.73	9.30	11.08	12.5
200 mM NaCl	3.07	5.65	7.42	8.93	10.64	12.2
300 mM NaCl	3.00	5.6	7.36	8.86	10.56	11.9

The sedimentation coefficients used in the analysis are not corrected to $s_{20,w}$, the values presented in this table have units of S (Svedberg), as 10^{-13} second.

Table 2: Model determination for ClpB assembly in buffer H with 300 mM NaCl.

Model	RMSD	$F_{calculated}$
1-6	3.374×10^{-2}	11.686
1-2, 1-6	1.055×10^{-2}	1.143
1-3, 1-6	1.201×10^{-2}	1.481
1-4, 1-6	1.672×10^{-2}	2.870
1-2, 1-4	1.748×10^{-2}	3.137
1-2, 1-3, 1-6	0.999×10^{-2}	1.024
1-2, 1-4, 1-6	0.987×10^{-2}	1
Isodesmic	2.172×10^{-2}	4.843

All models in the table represent stoichiometric binding reactions, e.g. 1-6 indicates monomers forming hexamers. An $F_{calculated}$ value larger than an $F_{critical}$ value = 1.005 indicates a significantly worse fit compared to the best fit. The "1-2, 1-4, 1-6" fit is significantly better than any other fit presented in this table. Thus, all other models were compared to this model. $F_{calculated}$ was determined as described in materials and methods. The value of $F_{calculated}$ =1 indicates the model that best describes the data.

Table 3-a: Global fitting results of sedimentation velocity experimental data at multiple ClpB concentrations in buffer H with the NaCl concentration indicated in the table.

	300 mM NaCl		200 mM NaCl		100 mM NaCl	
Reactions (1-2-4-6 model)	***RMSD*= 9.01 ×10^{-3}**		***RMSD*= 7.65 ×10^{-3}**		***RMSD*= 9.65 ×10^{-3}**	
	***K_n* (M^{-1})**	***Calculated $L_{n,0}$***	***K_n* (M^{-1})**	***Calculated $L_{n,0}$***	***K_n* (M^{-1})**	***Calculated $L_{n,0}$***
$2B \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} B_2$	**8.9 (8.8, 9.0) ×10^{4}**	**8.9 (8.8, 9.0) ×10^{4}**	**1.73 (1.68, 1.78) ×10^{5}**	**1.73 (1.68, 1.78) ×10^{5}**	**9.5 (6.0, 18.6) ×10^{4}**	**9.5 (6.0, 18.6) ×10^{4}**
$2B_2 \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} B_4$	**1.83 (1.75, 1.92) ×10^{5}**	**1.44 (1.36, 1.52) ×10^{15}**	**1.9 (1.8, 2.0) ×10^{6}**	**5.8 (5.4, 6.2) ×10^{16}**	**4.2 (1.0, 11.9) ×10^{8}**	**3.8 (-0.2, 13.9) ×10^{18}**
$B_2 + B_4 \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} B_6$	**8.6 (8.4, 8.9) ×10^{5}**	**1.10 (1.03, 1.17) ×10^{26}**	**2.0 (1.9, 2.1) ×10^{7}**	**2.0 (1.8, 2.2) ×10^{29}**	**1.6 (1.0, 2.3) ×10^{9}**	**5.8 (-2.2, 26) ×10^{32}**
Reactions (1-2, 1-4, 1-6 model)	***RMSD*= 9.87 ×10^{-3}**		***RMSD*= 8.02 ×10^{-3}**		***RMSD*= 9.97 ×10^{-3}**	
	***Calculated K_i* (M^{-1})**	***$L_{n,0}$***	***Calculated K_i* (M^{-1})**	***$L_{n,0}$***	***Calculated K_i* (M^{-1})**	***$L_{n,0}$***
$2B \overset{L_{2,0}}{\rightleftharpoons} B_2$	**9.7 (9.5, 9.8) ×10^{4}**	**9.7 (9.5, 9.8) ×10^{4}**	**1.47 (1.41, 1.52) ×10^{5}**	**1.47 (1.41, 1.52) ×10^{5}**	**0 (0, 7.8) ×10^{3}**	**0 (0, 7.8) ×10^{3}**
$4B \overset{L_{4,0}}{\rightleftharpoons} B_4$	**1.6 (1.5, 1.7) ×10^{5}**	**1.5 (1.4, 1.6) ×10^{15}**	**2.0 (1.8, 2.2) ×10^{6}**	**4.4 (4.2, 4.6) ×10^{16}**	**NA**	**0.9 (0.6, 1.1) ×10^{18}**
$6B \overset{L_{6,0}}{\rightleftharpoons} B_6$	**9 (8, 10) ×10^{5}**	**1.34 (1.31, 1.37) ×10^{26}**	**2.9 (2.7, 3.1) ×10^{7}**	**1.89 (1.84, 1.95) ×10^{29}**	**NA**	**3.9 (3.5, 4.2) ×10^{32}**

Note: Analysis using the "1-2, 1-4, 1-6" model employed the Newton-Raphson method to determine the concentration of each species by numerically solving the conservation of mass equation given by Equations 6. Whereas, analysis of data using the 1-2-4-6 employed the kinetic integrator (SEulEx) to numerically solve the system of coupled differential equations for the reactions. The step-wise equilibrium constants, K_n, were used as fitting parameters when the data were examined using the 1-2-4-6 model to describe the data. The stoichiometric equilibrium constants, $L_{n,0}$, were used as fitting parameters when the data were examined using the "1-2, 1-4, 1-6" model. The corresponding equilibrium constant, $L_{n,0}$ or K_n were calculated using Equation 12 – 14. The data are collected in buffer H with the NaCl concentration indicated in the table. The unit of $L_{n,0}$ is $M^{-(n-1)}$.

Table 3-b: Kinetic parameters determined from global NLLS analysis that are associated with 1-2-4-6 model presented in table 3-a.

	300 mM NaCl	**200 mM NaCl**	**100 mM NaCl**
Reactions (1-2-4-6 model)	***RMSD*= 9.01 ×10^{-3}**	***RMSD*= 7.65 ×10^{-3}**	***RMSD*= 9.65 ×10^{-3}**
	$k_{r,n}$ (s^{-1})	**$k_{r,n}$ (s^{-1})**	**$k_{r,n}$ (s^{-1})**
$2B \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} B_2$	**4.1***	**5.6 (5.1, 6.1) ×10^{-3}**	**0.2***
$2B_2 \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} B_4$	**0.01***	**6.4***	**0.04***
$B_2 + B_4 \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} B_6$	**0.7***	**5.4***	**1.3 (1.0, 1.6) ×10^{-3}**

Some rate constants floated to values equal to or larger than 0.01 s^{-1} (noted with a "*"), which is outside of the measureable range. Therefore, those rate constants were not allowed to float for F-statistic error determination and the errors for those parameters were not determined.

Table 4: Comparison of the goodness of fit for three replicates to the 1-2-3-6 vs. the 1-2-4-6 model

Replicates	1-2-3-6	1-2-4-6	$F_{Calculated}$*	F_{table}
	$RMSD_1$ of Fit	$RMSD_2$ of Fit	$RMSD_1^2/RMSD_2^2$	Confidence level: 68.3%
1	9.95×10^{-2}	9.01×10^{-3}	1.220	1.005
2	1.090×10^{-2}	1.082×10^{-2}	1.015	1.005
3	1.137×10^{-2}	1.111×10^{-2}	1.047	1.005

*For the calculation of $F_{Calculated}$, recall, $RMSD_1$ is defined as the larger standard deviation, see Materials and Methods.

Figures

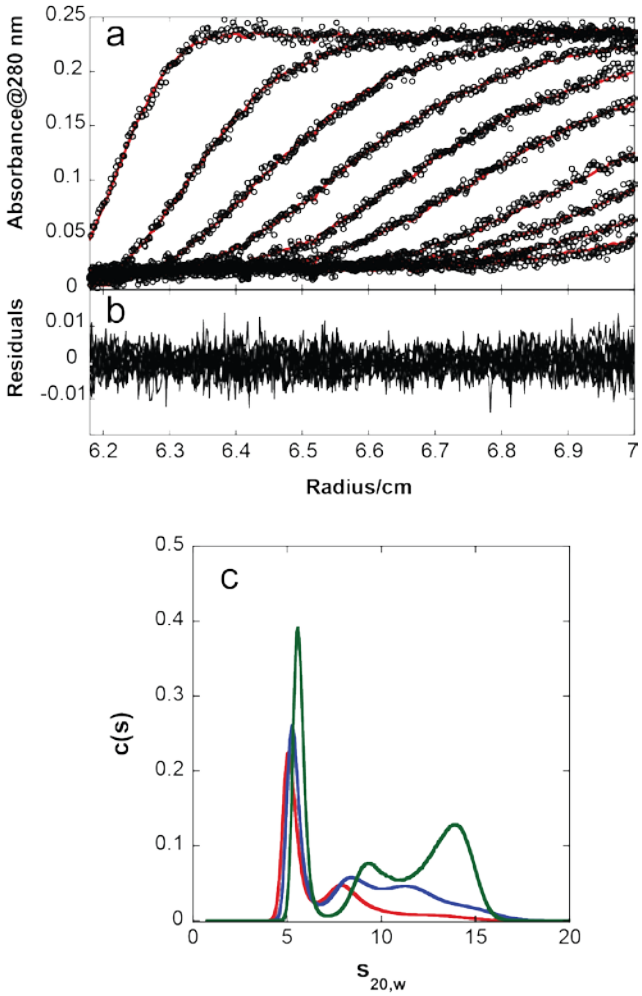

Figure 1. ClpB exhibits dynamic assembly in the absence of nucleotide (a) Raw sedimentation velocity scans as function of radial position for 6 μM ClpB in buffer H300 at 25 °C. The scans were collected every 4 min at 280 nm, every 8th scan is shown. Open circles are data and red solid lines are fits generated from c(s) analysis; (b) Residuals of fit. (c) c(s) distribution for ClpB assembly in buffer H supplemented with 300 mM NaCl at 6 (red), 9 (blue) and 15 (green) μM ClpB monomer.

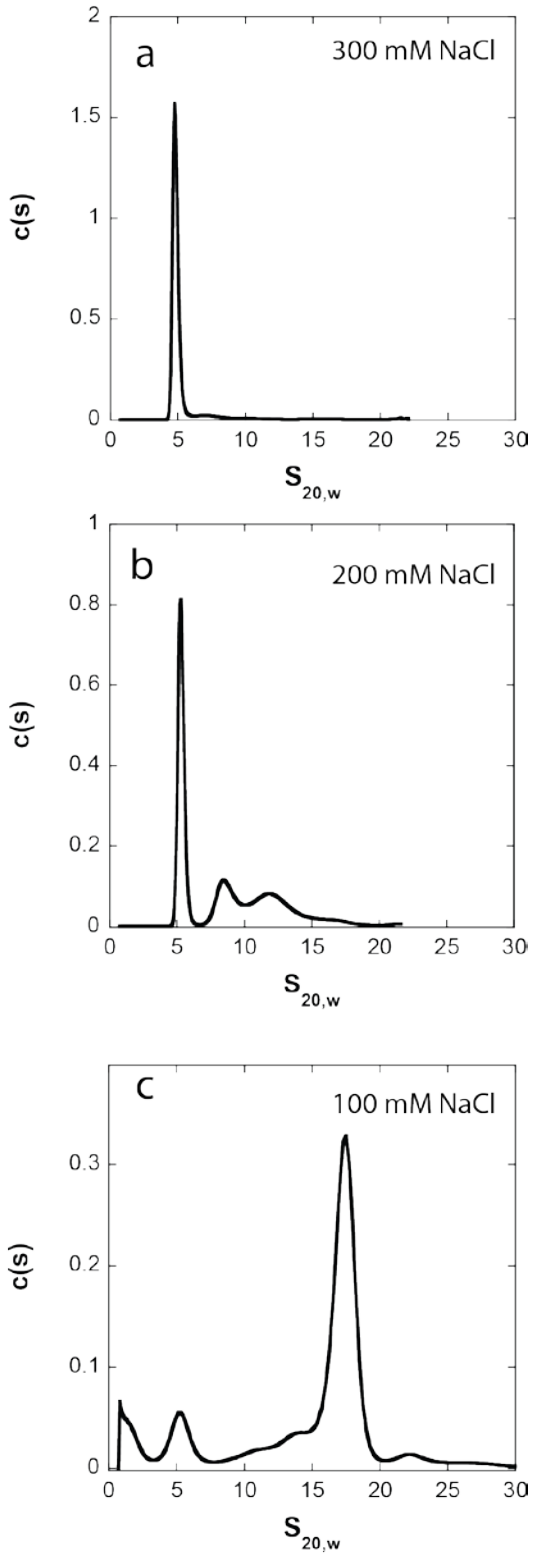

Figure 2: Salt effect on ClpB assembly. The c(s) distribution versus $s_{20,w}$ for 6 μM ClpB in Buffer H with NaCl concentration indicated in panel a, b and c.

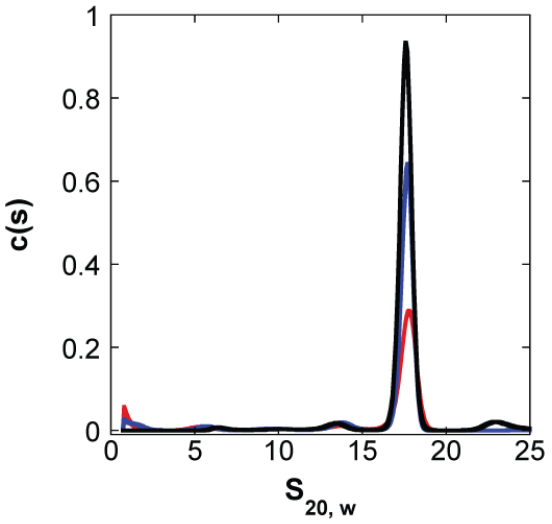

Figure 3: One ClpB oligomer is predominately populated in 100 mM NaCl. The c(s) distribution versus $s_{20,w}$ of 6 (red), 9 (blue) and 15 (black) μM ClpB in Buffer H with 100 mM NaCl at 25 °C.

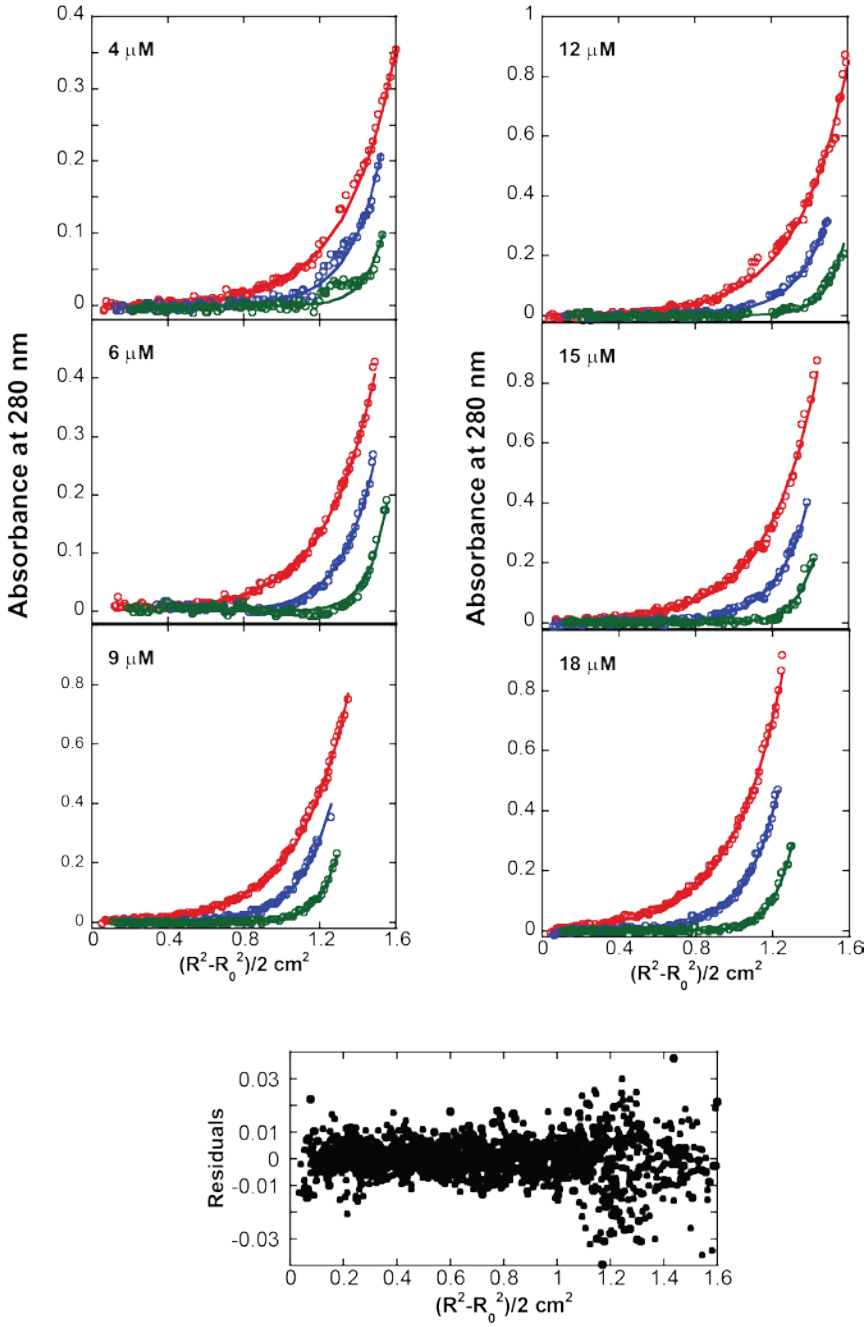

Figure 4. Global fitting of sedimentation equilibrium scans. Scans were collected with 4, 6, 9, 12, 15, and 18 μM ClpB, in 100 mM NaCl at 25 °C, at 8,000 (red), 10,000 (blue) and 13,000 (green) rpm. The radial position is presented as $(R^2 - R_0^2)/2$, where R is the radial position of each absorbance datum and R_0 is the radial positon of the meniscus of the sample. The data were subjected to NLLS analysis using Heteroanalysis using the single ideal species model. The open circles are the raw data and the solid lines are the fits. The residuals from the fits are shown in filled circles. The resulting molecular weight from the analysis is (578 ± 3) kDa. The fitting RMSD is 0.0090 ± 0.0003.

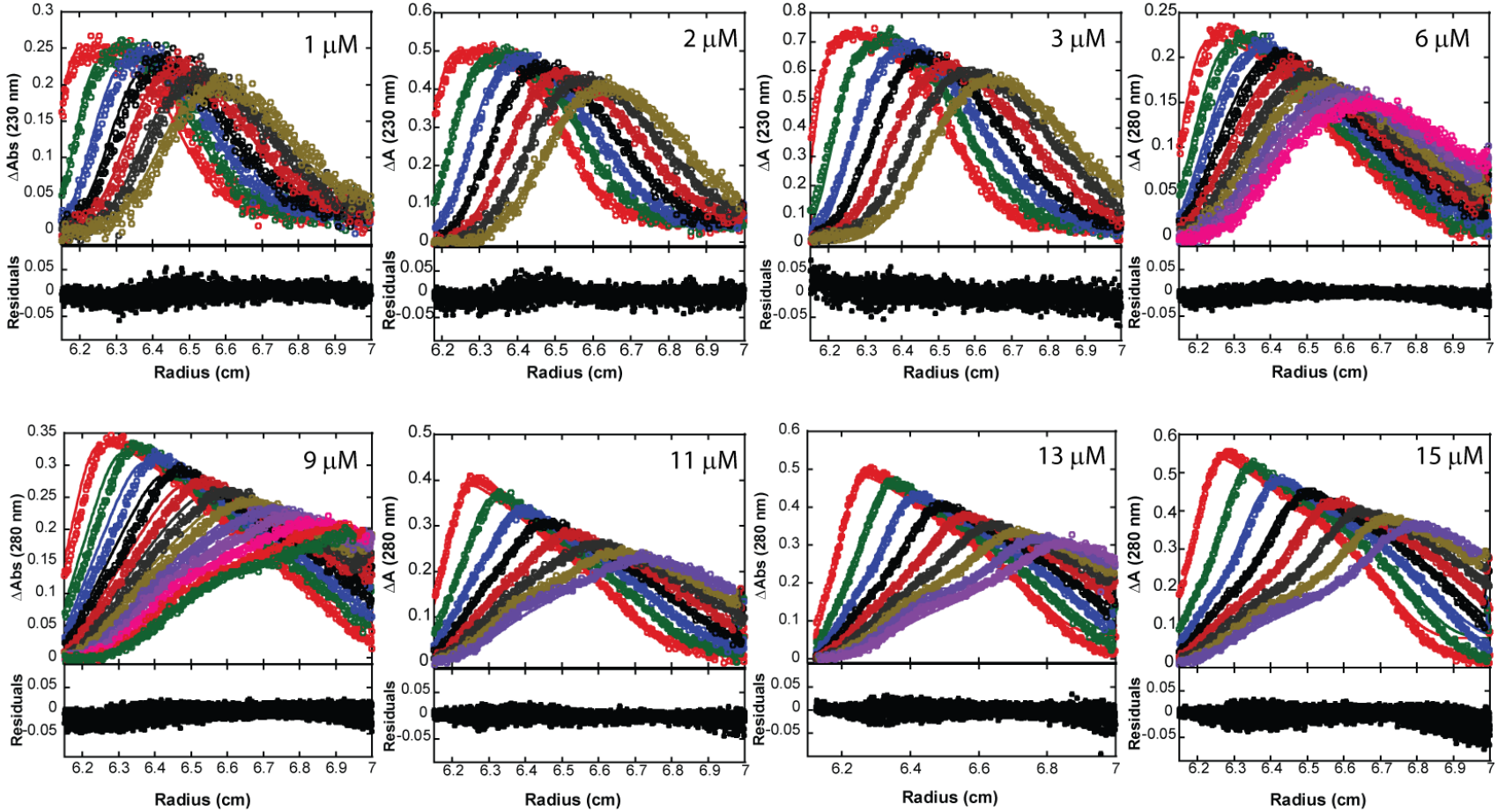

Figure 5: Time difference curves from Sedimentation velocity experiments performed on multiple ClpB concentrations in the presence of 300 mM NaCl. The loading concentrations of ClpB are indicated on the plots. For each ClpB concentration, the ΔAbsorbance time differences curves as a function of radial position are shown in the top panel and the residuals from the fit are shown in the bottom panel. The open circles are the raw data and the solid lines are the fit to the 1-2-4-6 model. The resultant parameters and RMSD are given in Table 3-a and b.

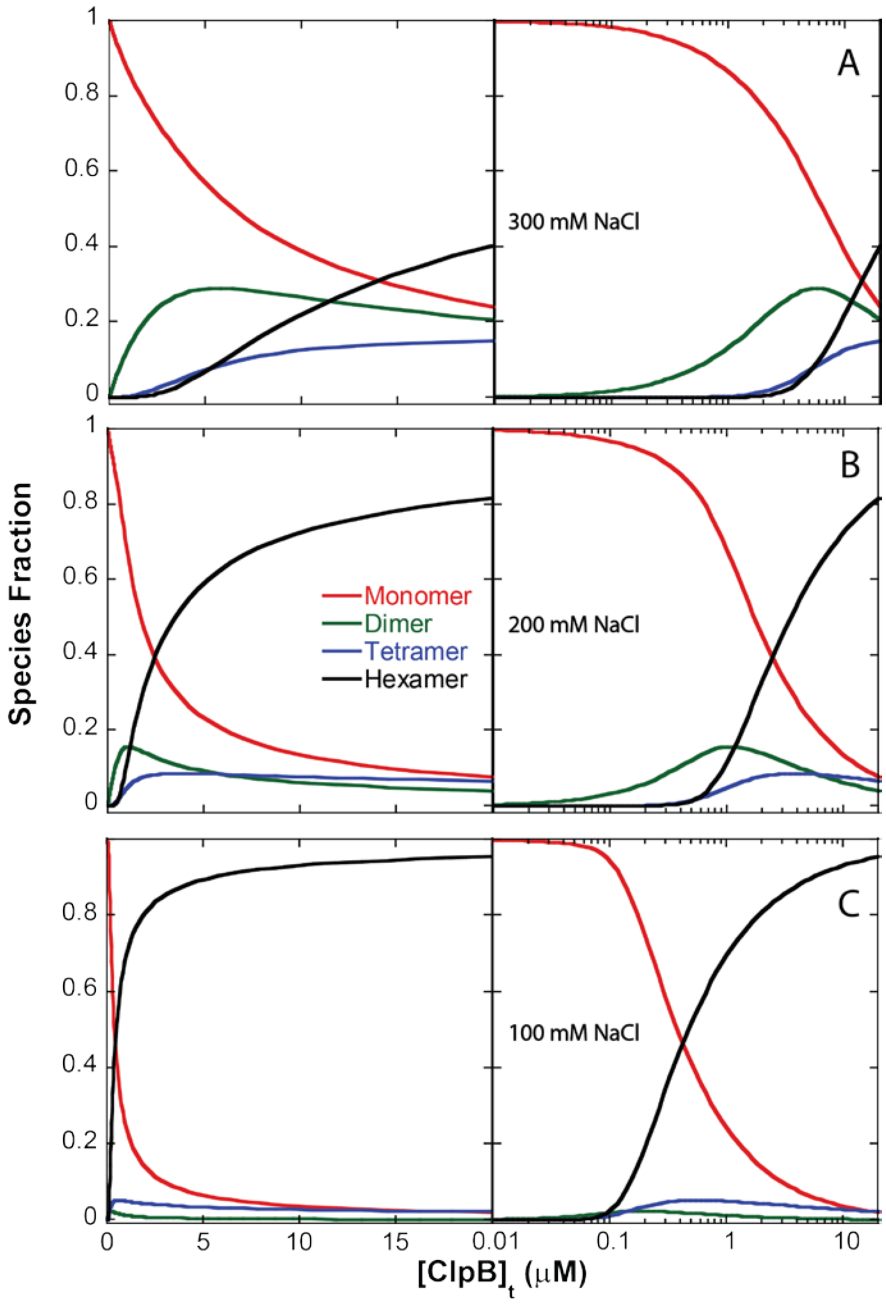

Figure 6: Species fractions simulations of ClpB. Simulations based on parameters determined in (A) 300 mM NaCl (B) 200 mM NaCl and (c) 100 mM NaCl, monomer (red), dimer (green), tetramer (blue) and hexamer (black). Every panel is composed of two plots: The X-axis is total ClpB monomer concentration, $[ClpB]_t$, and is presented in linear scale on the left and log scale on the right. The concentrations of ClpB oligomers are calculated using equilibrium constants presented in Table 3-a and b. The simulations were generated using the Micromath Scientist model provided in the supplemental materials. ClpB oligomer concentration divided by the total ClpB concentration gives us the species fraction of that oligomer. The species fractions of these oligomers at various [NaCl] are presented.

Supporting Information

Supporting Information

Impact of the presence of 10 % glycerol on sedimentation equilibrium experiments

The experiments reported here have been carried out in 10 % glycerol, which can lead to dynamic gradients leading to uncertainty in the partial specific volume. Thus, we asked the question; how much error in the partial specific volume would need to be present to result in an error in the molecular weight by as much as one protomer unit? In order to evaluate this error, we subjected the data in Figure 4 to NLLS analysis by allowing the parameter sigma to float, where sigma is given by Equation S1.[54]

$$\sigma = \frac{M\left(1-\bar{v}\rho\right)\omega^2}{RT} \qquad \text{S1}$$

The fitted value of sigma was then used to algebraically solve for the partial specific volume with the assumption that the molecular weight is a pentamer, hexamer, or heptamer. The determined sigma for one angular velocity of ω = 13,000 rpm is σ = 10.039 cm^{-2}, which is independent of any assumptions regarding the values of the parameters in Equation S1. Equation S1 was solved for the partial specific volume using a molecular weight M = 479 KDa, 575 KDa, and 671 KDa for pentamer, hexamer, or heptamer, respectively. The determined partial specific volumes were found to be $\bar{v}$ = 0.6969, 0.7422, and 0.7745 ml g^{-1}, respectively. The values of the partial specific volumes that would lead to an error of one protomer unit in the molecular weight determination of 0.6969 or 0.7745 ml g^{-1} are well outside of the range of what has been reported for most globular proteins. Such extreme values may be possible if posttranslational modifications such as phosphorylation or glycosylation were present.[55] However, no such modifications are known to be present on ClpB. Moreover, in order

for a glycerol gradient to increase the partial specific volume to a value of 0.7745 ml g^{-1} the glycerol would have to concentrate to ~93 % (v/v) based on Equation S1.

The analysis above shows that the error in the partial specific volume would have to be between 4 – 6 % in order for the error on the partial specific volume to lead to an error in molecular weight by one protomer unit. Although this seems unlikely, the question remains; does the presence of 10 % glycerol and thus dynamic gradients lead to such an uncertainty? This question is addressed below with global fitting of sedimentation velocity experiments.

Impact of the presence of 10 % glycerol on the final model determination

Inclusion of 10 % glycerol in these experiments can result in the formation of a dynamic gradient of glycerol. A glycerol concentration gradient will affect the partial specific volume and this could lead to errors in our model determination. To determine if the presence of 10 % glycerol is affecting our conclusions we performed a sedimentation velocity experiment with 10 % glycerol in water. The interference optical system was used to observe the sedimentation of the glycerol. The fringes as a function of radial position are shown in Figure S1. As can be seen in Figure S1, a gradient of glycerol is clearly formed.

ClpB monomer and hexamer have molecular weights of 95,817 and 574,902 Da, respectively and will sediment substantially faster than glycerol with a molecular weight of 92.09 Da. Thus, there is little concern about the gradient that forms at the top of the cell (~6 cm in Figure S1) since ClpB will be depleted in that region long before the gradient forms. However, the gradient at the bottom of the cell will be present as ClpB sediments.

To determine if the formation of the gradient is affecting the determination of the parameters for the data shown in Figure 5 and Table 3-a the sedimentation velocity data were truncated at 6.7 cm and subjected to global NLLS analysis using the 1-2-4-6 model. The choice to truncate at 6.7 cm was made based on the observation that the majority of the gradient is present between 6.7 – 7 cm. Moreover, truncation at about 6.7 cm is a strategy that has been employed by others.[56] The results of the analysis are shown in Figure S2. The resultant parameters determined for the truncated data are shown in Table S2. The comparison of the parameters determined from fitting the full data set and the truncated data set are shown in Table S3. The comparison reveals that, upon truncating the data, the parameters are all within ~1 % error of each other. This result indicates that the glycerol gradient formed is not impacting our model determination.

Since Figure S1 reveals that a glycerol gradient is clearly formed, the question remains; why does the gradient not affect the model determination? From the sedimentation of glycerol shown in Figure S1 we estimate that the highest glycerol concentration at the bottom of the cell would be ~ 12 %, and the lowest concentration at the top of the cell is ~ 5 %. Thus, the calculated partial specific volume based on equation 3 would be 0.7420 and 0.7443 ml g^{-1} for 5 % glycerol and 12 % glycerol, respectively. Keeping in mind that the partial specific volume for 10 % glycerol used in our analysis is 0.7436 ml g^{-1}, if a gradient from 5 – 12 % were formed this would result in a deviation in partial specific volume of less than 0.2 %, which is more than an order of magnitude lower than the required 4 – 6 % needed to be in error by one protomer unit discussed above. In conclusion, we find no evidence that 10 % glycerol is affecting the model determination.

Tables:

Table S1: Average of the fitting results of three replicates of sedimentation velocity experiments at multiple ClpB concentrations in buffer H supplemented with 300 mM NaCl with standard deviations shown.

Equilibrium Constants	Average of 3 repeats
$L_{2,0}$ (M^{-1})	$(7 \pm 2) \times 10^{4}$
$L_{4,0}$ (M^{-3})	$(9 \pm 5) \times 10^{14}$
$L_{6,0}$ (M^{-5})	$(8 \pm 3) \times 10^{25}$

Table S2: Analysis of sedimentation velocity data presented in Figure 5 by truncating the fitting range to 6.7 cm using SedAnal.

Fitting range: 6.1 cm to 6.7 cm	**300 mM NaCl**	
Reactions (1-2-4-6 model)	***RMSD*=8.73 ×10^{-3}**	
	K_i (M^{-1})	***Calculated $L_{n,0}$***
$2B \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} B_2$	1.02 (1.00, 1.03) ×10^{5}	1.02 (1.00, 1.03) ×10^{5}
$2B_2 \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} B_4$	1.4 (1.3, 1.6) ×10^{5}	1.5 (1.4, 1.6) ×10^{15}
$B_2 + B_4 \underset{k_{r3}}{\overset{k_{f3}}{\rightleftharpoons}} B_6$	1.0 (0.9, 1.1) ×10^{6}	1.5 (1.3, 1.8) ×10^{26}

This analysis was performed using the same analysis method when the fitting range is from 6.1 cm to 7 cm.

Table S3: Comparison of the fitting results between the analyses of sedimentation velocity data presented in Figure 5 and the analyses for the same experimental data with the fitting range truncated to 6.7 cm.

Fitting range: 6.1 cm to 6.7 cm ***Log($L_{n,0}$)***	**Fitting range: 6.1 cm to 7.0 cm** ***Log($L_{n,0}$)***	**Percentage of differences**
5.008 (5.000, 5.012)	4.949 (4.944, 4.954)	0.92%
15.17 (15.15, 15.20)	15.15 (15.13, 15.18)	Within error
26.17 (26.11, 26.24)	26.04 (26.01, 26.07)	0.15%

Micromath Scientist (Micromath Research, St. Louis MO) Model Used to Calculate Monomer, Dimer, Tetramer and Hexamer Concentrations of ClpB:

```
// Micromath Scientist Model File
IndVars: ClpBT
DepVars: Monomer, Dimer, Tetramer, Hexamer
Params: L20, L40, L60
ClpBT=ClpBf*(1+2*L20*ClpBf+4*L40*ClpBf^3+6*L60*ClpBf^5)
Monomer=ClpBf
Dimer=2*L20*ClpBf^2
Tetramer=4*L40*ClpBf^4
Hexamer=6*L60*ClpBf^6
0<ClpBf<ClpBT
```

Note: "ClpBT" is the total concentration of ClpB.

"ClpBf" is the free monomer concentration of ClpB.

"L20, L40 and L60" are the equilibrium constants, $L_{2,0}$, $L_{4,0}$ and $L_{6,0}$, as defined in Equation 10 - 12.

All concentrations are in monomeric unit.

The species fraction of an oligomer can be calculated using the concentration of the oligomer divided by the total ClpB concentration.

Figures

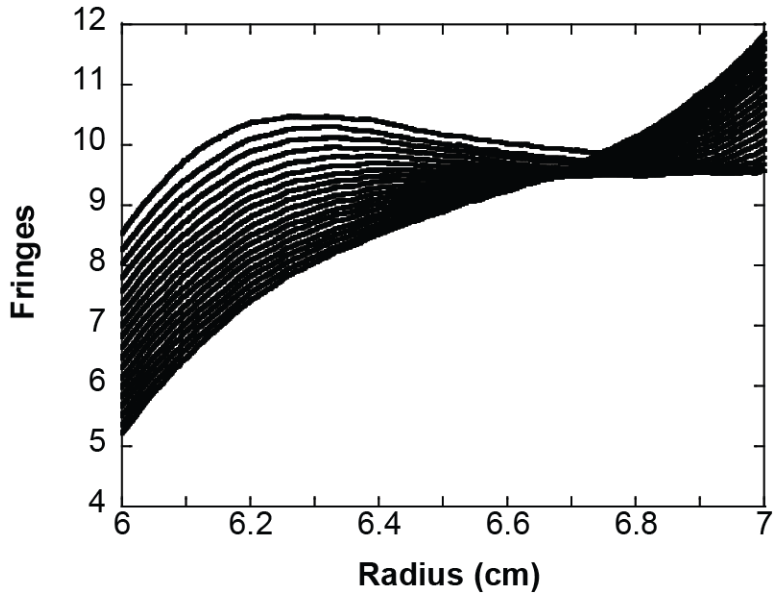

Figure S1: Sedimentation velocity scan on 10% glycerol using interference optics at 40,000 rpm, 25 °C. The scans were collected every 30 seconds. Every 16th scan is shown. The overall sedimentation time was the same as used for ClpB sedimentation velocity experiments.

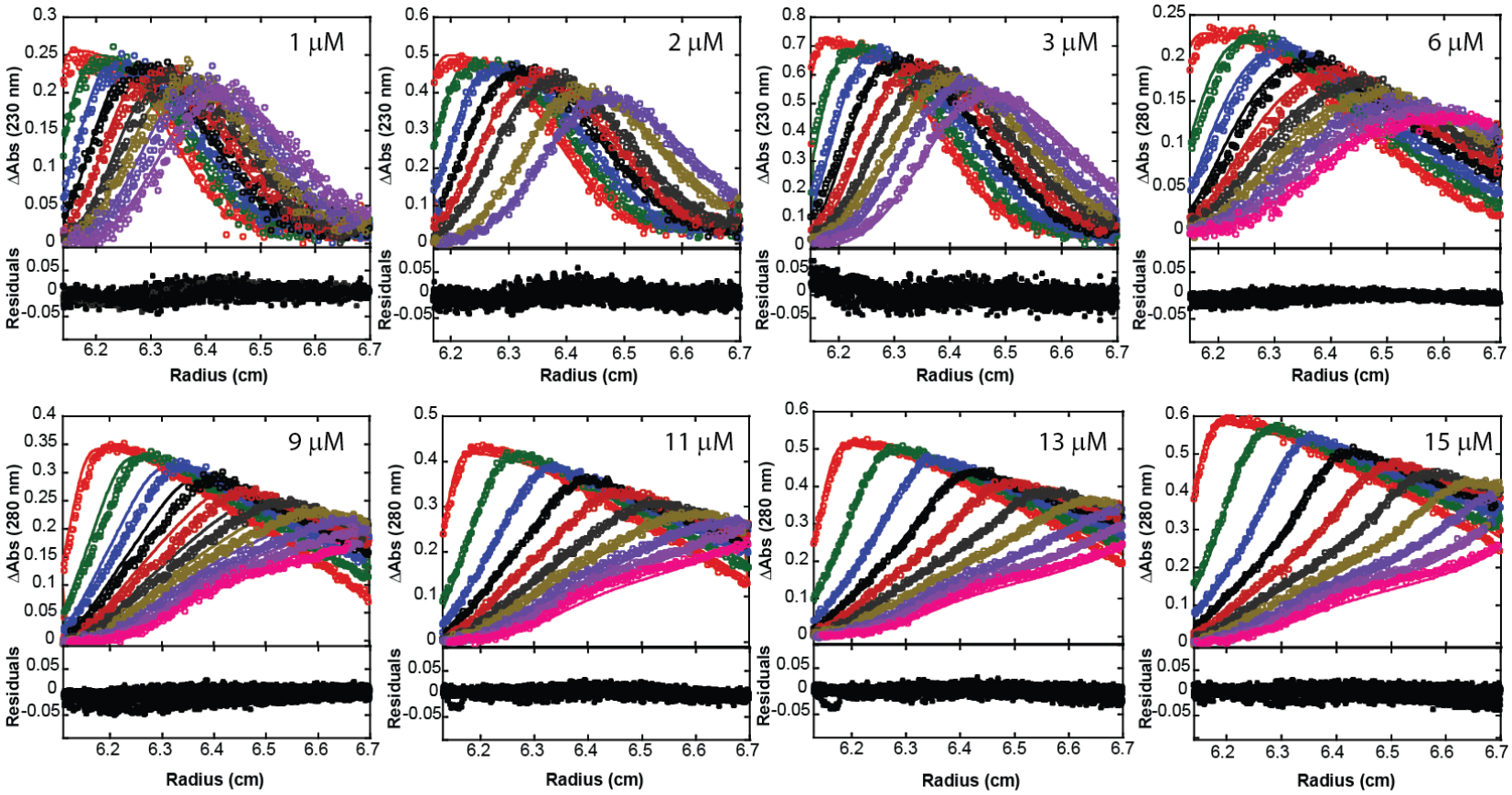

Figure S2: Time difference curves from Sedimentation velocity experiments shown in Figure 5 with the fitting range truncated to 6.7 cm. The same fitting strategy was applied as it was done for the analysis shown in Figure 5. The loading concentrations of ClpB are indicated on the plots. For each ClpB concentration, the ΔAbsorbance time differences curves as a function of radial position are shown in the top panel and the residuals of the fit are shown in the bottom panel. The open circles are raw data and the solid lines are the fits using "1-2-4-6" model.

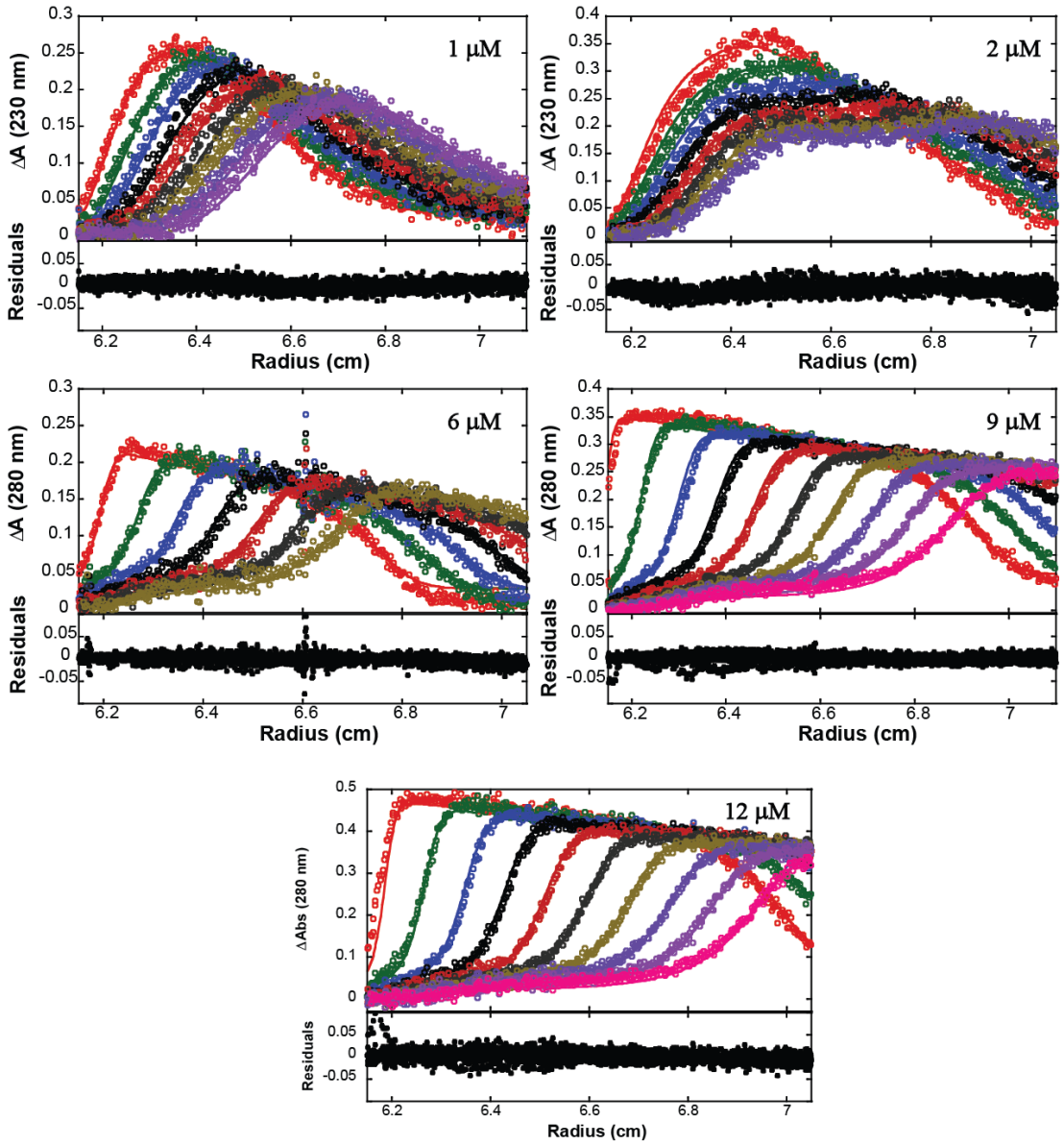

Figure S3: Time difference curves from Sedimentation velocity experiments performed on multiple ClpB concentrations in buffer H with 200 mM NaCl using 1-2-4-6 model. The loading concentrations of ClpB are indicated on the plots. For each ClpB concentration, the ΔAbsorbance time difference curves as a function of radial position are shown in the top panel and the residuals of the fit are shown in the bottom panel. The open circles are raw data and the solid lines are the fits.

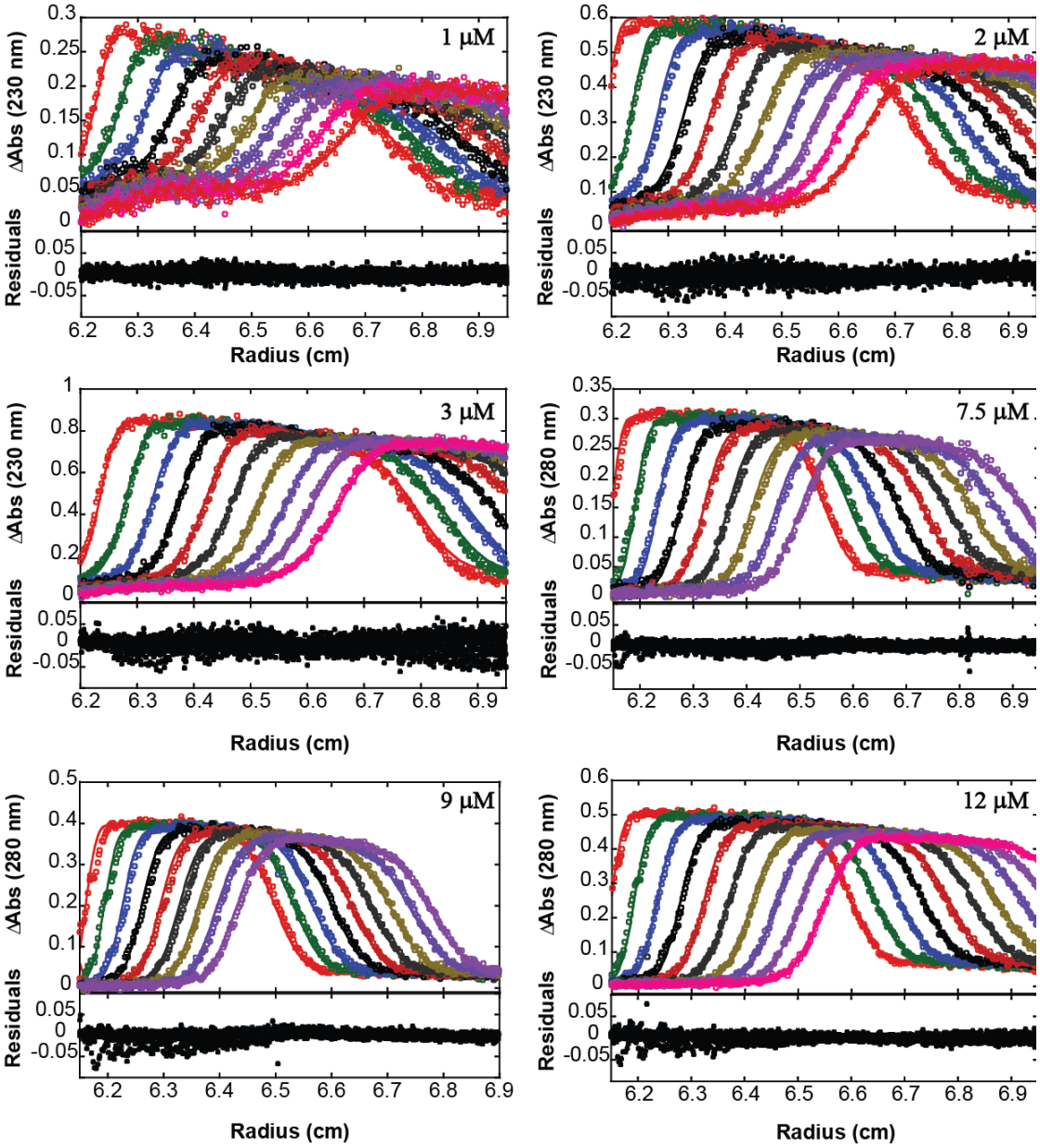

Figure S4: Time difference curves from Sedimentation velocity experiments performed on multiple ClpB concentrations in buffer H with 100 mM NaCl using 1-2-4-6 model. The loading concentrations of ClpB are indicated on the plots. For each ClpB concentration, the ΔAbsorbance time difference curves as a function of radial position are shown in the top panel and the residuals of the fit are shown in the bottom panel. The open circles are raw data and the solid lines are the fits.

EXAMINATION OF CLPB QUATERNARY STRUCTURE AND LINKAGE TO NUCLEOTIDE BINDING

by

JIABEI LIN and AARON LUCIUS

Submitted to *Biophysical Journal*

Format adapted for dissertation

Abstract

E. coli ClpB is a molecular chaperone with the unique ability to catalyze protein disaggregation in collaboration with the KJE system of chaperones. Like many AAA+ molecular motors, ClpB assembles into hexameric rings and this reaction is thermodynamically linked to nucleotide binding. Here we show that ClpB exists in a dynamic equilibrium of monomers, dimers, tetramers, and hexamers in the presence of both limiting and excess ATPγS. We find that ClpB monomer is only able to bind one nucleotide whereas all twelve sites in the hexameric ring are bound by nucleotide at saturating concentrations. Interestingly, dimers and tetramers exhibit stoichiometries of ~3 and 7, respectively. This is one fewer than the maximum number of binding sites, suggesting an open conformation for the intermediates based on the need for an adjacent monomer to fully form the binding pocket. We also report the protein-protein interaction constants for dimers, tetramers, and hexamers and their dependencies on nucleotide. These interaction constants make it possible to predict the concentration of hexamers present and able to bind to co-chaperones and polypeptide substrates. Such information is essential for the interpretation of many in vitro studies. Finally, the strategies presented here are broadly applicable to a large number of AAA+ molecular motors that assemble upon nucleotide binding.

Introduction

Escherichia coli ClpB (Caseinolytic peptidase B) is a member of the Hsp100 protein family, a sub family of AAA+ (ATPase Associated with various cellular Activities) superfamily [1]. Hsp100 proteins play important roles in cell survival under stress including assisting protein folding and removing unfolded or misfolded proteins by degradation or remodeling [2-5]. Most family members require nucleotide binding to assemble into their active oligomeric state.

ClpB and its eukaryotic homologue, Hsp104, can disaggregate denatured proteins with the assistance of Dnak/Hsp70 co-chaperone [6-8]. It has been suggested, by others, that DnaK interacts with the M-domain of ClpB to upregulate the chaperone activity of ClpB and that ATPγS can serve as a substitute of ATP for this interaction [9, 10]. Nevertheless, the mechanism of ClpB/Hsp104 catalyzed disaggregation and the role of the DnaK-ClpB protein-protein interaction remains unclear. To begin to elucidate the role of the protein-protein interactions in disaggregation, one not only needs the affinity for DnaK binding to ClpB but also requires a way to predict the concentration of hexamers present and available to be bound by DnaK. This is not trivial because hexamer formation is thermodynamically linked to nucleotide binding and we have shown that the system is polydisperse [11].

An accurate thermodynamic model for the linkage of ClpB assembly to nucleotide binding is essential for the examination of polypeptide binding and specificity. This is evidenced in our recent findings examining the polypeptide binding activity of ClpB. In that work we showed that in order to quantitatively describe binding isotherms

one must include the model for macromolecular assembly [12]. This was accomplished using the results reported in this paper.

Determining the binding affinities and stoichiometries for nucleotide binding to ClpB is further frustrated by the fact that there are two nucleotide binding domains (NBDs) per ClpB monomer and thus twelve NBDs per hexamer [13, 14]. Others have employed a mutagenesis strategy to study the binding affinity of each NBD and its contribution to ClpB assembly and chaperone activity [15-24]. Examination of nucleotide binding to those variants has simplified the examination of nucleotide binding to each individual NBD and much has been learned from that work. However, how the modifications on the primary structure of ClpB interrupt the protein assembly and their subsequent impact on nucleotide binding, ATP hydrolysis, and peptide recognition is not clear.

In this work, we quantitatively determined the energetics and kinetics of ATPγS binding to ClpB and the linkage of nucleotide binding to ClpB assembly. This was done using full length wild type ClpB and ATPγS, where ATPγS serves as a model for ATP. Our work shows that ClpB resides in a monomer-dimer-tetramer-hexamer equilibrium in the presence of ATPγS. We report binding constants for nucleotide binding to each of the oligomers and show that all twelve sites are bound in the hexameric ring of ClpB and all smaller oligomers are subsaturated. Moreover, we report the dependence of the protein-protein interaction equilibrium constants on the nucleotide concentration, which makes it possible to predict the concentration of hexamers and other oligomers present at any total ClpB concentration and total nucleotide concentration. Finally, our results show that ClpB is in rapid subunit exchange at nucleotide concentrations at or below 100 μM

ATPγS. However, at concentrations above 500 μM ATPγS the observed dissociation rate constants are slowed to a range consistent with a non-dissociating system. Thus, nucleotide binding affects both the thermodynamic state of the system and the kinetics.

Materials and Methods

Protein and buffer

The 95 kDa *E. coli* ClpB is purified as previously described [11]. All experimental buffers were prepared with reagent grade chemicals using deionized H_2O produced using the Purelab Ultra Genetic system (EVOQUA Water Technologies). Buffer H200 contains 25 mM HEPES pH 7.5 at 25 °C, 10 mM $MgCl_2$, 2 mM 2-mercaptoethanol, 10% Glycerol (v/v), and 200 mM NaCl.

Sedimentation velocity experiment using interference optics

Interference sedimentation velocity experiments were performed by using a Beckman ProteomeLab XL-I analytical ultracentrifuge. The experiments were carried out by loading 425 μL of the protein and 430 μL of the ATPγS reference sample into a double sector Epoxy charcoal-filled meniscus matching centerpiece. The sample was subjected to an angular velocity of 40,000 rpm. Interference scans were collected every 30 s at 25 °C.

Since interference is sensitive to the ATPγS concentration, a strategy to match the concentration of ATPγS in both the sample and reference sectors was developed. This was accomplished as follows. A stock concentration of ATPγS was prepared and the concentration was determined by measuring the absorbance. From this stock, identical volumes were put into two separate Eppendorf tubes. The volume in each Eppendorf tube

was verified by mass and the mass difference between these two aliquots was found to be less than 1 mg (≈ 1 μL). A sample of ClpB with the concentration determined from absorbance measurements was added into one of the two Eppendorf tubes to make the protein sample containing ATPγS. ClpB dialysate (H200) of the identical volume as the protein was added into the other Eppendorf tube to make the ATPγS reference. Using this strategy, the ATPγS was diluted by the identical volume and thus the two samples contained identical ATPγS concentrations. The dilution was further verified by mass.

The sample and reference solution for sedimentation velocity experiments were incubated at 25 °C for two hours before the first interference scan was taken. Sedimentation velocity experiments on 2, 3, and 6 μM ClpB in the presence of a fixed ATPγS concentration were performed in triplicate. Each experimental replicate was collected with freshly dialyzed ClpB samples no older than 3 days.

Sedimentation velocity experiment using absorbance optics

Sedimentation velocity experiments using absorbance optics were carried out by loading a 380 μL sample of protein with ATPγS and 400 μL of protein dialysate (H200), as the reference, into a double sector centerpiece. Absorbance scans were taken at 230 nm or 260 nm as indicated in the text. The samples were mixed and incubated for two hours before the first absorbance scan was collected. The sample was subjected to an angular velocity of 40,000 rpm and absorbance scans were collected every 4 minutes at 25 °C.

Analysis of sedimentation velocity data

REDATE (V. 0.1.7) was used to regenerate the sedimentation velocity data with corrected “elapse time” using the algorithm described by Zhao et al. [41]. *c(s)* analysis

using SedFit [42] and time difference curve analysis using SedAnal [43] were used to fit sedimentation velocity data as previous described [11]

The sedimentation coefficients used in SedAnal are not standardized to water solvent at 20 °C, $s_{20,w}$, and their values are constrained in the fits unless otherwise indicated. The sedimentation coefficient for ClpB monomer was obtained experimentally. A peak with sedimentation coefficient of s = (3.13 ± 0.06) S was observed in multiple *c(s)* distributions when ATPγS is in large excess, where the error is the standard deviation of the mean value of three independent observations. The sedimentation coefficients for ClpB dimer (5.6 S), trimer (7.4 S) and tetramer (8.9 S) were calculated using WinHydroPro [44] as previously described [11]. In the presence of 1 mM ATPγS the hexameric state dominates the population. Thus, the ClpB hexamer sedimentation coefficient was obtained from global fitting of sedimentation velocity experiments for 2, 3, 6, and 10 μM ClpB in the presence of 1 mM ATPγS. The resulting s value for the hexamer is (11.11 ± 0.06) S. The error was calculated from the standard deviation of three experimental replicates. This value was then used and constrained as the hexamer sedimentation coefficient for all subsequent fits.

Sedimentation velocity data were analyzed between the meniscus plus 0.01 cm and 6.7 cm to minimize the effect of gradients of glycerol and ATPγS that can form during sedimentation. This range was chosen because the isosbestic point (radial position where all scans have the same absorbance/fringes) is shown to be at ~6.7 cm for both glycerol [11] and ATPγS (see Fig. S1). This indicates that the dominant effect of the formed gradient will occur at radial positions above 6.7 cm since ClpB (95 kDa) sediments much faster than either ATPγS (547 Da) or glycerol (92 Da).

Global analysis of the time difference curves from sedimentation velocity experiments performed at various ClpB concentrations for a fixed ATPγS concentration

Experiments performed with 2, 3, 6, and 10 μM ClpB in the presence of one fixed [*ATPγS*] (≥ 100 μM) were globally analyzed using SedAnal and the 1-2-4-6 model described in Eqs. 1- 3,

$$2\{B_1\} \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} \{B_2\} \qquad 1$$

$$2\{B_2\} \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} \{B_4\} \qquad 2$$

$$\{B_2\} + \{B_4\} \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} \{B_6\} \qquad 3$$

where "$\{B_n\}$" represents ClpB n-mers at all ATPγS ligation states (including unligated). The stepwise equilibrium constants K_n for the formation of ClpB n-mer is related to the rate constants using Eq. 4.

$$K_n = \frac{k_{f,n}}{k_{r,\mathrm{n}}} \qquad 4$$

K_n , $k_{r,n}$ and the loading concentrations of ClpB were set to float in the global analysis. The details on setting up the analysis can be found in [11] and [26].

Analysis of the difference curves from sedimentation velocity data for a single ClpB concentration when the [ATPγS] is not in large excess over the [ClpB]

At a total [ATPγS] below 100 μM the free concentration, $[ATP\gamma S]_f$, cannot be assumed to be equal to the [ATPγS]t, $[ATP\gamma S]_t$. Under these conditions, each sedimentation velocity experiment performed at a total [ClpB] = 2, 3, or 6 μM at a fixed [ATPγS]t< 100 μM were individually fit to the 1-2-4-6 model. The free [ATPγS] was

experimentally determined for each sedimentation velocity experiment as described in Results.

Thermodynamic reversibility tests for ATPγS linked ClpB assembly

The ATPγS-linked ClpB assembly reaction was tested for thermodynamic reversibility and path independence. To address this, ATPγS was added to a large volume of ClpB to a final concentration of 3 μM ClpB and 500 μM ATPγS. The sample was incubated for two hours at 25 °C. It was then diluted with a stock of 3 μM ClpB to a final concentration of 200 μM and 100 μM ATPγS. All volumes were confirmed by mass. The diluted samples were then incubated for another four hours at 25 °C before the first sedimentation velocity interference scan was collected. The sedimentation velocity experiments were performed as described above in the "Sedimentation velocity experiment using interference optics" subsection. The data were analyzed as described in "Analysis of the difference curves from sedimentation velocity data for a single ClpB concentration" section.

Results:

ClpB exhibits dynamic assembly in the presence of nucleotide

Our previous studies have shown that ClpB resides in a monomer, dimer, tetramer, and hexamer equilibria in the absence of nucleotide [11]. To examine the impact of ATPγS binding on ClpB assembly, we first performed sedimentation velocity experiments at 2 μM ClpB in the presence of (A) 0, (B) 200, and (C) 1000 μM ATPγS as described in Material and Methods. The sedimentation boundaries were subjected to *c(s)* analyses and the results are shown in Fig. 1 with the [*ATPγS*] indicated in the plots.

Notably, in the absence of nucleotide (Fig. 1A), multiple *c(s)* peaks were observed, which we have reported to be the reaction boundaries for monomers, dimers, tetramers, and hexamers of ClpB [11]. When 200 μM ATPγS was added (Fig. 1B), the overall weighted-average sedimentation coefficient, $\overline{s}_{20,w}$, shifts to larger values, from ~8.7 S in the absence of ATPγS to ~10.2 S. This indicates that the presence of ATPγS shifts the equilibrium toward higher order ClpB oligomers. Consistently, at 1 mM ATPγS, where the [*ATPγS*] is considered to be in large excess over the [*ClpB*], there is still a distribution of multiple *c(s)* peaks but one peak at ~ 15.5 S is predominant (Fig. 1C). Under these conditions the weighted average sedimentation coefficient, $\overline{s}_{20,w}$ is increased to ~13.6 S. Clearly, the presence of excess ATPγS shifts the equilibrium toward higher order oligomers of ClpB. As shown in Fig. 1 panel B and C, the assembly of ClpB in the presence of ATPγS is likely a dynamic process as observed for the assembly in the absence of nucleotide [11]. To examine the linkage of ATPγS to ClpB assembly, we sought to determine the dependence of dimerization, tetramerization and hexamerization on the concentrations of ATPγS. To this end we examined ClpB assembly at multiple ClpB concentrations and multiple fixed ATPγS concentrations.

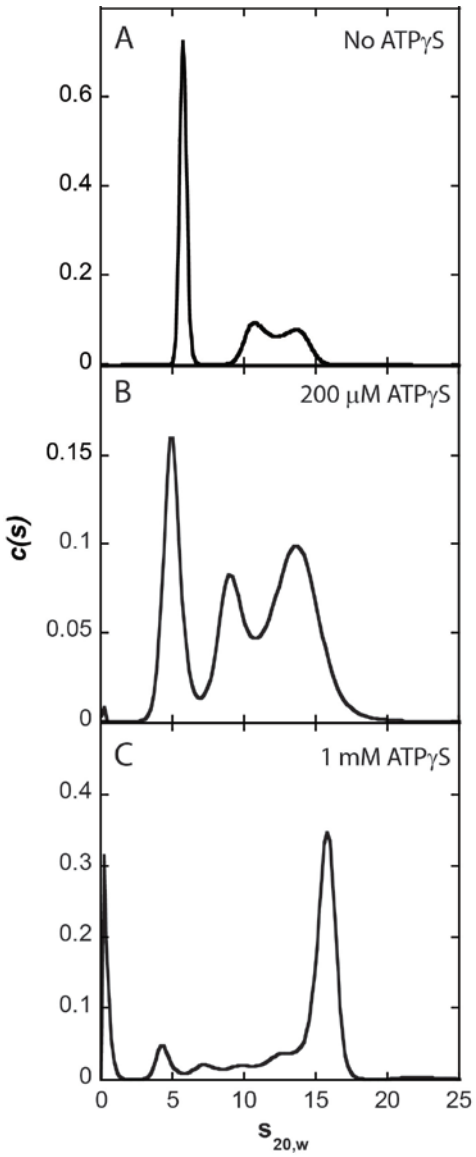

Figure 1: *c(s)* analysis of the sedimentation velocity data for ClpB assembly in the presence of various concentrations of ATPγS. The *c(s)* analysis results of 2 μM ClpB in the presence of (A) no ATPγS, (B) 200 μM ATPγS, and (C) 1 mM ATPγS.

ClpB resides in a monomer-dimer-tetramer-hexamer equilibrium in the presence of ATPγS

Sedimentation velocity experiments at 2, 3, 6, and 10 μM ClpB at fixed $[ATP\gamma S]_t$ were performed to examine the apparent equilibrium constants, $L_{n,app}$, for the oligomerization of ClpB n-mer. $L_{n,app}$ is defined by Eq. 5,

$$L_{n,app} = L_{n,0} \cdot \frac{P_n}{(P_1)^n} \qquad 5$$

where $L_{n,0}$ is the stoichiometric equilibrium constant for ClpB n-merization in the absence of nucleotide, P_n is the partition function for ATPγS binding to ClpB n-mer, and P_1 is the partition function for ATPγS binding to ClpB monomer. The partition function, P_n, is a function of the free ATPγS concentration ($[ATP\gamma S]_f$) and it has no [*ClpB*] dependence [25].

Sedimentation velocity experiments with 2, 3, 6, and 10 μM ClpB in the presence of 100, 200, 500 μM, or 1 mM ATPγS concentrations were performed using interference optics as described in Materials and Methods. The data collected from three experimental replicates were analyzed independently from the other replicates. Experimentally, when [*ATPγS*] is in large excess over the [*ClpB*], we consider $[ATP\gamma S]_f$ to be approximately equal to the total loading concentration. In this study we consider [*ATPγS*] of 100 μM, 200 μM, 500 μM, and 1 mM to be in large excess over [*ClpB*]. With this assumption in mind, the sedimentation velocity data at one fixed total [*ATPγS*] (e.g. 200 μM) with various ClpB concentrations were analyzed globally to obtain the stepwise assembly equilibrium constants, K_2, K_4, and K_6 defined in Eqs. 1 - 4 shown in Materials and Methods. Fig. S2 shows one example of the difference curves from a global analysis

performed using the monomer, dimer, tetramer, hexamer model (1-2-4-6 model) for 2, 3, 6, and 10 μM ClpB in the presence of a $[ATP\gamma S]_t$ of 200 μM.

The resulting stepwise equilibrium constants are converted to stoichiometric equilibrium constants, $L_{2,app}$, $L_{4,app}$, and $L_{6,app}$ using Eqs. 6 - 8. The averages and standard deviations from the three experimental replicates are reported in Table 1

$$L_{2,\text{app}} = \frac{\{B_2\}}{\{B\}^2} = K_2 \qquad 6$$

$$L_{4,\text{app}} = \frac{\{B_4\}}{\{B\}^4} = K_2^2 \cdot K_4 \qquad 7$$

$$L_{6,\text{app}} = \frac{\{B_6\}}{\{B\}^6} = K_2^3 \cdot K_4 \cdot K_6 \qquad 8$$

The rate constant for dissociation of each oligomer, $k_{r,n}$, as defined by Eqs. 1 - 3 was floated as a fitting parameter for each $[ATP\gamma S]_f$. The average from the three experimental replicates together with the standard deviation are shown in Table 2.

Table 1: The assembly equilibrium constants as a function of $[ATP\gamma S]_f$.

$[ATP\gamma S]_f$ (μM)	$\log(L_{2,app})$	$\log(L_{4,app})$	$\log(L_{6,app})$
0	5.24 ± 0.01	16.76 ± 0.03	29.30 ± 0.05
11.6 ± 0.5	4.9 ± 0.2	16.6 ± 0.1	27.6 ± 0.4
17.2 ± 0.3	5.1 ± 0.2	16.4 ± 0.2	27.5 ± 0.4
17.6± 0.2	5.2 ± 0.3	15.7 ± 0.7	27.7 ± 0.7
41.1 ± 0.2	5.2 ± 0.2	17.3 ± 0.3	28.4 ± 0.5
44.6 ± 0.6	4.5 ± 0.4	16.7 ± 0.3	27.8 ± 0.2
47 ± 1	4.6 ± 0.1	16.6 ± 0.2	27.7 ± 0.8
100	4.7 ± 0.4	17.3 ± 0.2	29.0 ± 0.2
200	4.7 ± 0.7	18.2 ± 0.1	30.2 ± 0.2
500	5.9 ± 0.2	19.5 ± 0.3	32.6 ± 0.5
1000	6.0 ± 0.3	20.1 ± 0.2	34.0 ± 0.2

Note: $L_{2,app}$ has units of M^{-1}, $L_{4,app}$ has units of M^{-3} and $L_{6,app}$ has units of M^{-5}.

At 100 μM ATPγS, the dissociation rate constants for dimer, tetramer, and hexamer are larger than 0.01 s^{-1}, which indicates that those oligomers undergo instantaneous dissociation on the time scale of sedimentation. As previously discussed, when the rate constants exceed ~ 0.01 s^{-1} the value is poorly constrained in the fitting and the corresponding uncertainty is large [11, 26].

Notably, the rate constant for ClpB hexamer dissociation appears to exhibit an ATPγS concentration dependence. As the ATPγS concentration is increased from 100 μM to 200 μM, the dissociation rate constants decreased to $(3 \pm 1) \times 10^{-3}$ s^{-1}. The dissociation rate constant is well constrained with a reasonably small standard deviation

Table 2: The dissociation rate constantans as a function of ATPγS.

$[ATP\gamma S]_f$ (μM)	k_{r2} **(s^{-1})**	k_{r4} **(s^{-1})**	k_{r6} **(s^{-1})**
11.6 ± 0.5	3.7 ± 3.2	0.7 ± 0.7	3 ± 1
17.2 ± 0.3	9.9 ±7.9	3.3 ± 4.7	4.9 ± 3.3
17.6± 0.2	0.48 ± 0.67	1.2 ± 0.9	1.3 ±1.2
41.1 ± 0.2	1.4 ±1.9	$(5 \pm 3) \times 10^{-4}$	5.3 ± 5.4
44.6 ± 0.6	3 ± 2	0.5 ± 0.6	4 ± 3
47 ± 1	9 ± 7	2.3 ± 1.7	2.6 ±0.9
100	**0.5 ± 0.6**	**0.2 ± 0.1**	**1.2 ± 1.4**
200	**2.4 ±3.3**	**$(8 \pm 7) \times 10^{-2}$**	**$(3 \pm 1) \times 10^{-3}$**
500	**4 ± 3**	**$(3.4 \pm 2.9) \times 10^{-5}$**	**$(2 \pm 1) \times 10^{-4}$**
1000	**$(6 \pm 3) \times 10^{-2}$**	**$(1 \pm 0.7) \times 10^{-5}$**	**$(5 \pm 1) \times 10^{-5}$**

because rate constants in a range between ~10^{-2} – 10^{-5} s^{-1} are in the measurable range for sedimentation velocity experiments [26]. As the concentration of ATPγS is further increased, the dissociation rate constant for ClpB hexamer continues to decrease. This observation suggests that as the extent of binding of nucleotide to ClpB hexamer is increased, so is the kinetic stability.

Measurement of the free ATPγS concentration in solution

For the experiments performed with 2, 3, or 6 μM ClpB in the presence of 20 μM or 50 μM ATPγS, respectively, the [*ATPγS*] are relatively low compared to the [*ClpB*]. Under these conditions $[ATP\gamma S]_t$ cannot be considered to be equal to $[ATP\gamma S]_f$. The $[ATP\gamma S]_f$ in solution will be different at each $[ClpB]_t$. Consequently, it is not possible to combine the sedimentation velocity data collected with 2, 3, and 6 μM ClpB in the presence of a total loading concentration of 20 or 50 μM ATPγS and subject the data to global NLLS analysis. Even though all three ClpB concentrations are examined at the same $[ATP\gamma S]_t$ each sample will contain a different $[ATP\gamma S]_f$ and thus a different ATPγS chemical potential.

To overcome this problem, we developed a sedimentation velocity strategy to determine the $[ATP\gamma S]_f$ at each $[ClpB]_t$ concentration for experiments performed with a $[ATP\gamma S]_t$ below 100 μM. We performed sedimentation velocity experiments as described above, but monitored absorbance at 260 nm. Fig. 2 panel A shows an example of absorbance scans from a sedimentation velocity experiment performed with 6 μM ClpB in the presence of 50 μM $[ATP\gamma S]_t$. The scans were collected as described in materials and methods. At the beginning of the run, the total absorbance at 260 nm is a sum of the absorbance of three components: (1) ClpB absorbance at 260 nm, $A_{ClpB,260}$, ATPγS absorbance at 260 nm, which is composed of (2) ATPγS that is bound to ClpB, $A_{ATP\gamma S,bound}$, and (3) free ATPγS that is not bound by ClpB, $A_{ATP\gamma S,free}$.

When the centrifugal force is applied, ClpB and ClpB bound by ATPγS will sediment faster than the free ATPγS. Thus, the two components will be separated. The

sedimentation boundaries shown in Fig. 2A were subjected to *c(s)* analysis. The observed *c(s)* from 3.8 S to 15.2 S (as shown in the black box of panel C) is considered to represent the absorbance at 260 nm for free ClpB and ATPγS bound by ClpB. Thus, the integrated area under the *c(s)* curve from 3.8 S to 15.2 S represents the contribution to the total absorbance signal of ClpB at 260 nm and ATPγS bound to ClpB. This contribution, ($A_{ClpB,\,260}$ +$A_{ATP\gamma S,bound}$), can be subtracted from the total absorbance at 260 nm, $A_{T,260}$, to yield the absorbance of the free ATPγS, $A_{ATP\gamma S,free}$. From this, $[ATP\gamma S]_f$ is determined at each $[ClpB]_t$ and $[ATP\gamma S]_t$.

These sedimentation velocity experiments were performed three times with 2, 3, and 6 μM ClpB at 20 and 50 μM total ATPγS. The average and standard deviation for each measured $[ATP\gamma S]_f$ is reported in Table 1. The resulting $L_{2,app}$, $L_{4,app}$, and $L_{6,app}$ are reported in Table 1 for the corresponding $[ATP\gamma S]_f$ determined as described in Materials and Methods.

Determination of the binding density of ClpB oligomers

To investigate the linkage of nucleotide binding to ClpB assembly, we first examined the binding stoichiometry of ClpB oligomers to ATPγS using a Wyman plot [27]. This strategy has been previously discussed by Timasheff and others [27]. By taking the natural log of Eq. 5 we arrive at Eq. 9.

$$ln\left(L_{\mathrm{n,app}}\right) = ln\left(L_{n,0}\right) + ln\left(P_n\right) - n \cdot ln\left(P_1\right) \qquad 9$$

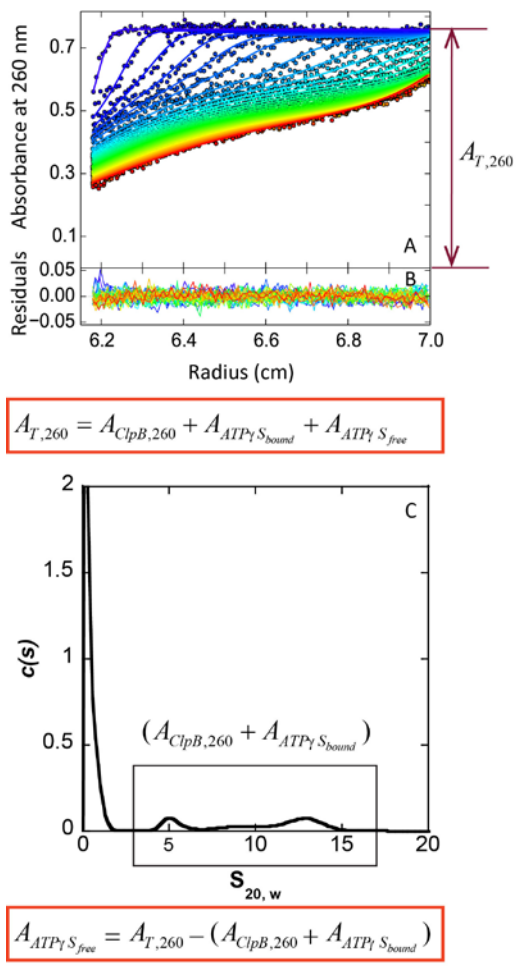

Figure 2: Sedimentation velocity experimental measurements of *[ATPγS]$_f$* for ClpB assembly in the presence of 20 and 50 μM *[ATPγS]$_t$*. (A) Sedimentation velocity experiment absorbance scans at 260 nm at 25 °C. The filled circles are data and solid lines are fits from *c(s)* analysis. (B) Residuals of *c(s)* analysis. (C) *c(s)* analysis of sedimentation velocity data shown in panel A.

Taking the first derivative of Eq. 9 with respect to $ln[ATP\gamma S]_f$ yields Eq. 10.

$$\frac{\partial ln\left(L_{\mathrm{n,app}}\right)}{\partial ln[\mathrm{ATP}\,\gamma\,\mathrm{S}]_f} = \frac{\partial ln\left(P_n\right)}{\partial ln[\mathrm{ATP}\,\gamma\,\mathrm{S}]_f} - n\frac{\partial ln\left(P_1\right)}{\partial ln[\mathrm{ATP}\,\gamma\,\mathrm{S}]_f} \qquad 10$$

The derivative of the natural log of a partition function for binding with respect to the natural log of the ligand concentration is the extent of binding, $\overline{X}$ = [ligand bound]/[Macromolecule]$_t$ [25]. Thus, the two terms on the right hand side of Eq. 10

represent the extent of binding to the n-mer ($\bar{X}_n$ = $[ATP\gamma S]_b/[ClpB_n]_t$) and the extent of binding to the monomer ($\bar{X}_1$ = $[ATP\gamma S]_b/[ClpB_1]_t$). Consequently, Eq. 10 tells us that the slope of the natural log of $L_{n,app}$ vs. the natural log of the $[ATP\gamma S]_f$ is equal to the difference between the extent of binding of nucleotide to the n-mer and n times the extent of binding to the monomer as given by Eq. 11.

$$\frac{\partial ln\left(L_{\mathrm{n,app}}\right)}{\partial ln[\mathrm{ATP}\,\gamma\,\mathrm{S}]_f} = \bar{X}_n - n\bar{X}_1 \qquad 11$$

With Eq. 11 in mind, the data in Table 1 were plotted as $ln(L_{n,app})$ vs. $ln([ATP\gamma S]_f)$. The linear region of $ln(L_{2,app})$, $ln(L_{4,app})$ and $ln(L_{6,app})$ as a function of $ln[ATP\gamma S]_f$ were subjected to NLLS analysis using an equation for a line as shown in Fig. 3. Notably, the linear regions of the Wyman plots were found to be at the high $[ATP\gamma S]$, which indicates that ClpB oligomers are likely approaching their maximum binding stoichiometry for ATPγS. Therefore, $\bar{X}_n$ can be a representation of the maximum binding stoichiometry of ClpB n-mer [27]. With the slopes determined from Fig. 3, Eq.11 can be written as:

$$\bar{X}_2 - 2\bar{X}_1 = 1.5 \pm 0.4 \qquad 12$$

$$\bar{X}_4 - 4\bar{X}_1 = 2.8 \pm 0.2 \qquad 13$$

$$\bar{X}_6 - 6\bar{X}_1 = 5.3 \pm 0.2 \qquad 14$$

Eqs. 12 - 14 represent a system of three linear equations with four unknowns. Thus, in order to solve Eqs. 12 - 14 for the binding stoichiometry to the dimer, tetramer, and hexamer we need to make assumptions about the binding of nucleotide to the

monomer. Structurally we know that there are two nucleotide binding sites per ClpB monomer; therefore, we must have $\bar{X}_1 \leq 2$.

Solving the system of linear equations given by Eq. 12 - 14 assuming that $\bar{X}_1$ is either 0, 1, or 2 yields the extent of binding for the other oligomers given in Table 3. If one assumes that both nucleotide binding sites are bound in the monomer, then Table 3 shows that the extent of binding to the dimers, tetramers, and hexamers is 5.5, 10.8, and 17.3, respectively. However, the dimer, tetramer, and hexamers have only 4, 8 and 12 binding sites, respectively. Because this analysis predicts a stoichiometry larger than the potential number of binding sites when both sites in the monomer are considered to be bound, we conclude that $\bar{X}_1$ cannot be equal to 2. Consistently, structural studies show that only one of the nucleotide binding sites in the monomer of ClpB is fully formed and the other is fully formed with an arginine finger from an adjacent monomer upon oligomerization [28].

Since a doubly ligated monomer is not likely, if we assume that the monomer does not bind or that the monomer binds one nucleotide, realistic numbers for the extent of binding to each oligomer are revealed, see Table 3. Thus, from this analysis we are left asking the question; does monomer bind only one nucleotide or none at all?

Table 3: Prediction for the extent of binding of ClpB dimer $\bar{X}_2$, tetramer $\bar{X}_4$ and hexamer $\bar{X}_6$ using the slopes determined from the Wyman Plots.

	$\bar{X}_1 = 0$	$\bar{X}_1 = 1$	$\bar{X}_1 = 2$
$\bar{X}_2$	**1.5 ± 0.4**	**3.5 ± 0.4**	**5.5 ± 0.4**
$\bar{X}_4$	**2.8 ± 0.2**	**6.8 ± 0.2**	**10.8 ± 0.2**
$\bar{X}_6$	**5.3 ± 0.2**	**11.3 ± 0.2**	**17.3 ± 0.2**

$\bar{X}_1$ is the extent of binding of ClpB monomer.

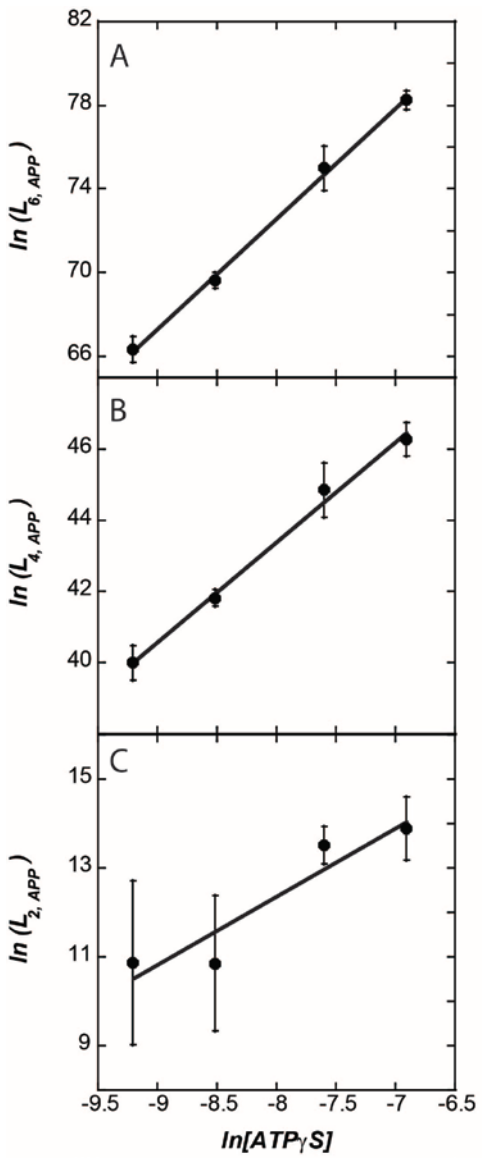

Figure 3: (A) *ln($L_{6,app}$)*, (B) *ln($L_{4,app}$)* and (C) *ln($L_{2,app}$)* as a function of *ln*[*ATPγS*]. The data are shown in black filled circles. A linear fit was applied to each plot. The slope and intercept for each line are: (A) Slope = 5.3 ± 0.2, intercept = 115 ± 1, (B) Slope = 2.8 ± 0.2, intercept = 66 ± 1, and (C) Slope = 1.5 ± 0.4, intercept = 25 ± 3.

Global analysis of $L_{n,app}$ data as a function of [ATPγS]$_f$

To resolve the question and to further investigate the linkage between ATPγS binding and ClpB assembly, the equilibrium constants in Table 1 are plotted in Fig. 4. In

all cases, the equilibrium constant is observed to decrease at low ATPγS concentration before increasing as the ATPγS concentration is increased.

Eq. 5 shows that $L_{n,app}$ is a function of the partition functions for nucleotide binding to the n-mer and the monomer. Thus, to analyze the plots shown in Fig. 4 A – C one must define a partition function for nucleotide binding to each oligomer. The simplest way to describe nucleotide binding to each of the sites in a ClpB oligomer is to assume that they are identical and independent, i.e. no cooperativity. Thus, the n-independent and identical sites model was used to examine the data. Eq. 15 is the partition function for the n-independent and identical sites model.

$$p_n = \left(1 + \kappa_n \cdot [\mathrm{ATP}\,\gamma S]_f\right)^{m_n} \qquad 15$$

Where κ_n represents the average step-wise equilibrium constant for binding to an n-mer and m_n is the number of binding sites in an n-mer [25].

The partition function for nucleotide binding to each oligomer, given by Eq.15, is substituted into Eq. 5 to yield Eq.16 - 18.

$$L_{2,app} = L_{2,0} \cdot \frac{\left(1 + \kappa_2 \cdot [\mathrm{ATP}\,\gamma S]_f\right)^{m_2}}{\left(\left(1 + \kappa_1 \cdot [\mathrm{ATP}\,\gamma S]_f\right)^{m_1}\right)^2} \qquad 16$$

$$L_{4,app} = L_{4,0} \cdot \frac{\left(1 + \kappa_4 \cdot [\mathrm{ATP}\,\gamma S]_f\right)^{m_4}}{\left(\left(1 + \kappa_1 \cdot [\mathrm{ATP}\,\gamma S]_f\right)^{m_1}\right)^4} \qquad 17$$

$$L_{6,app} = L_{6,0} \cdot \frac{\left(1 + \kappa_6 \cdot [\mathrm{ATP}\,\gamma S]_f\right)^{m_6}}{\left(\left(1 + \kappa_1 \cdot [\mathrm{ATP}\,\gamma S]_f\right)^{m_1}\right)^6} \qquad 18$$

Where, $L_{2,0}$, $L_{4,0}$, and $L_{6,0}$ represent dimerization, tetramerization, and hexamerization in the absence of nucleotide and have been previously reported by us in these solution conditions [11], κ_1, κ_2, κ_4, and κ_6 are the average step-wise equilibrium constants for nucleotide binding, m_1, m_2, m_4, and m_6 are the stoichiometries of binding to monomers, dimers, tetramers, and hexamers, respectively.

The data in Fig. 4 were subjected to global NLLS analysis using Eq.16 – 18. In this analysis, κ_1 and m_1 are global parameters, $L_{2,0}$, $L_{4,0}$, and $L_{6,0}$ are constrained to our previously reported values, κ_2, κ_4, κ_6, and m_2, m_4, and m_6 are local parameters and are allowed to float in the analysis. As discussed above, we have ruled out the possibility of two nucleotides binding to the monomer, $m_1 \neq 2$. Thus, we are left with $m_1 = 0$ or 1. However, if $m_1 = 0$ then the denominator in Eq. 16 – 18 collapses to 1 and the system of equations will not describe the decreasing equilibrium constant exhibited in Fig. 4 A – B at low ATPγS concentrations. As expected, when the data were subjected to NLLS analysis using Eq. 16 – 18 with $m_1 = 0$ the fit always has positive slope and the data are not well described (see Fig. S3).

When $m_1 = 1$, both κ_n and m_n floated to the values shown in Table 4 and the fits are able to describe both the descending and ascending region in the isotherms. The resulting parameters are presented in Table 4. The R-squared value for this fit is 0.99980, which indicates that the n-independent and identical sites model describes the data well (shown as solid lines in Fig. 4). The next level of complexity would be to include cooperativity between binding sites. However, the data do not suggest that any cooperativity is present, i.e. the isotherms are not sufficiently steep to suggest

cooperativity. Moreover, because of the quality of the fit to the n-independent and identical sites model, there is no justification for fitting the data to a more complex model.

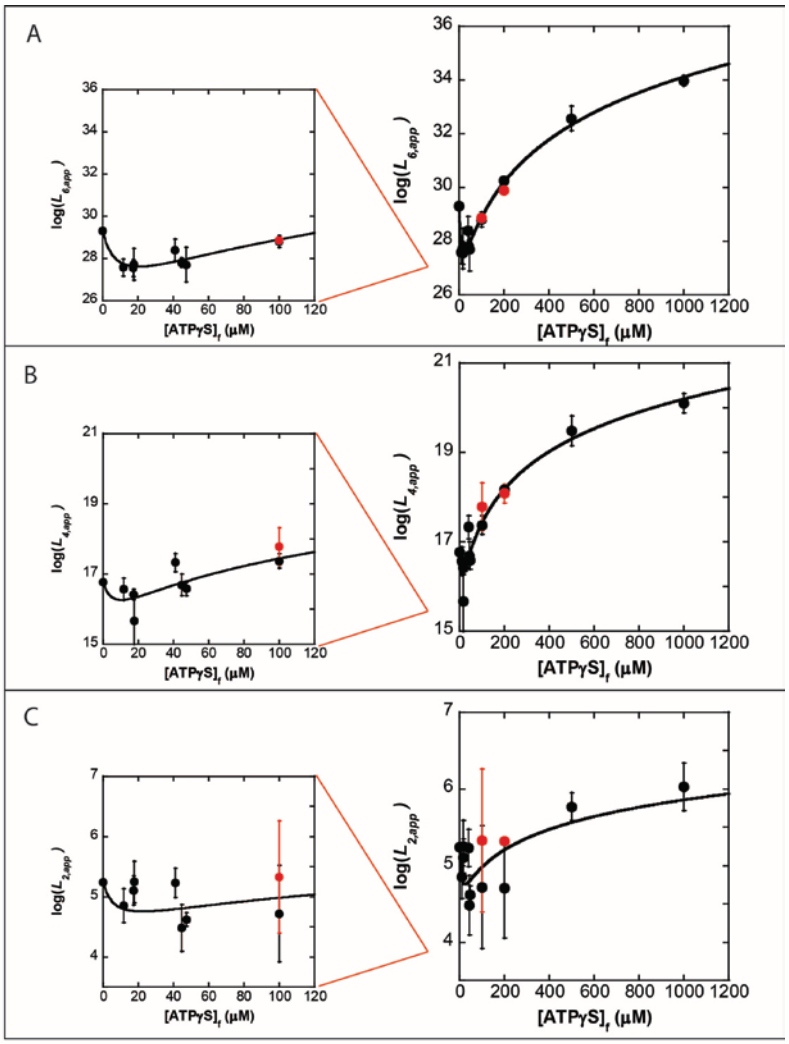

Figure 4: Global analysis of the (A) $L_{6,app}$, (B) $L_{4,app}$ and (C) $L_{2,app}$ as a function of $[ATP\gamma S]_f$ using n-independent identical model. Left panel is a zoom-in of the 0 to 120 μM $[ATP\gamma S]_f$ region of the right panel. The black filled circles are SedAnal fitting results of the sedimentation velocity data that are presented in Table 1. The solid lines are the fits from global analysis of the data shown in panel A, B, and C using n-independent identical model as described in materials and methods.

Table 4: ATPγS binding constants and binding stoichiometry determined from global analysis of $logL_{n,app}$ as a function of $[ATP\gamma S]_f$.

Floating parameters	With m_n floating R^2=0.99980	With m_n constrained R^2=0.99979
κ_1 (M^{-1})	$(2.1 \pm 0.5) \times 10^5$	$(2.4 \pm 0.4) \times 10^5$
κ_2 (M^{-1})	$(5 \pm 2) \times 10^4$	$(6 \pm 1) \times 10^4$
κ_4 (M^{-1})	$(6 \pm 2) \times 10^4$	$(7.0 \pm 0.8) \times 10^4$
κ_6 (M^{-1})	$(3.1 \pm 0.8) \times 10^4$	$(3.7 \pm 0.4) \times 10^4$
m_1	1*	1*
m_2	3.1 ± 0.3	3*
m_4	7.1 ± 0.3	7*
m_6	12.4 ± 0.6	12*

The values with the "*" notation are constrained in that fit.

Notably, when m_1 = 1, the resulting binding stoichiometry values, m_n, are consistent with the values determined using Wyman plots shown in Fig. 3 (compare m_n in Table 4 to values in Table 3 with $\bar{X}_1$ = 1). Moreover, the average binding constant for ATPγS binding to ClpB monomer is approximately an order of magnitude greater than that to ClpB hexamer. This indicates that the NBD that binds to ATPγS in the monomer has stronger ATPγS binding affinity than the average of the two NBDs in each monomer that resides in the hexamer.

As the numbers of binding sites for each oligomer should be integers, the resulting m_n were constrained in the analysis of the data in Fig. 5 to the closest integers, as m_1 = 1, m_2 = 3, m_4 = 7 and m_6 = 12. The R-squared value for the fit changed from 0.99980 to 0.99979 and the equilibrium constants are within error. This indicates no significant differences between the two fits (see Table 4).

ClpB monomer binds ATPγS

One might ask: is the observed decreasing self-association equilibrium constant between zero and ~ 11.6 μM $[ATP\gamma S]_f$ statistically significant? As seen in Table 1 and Fig. 4 the values of the equilibrium constants between zero and ~11.6 μM are not within error. However, the observation that the equilibrium constants decrease is surprising. Nevertheless, direct experimental evidence was acquired that supports this observation. Fig. 5 shows a *c(s)* plot for 2 μM ClpB in the presence and absence of 10 μM $[ATP\gamma S]_t$. Strikingly, in the absence of nucleotide, we observe a distribution of monomers and higher order oligomers. However, upon addition of 10 μM ATPγS the *c(s)* distribution does not exhibit any higher order oligomers and only monomers are present. Thus, binding of nucleotide at low concentration appears to disrupt the oligomers that form in the absence of nucleotide. We interpret this to be direct evidence of nucleotide binding to the monomer.

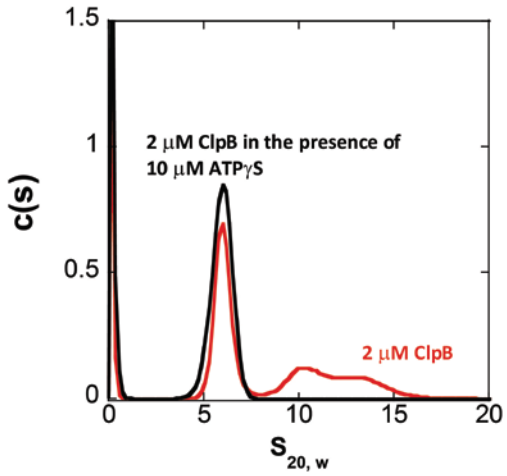

Figure 5: *c(s)* distributions for sedimentation velocity experiments on 2 μM ClpB in the absence (red curve) or presence of 10 μM $[ATP\gamma S]_t$ (black curve). The scans were collected at 230 nm, every 4 minutes at 25 °C.

ATPγS driven ClpB assembly reaction is thermodynamically reversible

The nucleotide driven assembly reaction was tested for thermodynamic reversibility. All of the experiments reported above were performed by adding ATPγS to ClpB. Thus, the sample proceeds from zero ATPγS to some final ATPγS concentration. To test for reversibility 500 μM ATPγS was added to a sample of ClpB at a final concentration of 3 μM. This sample was allowed to incubate for two hours. After the incubation, 3 μM ClpB was added to dilute the sample to either 200 or 100 μM total ATPγS. The sample was examined by sedimentation velocity experiments and the equilibrium constants for dimers, tetramer and hexamers were determined. As shown in Fig. 4, the equilibrium constants obtained from diluting the sample from a high concentration of ATPγS to 200 or 100 μM ATPγS (red filled circles) agree with the equilibrium constants obtained when going from no ATPγS to 200 or 100 μM ATPγS (black filled circles). This indicates that the ATPγS-linked ClpB assembly reactions are both thermodynamically reversible and path independent.

Discussion

This study, for the first time, quantifies the linkage between ATPγS binding and *E. coli* ClpB assembly. On the foundation of our previous study on ClpB self-assembly in the absence of nucleotide [11], here we show that ClpB resides in a dynamic equilibrium in the presence of ATPγS. Monomers, dimers, tetramers, and hexamers are present in solution and their populations and stabilities were measured at various ATPγS concentrations. The thermodynamic and kinetic stability of the ClpB hexamer is enhanced when the nucleotide binding domains of ClpB are saturated with ATPγS.

Moreover, the binding stoichiometry and binding affinity for ClpB oligomers to ATPγS are reported.

Using the assembly constants, $L_{n,0}$, and ATPγS binding constants, κ_n, found in Tables 2 and 4, the species fraction for each ClpB oligomer can be calculated at any given total [*ClpB*] and [*ATPγS*] using Eq. 16 – 18. Since this calculation requires numerically and implicitly solving a sixth order polynomial we have provided the Micromath Scientist model used to calculate the species fractions using the equilibrium constants in Tables 2 and 4 and Eq. 12 – 14 (see Supplemental Materials). As an example, ClpB species fraction as a function of ClpB total concentrations ranging from 10 nM to 30 μM in the presence of 2 mM, 200 μM, and 20 μM ATPγS are shown in Fig. S4 A-C.

ClpB resides in a dynamic equilibrium in the presence of ATPγS

Previous studies concluded that only hexamers were present for 20 and 30 μM ClpB in the presence of 2 mM ATPγS or ATP [29, 30]. This finding has subsequently been interpreted to mean that ClpB forms uniformly hexamer when ATPγS/ATP is in large excess [9, 10, 31-36]. Notably, the ClpB concentrations used in many studies to examine the activity are often as low as 20 nM to 0.2 μM. Since there is a large excess of nucleotide and often an ATP regenerating system it is typically assumed that the entire population of ClpB is hexameric. However, the results presented here indicate that, even in the presence of a large excess of ATP or ATPγS ClpB does not form uniformly hexamer. For example, Fig. S4 shows that for 0.2 μM ClpB in the presence of 2 mM ATPγS, only 45% of the total ClpB forms hexamer. This leads to a hexamer concentration of 15 nM (hexameric unit). This observation indicates that one cannot

consider only the relative ratio of nucleotide to protein to estimate the amount of hexamer present in solution. Rather, the self-association equilibrium constants are required for hexamer concentration prediction.

Being able to accurately predict the concentrations of ClpB oligomers is essential for the accurate interpretation of a variety of in vitro studies. Any quantitative examination of the interactions between ClpB and polypeptide substrate, ClpB and co-chaperones, or ClpB with nucleotide requires an accurate prediction of the concentrations of hexamers present in solution. For example, without $L_{n,app}$ determined in this work, quantitative analyses of the interaction of ClpB with polypeptide in our recent publication would not be possible [12]. In that work, only ClpB hexamer was observed to interact with polypeptide by Li et al in the presence of 1 mM ATPγS. However, it is essential to account for the population of other oligomers that are present in solution. In Li's study, the total concentration of ClpB was calculated using Eq. 19,

$$[ClpB]_T = \{\mathrm{B}_1\} + 2L_{2,\mathrm{app}}\{\mathrm{B}_1\}^2 + 4L_{4,\mathrm{app}}\{\mathrm{B}_1\}^4 + 6L_{6,\mathrm{app}}\{\mathrm{B}_1\}^6 \qquad 19$$

As dimer and tetramer do not interact with polypeptides, Li et al. constrained $L_{2,app}$ and $L_{4,app}$ to the values reported in the current work in the analysis of the ClpB - peptide biding isotherms. $L_{6,app}$ was allowed to float as a fitting parameter since it is possible that $L_{6,app}$ will also depend upon peptide. Interestingly, in the presence of polypeptide, the $log(L_{6,app})$ determined by Li et al. is 33.8 ± 0.3. This value agrees well with the log of the hexamerization equilibrium constant reported here to be 34.0 ± 0.2 (see Table 1), where peptide is absent.

There are two important points to be made about the ClpB - polypeptide binding studies. First, with the agreement between the $L_{6,app}$ reported here and the $L_{6,app}$

determined from fitting the peptide binding isotherms indicates that information on $L_{6,app}$ does reside within the binding isotherms and can be determined in that way. Second, the consistency between $L_{6,app}$ from the two approaches suggests that the peptide does not influence the ClpB hexamerization equilibrium for the polypeptide substrates examined. This is in stark contrast to what we observed for ClpA where we showed that the log of the hexamerization equilibrium constant was 33.1 ± 0.2 and 39.8 ± 0.3 for SsrA tagged substrates and α-casein substrates, respectively [37].

ClpB hexamer dissociation rate constant has an ATPγS concentration dependence

Previous studies reported stopped - flow FRET experiments designed to monitor the kinetic stability of ClpB oligomers [31, 38, 39]. What was termed the subunit exchange rate for ClpB hexamer was reported. In those experiments, a FRET donor-labeled ClpB sample was rapidly mixed with a FRET acceptor-labeled ClpB sample. The FRET signal change as a function of time was recorded and fitted using one or a sum of exponentials. The measured rate constants were reported to be 'subunit exchange rates'. However, it is important to note that the reported parameters are actually observed rate constants resulting from a combination of both dissociation and reassociation events. Consequently, they will exhibit a protein and nucleotide concentration dependence. However, this concentration dependence has not been examined. Further, even though these studies have been interpreted as hexamer exchange rates the FRET signal is potentially sensitive to changes in dimer, tetramer, and hexamer concentrations.

Apparent contradictions have been reported on the subunit exchange rate of ClpB in the presence of ATP using this FRET stopped-flow assay [31, 38]. Werbeck et al. reported that *T. thermophilus* ClpB hexamer resides in rapid subunit exchange using 0.2

μM FRET pair labeled ClpB in the presence of 200 μM ATP/ATPγS [31]. In contrast, Aguado et al. reported what they termed slow "subunit shuffling" when using 0.4 μM FRET pair labeled *E. coli* ClpB in the presence of 2 mM ATP [38]. The observed rate constant reported by Aguado et al. is two orders of magnitude smaller than the slowest rate constant reported by Werbeck et al.

The two studies were performed using enzymes from two different organisms and this may be the reason for the apparent discrepancy. However, the most noticeable difference that also needs to be considered, based on the results reported here, is that each study was performed at a different protein and nucleotide concentration and neither study examined the nucleotide or protein concentration dependence of the reaction. Here we have shown that the hexamer and tetramer are in slow subunit exchange at high nucleotide concentration and fast exchange at low nucleotide concentration while the dimer appears to be under rapid exchange at all nucleotide concentrations. Thus, when the protein and nucleotide concentration dependencies are taken into account the two studies may be in complete agreement.

Determination of the binding stoichiometry and affinity of ATPγS to each ClpB oligomer

In order to fully interpret ATP hydrolysis experiments one needs to know both the stoichiometry and the affinity for nucleotide binding. For an assembling system one also requires this knowledge for each oligomer. This study reveals, for the first time, the nucleotide binding affinity and binding stoichiometry for each oligomeric state of ClpB. Our results show that monomers, dimers, tetramers and hexamers of ClpB all bind ATPγS (see Table 4).

As shown from the model independent Wyman analysis (Fig. 4) hexameric ClpB exhibits a maximum binding stoichiometry of twelve. This observation is in direct contrast to previously published results using Isothermal Titration Calorimetry (ITC) [23]. In that work the binding isotherms were well described by the n-independent and identical sites model with n = 1. From that observation it was concluded that only NBD_2 in the hexamer is capable of binding ATPγS and thus the hexamer has a maximum stoichiometry of 6 nucleotides.

It is imperative to note that the derivation of the n-independent and identical sites model used in the analysis of the ITC data is derived under the express assumption that the macromolecule does not change its assembly state during the titration. At sufficiently high protein concentrations it may be reasonable to assume that everything is in the hexameric state and not changing during the titration and this was the assumption invoked in Fernandez-Higuero et al when using 28 μM ClpB [23]. However, to our surprise, we show here that the population of hexamers decreases upon nucleotide addition between 0 and ~100 μM ATPγS, which is a direct consequence of the fact that nucleotide binds tighter to the monomer than other oligomers. For example, using the self-assembly equilibrium constants in the absence of nucleotide that we recently reported [11], we predict that ~84 % of the $[ClpB]_t$ is in the hexameric state. At a total ClpB monomer concentration of 28 μM this equates to 3.9 μM hexameric ClpB. However, based on the dependence of these equilibrium constants on nucleotide concentration reported here we predict that from 0 - 100 μM ATPγS the concentration of hexamers drops from 3.9 μM to 2.1 μM. In other words, ~46% of the hexameric ClpB dissociated during the titration. This clearly violates the assumptions in the model used to

describe the data. Moreover, since the ITC experiments are performed over a range of ATPγS concentrations where we predict the maximum amount of change in the assembly states, there is most certainly a contribution to the heat due to breaking protein-protein interactions.

In contrast to monomers within a hexamer, we show that each ClpB monomer binds only one ATPγS. Its binding affinity is close to one order of magnitude stronger than the average binding affinity for hexamer (see Table 4). Recall, there are two nucleotide binding sites per monomer. Thus, our finding suggests that only one of the nucleotide binding domains is available for ATPγS binding when there is no adjacent subunit present. These results do not reveal if the binding is to D1 or D2. However, Yamasaki et al. performed mutations in the arginine finger of the adjacent subunit of ClpB for each NBD and showed that NBD_1 requires the arginine finger from an adjacent subunit to fully form the nucleotide binding site. In contrast, the arginine finger from the adjacent subunit is not required for NBD_2 [40]. Thus, we conclude that NBD_2 is most likely bound by nucleotide in the monomer.

Interestingly, the need for an adjacent monomer to complete the NBD_2 binding is consistent with our observed stoichiometries for the intermediate assembly states. The two intermediates (dimers and tetramers) both present stoichiometries that are one fewer than the maximum, ~3 and 7 for dimers and tetramers that have four and eight binding sites, respectively. Moreover, the hexamer exhibits a stoichiometry of 12, which is the maximum number of binding sites that would be present in a hexameric ring. This suggests that the dimer and tetramer are not closed ring structures because an open ring would leave one incomplete nucleotide binding site if adjacent arginines are required to

complete the binding pocket. Further, the hexamer is a closed ring and contains all twelve fully formed binding sites.

Previous studies using covalently truncated ClpB NBD_1-M-domain and NBD_2-C-terminal domain support our findings [16, 17, 21]. They have shown that both NBD_1-M-domain and NBD_2-C-terminal domain truncations cannot assemble to form hexamers. Thus, the individual domains can be considered to represent the binding to the individual domain in the monomeric state. They showed that the NBD_2-C-terminal domain truncation has nucleotide binding ability whereas the NBD_1-M-domain does not. This is consistent with our observation that only site 1 can bind to ATPγS in the full length ClpB monomer.

Our study, for the first time, determined the average binding affinity for each ClpB oligomer without performing any modifications to ClpB. Together with the binding stoichiometry and rigorous examination of the energetics and kinetics of ATPγS linked ClpB assembly, this work provides an essential foundation for quantitative studies on ClpB ATP hydrolysis and ClpB - protein interactions. It will allow us to better quantify how ClpB utilizes the energy from ATP hydrolysis to perform its disaggregation function. Moreover, the strategies employed here can serve as a model for studies on ligand linked assembly for other hexameric AAA+ motor proteins or any protein that exhibits self-association upon ligand binding.

Author Contributions

J. L designed and performed experiments, analyzed data, and wrote the manuscript. A.L.L. designed experiments, analyzed data, and wrote the manuscript.

Acknowledgements

We would like to thank Dr. Walter Stafford and BBRI for the XL-I analytical ultracentrifuge. Thanks to Elizabeth Duran, Clarissa Weaver, and Nate Scull for comments on the manuscript. This work was supported by NSF grant MCB-1412624 to ALL.

References:

[1] Vale RD. AAA proteins. Lords of the ring. J Cell Biol. 2000;150:F13-9.
[2] Arlt H, Tauer R, Feldmann H, Neupert W, Langer T. The YTA10-12 complex, an AAA protease with chaperone-like activity in the inner membrane of mitochondria. Cell. 1996;85:875-85.
[3] Neuwald AF, Aravind L, Spouge JL, Koonin EV. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome research. 1999;9:27-43.
[4] Lee S, Tsai FT. Molecular chaperones in protein quality control. Journal of biochemistry and molecular biology. 2005;38:259-65.
[5] Ogura T, Wilkinson AJ. AAA+ superfamily ATPases: common structure--diverse function. Genes Cells. 2001;6:575-97.
[6] Motohashi K, Watanabe Y, Yohda M, Yoshida M. Heat-inactivated proteins are rescued by the DnaK.J-GrpE set and ClpB chaperones. Proc Natl Acad Sci U S A. 1999;96:7184-9.
[7] Genevaux P, Georgopoulos C, Kelley WL. The Hsp70 chaperone machines of Escherichia coli: a paradigm for the repartition of chaperone functions. Molecular microbiology. 2007;66:840-57.
[8] Glover JR, Lindquist S. Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. Cell. 1998;94:73-82.
[9] Doyle SM, Shorter J, Zolkiewski M, Hoskins JR, Lindquist S, Wickner S. Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity. Nature structural & molecular biology. 2007;14:114-22.
[10] Rosenzweig R, Moradi S, Zarrine-Afsar A, Glover JR, Kay LE. Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction. Science. 2013;339:1080-3.
[11] Lin J, Lucius AL. Examination of the Dynamic Assembly Equilibrium for E. coli ClpB. Proteins. 2015.
[12] Li T, Lin J, Lucius AL. Examination of polypeptide substrate specificity for Escherichia coli ClpB. Proteins. 2015;83:117-34.
[13] Squires CL, Pedersen S, Ross BM, Squires C. ClpB is the Escherichia coli heat shock protein F84.1. Journal of bacteriology. 1991;173:4254-62.
[14] Woo KM, Kim KI, Goldberg AL, Ha DB, Chung CH. The heat-shock protein ClpB in Escherichia coli is a protein-activated ATPase. J Biol Chem. 1992;267:20429-34.
[15] Schlee S, Groemping Y, Herde P, Seidel R, Reinstein J. The chaperone function of ClpB from Thermus thermophilus depends on allosteric interactions of its two ATP-binding sites. J Mol Biol. 2001;306:889-99.
[16] Beinker P, Schlee S, Auvula R, Reinstein J. Biochemical coupling of the two nucleotide binding domains of ClpB: covalent linkage is not a prerequisite for chaperone activity. J Biol Chem. 2005;280:37965-73.
[17] Werbeck ND, Zeymer C, Kellner JN, Reinstein J. Coupling of oligomerization and nucleotide binding in the AAA+ chaperone ClpB. Biochemistry. 2011;50:899-909.

[18] Watanabe YH, Nakazaki Y, Suno R, Yoshida M. Stability of the two wings of the coiled-coil domain of ClpB chaperone is critical for its disaggregation activity. The Biochemical journal. 2009;421:71-7.
[19] Yamasaki T, Oohata Y, Nakamura T, Watanabe YH. Analysis of the Cooperative ATPase Cycle of the AAA+ Chaperone ClpB from Thermus thermophilus by Using Ordered Heterohexamers with an Alternating Subunit Arrangement. J Biol Chem. 2015;290:9789-800.
[20] Mogk A, Schlieker C, Strub C, Rist W, Weibezahn J, Bukau B. Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity. J Biol Chem. 2003;278:17615-24.
[21] Werbeck ND, Kellner JN, Barends TR, Reinstein J. Nucleotide binding and allosteric modulation of the second AAA+ domain of ClpB probed by transient kinetic studies. Biochemistry. 2009;48:7240-50.
[22] Watanabe YH, Motohashi K, Yoshida M. Roles of the two ATP binding sites of ClpB from Thermus thermophilus. J Biol Chem. 2002;277:5804-9.
[23] Fernandez-Higuero JA, Acebron SP, Taneva SG, Del Castillo U, Moro F, Muga A. Allosteric communication between the nucleotide binding domains of caseinolytic peptidase B. J Biol Chem. 2011;286:25547-55.
[24] Biter AB, Lee S, Sung N, Tsai FT. Structural basis for intersubunit signaling in a protein disaggregating machine. Proc Natl Acad Sci U S A. 2012;109:12515-20.
[25] Wyman J, Gill SJ. Binding and linkage : functional chemistry of biological macromolecules. Mill Valley, Ca.: University Science Books; 1990.
[26] Lin J, Lucius AL. Analysis of Linked Equilibria. Methods in Enzymology. 2015.
[27] Na GC, Timasheff SN. Velocity sedimentation study of ligand-induced protein self-association. Methods Enzymol. 1985;117:459-95.
[28] Lee S, Sowa ME, Watanabe YH, Sigler PB, Chiu W, Yoshida M, et al. The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. Cell. 2003;115:229-40.
[29] Akoev V, Gogol EP, Barnett ME, Zolkiewski M. Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB. Protein Sci. 2004;13:567-74.
[30] Zolkiewski M, Kessel M, Ginsburg A, Maurizi MR. Nucleotide-dependent oligomerization of ClpB from Escherichia coli. Protein Sci. 1999;8:1899-903.
[31] Werbeck ND, Schlee S, Reinstein J. Coupling and dynamics of subunits in the hexameric AAA+ chaperone ClpB. J Mol Biol. 2008;378:178-90.
[32] Doyle SM, Hoskins JR, Wickner S. Collaboration between the ClpB AAA+ remodeling protein and the DnaK chaperone system. Proc Natl Acad Sci U S A. 2007;104:11138-44.
[33] Doyle SM, Shastry S, Kravats AN, Shih YH, Miot M, Hoskins JR, et al. Interplay between E. coli DnaK, ClpB and GrpE during protein disaggregation. J Mol Biol. 2014.
[34] Seyffer F, Kummer E, Oguchi Y, Winkler J, Kumar M, Zahn R, et al. Hsp70 proteins bind Hsp100 regulatory M domains to activate AAA+ disaggregase at aggregate surfaces. Nature structural & molecular biology. 2012;19:1347-55.
[35] Mizuno S, Nakazaki Y, Yoshida M, Watanabe YH. Orientation of the amino-terminal domain of ClpB affects the disaggregation of the protein. The FEBS journal. 2012;279:1474-84.

[36] Haslberger T, Zdanowicz A, Brand I, Kirstein J, Turgay K, Mogk A, et al. Protein disaggregation by the AAA+ chaperone ClpB involves partial threading of looped polypeptide segments. Nature structural & molecular biology. 2008;15:641-50.
[37] Li T, Lucius AL. Examination of Polypeptide Substrate Specificity for E. coli ClpA. Biochemistry. 2013;52:4941-54.
[38] Aguado A, Fernandez-Higuero JA, Cabrera Y, Moro F, Muga A. ClpB dynamics is driven by its ATPase cycle and regulated by the DnaK system and substrate proteins. The Biochemical journal. 2015;466:561-70.
[39] DeSantis ME, Leung EH, Sweeny EA, Jackrel ME, Cushman-Nick M, Neuhaus-Follini A, et al. Operational plasticity enables hsp104 to disaggregate diverse amyloid and nonamyloid clients. Cell. 2012;151:778-93.
[40] Yamasaki T, Nakazaki Y, Yoshida M, Watanabe YH. Roles of conserved arginines in ATP-binding domains of AAA+ chaperone ClpB from Thermus thermophilus. The FEBS journal. 2011;278:2395-403.
[41] Zhao H, Ghirlando R, Piszczek G, Curth U, Brautigam CA, Schuck P. Recorded scan times can limit the accuracy of sedimentation coefficients in analytical ultracentrifugation. Analytical biochemistry. 2013;437:104-8.
[42] Schuck P. Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. Biophys J. 1998;75:1503-12.
[43] Stafford WF, Sherwood PJ. Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. Biophysical chemistry. 2004;108:231-43.
[44] Ortega A, Amoros D, Garcia de la Torre J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. Biophys J. 2011;101:892-8.

Supporting information

Scientist model

```
// Micromath Scientist Model File
IndVars: Mtotal
DepVars:Monomer,Dimer,Tetramer,Hexamer
Params: logxt, L20, L40, L60, k1, k2, k4, k6, n2, n4, n6

10^logxt=10^logxf+Mf*10^logxf*(k1+k2*L20*Mf*n2*(1+k2*10^logxf)^(-
1+n2)+k4*L40*(Mf^3)*n4*(1+k4*10^logxf)^(-
1+n4)+k6*(2.00E+29)*(Mf^5)*n6*(1+k6*10^logxf)^(-1+n6))

Mtotal=Mf*(1 + k1* 10^logXf + 2*L20*mf* (1 + k2* 10^logXf)^n2 + 4*L40* mf^3 *(1
+ k4 *10^logXf)^n4 + 6* L60*mf^5 *(1 + k6* 10^logXf)^n6)

Monomer=(Mf+Mf*k1*10^logxf)/Mtotal
Dimer=2*L20*mf^2* (1 + k2* 10^logXf)^n2/Mtotal
Tetramer= 4*L40* mf^4 *(1 + k4 *10^logXf)^n4/Mtotal
```

Hexamer= 6* L60*mf^6 *(1 + k6* 10^logXf)^n6/Mtotal

-100<logxf<logxt
0<Mf<Mtotal

Note: logxt is the log of the total ligand concentration, e.g. ATPgS. logxf is the log of the free ligand concentration

Mtotal is the total monomeric protein concentration, e.g. ClpB. Mf is the free monomeric protein concentration

L20, L40, and L60 are equilibrium constants of dimerization, tetramerization and hexamerization, respectively. Their values can be found in Table 1.

k1, k2, k4 and k6 are the average binding affinity of x to monomer, dimer, tetramer and hexamer, respectively. Their values can be found in Table 4.

n2, n4 and n6 are the binding stoichiometry of x to dimer, tetramer and hexamer, respectively. Their values can be found in Table 4.

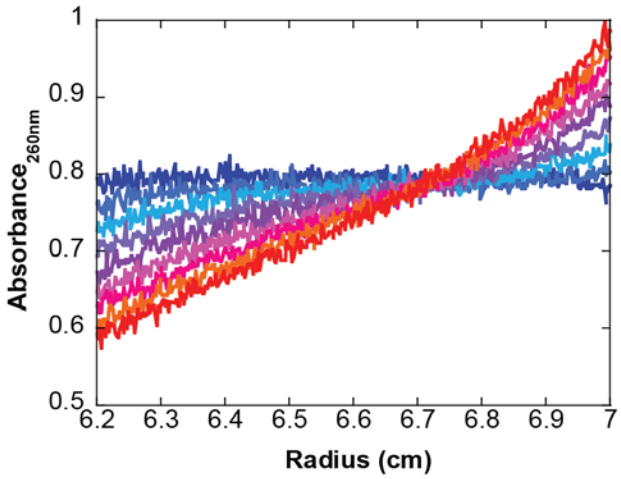

Figure S1: Sedimentation velocity experiment raw data for ATPγS. The scans were collected at 260 nm every 2 minutes at 25 °C. Every 40th scan is shown.

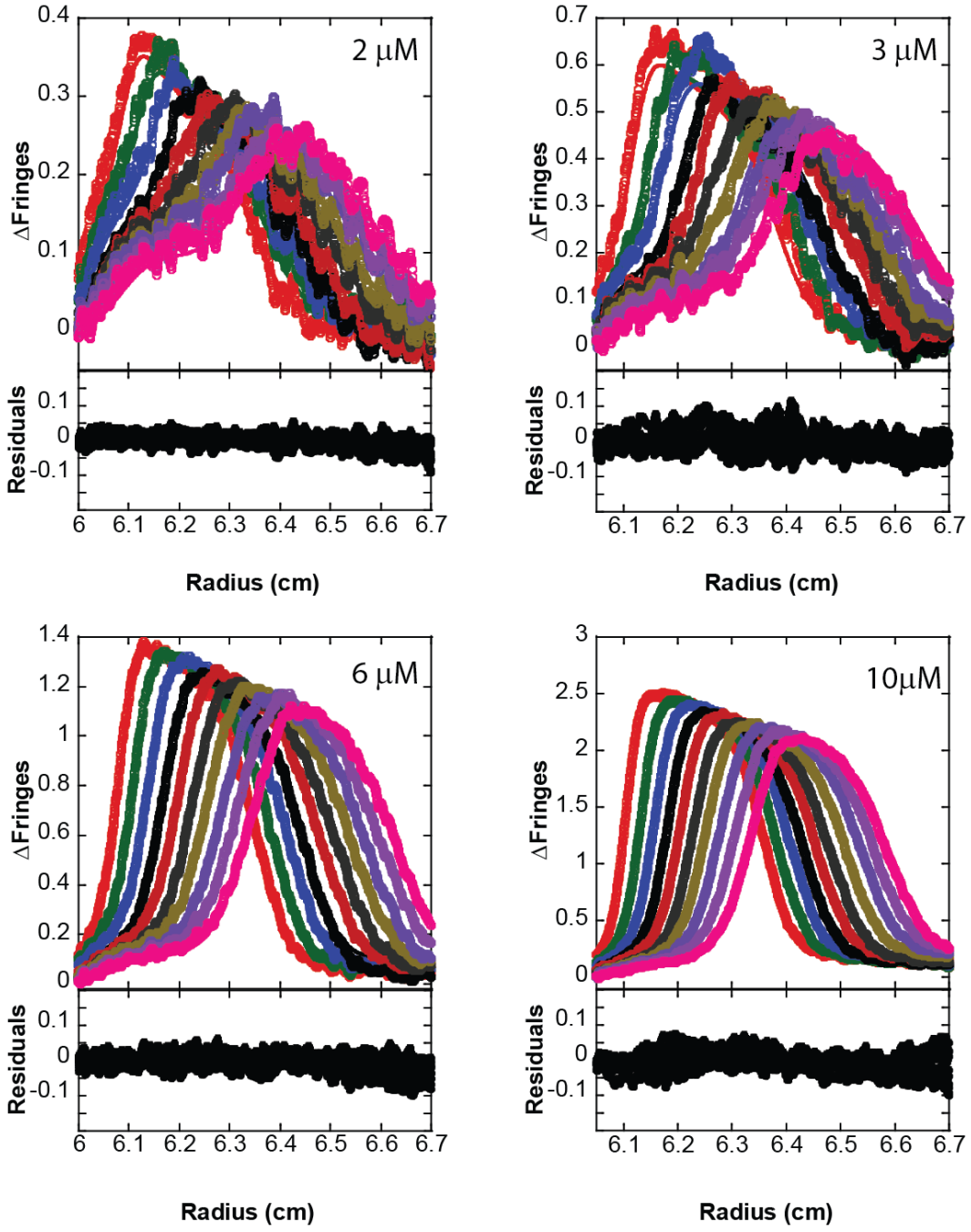

Figure S2: Global analysis of the difference curves for sedimentation velocity data for (A) 2, (C) 3, (E) 6 and (D) 10 μM ClpB in the presence of 200 μM ATPγS using "1-2-4-6" model. Every single scan that was collected in first two and half hours was subject into the fit. The scans were collected very 30 second. Every 5th difference curve is shown. The data are shown in open circles and fits are shown as solid lines. The fitting residuals are shown in panel B, D, F and H. The fitting RMSD is 0.0178. The residual data are distributed randomly around zero which indicates that the fit describes the data reasonably well.

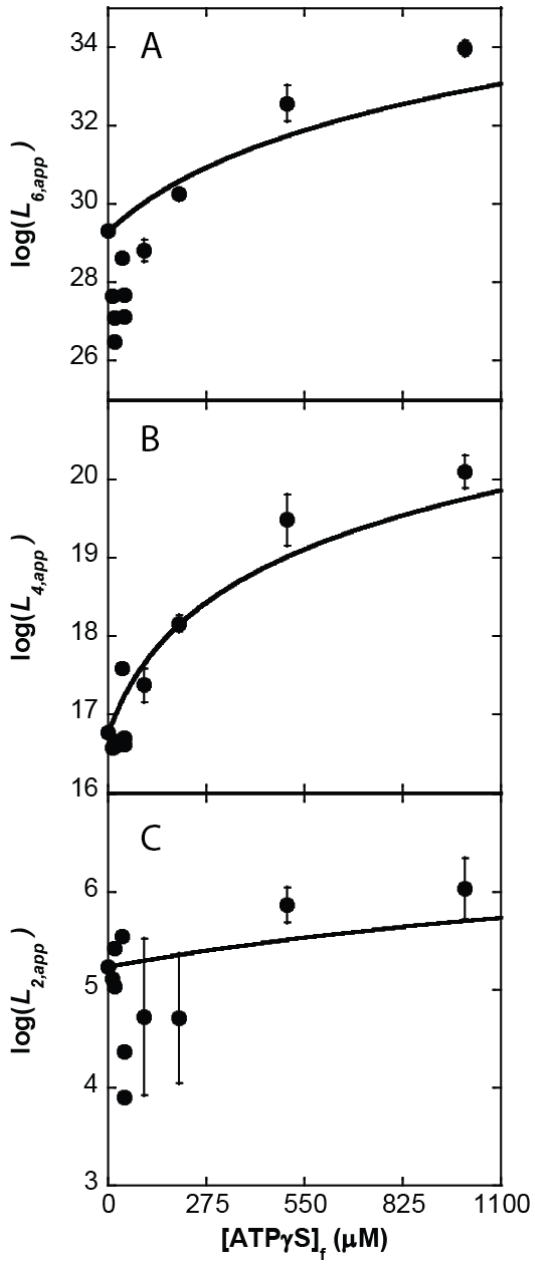

Figure S3: Global analysis of the (A) $L_{6,app}$, (B) $L_{4,app}$ and (C) $L_{2,app}$ as a function of $[ATP\gamma S]_f$ using n-independent identical model with m_1=0. The black filled circles are SedAnal fitting results of the sedimentation velocity data that are presented in Table 1. The solid lines are the fits from global fitting the data shown in panel A, B and C using n-independent identical model with m_1=0.

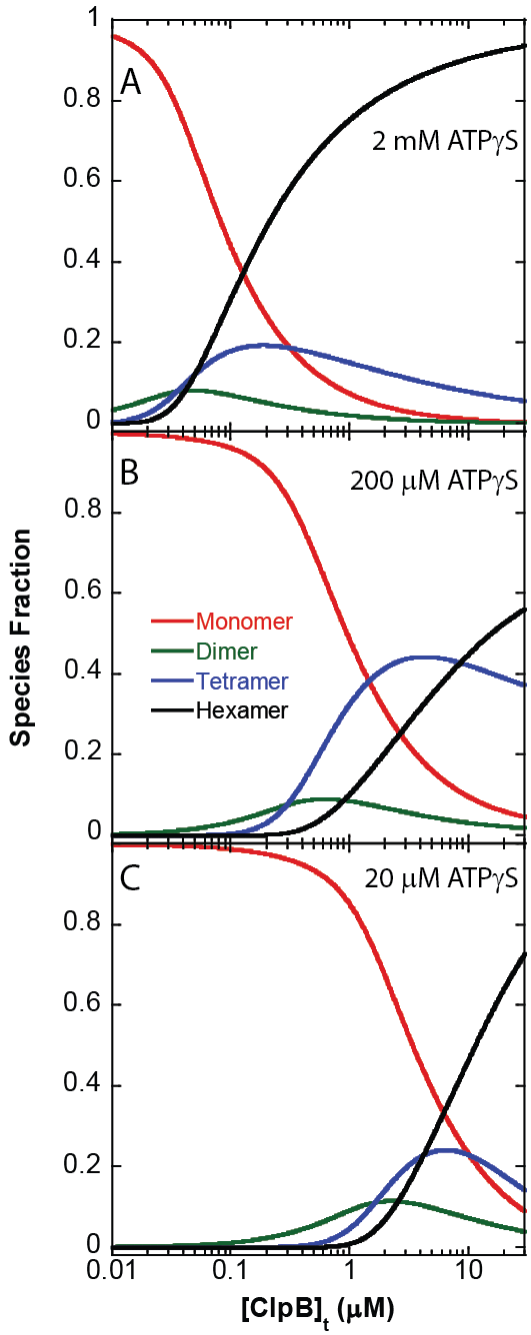

Figure S4: Species fraction for *[ClpB]$_{total}$* ranging from 10 nM to 30 mM in the presence of (A) 2 mM, (B) 200 μM and (C) 20 μM *[ATPγS]$_{free}$*. Monomer, dimer, tertramer and hexamer fractions are shown in red, green, blue and black curve, respectively. All concentrations used to generate these plots are in monomeric unit.

CONCLUSION

In this dissertation, we introduced a method to rigorously study the energetics and kinetics of a complicated ligand-linked monomer-dimer-tetramer-hexamer assembly system using analytical ultracentrifugation. In many cases, proteins in the AAA+ family assemble into hexamers or higher order oligomers in the presence ATP to perform their functions. The method developed in this work can be directly applied to investigate the energetics and kinetics for each intermediate step in the motor reaction for those proteins. This knowledge is important for understanding the mechanism of those proteins.

ClpB is a chaperone protein that is able to disaggregate denatured proteins in cell with the assistance of DnaKJE co-chaperones. As a member of the AAA+ superfamily, ClpB requires binding of ATP to form an active hexamer and perform its chaperon function. However, the linkage of ATP binding to ClpB assembly has not been determined. This work for the first time, examined ClpB assembly as a function of both ClpB and nucleotide concentrations. The linkage of nucleotide binding to assembly, together with nucleotide binding affinity and stoichiometry to each ClpB oligomer are determined without making any modification to the primary structure of protein. The knowledge provided by this work is essential for the quantitative examination of the mechanism of ClpB-DnaK and ClpB-protein aggregate interactions.

Analytical ultracentrifugation as a tool to study assembly equilibria

In chapter 2, we presented a method for using analytical ultracentrifugation to rigorously examine the energetics and kinetics of ligand linked assembly equilibria. In this chapter, we have focused on the discussions of monomer, dimer, tetramer, and hexamer equilibrium, because this is the assembly process that we have observed experimentally for *E.coli* ClpB. In order to investigate whether the energetics and kinetics of this complex system can be determined by NLLS analysis using SedAnal, sedimentation velocity data at various protein concentrations with given equilibrium constants and dissociation rate constants were simulated and then analyzed accordingly.

Three sets of dissociation rate constants, k_r, that are in the range of $k_r > 0.01\ s^{-1}$, $0.01\ s^{-1} > k_r > 10^{-5}\ s^{-1}$, and $k_r < 10^{-5}\ s^{-1}$ were used to simulate the data, respectively. $k_r > 0.01\ s^{-1}$ and $k_r < 10^{-5}\ s^{-1}$ are the empirical ranges for reactions that are considered to undergo instantaneous dissociation and that are non-dissociable on the time scale sedimentation, respectively. Accordingly, the data can be analyzed using a thermodynamic model if $k_r > 0.01\ s^{-1}$ and components analysis if $k_r < 10^{-5}\ s^{-1}$. If the dissociation rate constants of the reactions are in the empirical range of $0.01\ s^{-1} > k_r > 10^{-5}\ s^{-1}$, then they can be quantitatively determined. [10-13] By carefully analyzing the simulated data, our results show that the empirical bound of for measurable dissociation rate constants ($0.01\ s^{-1} – 10^{-5}\ s^{-1}$) may vary depending on the size of molecules and the rotor speed.

We observed that when both equilibrium constants and dissociation rate constants were allowed to float, the resulting values were most accurate. However, doing so requires more computational power and longer fitting time. To accelerate this process,

other fitting methods can be used to investigate the kinetic properties of the assembly to limit the assumptions on whether the system is under instantaneous dissociation. One way to do this is to perform *c*(*s*) analysis for the sedimentation velocity data and overlay the resulting *c*(*s*) distributions as a function of protein concentration.[11]

ClpB forms hexamer in the absence of nucleotide

Various techniques have been used to determine the oliogmeric states of ClpB in the absence of nucleotide. Contradicting conclusions were made due to the lacking of knowledge on the assembly energetics and kinetics of ClpB.[4, 19, 28, 29] In chapter 3, we presented a rigorous analysis of ClpB assembly as a function of [*ClpB*] and [*NaCl*] using sedimentation velocity experiments. At the conditions where one predominant *c*(*s*) peak was observed with a $s_{20,w}$ = ~17.6 S, the system presented as monodisperse. Sedimentation equilibrium experiments were then performed at multiple angular velocities to investigate the molecular weight of this 17.6 S oligomer. The determined molecular weight is (578 ± 3) kDa, which is in good agreement with the molecular weight of hexameric ClpB calculated from its primary structure (575 kDa).

ClpB exhibits a dynamic assembly

The study presented in Chapter 3 for the first time rigorously determined the assembly pathway for ClpB to form hexamers in the absence of nucleotide. Our results show that ClpB resides in a monomer-dimer-tetramer-hexamer equilibrium. Both equilibrium constants and dissociation rate constants for each of these oligomers were incorporated into the data analysis using the methods presented in Chapter 2. Our results show that both assembly energetics and kinetics of the assembly had a [NaCl] dependence. For example, decreasing [NaCl] from 300 mM to 100 mM, the assembly

equilibrium constant for hexamer increased from ~ 1.1 × 1026 to ~ 5.8 × 1032 M^{-5} and the dissociation rated constant decreased from > 0.01 s^{-1} to ~1.3 × 10-3 s^{-1}, see Table 3 in Chapter 3.

Notably, ClpB is assumed to form hexamer only in the presence of a large excess of nucleoside triphosphate in many studies.[5-8] This is because Zolkiewski et al. observed only hexameric ClpB using a sedimentation equilibrium approach at a single ClpB concentration.[3] Our study shows that ClpB still resides in a dynamic assembly in the presence of ATPγS. Monomer, dimer, tetramer, and hexamer were observed at various [*ClpB*] and [*ATPγS*]. In Chapter 4, we carefully examined ClpB assembly as function of [*ATPγS*]. The assembly energetics and kinetics presented a strong [*ATPγS*] dependence.

Interestingly, upon addition of ATPγS, the equilibrium constants for oligomerization, $L_{n,app}$, first decreased and then increased as a function of [*ATPγS*] as shown in Table 1 Chapter 4. The dissociation rate constants of oligomers are greater than 0.01 s^{-1} at low [*ATPγS*]. When increasing [*ATPγS*] to 200 μM, the dissociation rate constants of hexamer decreased to ~10^{-3} s^{-1}. At 1 mM [*ATPγS*], the dissociation rate constant of hexamer was then further decreased to ~ 10^{-5} s^{-1}. These results explain the discrepancy between Werbeck's[7] and Aguado's[21] findings, where Werbeck et al. observed rapid subunit exchange in the presence of 200 μM ATP and Aguado et al. observed slow exchange rates in the presence of 2 mM ATP.

ATPγS binding affinity and stoichiometry

In order to examine the nucleotide binding ability of ClpB, a mutagenesis strategy has been employed in previous studies.[15-17, 20, 22-27] However, how the modifications on

the primary structure of ClpB interrupt the protein assembly and their subsequent impact on nucleotide binding, ATP hydrolysis, and peptide recognition has not been examined. Instead, ClpB was once again assumed to form hexamers when the nucleotide concentration is in large excess. Based on the results of our study presented in Chapter 4, this assumption does not hold true when the protein concentration is small.

In our work, for the first time, the binding constants and stoichiometry for nucleotide binding to each of the oligomers were determined without performing any modifications to ClpB. Here we show that ClpB monomer, dimer, tetramer, and hexamer exhibit an ATPγS binding stoichiometry of 1, 3, 7 and 12 respectively. Compared to the full saturation of the hexameric ring, the binding stoichiometry of the smaller oligomers is one fewer than the maximum number of binding sites, which suggests an open conformation rather than a ring structure. Moreover, ClpB monomer has a stronger ATPγS average binding affinity than ClpB hexamer, which suggests one of the nucleotide binding domains binds ATPγS tighter than the other.

Summary

Our study provides a strategy to rigorously examine the energetics and kinetics of protein assembly and the linkage of ligand binding to assembly using analytical ultracentrifugation techniques. This strategy can be applied to the study of many AAA+ motor proteins. With the results determined by this type of study, the concentration of every protein oligomer can be determined for any given [*ATP*] and [*protein*]. Moreover, the number of ATP molecules bound per protein oligomer can be also predicted. Knowing the population of the active oligomers that is available in solution and the amount of ATP bound per oligomer is important for quantitative determination of how

those proteins utilize the energy provided by ATP hydrolysis to work as a protein machine.

REFERENCE FOR INTRODUCTION AND CONCLUSIONS

[1] Schirmer, E. C., Glover, J. R., Singer, M. A., and Lindquist, S. (1996) HSP100/Clp proteins: a common mechanism explains diverse functions, *Trends in biochemical sciences 21*, 289-296.

[2] Squires, C. L., Pedersen, S., Ross, B. M., and Squires, C. (1991) ClpB is the Escherichia coli heat shock protein F84.1, *Journal of bacteriology 173*, 4254-4262.

[3] Zolkiewski, M., Kessel, M., Ginsburg, A., and Maurizi, M. R. (1999) Nucleotide-dependent oligomerization of ClpB from Escherichia coli, *Protein Sci 8*, 1899-1903.

[4] Akoev, V., Gogol, E. P., Barnett, M. E., and Zolkiewski, M. (2004) Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB, *Protein Sci 13*, 567-574.

[5] Lee, S., Choi, J. M., and Tsai, F. T. (2007) Visualizing the ATPase cycle in a protein disaggregating machine: structural basis for substrate binding by ClpB, *Mol Cell 25*, 261-271.

[6] Schlieker, C., Weibezahn, J., Patzelt, H., Tessarz, P., Strub, C., Zeth, K., Erbse, A., Schneider-Mergener, J., Chin, J. W., Schultz, P. G., Bukau, B., and Mogk, A. (2004) Substrate recognition by the AAA+ chaperone ClpB, *Nature structural & molecular biology 11*, 607-615.

[7] Werbeck, N. D., Schlee, S., and Reinstein, J. (2008) Coupling and dynamics of subunits in the hexameric AAA+ chaperone ClpB, *J Mol Biol 378*, 178-190.

[8] Haslberger, T., Zdanowicz, A., Brand, I., Kirstein, J., Turgay, K., Mogk, A., and Bukau, B. (2008) Protein disaggregation by the AAA+ chaperone ClpB involves partial threading of looped polypeptide segments, *Nature structural & molecular biology 15*, 641-650.

[9] Li, T., Lin, J., and Lucius, A. L. (2015) Examination of polypeptide substrate specificity for Escherichia coli ClpB, *Proteins 83*, 117-134.

[10] Correia, J. J., Alday, P. H., Sherwood, P., and Stafford, W. F. (2009) Effect of kinetics on sedimentation velocity profiles and the role of intermediates, *Methods Enzymol 467*, 135-161.

[11] Dam, J., Velikovsky, C. A., Mariuzza, R. A., Urbanke, C., and Schuck, P. (2005) Sedimentation velocity analysis of heterogeneous protein-protein interactions:

Lamm equation modeling and sedimentation coefficient distributions c(s), *Biophys J 89*, 619-634.

[12] Stafford, W. F., and Sherwood, P. J. (2004) Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants, *Biophysical chemistry 108*, 231-243.

[13] Demeler, B., Brookes, E., Wang, R., Schirf, V., and Kim, C. A. (2010) Characterization of reversible associations by sedimentation velocity with UltraScan, *Macromolecular bioscience 10*, 775-782.

[14] del Castillo, U., Alfonso, C., Acebron, S. P., Martos, A., Moro, F., Rivas, G., and Muga, A. (2011) A quantitative analysis of the effect of nucleotides and the M domain on the association equilibrium of ClpB, *Biochemistry 50*, 1991-2003.

[15] Schlee, S., Groemping, Y., Herde, P., Seidel, R., and Reinstein, J. (2001) The chaperone function of ClpB from Thermus thermophilus depends on allosteric interactions of its two ATP-binding sites, *J Mol Biol 306*, 889-899.

[16] Watanabe, Y. H., Motohashi, K., and Yoshida, M. (2002) Roles of the two ATP binding sites of ClpB from Thermus thermophilus, *J Biol Chem 277*, 5804-5809.

[17] Watanabe, Y. H., Nakazaki, Y., Suno, R., and Yoshida, M. (2009) Stability of the two wings of the coiled-coil domain of ClpB chaperone is critical for its disaggregation activity, *The Biochemical journal 421*, 71-77.

[18] Mizuno, S., Nakazaki, Y., Yoshida, M., and Watanabe, Y. H. (2012) Orientation of the amino-terminal domain of ClpB affects the disaggregation of the protein, *The FEBS journal 279*, 1474-1484.

[19] Woo, K. M., Kim, K. I., Goldberg, A. L., Ha, D. B., and Chung, C. H. (1992) The heat-shock protein ClpB in Escherichia coli is a protein-activated ATPase, *J Biol Chem 267*, 20429-20434.

[20] Mogk, A., Schlieker, C., Strub, C., Rist, W., Weibezahn, J., and Bukau, B. (2003) Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity, *J Biol Chem 278*, 17615-17624.

[21] Aguado, A., Fernandez-Higuero, J. A., Cabrera, Y., Moro, F., and Muga, A. (2015) ClpB dynamics is driven by its ATPase cycle and regulated by the DnaK system and substrate proteins, *The Biochemical journal 466*, 561-570.

[22] Beinker, P., Schlee, S., Auvula, R., and Reinstein, J. (2005) Biochemical coupling of the two nucleotide binding domains of ClpB: covalent linkage is not a prerequisite for chaperone activity, *J Biol Chem 280*, 37965-37973.

[23] Werbeck, N. D., Zeymer, C., Kellner, J. N., and Reinstein, J. (2011) Coupling of oligomerization and nucleotide binding in the AAA+ chaperone ClpB, *Biochemistry 50*, 899-909.

[24] Yamasaki, T., Oohata, Y., Nakamura, T., and Watanabe, Y. H. (2015) Analysis of the Cooperative ATPase Cycle of the AAA+ Chaperone ClpB from Thermus thermophilus by Using Ordered Heterohexamers with an Alternating Subunit Arrangement, *J Biol Chem 290*, 9789-9800.

[25] Werbeck, N. D., Kellner, J. N., Barends, T. R., and Reinstein, J. (2009) Nucleotide binding and allosteric modulation of the second AAA+ domain of ClpB probed by transient kinetic studies, *Biochemistry 48*, 7240-7250.

[26] Fernandez-Higuero, J. A., Acebron, S. P., Taneva, S. G., Del Castillo, U., Moro, F., and Muga, A. (2011) Allosteric communication between the nucleotide binding domains of caseinolytic peptidase B, *J Biol Chem 286*, 25547-25555.

[27] Biter, A. B., Lee, S., Sung, N., and Tsai, F. T. (2012) Structural basis for intersubunit signaling in a protein disaggregating machine, *Proc Natl Acad Sci U S A 109*, 12515-12520.

[28] Kim, K. I., Cheong, G. W., Park, S. C., Ha, J. S., Woo, K. M., Choi, S. J., and Chung, C. H. (2000) Heptameric ring structure of the heat-shock protein ClpB, a protein-activated ATPase in Escherichia coli, *J Mol Biol 303*, 655-666.

[29] Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T., and Baker, T. A. (2000) Dynamics of substrate denaturation and translocation by the ClpXP degradation machine, *Mol Cell 5*, 639-648.