2016

Exploring the Impact of the E. coli Proteostasis Network on the Folding Fate of Proteins with Different Intrinsic Biophysical Properties

Kristine Faye R. Pobre

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Biochemistry Commons, Molecular Biology Commons, and the Systems Biology Commons

Recommended Citation
https://scholarworks.umass.edu/dissertations_2/595

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
EXPLORING THE IMPACT OF THE E. COLI PROTEOSTASIS NETWORK ON THE FOLDING FATE OF PROTEINS WITH DIFFERENT INTRINSIC BIOPHYSICAL PROPERTIES

A Dissertation Presented

by

KRISTINE FAYE R. POBRE

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2016

Molecular and Cellular Biology
EXPLORING THE IMPACT OF THE E. COLI PROTEOSTASIS NETWORK ON THE FOLDING FATE OF PROTEINS WITH DIFFERENT INTRINSIC BIOPHYSICAL PROPERTIES

A Dissertation Presented

by

KRISTINE FAYE R. POBRE

Approved as to style and content by:

Lila M. Gierasch, Chair

Evan T. Powers, Member

Peter Chien, Member

Jeffrey L. Blanchard, Member

Elizabeth R. Dumont, Director
Interdepartmental Graduate Programs
DEDICATION

This work is lovingly dedicated to my family…
ACKNOWLEDGMENTS

This work is a masterpiece, where the heart and mind of the people I treasure have been imprinted. I am very grateful to Lila Gierasch. She is a great mentor. She has guided me with this journey, and has helped me reach the finish line with joy and fulfillment. I also thank her for the trust she gave me as I combined computation with wet lab experiments. To think critically is what she instilled in my mind. She has inspired me in many ways. Indeed, she is worthy of emulation. I am thankful to Evan Powers, who has been a key in honing my computational skills. I thank him for teaching me FoldEco, and for sharing me his computing knowledge. His help with my data analysis and interpretation is greatly appreciated. I thank Peter Chien for his pieces of advice and comments about my thesis work, especially on protein degradation. I appreciate Jeffrey Blanchard for giving me the opportunity to rotate in his lab where I learned Perl and R programming. I acknowledge the patience and understanding of my committee members, and thank them for their helpful insights towards the implementation of my dissertation.

I thank all the Gierasch Lab members for their support through the years. To Eugenia Clerico, my first mentor in the lab. She taught me how to run SDS-PAGE gels, do Western blotting, and purify proteins. To Mylene Ferrolino, my friend and labmate. We had great adventures together inside and outside the lab. We had fun playing around with CRABP1 (aggregated, folded, degraded). To Ivan Budyak for teaching me how to measure folding kinetics using the stopped
flow instrument, and for the travels to different places with Olga and Elena. To Anne Gershenson for her critical comments during group meetings, especially on kinetics and thermodynamics. To Karan Hingorani for giving an extra hand in the lab, and for his help in reading the first draft of Chapter 1. I also appreciate Anastasia Zhuravleva, Rob Smock, Mandy Blackburn, Ilene Magpiong, Weiwei Kuo, Abhay Thakur, Wenli Meng, Charles English, Joe Tilitsky for providing a friendly environment in the lab. I thank Ellen Kalt for her assistance in my RA or TA paperwork while working in the lab, and for her smiles everyday.

I want to thank the Kelly Lab at The Scripps Research Institute (La Jolla) for their warm hospitality when I went there to work with our collaborators. To Younhee Cho and Xin Zhang who had both helped me with my experiments with the whole proteostasis network. It was nice working with them at Scripps, and even when I went back to UMass. Thanks to Jeff Kelly for his suggestions regarding our project. To Gina Dendle, Jenn Catacutan, Michael Saure, Gareth Morgan, Aleksandra Baranczak for making my stay in La Jolla more convenient and more fun. Thanks are also due to the LSL 3rd floor labs (Chien, Vierling, and Hebert Labs). I appreciate the times when I had to use their equipment and they were kind to offer help. I thank the Hardy Lab for allowing me to work at their space while waiting for our new lab at LSL to be set up after the move (from LGRC to LSL).

I want to extend my great appreciation and thanks to the Molecular and Cellular Biology (MCB), Chemistry-Biology Interface (CBI), and to the Institute for
Cellular Engineering - Integrative Graduate Education and Research Traineeship (ICE-IGERT) programs for the supportive communities that they provided and had helped me become a better researcher. I had learned a lot from interactions with researchers from other fields. I am also indebted to Sarah Czerwonka for assisting me with requirements necessary for PhD completion. I thank the National Institutes of Health (NIH) for the funding of my research.

Graduate school life had been easier with friends around. To Amherst Pals—Mylene and JB Ferrolino, Abigail Merriam, Tina Martin, Princess Wilson, Kevin and Dolly Dagbay, Risha and Lem de Leon, Phey and Ron Lerum, Vanessa Gill, Bay and Mahalia Serrano, Bernard and Aileen, Julius Campeciño, and Rae Recto, for the good memories here in Amherst and our wonderful trips. To Cindy Utama and Hyeseung Seo, my music partners at church and UMass chorale. To First Baptist Church of Amherst and UMass Intervarsity Graduate Fellowship for keeping me sound in both science and faith. I want to thank Anushka Dongre, my roommate and friend for many years, for our fabulous adventures as we pursued our PhD. I am grateful to Ron and Nancy Loescher for being so kind and generous in giving me rides every Sunday. I give my sincerest thanks to Evie Hopkins for allowing us to stay at her house as I was finishing the last lap of my PhD.

With a grateful heart, I thank my family in the US who has taken care of me since I first came to the US. I thank my grandmother (Mommy Caridad Pobre) who had inspired me ever since and had encouraged me to do my best in my
PhD studies. To my aunt (Ellen Pobre) who has set an example to dream big and to go the distance. To Uncle Edwin and Auntie Anita Pobre, Dennis and Weng Pobre, Joel and Jeanette Pobre, Eric and Gerrie Gaugler. I thank Uncle Boyet and Tita Monette Alegre for all those relaxing moments with them.

Special thanks to Paoay Faith Baptist Church. To my best friend, Rev Pacol. To the Poor And Rich Are Happy family.

No words can express my gratitude to my parents, Papa Wilson and Mama Julie Pobre for being my inspiration. They have instilled in me virtues of love, patience, humility, and kindness. They have sacrificed a lot for my education. To be God-fearing is also imprinted in my heart and mind. Many thanks to my siblings, Manang Elaine Joy, Clifford Immanuel (and Criscelle), Glaiza May (and Christopher), and Lindsey Lorraine. Despite the distance that we had, we had kept in touch, and they had given me joy while I was away from home. To my in-laws, Papa Adriano and Mama Josie Piza, Sharon Faye, Eunice Joy, and Paul Shelton, thank you.

With all my heart, I want to thank my husband, Francis Alan Piza for all the love and care he has shown me. He has sacrificed a lot for our family. He has reminded me to always think positive. I also appreciate his humility and kindness. His cooking is the best and greatly appreciated. I thank my daughter, Francine Adriana, for giving me joy beyond compare. She has given me company while I was writing and finalizing my thesis.

Finally, I thank God for everything.
ABSTRACT

EXPLORING THE IMPACT OF THE E. COLI PROTEOSTASIS NETWORK ON THE FOLDING FATE OF PROTEINS WITH DIFFERENT INTRINSIC BIOPHYSICAL PROPERTIES

FEBRUARY 2016

KRISTINE FAYE R. POBRE

B.S., UNIVERSITY OF THE PHILIPPINES DILIMAN

PhD., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Lila M. Gierasch

The three-dimensional (3D) native structure of most proteins is crucial for their functions. Despite the complex cellular environment and the variety of challenges that proteins experience as they fold, proteins can still fold to their native states with high fidelity. The reason for this is the presence of the cellular proteostasis network (PN), consisting of molecular chaperones and degradation enzymes, that collaborates to maintain proteostasis, in which the necessary levels of functional proteins are optimized. Although extensive research has been carried out on the mechanisms of individual components of the proteostasis network, little is known about how these components contribute to the functioning of the network as a whole. A new protein can have three folding fates: natively folded, aggregated, or degraded. The fate is determined by both a protein’s
intrinsic biophysical properties and the cellular proteostasis network through kinetic partitioning. To understand the interplay between a protein’s intrinsic biophysical properties and the cellular proteostasis network, an integrated computational and experimental approach was used. The folding fates of model proteins with different intrinsic biophysical properties under varying conditions of the proteostasis network were determined. Using FoldEco, the effects of the kinetic and thermodynamic properties of proteins on their folding fates were investigated systematically, and predictions were consistent with wet lab experiments. The folding fate of a protein is under a thermo-kinetic limitation, which indicates that the fate depends on either the kinetics or thermodynamics, but (for the most part) not on both at the same time. Different proteins behave according to the values of their limiting properties. Furthermore, up-regulation of the entire proteostasis network through the $\sigma^{32}$ transcription factor has beneficial effects on model proteins with low stabilities and high aggregation propensities. However, the effects of up-regulation of individual chaperones or the major degradation enzyme, Lon are substrate-dependent and are related to their biophysical properties. Furthermore, KJE, GroELS, and Lon form an efficacious triad for maintaining proteostasis, and their contributions depend on the biophysical properties of their substrates, and on the concentrations of these PN components and substrates at any given time.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>xvi</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xvii</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Protein folding in the cell</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Proteostasis</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Importance of proteostasis</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Intrinsic and extrinsic factors determine protein folding fate</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>Biophysical properties of proteins</td>
<td>5</td>
</tr>
<tr>
<td>1.5</td>
<td>The proteostasis network</td>
<td>6</td>
</tr>
<tr>
<td>1.6</td>
<td>Processes of the proteostasis network</td>
<td>7</td>
</tr>
<tr>
<td>1.6.1</td>
<td>Biogenesis</td>
<td>7</td>
</tr>
<tr>
<td>1.6.2</td>
<td>Conformational maintenance</td>
<td>8</td>
</tr>
<tr>
<td>1.6.3</td>
<td>Protein clearance</td>
<td>10</td>
</tr>
<tr>
<td>1.6.4</td>
<td>Interplay among synthesis, folding, chaperoning and degradation</td>
<td>11</td>
</tr>
<tr>
<td>1.7</td>
<td>Molecular chaperones of the proteostasis network (PN)</td>
<td>12</td>
</tr>
<tr>
<td>1.7.1</td>
<td>Ribosome-associated chaperones</td>
<td>12</td>
</tr>
<tr>
<td>1.7.2</td>
<td>Chaperones acting downstream of the ribosome</td>
<td>13</td>
</tr>
<tr>
<td>1.8</td>
<td><em>E. coli</em> as a model organism</td>
<td>15</td>
</tr>
<tr>
<td>1.9</td>
<td>Proteostasis network in <em>E. coli</em></td>
<td>15</td>
</tr>
<tr>
<td>1.9.1</td>
<td>Trigger factor (TF)</td>
<td>15</td>
</tr>
<tr>
<td>1.9.2</td>
<td>DnaK/DnaJ/GrpE (KJE) family</td>
<td>17</td>
</tr>
<tr>
<td>1.9.3</td>
<td>GroEL/GroES (GroELS) family</td>
<td>20</td>
</tr>
<tr>
<td>1.9.4</td>
<td>Disaggregation family, ClpB/DnaK/DnaJ/GrpE</td>
<td>22</td>
</tr>
<tr>
<td>1.9.5</td>
<td>IbpA/IbpB</td>
<td>23</td>
</tr>
</tbody>
</table>
1.9.6 HtpG ........................................................................................................... 24
1.9.7 Degradation enzymes .................................................................................. 24

1.9.7.1 Lon ............................................................................................................. 25
1.9.7.2 ClpAP/XP ................................................................................................. 27
1.9.7.3 ClpYQ/HslUV ......................................................................................... 27
1.9.7.4 FtsH ........................................................................................................... 28

1.10 Cooperation among chaperones and degradation enzymes in E. coli ........................................................................ 29
1.11 Heat shock response in E. coli ...................................................................... 30
1.12 Protein’s intrinsic biophysical properties and the cellular PN dictate protein folding fate through kinetic partitioning ............................................ 31
1.13 FoldEco model of proteostasis in E. coli ............................................................. 34

1.13.1 Parts of FoldEco .......................................................................................... 34
1.13.2 Implementing and parametrizing FoldEco ..................................................... 38

1.14 Statement of thesis .......................................................................................... 41

2 INFLUENCE OF INTRINSIC BIOPHYSICAL PROPERTIES ON
PROTEIN FOLDING FATE IN E. COLI UNDER THE BASAL
PROTEOSTASIS NETWORK ..................................................................................... 56
2.1 Introduction ....................................................................................................... 56
2.2 Results ............................................................................................................... 60

2.2.1 Comprehensive examination of protein folding fate versus biophysical properties using FoldEco ................................................................. 60

2.2.1.1 Case 1: No aggregation, no misfolding .................................................... 62

2.2.1.1.1 Partitioning between folding and degradation depends exclusively on rate and equilibrium constants .......................................................... 62
2.2.1.1.2 Rate and equilibrium constants affect the fraction of soluble protein similarly ................................................................. 63

2.2.1.2 Case 2: Misfolding, but no aggregation ...................................................... 64

2.2.1.2.1 Without aggregation, misfolding increases degradation ....................... 64
2.2.1.2.2 Misfolding penalty” is also thermo/kinetically-limited ...................... 65
2.2.1.3 Case 3: Misfolding and aggregation

2.2.1.3.1 Aggregation following misfolding further suppresses native folding

2.2.1.3.2 Aggregation penalty” and fraction aggregated protein are also thermo/kinetically limited

2.2.2 Investigating the effects of biophysical properties on protein folding fates and testing FoldEco predictions using wet lab experiments

2.2.2.1 The folding fates of CRABP1(M9A), CRABP1(V67A), and CRABP1(L118V) are similar to CRABP1(WT*) while the folding fate of CRABP1(Y133S) is different from the others

2.2.2.1.1 Low levels of total protein for Y133S is due to the presence of Lon

2.2.2.2 Experimental data are consistent with computational expectations

2.2.3 Preliminary analyses of the effects of biophysical properties on protein folding fates of proteins using FoldEco

2.3 Discussion

2.4 Conclusion

2.5 Materials and Methods

2.5.1 E. coli strains and plasmids

2.5.2 Protein expression in E. coli

2.5.3 Quantitative Coomassie staining and Western blotting

2.5.4 Measurement of relative levels of PN components

2.5.5 FoldEco simulations

2.5.5.1 Preliminary FoldEco simulations

2.5.6 Protein sequences

2.5.7 Protein purification

2.5.8 Folding kinetics measurements

3 DISSECTING THE IMPACT OF THE PROTEOSTASIS NETWORK ON PROTEIN FOLDING FATE
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>109</td>
</tr>
<tr>
<td>3.2</td>
<td>Results</td>
<td>113</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Enhancement of the whole PN through 154N $\sigma^{32}$ overexpression reduces the fraction of aggregated protein</td>
<td>113</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Up-regulation of PN components decreases CRABP1(Y133S) aggregation to different extents</td>
<td>114</td>
</tr>
<tr>
<td>3.2.2.1</td>
<td>Up-regulation of KJE, GroELS, or Lon individually decreases CRABP1(Y133S) aggregation</td>
<td>115</td>
</tr>
<tr>
<td>3.2.2.2</td>
<td>The beneficial effects of KJE, GroELS, and Lon in pairs are largely additive</td>
<td>117</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Up-regulation of PN components decreases the aggregation of other model proteins</td>
<td>118</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Low levels of expression of CRABP1(Y133S)</td>
<td>120</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Analyses of CRABP1(Y133S) protein folding fates</td>
<td>125</td>
</tr>
<tr>
<td>3.2.5.1</td>
<td>Phenomenological models</td>
<td>125</td>
</tr>
<tr>
<td>3.2.5.2</td>
<td>Analysis of protein folding fate of CRABP1(Y133S) using FoldEco</td>
<td>127</td>
</tr>
<tr>
<td>3.3</td>
<td>Discussion</td>
<td>130</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusion</td>
<td>135</td>
</tr>
<tr>
<td>3.5</td>
<td>Materials and Methods</td>
<td>136</td>
</tr>
<tr>
<td>3.5.1</td>
<td><em>E. coli</em> strains and plasmids</td>
<td>136</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Protein expression in <em>E. coli</em></td>
<td>137</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Quantitative Western blotting</td>
<td>138</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Measurement of relative levels of PN components</td>
<td>140</td>
</tr>
<tr>
<td>3.5.5</td>
<td>Protein sequences</td>
<td>140</td>
</tr>
<tr>
<td>3.5.6</td>
<td>Fits of Equations (1) and (2) and FoldEco to <em>in vivo</em> folding fate data</td>
<td>141</td>
</tr>
<tr>
<td>4</td>
<td>CONCLUSIONS</td>
<td>160</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary</td>
<td>160</td>
</tr>
<tr>
<td>4.2</td>
<td>Significance of findings</td>
<td>161</td>
</tr>
<tr>
<td>4.3</td>
<td>Future directions</td>
<td>162</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Biophysical properties</td>
<td>162</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Proteostasis network</td>
<td>164</td>
</tr>
</tbody>
</table>
APPENDICES

A. SUPPLEMENTARY MATERIAL FOR INFLUENCE OF INTRINSIC BIOPHYSICAL PROPERTIES ON PROTEIN FOLDING FATE IN *E. COLI* UNDER THE BASAL PROTEOSTASIS NETWORK .............................................. 165

B. SUPPLEMENTARY MATERIAL FOR DISSECTING THE IMPACT OF THE PROTEOSTASIS NETWORK ON PROTEIN FOLDING FATE ........................................................................................................ 183

BIBLIOGRAPHY ....................................................................................................................................................... 191
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Concentrations of some PN components</td>
<td>55</td>
</tr>
<tr>
<td>2.1</td>
<td>Thermodynamic and kinetic parameters of CRABP1 variants measured <em>in vitro</em></td>
<td>92</td>
</tr>
<tr>
<td>2.2</td>
<td>Biophysical properties of CRABP1 variants and their folding fates in <em>E. coli</em> after 3 h of protein expression under the adapted-basal PN condition</td>
<td>107</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>1.10</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>1.11</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>1.12</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>98</td>
<td></td>
</tr>
</tbody>
</table>
2.8 Comparison of fraction soluble vs. $k_f$ and $K_f$ of proteins that misfold and do not aggregate, and those that misfold and aggregate ........................................... 99

2.9 Effect of aggregation .................................................................................................................. 100

2.10 The behavior of aggregation-prone proteins .............................................................................. 101

2.11 Folding fate of CRABP1(WT*) upon expression in *E. coli* under adapted-basal conditions .......................................................... 102

2.12 Folding fate of CRABP1(M9A) upon expression in *E. coli* under adapted-basal conditions ........................................................................................................ 103

2.13 Folding fate of CRABP1(V67A) upon expression in *E. coli* under adapted-basal conditions ........................................................................................................ 104

2.14 Folding fate of CRABP1(L118V) upon expression in *E. coli* under adapted-basal conditions ........................................................................................................ 105

2.15 Folding fate of CRABP1(Y133S) upon expression in *E. coli* under adapted-basal conditions ........................................................................................................ 106

2.16 Biophysical properties of CRABP1 variants in terms of thermo-kinetically limited parameters ........................................................................................................ 108

3.1 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 2 h at 30 °C under adapted-basal PN conditions and up-regulation of PN components through overexpression of I54N $\sigma^{32}$ ........ 143

3.2 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 2 h at 30 °C under adapted-basal conditions and with up-regulation of individual PN components .......................................................... 145

3.3 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 2 h at 30 °C under basal conditions and with the pairwise up-regulation of PN components ........................................................................ 147

3.4 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 1 h at 30 °C under adapted-basal conditions and on conditions with overexpression of I54N $\sigma^{32}$ ........................................................................ 149

3.5 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 1 h at 30 °C under adapted-basal conditions and with up-regulation of individual PN components ........................................................................ 151
3.6 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 1 h at 30 °C under adapted-basal conditions and with the pairwise up-regulation of PN components ................................................................. 153

3.7 Dependence of CRABP1(Y133S) protein folding fates on different PN components based on phenomenological fits of *in vivo* folding fate data ................................................................. 155

3.8 Results of FoldEco-derived fit of experimental (*in vivo*) folding fate data from the overexpression of CRABP1(Y133S) for 2 h ......................................................... 157

3.9 Schematic for experimental methods ................................................................. 159
CHAPTER 1

INTRODUCTION

Proteins have key roles in almost all biological processes. The three-dimensional (3D) native structure of most proteins is crucial for their functions. How a protein folds inside the cell, and how the cellular environment maintains levels of functional proteins are important problems in biology, and my thesis research has been directed towards answering these questions. This chapter covers the following background topics to my thesis work: 1) challenges that proteins encounter as they fold inside the cell, 2) protein folding fates determined by both the intrinsic biophysical properties of proteins and the cellular proteostasis network (PN) through kinetic partitioning, and 3) the use of FoldEco model of proteostasis in understanding the mechanisms of folding in the cell.

1.1 Protein folding in the cell

Our knowledge of protein folding is mostly based on in vitro experiments where proteins are denatured and are being refolded under optimized conditions and high dilution in order to prevent aggregation, as aggregation is concentration-dependent. With these conditions, some proteins can refold efficiently. Their native structure is dependent on their amino acid sequence, which encodes the information about their fold (Anfinsen, 1973). However, other proteins (usually large, multidomain) cannot refold after being denatured in vitro (Fersht, 1999).
Thus, protein folding models have been based only on “well-behaved” proteins that can be denatured and refolded in vitro (Clark, 2004). These in vitro models provide a picture of individual proteins that fold in isolation, which is different from the actual folding of proteins inside the cell where proteins fold in the presence of a variety of macromolecules (Vendruscolo, 2012).

There are many differences between folding in vitro and in vivo (Hingorani and Gierasch, 2014). First, in vitro, the solutions and buffers are highly dilute, while in vivo, the cellular environment is concentrated and crowded (20-30% of cell volume or ~300-400 mg/mL is occupied by macromolecules) (Zimmerman and Trach, 1991), which makes the unfolded polypeptide prone to interact with macromolecules (Zhou et al., 2008). Macromolecular crowding drives the compaction of folding intermediates or partially folded states, which means that it can enhance the formation of native states, oligomeric structures and aggregates (Cheung et al., 2005; Ellis, 2001a, b; McGuffee and Elcock, 2010; Minton, 2000; Minton and Wilf, 1981; Zhou et al., 2008). Second, in vitro, a full length chain of unfolded polypeptide is allowed to fold, while in vivo, folding can start while the polypeptide is still being made (co-translational folding), and the polypeptide can also bind to ribosome-associated chaperones (Gershenson and Gierasch, 2011). Third, in vitro, folding reactions happen in a homogeneous milieu, while in vivo, folding reactions take place in inhomogeneous environment that are compartmentalized and spatially organized (Gershenson and Gierasch, 2011). Fourth, in vitro, a protein can fold on its own, while in vivo, some proteins fold
with the help of chaperones (Kim et al., 2013). Despite this complex cellular environment (Fig. 1.1) and the variety of challenges that proteins experience as they fold, proteins can still fold to their native states with high fidelity even in concentrations higher than the threshold for aggregation in vitro refolding experiments. The reason for this is the presence of the cellular protein homeostasis ("proteostasis") network (PN), consisting of different processes (biogenesis, conformational maintenance, and protein clearance) as well as various players (molecular chaperones and degradation enzymes), that works together to maintain levels of functional proteins inside the cell. Our goal is to understand how a protein folds in the context of the cellular environment because this is where folding actually occurs (Gershenson and Gierasch, 2011; Hartl et al., 2011; Vendruscolo, 2012).

1.2 Proteostasis

A newly synthesized polypeptide (unfolded, U) can have three folding fates. First, it can fold to its native state (N) and become functional. Second, it can misfold (M) and when these misfolded states self-associate they can form aggregates (A). Third, it can be degraded (Cho et al., 2015) (Fig. 1.2). Aggregated and degraded proteins are not functional. Proteostasis is achieved when levels of proteins are sufficient to perform their functions, and levels of misfolded and aggregated proteins are minimized to avoid toxicity (Balch et al., 2008; Powers et al., 2009).
Aggregates can be amorphous, oligomers rich in β-sheets, or amyloid fibrils (Jahn and Radford, 2005; Mitraki and King, 1989; Wetzel, 1996). Amyloid fibrils are characterized by cross β-structure, in which β-strands run perpendicular to the long axis of the fibril (Tycko, 2004). Aggregation is concentration-dependent (Harper and Lansbury, 1997). Aggregates can be formed via nucleation-dependent polymerization, in which monomers form an aggregation nucleus that grows rapidly to form aggregates. When the monomer concentration of partially folded or misfolded states exceeds a certain level known as the critical concentration, polymerization starts. As protein concentration increases, polymer concentration also increases but the monomer concentration remains constant (Harper and Lansbury, 1997).

Degradation is the elimination of abnormal proteins or those that are not needed anymore. “Abnormal” proteins are those that are incomplete, misfolded, or damaged by the environment (postsynthetic denaturation or chemical damage) (Goldberg, 1972, 2003).

1.2.1 Importance of proteostasis

Failure of proteostasis can cause disease. For example, in humans, conformational diseases arise if proteostasis is not maintained. Loss-of-function diseases such as cystic fibrosis (Riordan, 1999) and phenylketonuria (Waters et al., 2000) are due to insufficient amounts of functional proteins caused by variants that inefficiently fold and are rapidly degraded. Gain-of-toxicity diseases
such as Alzheimer’s disease (AD), Parkinson’s disease (PD), Huntington’s disease (HD), and amyotrophic lateral sclerosis (ALS) are associated with the formation of intracellular aggregates with amyloid-like properties as well as extracellular amyloid fibrils (usually toxic to the cells) (Brehme et al., 2014; Chiti and Dobson, 2006; Gregersen et al., 2005; Gregersen et al., 2000; Hipp et al., 2014; Labbadia and Morimoto, 2015; Ross and Poirier, 2004; Serpell and Smith, 2000). Understanding how proteostasis is maintained in the cell helps us design ways to ameliorate and/or prevent such diseases (Balch et al., 2008; Hopkins, 2008; Powers et al., 2009).

1.3 Intrinsic and extrinsic factors determine protein folding fate

An overarching question is, “What determines the folding fate of a protein?”. As mentioned earlier, an unfolded polypeptide can be folded, aggregated, or degraded. The folding fate of a protein is determined by both intrinsic and extrinsic factors, which are the protein’s intrinsic biophysical properties (Powers et al., 2012) and the cellular proteostasis network (PN), respectively (Cho et al., 2015).

1.4 Biophysical properties of proteins

Each protein has its own intrinsic biophysical (e.g. kinetic and thermodynamic) properties dictated by its amino acid sequence. These include the folding rate constant ($k_f$) and folding equilibrium constant ($K_f$), misfolding rate
constant \((k_m)\) and misfolding equilibrium constant \((K_m)\), aggregation rate constant \((k_a)\), aggregation equilibrium constant \((K_a)\) or the inverse of the critical concentration for aggregation \(c_{crit}\) and many others (Powers et al., 2009; Powers et al., 2012). The rate constant is a kinetic property that determines the rate of folding, misfolding, or aggregation. The equilibrium constant is a thermodynamic property, which is the ratio of forward and reverse rate constants at equilibrium. These values are measured \textit{in vitro} or can be estimated using \textit{in silico} techniques that are also based on \textit{in vitro} data.

An unfolded polypeptide explores a funnel-shaped energy landscape as it folds towards its native state (Hartl and Hayer-Hartl, 2009; Jahn and Radford, 2005) (Fig. 1.3). The native state, formed via intramolecular contacts, is the conformation that is usually the most thermodynamically stable (lowest energy). Partially folded or misfolded states can be kinetically trapped due to high free-energy barriers that they need to cross in order to reach a favorable downhill path. When these partially folded or misfolded states form intermolecular interactions, aggregates (amorphous, oligomers, or amyloids) are formed (Jahn and Radford, 2005).

### 1.5 The Proteostasis network (PN)

The cellular proteostasis network (PN), consisting of different processes (biogenesis, conformational maintenance, and protein clearance) as well as various players (molecular chaperones and degradation enzymes), modulates a
protein’s folding fate. During its lifetime, a protein interacts with chaperones and/or degradation enzymes for help in its folding or degradation to minimize aggregation. Chaperones and degradation enzymes minimize aggregation by interacting with partially folded or misfolded states and then can perform the following: 1) help them fold to native, 2) hold them to prevent intermolecular interactions, or 3) degrade them, thereby lowering their concentrations.

1.6 Processes of the proteostasis network (PN)

There are three main processes of the PN: 1) biogenesis, 2) conformational maintenance, and 3) clearance (Hipp et al., 2014).

1.6.1 Biogenesis

Biogenesis consists of protein synthesis and folding. The synthesis (translation) of new proteins is a complicated process. Aside from the critical steps a ribosome performs in order to synthesize a new polypeptide chain, this new chain has to be folded, processed, or targeted to its proper location (Kramer et al., 2009). Translation is carefully managed inside the cell. The production of new proteins is time-dependent in terms of the levels of proteins produced at a time and how the different domains of a particular protein fold. Translation rates affect folding (Komar et al., 1999; Meriin et al., 2012; Tsai et al., 2008). Variations in rates can be due to local mRNA stability and codon sequence (Tsai et al., 2008; Yu et al., 2015). In *E. coli*, for example, when the 16 consecutive rare
codons in chloramphenicol acetyltransferase (CAT) gene are replaced by frequent codons, the translation rate of CAT protein is increased; however, the specific activity of the protein is reduced (Komar et al., 1999). If the rate was too fast, it can cause an accumulation of misfolded states. By contrast, if the rate was too slow, it can cause the production of lower amounts of proteins that are insufficient to carry out their functions. Mild inhibition of synthesis or moderately slowing down the rate, though, improves the folding of nascent chains and recovers the function of a mutant protein (Meriin et al., 2012). In eukaryotes, synthesis rate is slower (~4 amino acids per second) compared to that in prokaryotes (~20 amino acids per second) (Hartl and Hayer-Hartl, 2009). The slower rate in eukaryotes may be due to bigger proteins (multi-domains) that are ~52 kDa (in bacteria, only ~35 kDa). These domains may fold co-translationally because of their slow folding kinetics (Netzer and Hartl, 1998). In E. coli, in vivo translation kinetics (and not thermodynamics) of the ribosome nascent chain (RNC) influences co-translational folding. Faster rates delay co-translational folding causing longer nascent chains (Ciryam et al., 2013a).

1.6.2 Conformational maintenance

Inside the cell, a protein’s conformation (either folded or unfolded, as long as it does not self-associate and aggregate) is maintained throughout its lifetime. This crucial task is performed by different groups of molecular chaperones and/or degradation enzymes by helping the proteins fold, refold, and solubilize
aggregated proteins. As mentioned earlier, the consequence of self-association of partially folded or misfolded states is aggregation. A fundamental feature of folding in the cell is the competition of folding and aggregation (Clark, 2004; Niwa et al., 2009; Pechmann et al., 2009; Tartaglia and Vendruscolo, 2010). Studies have shown how aggregation may happen in the cell and how it is prevented. Aggregation is a widespread phenomenon in the cell (David et al., 2010; Narayanaswamy et al., 2009; Tartaglia et al., 2007). Proteins are at high risk of aggregation either in normal or stressed conditions. Aggregation is a serious problem in the cell because it can cause gain-of-toxicity, or it may be a last-line of defense to cells under stress (Miller et al., 2015). In humans, expression levels of proteins in vivo are anti-correlated with their aggregation rates in vitro (Tartaglia et al., 2007). Widespread aggregation in neurodegenerative diseases and aging is also linked with supersaturation (Ciryam et al., 2013b). In yeast and mammals, misfolded proteins are sequestered in subcellular deposition sites such as the insoluble protein deposit (IPOD), intranuclear quality control compartment (INQ), and cytoplasm quality control compartment (CytoQ) (Kaganovich et al., 2008; Miller et al., 2015), while in E. coli, aggregates are typically localized to the cell poles (Rokney et al., 2009; Winkler et al., 2010). These deposits are being repaired by disaggregation chaperones and/or degraded by degradation enzymes, and also allows asymmetric distribution of aggregates during cell division (Miller et al., 2015). Some proteins can also co-aggregate with aberrant polypeptides such as amyloid-forming polypeptides (Gidalevitz et al., 2006) or
polyglutamine expansions (Olzscha et al., 2011). Such damaged proteins deregulate the proteostasis network by competing for the need for chaperones and degradation enzymes. In order to minimize aggregation, chaperones and degradation enzymes interact with partially folded or misfolded states, protecting the proteins from intermolecular interactions. Up-regulation of chaperones through heat shock or oxidative stress increases the capacity of the cell to prevent aggregation (Guisbert et al., 2008; Lindquist, 1986; Zhang et al., 2014).

1.6.3 Protein clearance

Protein clearance, either by degradation or autophagy, has a critical role in proteostasis. In eukaryotes, the ubiquitin-proteasome system (UPS) is mainly responsible for the clearance of proteins (Hershko and Ciechanover, 1998; Varshavsky, 2012). Ubiquitin, a small protein with 76 residues, is enzymatically conjugated to proteins that have degrons (or tags for degradation). Then, this ubiquitin-protein complex is recognized and degraded by the 26S proteasome, a processive ATP-dependent protease. Since the proteasome complexes are located in the nucleus and cytosol, proteins in the endoplasmic reticulum (ER) that are targeted for degradation need to be retrotranslocated to the cytosol first before they are recognized by the UPS (Smith et al., 2011). About ~1% of the ribosome-bound nascent chains and ~0.5% of newly synthesized polypeptides are ubiquitinated, and then are degraded by the proteasome in vivo (Duttler et al., 2013). In autophagy, large aggregates or insoluble deposits are engulfed by
double-membrane vesicles (autophagosomes) and degraded by lysosomal components (Kundu and Thompson, 2008).

In prokaryotes, degradation is mediated by different proteases such as Lon, ClpXP/ClpAP, HslUV, and FtsH, which belong to AAA+ family of proteins (ATPases associated with diverse cellular activities) (Gottesman, 2003; Gur et al., 2011). Generally, they consist of AAA+ ring and an interior chamber with proteolytic active sites (Gur et al., 2011; Sauer and Baker, 2011). First, the substrate binds to the ATPase domain, and then cycles of ATP binding/hydrolysis cause conformational change to unfold the substrate and degrade it through the proteolytic chamber (Gottesman, 2003).

1.6.4 Interplay among synthesis, folding, chaperoning and degradation

Is there interplay among synthesis and folding, chaperoning, and degradation? A balance among these processes is important for the efficiency of the PN. Chaperones and degradation enzymes work together (Sakr et al., 2010; Tomoyasu et al., 2001). The functional and physical features of chaperones and degradation enzymes overlap (Wickner et al., 1999). Some chaperones facilitate degradation (Arndt et al., 2007; Jubete et al., 1996; Kandrør et al., 1999; Kettern et al., 2010; Rottgers et al., 2002; Sakr et al., 2010; Savel'ev et al., 1998; Wagner et al., 1994). An example is CHIP, a cochaperone and ubiquitin ligase, binds to Hsp70 and Hsp90 to initiate the destruction of abnormal proteins (Arndt et al., 2007). Likewise, some degradation enzymes have chaperone activities.
(Gottesman et al., 1997; Sauer and Baker, 2011; Suzuki et al., 1997). While these studies suggest that molecular chaperones and degradation enzymes work synergistically, they do not include the mechanisms by which they do it.

1.7 Molecular chaperones of the proteostasis network (PN)

Chaperones have diverse roles in maintaining proteostasis, including de novo folding and refolding of proteins, disaggregation, and degradation (Hartl et al., 2011; Saibil, 2013). Some chaperones are present in high levels in the cell; others are in low amounts (Finka and Goloubinoff, 2013; Powers et al., 2009; Powers et al., 2012), depending on cell types. When the cell is stressed, their levels increase to different extents. The first chaperones that were identified were named “heat shock proteins” (Hsp) because they were induced by heat (Tissieres et al., 1974). Chaperones are grouped depending on their sequence homology, and are named according to their molecular weight. The four major groups are: Hsp60s, Hsp70s, Hsp90s, and Hsp100s. In the cytosol, the general pathway for folding is conserved among bacteria, archaea, and eukarya (Fig. 1.4) (Kim et al., 2013).

1.7.1 Ribosome-associated Chaperones

The ribosome-associated chaperones, including trigger factor (TF) in bacteria, and specialized Hsp70s such as ribosome-associated complex (RAC, in Saccharomyces cerevisiae, MP11 and HSP70L1 in mammals, and nascent
chain-associated complex (NAC) in archaea and eukarya are the first to interact with newly synthesized polypeptides. About ~70% of nascent chains interact with TF in *E. coli* (Fig. 1.4) (Kim et al., 2013). A nascent polypeptide is restricted from folding until a complete protein domain or segments (~50-300 amino acid residues) is synthesized and emerges from the ribosome (Elcock, 2006; Kaiser et al., 2011). As the nascent chain is emerging, ribosome-associated chaperones bind to its hydrophobic sequences to prevent intramolecular or intermolecular interactions. In fungi, the ribosome-associated complex (RAC), which consists of Ssz1 (a specialized Hsp70) and zuotin (Hsp40), together with the Hsp70 isoform, Ssb, assist nascent chain folding (Kim et al., 2013).

1.7.2 Chaperones acting downstream of the ribosome

Members of the Hsp70 family (DnaK/DnaJ/GrpE in bacteria, Hsp70/Hsp40/Nucleotide exchange factors in eukaryotes), mediate co- and post-translational folding (Frydman, 2001; Hartl et al., 2011) by interacting with nascent and newly synthesized polypeptides (Calloni et al., 2012; Niwa et al., 2012) (Fig. 1.4). ATP-regulated cycles modulate the binding and release of substrates (by kinetic partitioning). In most species of archaea, instead of Hsp70, prefoldin (also known as the Gim complex, GimC) binds to certain nascent chains in an ATP-independent manner (Hartl and Hayer-Hartl, 2002). These substrates are then transferred to downstream chaperones.
The downstream chaperones include the chaperonins (Hsp60s): GroEL/GroES in bacteria, thermosome in archaea, and tailless complex polypeptide-1 (TCP-1) ring complex (TriC)/chaperonin-containing TCP-1 (CCT) in eukaryotes. Chaperonins promote folding by encapsulating proteins inside a large double-ring complex (7-9 rings, \( \sim 60 \text{ kDa} \) subunits per ring) in a central cavity (Hartl, 1996), providing the protein an environment that is less crowded than the cell and avoids aggregation. About 10-15% newly synthesized proteins interact with chaperonins in bacteria and archaea (Kerner et al., 2005) while 5-10% in eukarya (Yam et al., 2008). In eukaryotes, the Hsp90 system mediates folding of substrates from the heat shock cognate 70 (Hsc70) (Taipale et al., 2010). In yeast and other fungi, Hsp70 cooperates with AAA+ chaperone Hsp104 in untangling and resolubilizing aggregated proteins. In bacteria, ClpB (homologous to Hsp104) works with DnaK in protein disaggregation (Haslberger et al., 2010).

Small heat shock proteins (sHsp) are ubiquitous chaperones that can prevent irreversible aggregation. These chaperones form a complex with non-native proteins in an ATP-independent manner (Haslbeck and Vierling, 2015). Their sizes range from 12 to 42 kDa, and they form large oligomeric ensembles. Their functions rely on the changes in the distribution in tertiary or quarternary structures of the ensembles (Haslbeck and Vierling, 2015).
1.8 *E. coli* as a model organism

*E. coli* is used as a simple model organism to study proteostasis in the cell. It only has ~4000 genes, and only about 1 femtoliter volume. It has two primary chaperone systems, DnaK (Hsp70) and its co-chaperones DnaJ (Hsp40) and GrpE (nucleotide exchange factor), and GroEL (Hsp60) and its co-chaperone GroES. It also has a ribosome-associated chaperone (trigger factor), a protein disaggregation chaperone system (ClpB/DnaK/DnaJ/GrpE), and degradation machinery (Lon, ClpXP, ClpAP, HslUV, FtsH). A wealth of biochemical and genetic data is also available for these chaperones and degradation enzymes (Powers et al., 2012).

1.9 Proteostasis network in *E. coli*

1.9.1 Trigger factor (TF)

The role of TF in protein folding *in vivo* was first found out by genetic studies in which Δ*tig* mutants (lacking TF) were viable and had no major defects on the folding of newly synthesized polypeptides. However, when the *E. coli* Hsp70 chaperone (DnaK) was deleted, (Δ*dnaK*), at temperatures above 30 °C, the cells (with combined Δ*tig* and Δ*dnaK* mutations) were not viable (synthetic lethality) (Deuerling et al., 1999). Deuerling and colleagues investigated the cause of synthetic lethality by depletion of DnaK and DnaJ in cells with trigger factor (tig*) and without (Δ*tig*). The tig* cells grew at 30 and 37 °C but not at 15 and 42 °C (similar results for the cold-sensitive and heat-sensitive Δ*dnaK*
mutants) while Δtig cells did not grow at all the temperatures tested (15, 30, 37 and 42 °C). In addition, when DnaK and DnaJ were depleted, aggregated newly synthesized proteins increased (~40 proteins) in Δtig cells compared to tig+ cells. These results show that TF is associated with the folding of newly synthesized proteins in cooperation with DnaK (Deuerling et al., 1999; Teter et al., 1999). Cells can tolerate individual deletions of TF or DnaK, suggesting that the two are functionally redundant (Bukau et al., 2000; Calloni et al., 2012; Genevaux et al., 2004). TF serves like a cradle for folding space for most nascent chains to prevent them from being degraded or aggregated (Ferbitz et al., 2004; Hartl and Hayer-Hartl, 2002).

TF is a ~50 kDa protein that has an elongated structure with three domains: an N-terminal ribosome-binding domain (contains the ribosome-binding loop, Phe-Arg-Lys), a peptidyl-prolyl cis/trans isomerase (PPIase) domain, and a C-terminal domain (Fig. 1.5) (Ferbitz et al., 2004). The binding of monomeric TF to the ribosome (at subunit L23 and L29) is important for interaction with the nascent chain (Baram et al., 2005; Kramer et al., 2002; Schlunzen et al., 2005). TF binds to hydrophobic segments on nascent chains, and helps folding through ATP-independent cycles of binding and release from both ribosome and nascent chain. TF is released from the chain when the bound peptide starts to bury its hydrophobic sequences. Sometimes, TF may remain bound to the polypeptides even after their release from the ribosome, functioning as a holdase (binds to substrates and hold them, preventing them from aggregating) (Martinez-Hackert
and Hendrickson, 2009). Furthermore, hydrophobic chain collapse is slowed down and co-translational folding is delayed by TF (Agashe et al., 2004; Gupta et al., 2010; Hoffmann et al., 2012). TF can also adapt to and help different kinds of nascent polypeptides (Merz et al., 2008). TF also interacts with small basic proteins such as ribosomal proteins (Calloni et al., 2012; Lakshmipathy et al., 2010; Martinez-Hackert and Hendrickson, 2009).

### 1.9.2 DnaK/DnaJ/GrpE (KJE) family

DnaK is the major Hsp70 chaperone in bacteria. It is constitutively expressed and also induced during stress. It is not essential for viability under nonstress conditions at intermediate temperatures (at 30 °C), but it is absolutely necessary at high temperatures (42 °C) (Bukau and Walker, 1989). It is a ~70 kDa protein that consists of N-terminal nucleotide-binding domain (NBD) and a C-terminal substrate-binding domain (SBD) connected by a highly conserved interdomain linker (Fig.1.6) (Bertelsen et al., 2009). ATP-binding to the NBD allosterically controls substrate affinity to the SBD. Higher affinity for substrate when ADP is bound (ADP-substrate-bound state) compared to when ATP is bound (ATP-bound state). In the ADP-bound state, NBD and SBD are independent of each other and they retain the structures that they adopt as separated domains joined by a flexible linker (Bertelsen et al., 2009). When ATP is bound (ATP-bound state), the two domains are docked; the hydrophobic linker and the α-helical lid of the SBD are attached to the NBD (Kityk et al., 2012; Qi et
al., 2013; Swain et al., 2007). The DnaJ (co-chaperone of DnaK) and GrpE (nucleotide exchange factor in *E. coli*) regulate the DnaK system. DnaJ delivers substrates to DnaK and activates ATPase activity on DnaK (Laufen et al., 1999) while GrpE catalyzes exchange of ADP for ATP (Packschies et al., 1997).

What is the basis of substrate recognition? For many years, peptide models have been used to determine how Hsps (in general) recognize their substrates (Clerico et al., 2015). Peptide studies on DnaK allow us to determine some preferences of DnaK, but the binding motifs are not very clear. However, based on these studies, predictive algorithms have been developed to identify potential DnaK binding sites (Rudiger et al., 1997; Van Durme et al., 2009). DnaK recognizes short polypeptides, which consists of a hydrophobic core of four to five residues that are enriched in Leu, Ile, Val, Phe, and Tyr and the two flanking regions are mostly basic residues (Rudiger et al., 1997). Does the mode of binding to peptides similar to that of the full-length substrates? One particular study shows that the preferred binding sequence of DnaK within the *E. coli* σ^{32} transcription factor when peptide (by peptide scanning) is used is the same sequence as when full-length protein (by proteolysis footprinting) is analyzed (Clerico et al., 2015; Rodriguez et al., 2008). Recently, the interaction of DnaK with full-length proteins has been explored. Lee and colleagues showed that when the full-length SRC homology 3 domain (SH3) was bound to DnaK, it had multiple, slowly interconverting disordered conformations on different sequences that were distinct from the unfolded state without DnaK. The bound client has
structural and dynamical heterogeneity revealing that proteins may undergo
conformational sampling when bound to DnaK (Lee et al., 2015). On the other
hand, Sekhar and colleagues found that the conformation of human telomere
repeat binding factor 1 (hTRF1) did not change when it was bound to ATP-, ADP-
and nucleotide-free DnaK. They also showed that the bound hTRF1 retained its
intrinsic structural propensities; when bound to DnaK, it acquired secondary
structures determined by the intrinsic properties of the substrate. The nucleotide
state of DnaK does not affect the overall conformation of the bound client
(Sekhar et al., 2015). Using rhodanese as a model protein, Kellner and
colleagues showed that the substrate expanded ~30 fold when bound to DnaK.
This may prevent unwanted interactions within or between substrates (Kellner et
to the same site on their substrate? With the σ^{32} transcription factor as the
substrate, researchers have found that DnaK and DnaJ bind to different sites on
σ^{32}. Their binding causes conformational changes that destabilize the protein and
prepare it for degradation (Rodriguez et al., 2008).

Using quantitative proteomics, it is found that DnaK interacts with ~700
proteins of E. coli, mostly are cytosolic (Calloni et al., 2012). In addition, ~180 are
aggregation-prone that use DnaK in their folding. Furthermore, DnaK interacts
with newly synthesized and pre-existent proteins. The deletion of TF or GroEL
changes the DnaK interactome. The deletion of TF causes more proteins to bind
to DnaK, suggesting a functional redundancy of TF and DnaK (Calloni et al.,
The depletion of GroEL causes an accumulation of GroEL substrates, and DnaK and GroEL have \( \sim 119 \) overlapping substrates. Thus, DnaK cooperates with the upstream TF and downstream GroEL (Calloni et al., 2012).

What kinds of substrates do DnaK/DnaJ/GrpE help? What are their properties? DnaK is both a holdase and an unfoldase (unfolds substrates). It is a holdase when it binds to newly synthesized proteins, protecting them from intramolecular and intermolecular interactions, and hands it off to downstream chaperones that help the proteins fold. It is also an unfoldase since it binds misfolded proteins and unfolds them. These proteins can then refold, or again misfolds and rebinds to the chaperone. For example, in vitro, DnaK binds to misfolded luciferase species, and releases them as unfolded proteins, which can refold to its native state (Sharma et al., 2010). Misfolding-prone proteins benefit the most from the KJE system as shown by a simulation with FoldEco (Powers et al., 2012).

### 1.9.3 GroEL/GroES (GroELS) family

GroEL is a 15-nm long cylindrical structure with two large rings (7-9 subunits, \( \sim 60 \) kDa each) (Braig et al., 1994). Each subunit contains an equatorial ATPase domain, intermediate hinge domain, and apical domain (Mayer, 2010). The apical domain provides the entrance to the cavity, and is covered with hydrophobic residues. GroES is a ring (7 units, \( \sim 10 \) kDa each) that binds the
GroEL apical domains (Fig. 1.7). A substrate or folding intermediate binds to the apical domain, which causes 1 ATP to bind to each subunit. This causes a conformational change in the apical domains that expose GroES binding sites. The substrate is then encapsulated with the GroES cap in the cis complex. Upon ATP hydrolysis and GroES dissociation, the substrate leaves the chamber in the trans ring. Substrates that are still misfolded can rebind and refold again in GroEL even without GroES (Horwich et al., 2007).

The GroEL/ES is the only *E. coli* chaperone that is essential for viability (Fayet et al., 1989; Horwich et al., 1993). It is shown to act downstream of TF and DnaK (Ewalt et al., 1997; Houry et al., 1999). Using proteome-wide analysis of GroEL/ES-dependent folding in *E. coli*, ~250 proteins interact with GroEL. Some of these proteins also interact with TF and DnaK (Kerner et al., 2005). GroEL substrates are categorized to classes based on their dependency to GroEL: I) proteins that fold with GroEL/ES-independent pathway, II) proteins that fold with a partial help from GroEL/ES, and III) proteins that are obligate substrates and highly-dependent on GroEL/ES (Ewalt et al., 1997). Approximately 85 proteins are in class III and this contributes to 75-80% of the total mass of GroEL (Kerner et al., 2005). Which particular protein fold does GroEL help for folding? Using a homology-based fold assignment and using SCOP database of structural domains, the GroEL interactors are found to be TIM-barrel fold (enriched in (βα)₈). A more extensive search for a more detailed structural feature of class III substrates has been done; however, it is
unproductive. Researchers speculate that the final structure is not the main determinant for GroEL interaction; instead, it could be the folding intermediates (Kerner et al., 2005). Global aggregation of newly synthesized proteins has been observed when GroEL is deficient, suggesting that GroEL helps a majority of newly synthesized proteins. Here, Chapman and others used a temperature-sensitive lethal E. coli mutant of GroEL and found that most newly made proteins aggregate (Chapman et al., 2006). GroEL/ES dependency is also correlated with aggregation propensity (Ishimoto et al., 2014). In silico, GroEL/ES is predicted to help slow-folding proteins (Powers et al., 2012). Another in silico work describes that the physicochemical properties of proteins determine their chaperone requirements (Tartaglia et al., 2010). Here, class III obligate substrates of GroEL are aggregation-prone, have poor solubility, and have low levels of both mRNA and protein expression.

1.9.4 Disaggregation family, ClpB/DnaK/DnaJ/GrpE

Disaggregation is crucial in reversing protein aggregation by untangling and re-solubilizing aggregates (Mogk et al., 2003). For smaller aggregates, KJE can perform this function (Diamant et al., 2000), but for larger aggregates, ClpB is necessary (Mogk et al., 2003; Weibezahn et al., 2004). ClpB is a hexameric AAA+ protein. It consists of ATP domain with characteristic walker motifs (Walker A and Walker B) as well as sensor 1 and sensor 2 and Arg finger motifs (Doyle et al., 2013). The crystal structure from Thermus thermophilus shows that ClpB has
N-domain, nucleotide-binding domain 1 (NBD1), nucleotide-binding domain 2 (NBD2) and a middle domain (M) that composed of four helices and coiled coil (Fig. 1.8) (Lee et al., 2003). ClpB is important in thermotolerance; its absence (clpB null mutation) causes the cells to have lower growth rate at 44 °C and have a higher death rate at 50°C (Squires et al., 1991). ClpB coordinates with KJE to untangle protein aggregates, thereby preventing aggregation (Mogk et al., 1999). DnaK couples aggregates to ClpB by binding to the M-domain motif 2 and this stimulates the ATPase activity of ClpB (Rosenzweig et al., 2013; Seyffer et al., 2012). Which part of DnaK binds to the M domain? The GrpE binding site on the NBD of DnaK is the same site that interacts with the M domain. GrpE and M compete for DnaK binding. Thus, the proposed model is: DnaK/DnaJ acts on aggregates. Then, DnaK presents aggregates to ClpB by binding to M domain. Once bound to the M-domain, ATPase power of ClpB is increased. ClpB unfolds aggregates. Once unfolded, the substrates can refold or recovered by KJE (Rosenzweig et al., 2013).

1.9.5 IbpA/IbpB

Inclusion body-binding protein A (IbpA) and IbpB in E. coli are associated with inclusion bodies and aggregates during heat stress (Allen et al., 1992; Laskowska et al., 1996). They are present in the cell under physiological conditions, and also upon heat stress, their levels increase up to ~300-fold more (Richmond et al., 1999). Their multimeric sizes can be as large as 3 MDa. The
two proteins physically interact and they cooperate with DnaK/DnaJ and ClpB in resolubilizing aggregation, synergistically (Matuszewska et al., 2005).

1.9.6 HtpG

High temperature protein G (HtpG) is the *E. coli* chaperone that is homologous to the eukaryotic Hsp90 (Bardwell and Craig, 1987, 1988). Despite the extensive studies about Hsp90, the function of HtpG remains unclear. It is an ATP-dependent chaperone that is abundant in *E. coli*, and also up-regulated during heat shock (Grudniak et al., 2013). It is not essential for viability; however, its deletion causes slower growth at higher temperatures (Bardwell and Craig, 1988) and increase in protein aggregation at 42 °C (Thomas and Baneyx, 2000). The only client protein that has been identified to date is the ribosomal protein L2 (Motojima-Miyazaki et al., 2010), which activates the ATPase activity of HtpG. In addition, HtpG regulates the clustered regularly interspaced short palindromic repeats (CRISPR) system by maintaining functional levels of Cas3 (a protein essential for the activity of CRISPR) (Yosef et al., 2011).

1.9.7 Degradation enzymes

Protein degradation in *E. coli* is crucial in minimizing aggregation (Rosen et al., 2002). In *E. coli*, there are four classes of energy-dependent proteases that have been found (Lon, FtsH, ClpAP/XP, HslUV), and these have also counterparts in the mitochondria and chloroplasts. A single polypeptide encodes
both ATPase and proteolytic sites for Lon and FtsH, while two subunits encode the ATPase and proteolytic sites of ClpAP/XP and HslUV (Gottesman, 2003). FtsH is the only essential protease in *E. coli*. The four classes share the following properties: 1) they depend on ATP-hydrolysis for degradation, 2) two separate domains for ATPase and proteolytic sites; ATPase domain affects substrate specificity, 3) called ‘self-compartmentalizing’ proteases because of their organization as a multi-subunit complex where the access to the proteolytic active is gated by the ATPase domain, and 4) products are short peptides (10-15 amino acids) (Gottesman, 2003). The overall mechanism is: first, the substrate binds to the ATPase domain, and then after several cycles of ATP hydrolysis, it undergoes conformational change, and then is unfolded (sometimes), and finally is degraded in the proteolytic chamber (Gottesman, 2003). Proteases may share the same substrates depending on the substrate’s recognition motifs and conditions (Smith et al., 1999; Tsilibaris et al., 2006; Vera et al., 2005; Wu et al., 1999).

1.9.7.1 Lon

The *E. coli* Lon protease is the first ATP-dependent protease to be studied (Chung and Goldberg, 1981; Goldberg et al., 1994; Swamy and Goldberg, 1981). Mutations in the *lon* gene cause an increase in the levels of abnormal proteins, and a decrease in the elimination of these proteins (Goff and Goldberg, 1987). This suggests that Lon is the predominant protease in *E. coli* (Gottesman, 2003).
Lon is an oligomer of identical six subunits (~87 kDa each), with a serine-lysine catalytic dyad (Botos et al., 2004). There is no detailed structure of an intact Lon from *E. coli* yet, but a crystal structure of Lon shown in Fig. 1.9 from *Thermococcus onnurineus* (*TonLon*) (Cha et al., 2010) gives us insight into the general structure of the protease. It forms a hexagonal cylinder with a large sequestered chamber. It consists of: 1) an amino-terminal domain (in *TonLon* is the insertion domain (I) that contains the membrane anchor), which is important for substrate recognition, allosteric regulation, and chaperone activity (Wohlever et al., 2014); 2) AAA+ ATPase domain (A) that contains the ATP-binding motif for substrate unfolding; 3) and a protease domain (P) that has the degradation chamber.

*E. coli* Lon degrades naturally unstable and misfolded or abnormal proteins (Christensen et al., 2004; Dervyn et al., 1990; Goff and Goldberg, 1987; Griffith et al., 2004; Langklotz and Narberhaus, 2011; Mizusawa and Gottesman, 1983; Shah and Wolf, 2006; Stout et al., 1991; Van Melderen et al., 1996). Little is known about its substrate preference. It can degrade a folded protein like the α-crystallin domain of IbpA and IbpB (Bissonnette et al., 2010). It also recognizes either the C or N terminal regions of a protein like the His at the C-terminal region of SulA (Ishii and Amano, 2001) and the first 24 N-terminal amino acids of UmuD (Gonzalez et al., 1998), as well as the first 21 N-terminal residues of SoxS (Shah and Wolf, 2006). What Lon recognizes in abnormal/damaged proteins is not clear. It could access target sequences that are exposed in unfolded polypeptides but
inaccessible in the native structure. In the unstructured β-galactosidase, Lon recognizes specific sequences, mostly aromatic residues (Gur and Sauer, 2008).

1.9.7.2 ClpAP/XP

ClpA is the ATPase domain while ClpP is the proteolytic domain of ClpAP. ClpX also associates with ClpP and it has another substrate specificity compared to ClpA. ClpP is a serine protease whose active site is at Ser-111 (Maurizi et al., 1990; Maurizi et al., 1998). ClpX and ClpA both assemble as hexameric rings and interact with the seven-subunit rings of ClpP. At the entrance of the proteolytic chamber are the ATPase rings. ClpA and ClpX may also act as chaperones, just by unfolding substrates without proteolysis (Singh et al., 2000; Wickner et al., 1994).

1.9.7.3 ClpYQ/HslUV

ClpY (HslU) is the ATPase domain, which is similar to ClpX. ClpQ (HslV) is the proteolytic domain with a threonine active site. ClpQ is a double-ring, with six identical subunits. ClpY has also six subunits. Unlike ClpAP and ClpXP with ring symmetry mismatch, ClpYQ rings do match (Kessel et al., 1996).
1.9.7.4 FtsH

FtsH is a cytoplasmic Zn metalloprotease that is anchored to the inner membrane. It degrades the heat-shock transcription factor, σ^{32} (Tomoyasu et al., 1995).

Degradation takes part in kinetic partitioning of substrates, as discussed earlier (Fig. 1.3). When are proteins being degraded instead of being folded or aggregated? How do degradation enzymes modulate the fate of proteins with different properties? Many substrates have signals for degradation by a particular protease. Usually, the hydrophobic patches on a protein are being recognized by proteases such as Lon. The N-terminal amino acid is recognized by some proteases (N-end rule) (Varshavsky, 1992). But other proteases bind anywhere on the protein depending on the location of the recognition sequence. The ClpAP system degrades repA and the recognition sequences are between amino acids 1 and 15 (Hoskins et al., 2000). The ClpXP system degrades lambda O protein and the sequences are in the first 18 amino acids (Gonciarz-Swiatek et al., 1999). In bacteria, some substrates are tagged with the tmRNA or SsrA tag, which is the 11-amino acid tag (AANDENYALAA) added to the C-terminus of polypeptides bytmRNA (Keiler et al., 1996; Withey and Friedman, 2003). Most proteins with the tag are degraded by ClpXP, and to a lesser extent by ClpAP (Gottesman et al., 1998). Another tagging system in bacteria is the binding of polyphosphate to ribosomal proteins and Lon, stimulating Lon degradation (Kuroda et al., 2001). But little is known about the mechanism of stimulation. Adaptor proteins also
regulate proteolysis by either facilitating degradation or preventing degradation of specific proteins (Dougan et al., 2002).

1.10 Cooperation among chaperones and degradation enzymes in *E. coli*

As discussed earlier, chaperones and degradation enzymes of the prokaryotic and eukaryotic PN cooperate to maintain levels of functional proteins in the cell. Specifically, in *E. coli*, studies have shown a variety of important partnerships among chaperones and degradation enzymes. However, most of these studies are only functional relationships. The mechanisms by which they collaborate are unclear.

TF and DnaK cooperate to aid the folding of newly synthesized proteins (as discussed earlier) (Deuerling et al., 1999). In addition, TF and GroEL function together to promote the degradation of the fusion protein, CRAG. High levels of TF enhanced the capacity of GroEL to bind to CRAG, and, thus, increasing degradation. Low levels of TF reduced degradation. In addition, high levels of TF stimulated GroEL’s ability to bind and unfold other proteins (Kandror et al., 1997; Kandror et al., 1995). Furthermore, GroEL has overlapping substrates with TF and DnaK. In cells without TF and DnaK, GroEL interactors increased, but the amounts of Class III proteins was reduced. In addition, TF and DnaK can help in the folding of class I and II. They also facilitate GroEL to efficiently fold the class III substrates by removing the load of class I and II substrates on GroEL (Kerner et al., 2005). The overproduction of both DnaK/DnaJ and GroEL/ES systems
prevented the aggregation of proteins in *E. coli* rpoH mutants that were deficient in heat shock proteins (Gragerov et al., 1992). It has also been shown that DnaK and Lon act synergistically to prevent aggregation (Tomoyasu et al., 2001). IbpA/B are substrates of Lon (Bissonnette et al., 2010), and it is suggested that probably Lon degrades the IbpA/B (with its substrates) or IbpA/B (only, no substrates). Small heat shock proteins of *E. coli* (IbpA/B) cooperate with ClpB, and DnaK system in reversing aggregation *in vitro* and *in vivo* (Mogk et al., 2003). DnaJ is required for the degradation of PhoA (Huang et al., 2001). DnaJ protects presecretory proteins from being degraded by Lon (Sakr et al., 2010). HtpG, together with ClpB, helps in the de novo folding of proteins in stressed cells (Thomas and Baneyx, 2000). Lon and GroEL/ES compete for binding to folding intermediates of *E. coli* dihydrofolate reductase (DHFR) (Bershtein et al., 2013). TF + KJE and TF + GroELS cause a slight increase in solubility, whereas KJE + GroELS is more effective in increasing solubility. TF has only a marginal effect on solubility (Niwa et al., 2012).

### 1.11 Heat shock response (HSR) in *E. coli*

The transcription of the cytosolic proteostasis network is regulated by the heat shock response signalling pathway (Lindquist, 1986; Zhang et al., 2014). The heat shock response (HSR) is a cellular response to temperature increase. In *E. coli*, it is regulated by σ^{32} transcription factor that causes the up-regulation of most components of the PN σ^{32} regulon. There are three major modes of
regulation for $\sigma^{32}$ (Guisbert et al., 2008) (Fig. 1.10). First, at high temperatures, translation of $\sigma^{32}$ is increased. This causes the up-regulation of chaperones and degradation enzymes. When unbound (or excess) DnaK/DnaJ and GroEL/ES chaperones bind to $\sigma^{32}$, $\sigma^{32}$ is inactivated. Second, $\sigma^{32}$ is degraded by FtsH protease with the help of other chaperones. Third, a decrease in temperature alters the rate of translation for $\sigma^{32}$ (Guisbert et al., 2008). Many proteins in E. coli (mostly, chaperones) are under the control of $\sigma^{32}$. During normal conditions, the concentrations of the PN components are sufficient to aid the folding of proteins. However, during stress (e.g. heat) or when more proteins are synthesized in the cell, the basal levels of the PN components are insufficient to aid folding, which may cause aggregation. Some chaperones are limiting (Heldens et al., 2010), and this causes them to be saturated. In order to increase proteostasis capacity, the induction of HSR helps upregulate PN components. Each component is up-regulated to different extents (Zhang et al., 2014).

1.12 Protein’s intrinsic biophysical properties and the cellular PN dictate protein folding fate through kinetic partitioning

As discussed earlier, a newly synthesized polypeptide (U) can have three fates: it can fold to native (N) and be functional, it can misfold (M) and/or aggregate (A), or it can be degraded (Fig. 1.2). The fate is dictated by both the proteins’ intrinsic biophysical properties and the cellular PN (Cho et al., 2015) (Fig. 1.11). Kinetic partitioning exists among the processes of folding, misfolding,
aggregation, degradation, and chaperone interaction (Kim et al., 2013; Wickner et al., 1999). This determines whether a protein is folded, aggregated, or degraded at a given time. For example in *E. coli*, DnaK, DnaJ, and GrpE (the *E. coli* Hsp70/Hsp40/nucleotide exchange factor) decrease the concentrations of misfolded states by unfolding these states (Calloni et al., 2012; Mayer and Bukau, 2005; Sharma et al., 2010). GroEL and GroES (the *E. coli* Hsp60/Hsp10) encapsulate unfolded states to promote their folding (Chapman et al., 2006; Horwich and Fenton, 2009). DnaK, DnaJ, GrpE with ClpB untangle and solubilize aggregates (Doyle et al., 2013). Lon protease prevents aggregation by degrading misfolded proteins (Gur and Sauer, 2008).

Previous studies have investigated the relationships of properties of client proteins and their folding fates, as well as their sensitivities to chaperones. The correlation between thermodynamic stability and aggregation using *E. coli* as host organism has been tested. Destabilized variants of single point mutants of HypF-N, an *E. coli* protein domain, invariably aggregated in cells (Calloni et al., 2005). The extent of destabilization (as a result of mutations) is only roughly correlated with the aggregation propensity of cellular retinoic acid-binding protein-1 (CRABP1) (Ferrolino et al., 2013). On the other hand, the levels of folded recombinant p53 in *E. coli* are correlated with its intrinsic thermodynamic stability (Mayer et al., 2007); however, it is not clearly shown how the levels of aggregates are affected. By contrast, the *in vitro* thermodynamic stability and *in vivo* solubility of SH3 domains have a strong correlation (Castillo et al., 2010).
They conclude that intracellular aggregation propensity is under thermodynamic rather than kinetic control. Using Foldeco model of protein folding in *E. coli*, predictions of substrate preferences of KJE and GroELS have been studied for clients with different biophysical profiles, and found that misfolding-prone proteins benefit the most from the KJE system while slow-folding proteins benefit the most from GroELS (Powers et al., 2012). Furthermore, integrating folding energetics with the PN capacity should be considered since the two are interdependent of each other (Powers et al., 2009). The FoldEx model captures how the inherent energetics of polypeptide chains are influenced by the PN components in the endoplasmic reticulum (Wiseman et al., 2007). The model is used to define “minimal export threshold” which is a boundary in the 3D space defined by protein stability, folding rate, and misfolding rate. The boundary depends on the proteostasis network capacity. Proteins with energetics that are within the boundary are exported, but those whose energetics outside the boundary are not. In another study using computation, protein unfolding rates correlate with folding rates, which means that proteins that fold quickly also unfold quickly and those that fold slowly also unfold slowly (Broom et al., 2015).

It is not clear how PN modulates the folding fate of different proteins with intrinsic biophysical properties. How does the PN work together as a whole? How do protein’s biophysical properties affect its fate inside the cell? These are the questions that we investigated in this thesis. We used both experimental and computational approach to address these questions.
1.13 FoldEco model of proteostasis in *E. coli*

We use FoldEco, a computational model of the proteostasis network (PN) of the *E. coli* cytoplasm (Fig. 1.11), to understand how the PN modulates the folding fates of proteins with different biophysical properties. It is a holistic approach to understand proteostasis, in which the cooperation and competition of the different processes and components of the PN are investigated when they are working simultaneously. Here, the folding fates of proteins-of-interest (“client proteins”) are tracked as they are produced in the cytoplasm.

This section on FoldEco is based on the work of Powers and colleagues (Powers et al., 2012).

1.13.1 Parts of FoldEco

FoldEco consists of five systems: 1) synthesis and folding, 2) chaperoning by DnaK/DnaJ/GrpE or KJE system, 3) chaperoning by GroEL/GroES or GroELS system, 4) disaggregation by ClpB/DnaK/DnaJ/GrpE or B+KJE system, and 5) degradation (Powers et al., 2012).

Protein synthesis and folding (Fig. 1.12, light red area) start with a nascent client protein (protein “*i*”) that is unfolded (*U*$_i$) coming off from the ribosomal exit tunnel of a translationally active ribosome (R$_{a,i}$) to form a ribosome:unfolded protein complex (R$_{a,i}:U_i$). This complex can also bind to trigger factor (R$_{a,i}:U_i:T$). When the protein is completely synthesized, it is released from the ribosome as
an unfolded protein (U_i) or still bound to trigger factor (U_i:T). The unfolded protein then reversibly folds to native (N_i) state or misfolds to misfolded (M_i) state. Misfolded states can either self-associate to form aggregates, A_{i,j} (where “j” are the number of monomers in the aggregate) through nucleated polymerization or bind to chaperones (Powers et al., 2012).

The KJE system provides a channel for misfolded states (M) to unfold (Fig. 1.12, light blue area). Here, unfolded (U) or misfolded (M) states enter the system and U is released. First, U_i or M_i binds to dimeric DnaJ (J_2) or ATP-bound DnaK (K_T) to form a binary complex (J_2:U_i/M_i or K_T:U_i/M_i, “/” indicates “either-or”). This complex then binds to either J_2 or K_T to form a ternary complex (K_T:U_i/M_i:J_2). U or M preferentially binds to DnaJ because K_T binds weakly to substrates (Mayer et al., 2000). Upon DnaJ binding, ATP hydrolysis is stimulated by DnaK, which causes a conformational change within DnaK. This change in DnaK is transmitted to the bound client protein, and causes the client protein to unfold. DnaJ is released from the K_D:U_i:J_2 complex (K_D indicates the ADP-bound DnaK) to give K_D:U_i. Upon dimeric GrpE (E_2) binding to the K_D:U_i complex, ADP is exchanged to ATP to give K_T:U_i:E_2. When GrpE is dissociated, K_T:U_i complex is left. Upon dissociation of K_T, the unfolded protein, U_i, is released. In FoldEco, the total number of KJE cycles depends on the relative rates of entry of client proteins to the KJE cycle and folding. The model is partly based on that of Hu and others (Hu et al., 2006; Powers et al., 2012).
The GroEL system (Fig. 1.12, light green area) promotes folding by encapsulating unfolded (U) or misfolded (M) states providing an isolated environment for the protein to fold (Horwich et al., 2006). The cycle starts with either U or M binding to the cis ring of GroEL (either ATP-bound, GrLT or not, GrL) to give GrLT:U/M or GrL:U/M. There should only be a few ATP-free GroEL, given the typical ATP concentrations in vivo (Tyagi et al., 2009). Upon ATP binding, GrL:U and GrL:M are both forcibly unfolded to give GrLT:U (Lin et al., 2008). GroES then caps the GrLT:U/M complex, which causes the client protein to be released into the cis cavity (GrLT:U/M:GrS) to fold or misfold (GrLT:U/M/N:GrS). Upon ATP hydrolysis in the cis ring, there are two possibilities for GrLD:U/M/N:GrS: 1) binding of ATP in the trans ring yields GrLD:U/M/N:GrS:GrLT and causes the release of ADP, GrS, and the client protein from the cis ring, and 2) binding of U or M client protein in the trans ring forms GrLD:U/M/N:GrS::GrL:Uk/Mk (“k” indicates that the client protein can be different from the one in the cis cavity), and upon ATP binding, ADP, GrS, and client protein is released from the cis ring. GrLT:Uk is left and ready to reenter the cycle. The total number of GroELS cycle that a client protein experiences depends on the relative rates of entry into the cycle and folding (Jewett and Shea, 2008; Powers et al., 2012; Tehver and Thirumalai, 2008).

The ClpB/DnaK/DnaJ/GrpE (B+KJE) system (Fig. 1.12, light yellow area) is important in reversing protein aggregation. In FoldEco, the model for the disaggregation system is based on studies that suggest that aggregates are
prepared by KJE for B (Acebron et al., 2008; Acebron et al., 2009) and then B removes monomers from aggregates. The cycle begins with the binding of J₂ or ATP-bound DnaK to aggregates (Aᵢ,j) to form the complex, J₂:Aᵢ,j or Kₜ:Aᵢ,j (the stoichiometry of chaperone:aggregate is assumed to be 1:1). This complex binds to either Kₜ or J₂ to form the ternary complex, Kₜ:Aᵢ,j:J₂. ATP hydrolysis yields the complex K₉:Aᵢ,j*:J₂, in which the aggregate is prepared for ClpB binding (as indicated by *). J₂ is released to give K₉:Aᵢ,j*, which has two possible pathways. First, GrpE binds to K₉:Aᵢ,j* and induces nucleotide exchange to yield Kₜ:Aᵢ,j:E₂.

Upon the release of GrpE, either a monomer is lost to give Kₜ:Aᵢ,j⁻¹ and Uᵢ (for small oligomers, j < 4 in FoldEco) or no monomer is lost to yield Kₜ:Aᵢ,j (for larger aggregates, j > 4). Kₜ:Aᵢ,j can enter the cycle again or can dissociate. Second, ClpB binds to K₉:Aᵢ,j* to produce K₉:Aᵢ,j*:B. Upon GrpE binding, E₂:Kₜ:Aᵢ,j*:B is formed. When E₂ and Kₜ dissociate, Aᵢ,j*:B is left. The final step involves the translocation of a monomer through the central pore of B and the dissociation of the complex to yield Aᵢ,j⁻¹, Uᵢ and B (Powers et al., 2012).

FoldEco consists of two degradation pathways-- the Lon and ClpAP-type proteases (Fig. 1.12, light purple). In the pathway with Lon, it starts with the interaction of Lon with U or M states to form a reversible complex, Lon:Uᵢ/Mᵢ. The client protein is then transferred to the proteolytic chamber (for misfolded protein, it is forcibly unfolded first) to yield Lon:Uᵢ* (* indicates that the substrate is committed to be degraded). Here, degradation is processive, in which the ATPase domains of Lon feed the protein into the proteolytic chamber. In the
second pathway, ClpAP-type proteases (Dn) degrade native proteins (N) that have degradation signals or degrons. First, Dn binds to N (client protein may be delivered through an adaptor such as ClpS) to form the reversible complex, Dn:N. Then the client protein is forcibly unfolded and transferred to the proteolytic chamber to form Dn:N+. Degradation is also processive (Powers et al., 2012).

1.13.2 Implementing and parametrizing FoldEco

FoldEco consists of ordinary differential equations describing the time-dependent concentrations of each species (e.g. concentration of unfolded U, misfolded M, or native N states). In order to solve for the systems of equations, model parameters such as rate constants and initial concentrations are required. The parameters used in FoldEco are estimated from literature data, experiments, or models. Rate constants and folding parameters as well as the binding affinities of substrates to chaperones and on/off rates, which are protein-specific are estimated from literature. Initial concentrations are also derived from literature. Powers and colleagues converted primary data (from previous studies) of protein abundances for the major chaperones and degradation enzymes in *E. coli* to concentrations, and calculated the geometric mean (Table 1.1). Chaperone levels vary. They also differ in their active forms. Trigger factor (TF) and DnaK are active as a monomer, DnaJ and GrpE are dimer, GroEL is 14-mer, GroES is 7-mer, Lon and ClpB are 6-mer. The concentrations of the active PN components are: [Tf] = 20 μM, [DnaK] = 30 μM, [DnaJ] = 0.5 μM, [GrpE] = 7.5 μM, [GroEL] =
In running FoldEco, users can change these parameter values depending on the proteins-of-interest. The outputs are the concentrations of all species at a given time (Powers et al., 2012).

Like any biological network model, FoldEco has some approximations and assumptions (Powers et al., 2012). First, FoldEco is a deterministic model (not stochastic), in which the average state of proteostasis in populations of *E. coli* (not individual cells) is assessed. This captures some results of *in vivo* experiments, in which variables of interest are measured through the average of populations rather than individual cells. Second, FoldEco assumes that *E. coli* is a single compartment (the cytoplasm), in which components are mixed freely and their concentrations are the same across the cell. This is to simplify the model and not complicating it with spatial heterogeneous concentrations that we do not have estimates yet. Third, only the PN components of the cytoplasm, and not membrane or periplasmic proteins, are being studied. Fourth, FoldEco uses parameters measured *in vitro* to account for processes *in vivo*. Some *in vitro* and *in vivo* parameters (e.g. kinetic and thermodynamic properties) differ, while others are similar. In FoldEco, a conservative approach is used such that *in vitro* data without correction are utilized. These values will be replaced in FoldEco when specific information on rate and equilibrium constants become available. Fifth, in FoldEco, bacterial growth is not considered, which means that bacteria are not growing. This assumption may reflect bacteria that are in the lag or
stationary phase or when overexpressing heterologous proteins *in vivo*, in which growth is strongly inhibited. Sixth, the background proteome, which could compete with a client protein of interest for PN machineries, is absent in FoldEco. Given the FoldEco condition in which bacteria are not growing, only few proteins are produced and do not outcompete client proteins for PN components. Seventh, stress response (such as heat shock) is not included in FoldEco, but it can be estimated by increasing the concentrations of the relevant PN components. Further, in FoldEco, ATP concentration is assumed to be constant, and the concentrations of mRNA are supposed to be at a steady state. In addition, a simplified model of translation is used, in which translation is a three-step process: initiation of translation, elongation of nascent chain until it emerges from the ribosomal exit tunnel, and final translation of the remaining mRNA. FoldEco does not account for co-translational folding and misfolding. Furthermore, free aggregates composed of two or more monomers as well as those in complex with chaperones are considered as individual species. Lastly, some PN components (such as IbpA/IbpB and HtpG) are not included in FoldEco. However, they do not have a very significant effect on proteostasis (Bardwell and Craig, 1988). In the future, these will be incorporated when more mechanistic information will be available (Powers et al., 2012).

FoldEco is a powerful tool to rationalize experimental observations and to generate hypotheses. We utilize FoldEco to do simulations with conditions similar to those in the lab, and we use these to guide other conditions to test
experimentally. We also input our *in vivo* data to the FoldEco program, and the program gives us the folding properties of the proteins. This integration of knowledge from computational and experimental approaches helps us understand the proteostasis network better (Powers et al., 2012).

### 1.14 Statement of thesis

Despite our knowledge about the main functions of the individual components of the PN, little is known about how the whole PN functions as a system. In addition, it is not clear how the biophysical properties of a protein affect its folding fate in the context of the whole PN. How does the PN, consisting of molecular chaperones and degradation enzymes, modulate the folding fate of proteins with different intrinsic biophysical properties? We hypothesized that molecular chaperones and degradation enzymes modulate the folding fate of a protein by working together to maintain proteostasis. The sensitivity of a protein to a particular chaperone or degradation enzyme depends on the protein’s intrinsic biophysical properties, and on the concentrations of the protein and chaperones and degradation enzymes at any given time.

Understanding protein folding in the cell requires a holistic approach. We need to understand the cooperation and/or competition among the PN components when they are present and operating simultaneously. In order to understand the regulation of protein folding fate, we utilize an integrated computational and experimental approach. We use *E. coli* because it is a simple
model organism to understand protein folding in the cell. We perform wet lab experiments in *E. coli* by expressing proteins with different biophysical properties and determining their folding fates *in vivo*. In addition, we use FoldEco to provide us a better understanding about the mechanisms of the PN as a whole.

In Chapter 2 of this thesis, we investigate how the intrinsic biophysical properties of proteins (e.g. rate and equilibrium constants) dictate folding fate under basal levels of the PN components (molecular chaperones and degradation enzymes). We express in *E. coli* cells several variants of a model protein that have unique biophysical properties, and determine their fates at a given time. In combination with FoldEco, we are able to determine the effects of rate and equilibrium constants on folding fates of proteins. In Chapter 3, we explore how varying levels of some of the major PN components affect the folding fates of model proteins with low stabilities and high aggregation propensities. We express model proteins in *E. coli* cells and determine their folding fates under conditions of varying concentrations of individual chaperones and degradation enzymes or in combinations. Using FoldEco, we are able to determine the individual and collective contributions of chaperones and degradation enzymes to proteostasis. We have used FoldEco to perform simulations on the folding fates of proteins-of-interest given their biophysical properties, and we have also elucidated the biophysical properties of proteins given their folding fates *in vivo* (Powers et al., 2012).
Figure 1.1 Schematic depiction of the complex environment inside the *E. coli* cell where protein folding takes place

Despite the complex environment, proteins can still fold very well inside the cell. The reason for this is the presence of the proteostasis network, consisting of molecular chaperones and degradation enzymes. The protein of interest (in orange) as it is being made from the ribosome (ribosomal proteins are purple, all RNA in salmon) is helped by molecular chaperones (GroEL in green, DnaK in red, and trigger factor in yellow). (Reproduced from (Gershenson and Gierasch, 2011)).
Figure 1.2 Schematic diagram of folding fates of proteins

An unfolded polypeptide when it comes off from the ribosome can have three fates: 1) folded and become functional, 2) misfolded and/or aggregated, and 3) degraded.
Figure 1.3 A protein’s energy landscape of folding and aggregation

As an unfolded polypeptide folds to its native state (purple area) via intramolecular contacts, it explores a funnel-shaped energy landscape. The protein adopts different conformations such as folding intermediates, partially folded states, or misfolded states. Species (usually the partially folded or misfolded states) can be kinetically trapped and they need to cross free-energy barriers to reach a favorable downhill path. When these trapped species form intermolecular interactions, aggregates (amorphous aggregates, oligomers, or amyloid fibrils) are formed (pink area). (Reproduced from (Hartl et al., 2011)).
Figure 1.4 Chaperone pathways in the cytosol

The general pathway of protein folding in the cytosol of a) bacteria, b) archaea, and c) eukarya is conserved. Newly synthesized polypeptides first interact with ribosome-associated chaperones (e.g. trigger factor (TF) in bacteria, nascent-chain-associated complex (NAC) in archaea and eukarya) that protect the new polypeptide from intramolecular or intermolecular interactions. Longer nascent chains interact with Hsp70 (DnaK in bacteria, Hsp70 in eukarya), Hsp40 (DnaJ in bacteria) and nucleotide exchange factors (GrpE in bacteria) for co- and post-translational folding. In archaea, prefoldin (PFD) acts downstream of NAC. Then, these partially folded states are transferred to the chaperonins (GroEL/GroES in bacteria, thermosome in archaea, and tailless complex polypeptide-1 (TCP1) ring complex (TriC)/chaperonin-containing TCP-1 (CCT) in eukarya. Hsp90 system (with additional co-factors) mediates folding of substrates from the heat shock cognate 70 (Hsc70). The insert in panel c shows the ribosome-associated complex (RAC) in fungi, which consists of Ssz1 (a specialized Hsp70) and zuotin (Hsp40). RAC and the Hsp70 isoform (Ssb) assist nascent chain folding. The estimated protein flux is shown as percentages. (Reproduced from (Kim et al., 2013)).
**Figure 1.5 Structure of *E. coli* trigger factor** (PDB 1W26, (Ferbitz et al., 2004))

TF consists of three domains: 1) N-terminal domain for ribosome binding, 2) peptidyl-prolyl isomerase (PPIase) domain for nascent chain binding, and 3) C-terminal domain for primary binding site of nascent chain.
Figure 1.6 Structure of Hsp70

A. In the ADP-bound state (PDB 2KHO, (Bertelsen et al., 2009)) of *E. coli* DnaK, the nucleotide-binding domain (NBD) and substrate-binding domain (SBD) are independent of each other, linked by a linker. B. A representation of the ATP-bound state (PDB 2QXL, (Liu and Hendrickson, 2007)) for yeast Sse1 shows that the two domains are docked.
Figure 1.7 Structures of GroEL and GroES (PDB 1PF9, (Chaudhry et al., 2003))

GroEL consists of two large rings (7 subunits each). Each subunit has an apical domain, hinge domain, and ATPase domain. GroES also forms a ring with 7 subunits each. It binds to the apical domains of GroEL.
Figure 1.8 Structure of ClpB from *Thermus thermophilus* (PDB 1QVR, (Lee et al., 2003))

ClpB consists of an amino-terminal domain (N-domain), two nucleotide-binding domains (NBD1 and NBD2), and a middle domain.
Lon is a hexamer of identical subunits (~87 kDa each). **A.** Each subunit consists of an insertion domain (I), AAA+ ATPase domain (A), and a protease domain (P). **B.** Lon shown as a hexamer. (Image from the RCSB PDB [www.rcsb.org](http://www.rcsb.org) of PDB ID 3K1J).

**Figure 1.9 Structure of Lon from Thermococcus onnurineus** (PDB 3K1J, (Cha et al., 2010)}
Figure 1.10 Three modes of regulation of $\sigma^{32}$

1) Translation increased at high temperatures, 2) Degradation by FtsH and other chaperones, 3) Inactivation mediated by chaperones DnaK/DnaJ and GroEL/GroES. (Reproduced from (Guisbert et al., 2008))
When an unfolded (U) polypeptide is synthesized from the ribosome (Synthesis), it can have different folding fates: It can fold into native (N), it can misfold (M), and if M self-associates, it can aggregate (A), or it can be degraded (D). The fate of the protein is dictated by its intrinsic biophysical properties (red text). In addition, the PN components modulate its fate in vivo. Kinetic partitioning affects folding. For example, In E. coli, GroEL and GroES (Hsp60/Hsp10) promote folding by encapsulating unfolded states and helping them fold to native. DnaK, DnaJ, and GrpE (Hsp70/Hsp40/Nucleotide exchange factor) rescue misfolded states by binding to M and helping them to become unfolded. ClpB and KJE collaborate to untangle and solubilize aggregates. (Reproduced from (Cho et al., 2015)).
Figure 1.12 The FoldEco model of proteostasis network (PN) in *E. coli*

The five parts of FoldEco are: 1) synthesis and folding (light red), 2) DnaK/DnaJ/GrpE chaperone system (light blue), 3) GroEL/GroES chaperone system (light green), 4) ClpB/DnaK/DnaJ/GrpE disaggregation system (light yellow), and 5) degradation (light purple). The free states (bold fonts) are denoted as: unfolded (U<sub>i</sub>), misfolded (M<sub>i</sub>), native (N<sub>i</sub>), and aggregated (A<sub>i,j</sub>). Complexes between PN components and client proteins are separated by colons (key in upper right). The subscripts “T” and “D” are for ATP- and ADP-bound states, respectively. (Reproduced from (Powers et al., 2012)).
Table 1.1 Concentrations of some PN components

The concentrations are converted from primary data of five databases. The values are when they are monomer or active. (Modified from (Powers et al., 2012)).

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration (μM), monomer</th>
<th>Concentration (μM), active</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger factor, active as monomer</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>DnaK, active as monomer</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>DnaJ, active as dimer</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>GrpE, active as dimer</td>
<td>15</td>
<td>7.5</td>
</tr>
<tr>
<td>GroEL, active as 14-mer</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>GroES, active as 7-mer</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>ClpB, active as 6-mer</td>
<td>1.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Lon, active as 6-mer</td>
<td>1.8</td>
<td>0.3</td>
</tr>
</tbody>
</table>
CHAPTER 2

INFLUENCE OF INTRINSIC BIOPHYSICAL PROPERTIES ON PROTEIN FOLDING FATE IN E. COLI UNDER THE BASAL PROTEOSTASIS NETWORK

This chapter is the result of collaboration with Evan T. Powers and Lila M. Gierasch.

This chapter investigates how the kinetic and thermodynamic properties of proteins determine their folding fates in the cell, using the FoldEco model and testing the predictions with wet lab experiments.

2.1 Introduction

A newly synthesized polypeptide (U) can have different folding fates: 1) it can fold to native (N) and be functional, 2) it can misfold (M) and/or aggregate (A), or it can be degraded. Aggregated and degraded proteins are non-functional (Fig. 2.1). Proteostasis is achieved when necessary levels of functional proteins are maintained, and non-functional proteins are minimized (Balch et al., 2008). A protein’s intrinsic biophysical properties and the cellular proteostasis network (PN) determine the folding fate of a protein inside the cell (Cho et al., 2015) (Fig. 2.1). The intrinsic biophysical properties of proteins influence folding, misfolding, and aggregation (Powers et al., 2009; Powers et al., 2012). PN components interact with unfolded, partially or misfolded proteins and help them fold to native, hold them to prevent their self-association and aggregation, or degrade them to lower their concentrations (Cho et al., 2015). Kinetic partitioning occurs among
the processes of folding, misfolding, aggregation, and degradation. The PN, consisting of different processes (biogenesis, conformational maintenance, and protein clearance) as well as various players (molecular chaperones and degradation enzymes), collaborates to maintain proteostasis (Powers et al., 2009; Powers et al., 2012).

While several attempts have been made to explore the correlation of intrinsic properties of proteins (e.g. stability) to their solubility and/or aggregation in *E. coli* (Calloni et al., 2005; Castillo et al., 2010; Ferrolino et al., 2013; Mayer et al., 2007), it is still not clear how intrinsic biophysical properties dictate protein folding fates *in vivo*. What are the limits of proteostasis for an unperturbed proteostasis network? Or in other words, what biophysical properties of proteins can the basal levels of the PN components tolerate? We utilized an integrated computational and experimental approach to address these questions.

We used *E. coli* as a model organism because it is a simple organism, with a PN consisting of the major chaperones (Trigger factor, DnaK/DnaJ/GrpE, GroEL/ES, ClpB/DnaK/DnaJ/GrpE) and proteases (Lon, ClpXP, ClpAP, HslUV), and because of the availability of the FoldEco model of *E. coli*’s PN (Powers et al., 2012). We used the *E. coli* strain, HMS174(DE3), which is under a K-12 background (Novagen) and has Lon protease (Blattner et al., 1997; Marisch et al., 2013; Studier et al., 2009). We utilized FoldEco to predict the folding fates of proteins (i.e. how they partition between folding, aggregation, and degradation) based on our current understanding of how the *E. coli* PN components work. We
performed wet lab experiments to determine the folding fates of proteins with different biophysical properties and compared their behavior to the predictions from FoldEco. The partitioning of a protein among its potential folding fates were measured in terms of the concentrations of soluble, aggregated, and degraded proteins at any given synthesis rate and time. It is assumed that soluble proteins include proteins that are natively folded (N) and those unfolded (U) and misfolded (M) proteins that are free or bound to chaperones making them soluble. Aggregated proteins (A) consist of proteins that are insoluble. Degraded proteins are from U and M states.

Knowledge about the levels of PN components inside the cell is necessary to understanding how the PN works. Inside the cell, the levels of PN components are different (Powers et al., 2009). For example, DnaK levels are high relative to DnaJ and GrpE despite the fact that the three work together as an Hsp70 ‘team’. ClpB works with KJE, but its concentration is also low relative to that of DnaK. GroEL and GroES are present at similar levels (Lu et al., 2007; Powers et al., 2012). Furthermore, PN components differ in their active forms. Trigger factor (TF) and DnaK are active as monomers, DnaJ and GrpE are dimers, GroEL is a 14-mer, GroES is a 7-mer, and Lon and ClpB are 6-mers. The concentrations of the components in an unperturbed E. coli PN that were used in our calculations were taken from Cho et al. (2015) (Table A.1). Here, “unperturbed” or “basal” PN refers to the “adapted-basal” PN, which is the PN in the presence of heterologous expression of proteins (Cho et al., 2015; Gasser et al., 2008; Hoffmann and
In our study, unperturbed or basal PN was also referred to as the “adapted-basal” PN.

To determine the levels of the major chaperones (DnaK and GroEL) and Lon protease in cells expressing the test proteins, we compared levels of DnaK, GroEL, and Lon in cells expressing the test proteins and cells not expressing the test proteins (or cells with empty vector). The levels of DnaK, GroEL, and Lon in cells expressing CRABP1 at 1, 2, and 3 h of induction were ~10-30% lower or ~10-20% higher than those in cells without expressing CRABP1 (Fig. A.1). The PN as a result of these slight perturbations due to expression of test proteins is referred to as the “adapted-basal” PN.

We used a model protein, murine cellular retinoic acid-binding protein-1 (CRABP1) and variants that have different biophysical properties (Budyak et al., 2013). CRABP1 is a member of the intracellular lipid binding-protein family (iLBP) that binds and transports hydrophobic ligands. For CRABP1, its ligand is retinoic acid. CRABP1 is a ten-stranded β-barrel with 136 amino acids (Fig. 2.2) and a molecular weight of ~15 kDa. Our lab has generated a number of single-residue substitutions at different sites of the protein. The pseudo wild-type CRABP1 we have used has a stabilizing mutation, R131Q, and is termed WT* (Zhang et al., 1992). WT* was used as the template for the other CRABP1 variants (Budyak et al., 2013). Their stabilities, folding, and unfolding kinetics have been studied in vitro (Budyak et al., 2013). In addition, the aggregation propensities of regions in the sequences have been determined (Ferrolino et al., 2013). We chose five
variants for our studies: WT*, M9A, V67A, L118V, and Y133S (Fig. 2.2). These variants have distinct properties (Table 2.1). CRABP1(WT*) is stable, with a $\Delta G^\circ$ of $-10.0 \pm 0.2$ kcal/mol at 25 °C ($K_f = 2.17 \times 10^7$). It folds relatively fast ($k_f = 0.77$ s$^{-1}$) and unfolds relatively slowly ($k_u = 1.67 \times 10^{-5}$ s$^{-1}$). CRABP1(M9A) is slightly destabilized ($\Delta \Delta G^\circ = -1.8 \pm 0.2$ kcal/mol). It is a slow folder ($k_f = 0.31$ s$^{-1}$, ~2.5 times slower than WT*) and a slow unfolder ($k_u = 1.01 \times 10^{-5}$ s$^{-1}$). By contrast, CRABP1(V67A) is highly destabilized ($\Delta \Delta G^\circ = -3.3 \pm 0.2$ kcal/mol), but folds fast ($k_f = 0.57$ s$^{-1}$). It unfolds faster than WT* by ~2 orders of magnitude ($k_u = 1.50 \times 10^{-3}$ s$^{-1}$). Furthermore, CRABP1(L118V) is moderately destabilized ($\Delta \Delta G^\circ = -2.3 \pm 0.4$ kcal/mol), slow folder ($k_f = 0.41$ s$^{-1}$), and unfolds 3 times faster than WT* ($k_u = 5.02 \times 10^5$). Lastly, Y133S is highly destabilized ($\Delta \Delta G^\circ = -4.8 \pm 0.2$ kcal/mol). It is the slowest folder ($k_f = 0.08$ s$^{-1}$, 10 times slower than WT*). It unfolds 5 times faster than WT* ($k_u = 8.96 \times 10^5$). The $k_f$ and $k_u$ values for WT* and Y133S were also determined at 30 °C (Table 2.1B). Also at this temperature, Y133S folds 10 times slower than WT*.

2.2 Results

2.2.1 Comprehensive examination of protein folding fate versus biophysical properties using FoldEco

To determine the effect of intrinsic biophysical properties on protein folding fates under the adapted-basal PN, FoldEco simulations were performed. Most of the parameters in FoldEco are independent of, or weakly dependent on, the
nature of the protein substrate (Powers et al., 2012). Only the folding parameters, including the folding rate and equilibrium constants ($k_f$ and $K_f$), misfolding rate and equilibrium constants ($k_m$ and $K_m$), and aggregation rate and equilibrium constants ($k_a$ and $K_a$), and initial concentrations were varied in running simulations. The values for the other parameters were those from literature (Cho et al., 2015; Powers et al., 2012). The synthesis rate was set to 135 μM h$^{-1}$ (a typical protein expression rate in E. coli) (Fig. A.2). Combinations of folding parameters were generated to cover the following ranges (that are physically reasonable): $k_f$: $10^{-3}$ to $10^3$ s$^{-1}$; $K_f$: $10^2$ to $10^8$; $k_m$: $10^{-3}$ to $10^3$ s$^{-1}$; $K_m$: $10^{-3}$ to $10^3$; $k_a$: $10^{-3}$ to $10^3$ μM$^{-1}$ s$^{-1}$; $K_a$: $10^{-3}$ to $10^3$ μM$^{-1}$ (Fig. 2.3). $K_a$ is the inverse of critical concentration. Each was varied in 10-fold steps to make 117,649 ($7^6$). The initial concentrations for the chaperones and degradation enzymes were their levels in the “adapted-basal” PN (Cho et al., 2015) (Table A.1). Degradation rate was set as constant. The model was solved for each set of parameters. The concentrations of soluble, aggregated, and degraded proteins were compared after 3 h (~405 μM of total protein synthesized) as a function of their intrinsic biophysical properties. Results of FoldEco modeling across parameter space are shown in the following sections.
2.2.1.1 Case 1: No aggregation, no misfolding

Fifty-nine percent (59%) of all parameter sets are proteins that do not misfold and do not aggregate, which requires \( k_m < 0.1 \) or \( K_m < 1 \). In these cases, folding only competes with degradation (Fig. 2.4A).

2.2.1.1.1 Partitioning between folding and degradation depends exclusively on rate and equilibrium constants

The partitioning between folding and degradation depends exclusively on the folding rate constant \( (k_f) \) and equilibrium constant \( (K_f) \) (Fig. 2.4B). The contour plots of the fraction of soluble protein (amount of soluble protein/total amount of protein synthesized) as a function of \( k_f \) and \( K_f \) are L-shaped, which indicates that the fraction soluble depends on either the kinetics or thermodynamics of folding, but (for the most part) not on both at the same time. We describe this behavior as “thermo/kinetically limited behavior.” Fraction soluble depends on the limiting aspect of the protein’s biophysical properties. For example, if initially \( k_f = 0.001 \, \text{s}^{-1} \) and \( K_f = \sim 10^{3.5} \), as \( k_f \) increases (from 0.001 to <0.1 \, \text{s}^{-1}) with constant \( K_f \), the fraction of soluble protein increases (Fig. 2.4C); the partitioning into soluble or degraded protein is under kinetic control. Further increase of \( k_f \) (0.1 to <100 \, \text{s}^{-1}) with constant \( K_f \) does not increase the fraction soluble; fraction soluble is independent of \( k_f \). As \( K_f \) increases \((10^{3.5} \, \text{to} \, 10^6)\) and with constant \( k_f \) (100 \, \text{s}^{-1}), the fraction of soluble protein also increases; at this point the fraction soluble or degraded is under thermodynamic control.
2.2.1.1.2 Rate and equilibrium constants affect the fraction of soluble protein similarly

The shapes of fraction soluble vs. $k_f$ (at high $K_f$) or $K_f$ (at high $k_f$) plots are very similar. At the maximum value of $k_f$ or $K_f$ tested, the fraction of soluble protein (or fraction of non-degraded protein) depends on the other parameter that is not maximum. For example, at maximum $K_f$ ($10^8$), as $k_f$ increases from 0.001 s$^{-1}$ to 1000 s$^{-1}$, the fraction of soluble protein increases from 0.1 to 0.9. At maximum $k_f$ (1000 s$^{-1}$), as $K_f$ increases from $10^2$ to $10^8$, the fraction of soluble protein also increases (Fig. A.3). The folding rate and equilibrium constants affect the fraction of soluble protein similarly as shown in the graphs of fraction of soluble protein vs. $k_f$ or $K_f$ (Fig. 2.5A). When plotted in one graph, the equilibrium constant needs to be offset by ~4.5 to 4.6 log units, or a factor of ~38,000, to match the rate constant (Fig. 2.5B). This result implies that there is a line along which the rate and equilibrium constants each give the same fraction soluble. We can map this line on the contour plot of fraction of soluble protein vs. $k_f$ and $K_f$ (Fig. 2.5C). The line is called the “line of thermo-kinetic crossover”. This line separates rate- and equilibrium-limited regions. Above the line, fraction soluble depends on $k_f$ alone, and below the line the fraction soluble depends on $K_f$ alone. Along the line, fraction soluble depends on both $k_f$ and $K_f$. The line of thermo-kinetic crossover is: $\log_{10} k_f = \log_{10} K_f - 4.58$. In the contour plot, using the line of thermo-kinetic crossover, we can collapse the two folding parameter dimensions into one by projecting points above the line and down to the $k_f$ axis and by
projecting points below over to the $K_f$ axis and then offset. When plotted on one axis, the result is the Limiting Folding Parameter (LFP), in which $\log_{10} LFP = \text{Min}[\log_{10} k_f, \log_{10} K_f - 4.58]$. The best fit to the data is:

$$F_s = \frac{1}{1 + 0.23 \times 10^{-0.57 \times \text{Min}[\log_{10} k_f, \log_{10} K_f - 4.58]}}$$

The pre-factors (0.23 and -0.57) in this equation likely reflect the effects of the degradation rate and any other PN components or processes on the folding vs. degradation partitioning.

2.2.1.2 Case 2: Misfolding, but no aggregation

Proteins that misfold but do not aggregate are characterized by $k_m \geq 0.1$ s$^{-1}$ and $K_m \geq 1$, $k_a < 10$ $\mu$M$^{-1}$ s$^{-1}$ or $K_a < 10$ $\mu$M$^{-1}$. “No aggregation” means that less than 5 $\mu$M of aggregates are produced at 3 h. Thirty-eight percent (38%) of parameter sets are in this category (Fig. 2.6A).

2.2.1.2.1 Without aggregation, misfolding increases degradation

For proteins that misfold but do not aggregate, folding still only competes with degradation, but degradation is increased because the M state is susceptible to degradation. An example is shown in Fig. 2.6B. The fraction soluble for proteins that have higher propensity to misfold ($k_m = 100$ s$^{-1}$ and $K_m = 100$) is lower compared to those that do not misfold ($k_m < 0.1$ s$^{-1}$ and $K_m < 1$) at any given pair of $k_f$ and $K_f$. The shapes of the contours are similar, but the positions differ. In this case, the contours for proteins that misfold “shrink”; that is, they shift
upward and to the right, which means that more of the protein is degraded,
leaving less soluble protein at the 3 h time point. Proteins with high misfolding
propensity accumulate M states that are being degraded.

2.2.1.2.2 “Misfolding penalty” is also thermo/kinetically-limited

The “misfolding penalty” refers to the change in the fraction of soluble
protein (Δ $F_s$) due to misfolding. For a given $k_f$ and $K_f$ (e.g. $k_f = 1 \text{ s}^{-1}$ and $K_f = 10^5$), the misfolding penalty exhibits thermo/kinetically-limited behavior similar to
folding (Fig. 2.7A). It is controlled by the limiting aspect of the misfolding
parameters. The line of thermo-kinetic crossover separates the kinetic vs.
thermodynamic control regions. The misfolding penalty depends on $k_m$ above the
line, on $K_m$ below the line, or on both along the line. The effect of misfolding can
be expressed using a single parameter, the Limiting Misfolding Parameter (LMP)
(Fig. 2.7B). The fraction of soluble proteins for proteins that misfold but do not
aggregate depends on two parameters, the LFP and LMP (Fig. 2.7B). The best fit
to the data is:

$$F_s = \frac{1}{1 + 0.59 \times 10^{0.38 \times \min[\log_{10} k_m, \log_{10} K_m - 0.48] - 0.63 \times \min[\log_{10} k_f, \log_{10} K_f - 4.30]}}$$

As above, the parameters in the equation are a function of the PN components.
The effect of misfolding, which contributes a positive term in the exponential,
counters the effect of folding, which contributes a negative term.
2.2.1.3 Case 3: MIsfolding and aggregation

Only 3% of proteins described by the parameter sets tested misfold and aggregate substantially. Aggregation requires very specific combinations of parameters. By “substantial aggregation”, we mean that the concentration of aggregated protein is greater than or equal to 5 μM at the 3 h time point. Proteins that misfold and aggregate require $k_m \geq 0.1 \text{ s}^{-1}$ and $K_m \geq 0.1$, $k_a \geq 10 \mu\text{M}^{-1} \text{s}^{-1}$ and $K_a \geq 10 \mu\text{M}^{-1}$. Proteins with an unstable misfolded state, an unstable aggregated state, or those with low aggregation rate do not aggregate (Fig. 2.8A).

2.2.1.3.1 Aggregation following misfolding further suppresses native folding

Aggregation competes with folding and degradation. Aggregation following misfolding further suppresses the folding of native states. For example, the fraction soluble of proteins that misfold and aggregate (e.g., $k_m = 100 \text{ s}^{-1}$, $K_m = 100$, $k_a = 10 \mu\text{M}^{-1} \text{s}^{-1}$ and $K_a = 1000 \mu\text{M}^{-1}$) is generally lower than those of proteins that misfold and do not aggregate ($k_m = 100 \text{ s}^{-1}$, $K_m = 100$, $k_a = 10 \mu\text{M}^{-1}$ s$^{-1}$ and $K_a = 1 \mu\text{M}^{-1}$) at any given pair of $k_f$ and $K_f$ (Fig. 2.8B). The contours have the same shapes, but the positions differ. In this case, the contours shift to the right, which means that there is a smaller fraction of soluble proteins.
2.2.1.3.2 “Aggregation penalty” and fraction of aggregated protein are also thermo/kinetically limited

“Aggregation penalty” refers to the change in the fraction of soluble protein (ΔFs) due to aggregation. For a given kf, Kf, km and Km (e.g. kf = 1 s\(^{-1}\), Kf = 10\(^5\), km = 100 s\(^{-1}\), Km = 100), the aggregation penalty exhibits thermo/kinetically limited behavior similar to folding and misfolding (Fig. 2.9A). It can also be expressed in terms of a single parameter, the Limiting Aggregation Parameter (LAP). In addition, the fraction of aggregated protein (aggregated protein/total protein synthesized) is also thermo/kinetically limited (Fig. 2.9B).

The fractions of both soluble (Fs) and aggregated (Fa) proteins are important in describing the behavior of aggregation-prone proteins. These quantities can be estimated by solving equations using three parameters: LFP, LMP, and LAP (Fig. 2.10). The best fits to the data are:

\[
Fs = \frac{1}{1 + 0.026 \times 10^{-0.39xMin[\log_{10}Kf,\log_{10}Km]+0.07} + 0.69xMin[\log_{10}Km,\log_{10}Km]+0.40} - 0.85xMin[\log_{10}Kf,\log_{10}Km]+4.40}
\]

\[
Fa = \frac{1}{1 + 1600 \times 10^{0.90xMin[\log_{10}Km,\log_{10}Km]+0.3}] - 0.48xMin[\log_{10}Km,\log_{10}Km]+0.47} + 0.18xMin[\log_{10}Kf,\log_{10}Kf]+4.64}
\]

Fig. 2.10 shows the fraction of soluble and aggregated proteins as indicated by the points on the contour surfaces. The mostly soluble proteins (Fs ~>0.9, at lower-left-back) are with high LFP, low LMP, and low LAP. The mostly aggregated proteins (Fa ~>0.7, at upper-right-front) are with high LAP, high LMP, and low LFP. The mostly degraded proteins (Fa ~>0.1, at lower-right-front) are with high LMP, low LAP, and low LFP.
Our collaborator, Evan T. Powers, performed the simulations that I just discussed in this section. He extended my initial observations, which were based on FoldEco simulations (which will be discussed briefly in section 2.2.3) and wet lab experiments that I had performed.

### 2.2.2 Investigating the effects of biophysical properties on protein folding fates and testing FoldEco predictions using wet lab experiments

To determine the effects of biophysical properties on protein folding fates and to test for thermo-kinetic limiting behavior in vivo under adapted-basal PN condition, we expressed CRABP1 variants with different biophysical properties in *E. coli*. It is assumed that mutations affect folding parameters ($k_f$ and $K_f$) more than misfolding ($k_m$ and $K_m$) or aggregation ($k_a$ and $K_a$) parameters. Mutations can affect aggregation propensities, but the native state should be more sensitive to changes in amino acids. The variants that were used were WT*, M9A, V67A, L118V, and Y133S (Fig. 2.2). Their folding parameters were shown on Table 2.1.

#### 2.2.2.1 The folding fates of CRABP1(M9A), CRABP1(V67A), and CRABP1(L118V) are similar to CRABP1(WT*) while the folding fate of CRABP1(Y133S) is different from the others

To determine the folding fates of different CRABP1 variants under the adapted-basal PN of *E. coli*, the variants were expressed in *E. coli* K12 HMS174 (DE3) cells by induction of plasmid under the control of a T7 promoter with
isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1, 2, and 3 h at 30 °C. Cells
were lysed and partitioned into total, soluble and aggregated fractions. Soluble
(i.e. supernatant) and aggregated (i.e. pellet) fractions were separated by
centrifugation. Fractions were run on SDS-PAGE and analyzed through
quantitative Coomassie staining. In the wet lab experiments, “folding fate” is
determined by quantitating the total, soluble, and aggregated proteins. These are
the existing proteins after 1, 2, and 3 h of protein expression. It is assumed that
soluble proteins are natively folded proteins and may also include unfolded or
misfolded proteins that are free or bound to chaperones. It is also assumed that
the pellet consists of the aggregated proteins. Absolute amount of proteins were
quantitated by using a standard curve constructed with purified recombinant
protein (Fig. A.4; Methods).

Fig. 2.11 A-C shows the gel for Coomassie blue stain for the expression of
CRABP1(WT*), and graphs for cytoplasmic concentration and fraction of soluble
and aggregated proteins at different induction times. The total protein
concentrations increased with longer induction times: 95 ± 15 μM, 202 ± 25 μM,
and 275 ± 28 μM after 1, 2, and 3 h of induction, respectively. The soluble protein
concentrations were also increasing: 95 ± 15 μM, 192 ± 22 μM, and 251 ± 19 μM
after 1, 2, and 3 h of induction, respectively. Furthermore, the aggregated protein
concentrations were increasing: 0 ± 0, 10 ± 5 μM, and 24 ± 9 μM after 1, 2, and 3
h of induction, respectively. The proteins were not aggregated after 1 h of
induction. At 2 h, the proteins were 5% \( \pm \) 2% aggregated. After 3 h, the proteins were 8% \( \pm \) 3% aggregated (Table A.2).

Fig. 2.12 A-C shows the Coomassie blue stain for the expression of CRABP1(M9A), its cytoplasmic concentration, and fraction of soluble and aggregated proteins at different induction times. The levels of total, soluble, and aggregated proteins increased with longer induction times. The levels of total protein were: 115 \( \pm \) 17 \( \mu \)M, 202 \( \pm \) 27 \( \mu \)M, and 281 \( \pm \) 36 \( \mu \)M after 1, 2, and 3 h of induction, respectively. The levels of soluble protein were: 111 \( \pm \) 14 \( \mu \)M, 189 \( \pm \) 24 \( \mu \)M, and 240 \( \pm \) 29 \( \mu \)M after 1, 2, and 3 h of induction, respectively. In addition, the levels of aggregated proteins were: 4 \( \pm \) 4 \( \mu \)M, 13 \( \pm \) 3 \( \mu \)M and 41 \( \pm \) 6 \( \mu \)M after 1, 2, and 3 h of induction, respectively. At 1 h, the proteins were 3% \( \pm \) 3% aggregated, at 2 h, 6% \( \pm \) 1%, and at 3 h, 15% \( \pm \) 0% (Table A.2).

Fig. 2.13 A-C shows the Coomassie blue stain for the expression of CRABP1(V67A), its cytoplasmic concentration, and fraction of soluble and aggregated proteins at different induction times. The concentrations of total protein were: 125 \( \pm \) 5 \( \mu \)M, 189 \( \pm \) 11 \( \mu \)M, and 294 \( \pm \) 32 \( \mu \)M after 1, 2, and 3 h of induction, respectively. The soluble protein concentrations were: 108 \( \pm \) 6 \( \mu \)M, 159 \( \pm \) 11 \( \mu \)M, and 240 \( \pm \) 27 \( \mu \)M after 1, 2, and 3 h of induction, respectively. The aggregated protein concentrations were: 17 \( \pm \) 2 \( \mu \)M, 30 \( \pm \) 5 \( \mu \)M, and 54 \( \pm \) 5 \( \mu \)M after 1, 2, and 3 h of induction, respectively. The fractions of aggregated proteins were 14% \( \pm \) 2% at 1 h of induction, 16% \( \pm \) 2% at 2 h, and 18% \( \pm \) 0% at 3 h (Table A.2).
Fig. 2.14 A-C shows the Coomassie blue stain for the expression of CRABP1(L118V), its cytoplasmic concentration, and fraction of soluble and aggregated proteins at different induction times. The [total protein] were: 94 ± 11 μM, 182 ± 44 μM, and 257 ± 86 μM after 1, 2, and 3 h of induction, respectively. The [soluble protein] were: 82 ± 5 μM, 156 ± 35 μM, and 212 ± 64 μM after 1, 2, and 3 h of induction, respectively. The [aggregated protein] were: 12 ± 7 μM, 26 ± 13 μM, and 45 ± 23 μM after 1, 2, and 3 h of induction, respectively. The fractions of aggregated proteins were 12% ± 6% at 1 h induction, 13% ± 7% at 2 h, and 16% ± 5% at 3 h (Table A.2).

Fig. 2.15 A-C shows the Coomassie blue stain for the expression of CRABP1(Y133S), its cytoplasmic concentration, and fraction of soluble and aggregated proteins at different induction times. The total, soluble, and aggregated protein concentrations increased with longer induction times. The total protein concentrations were: 67 ± 5 μM, 109 ± 13 μM, and 125 ± 25 μM after 1, 2, and 3 h of induction, respectively. The soluble protein concentrations were: 22 ± 1 μM, 34 ± 2 μM, and 35 ± 7 μM after 1, 2, and 3 h of induction, respectively. The aggregated protein concentrations were: 45 ± 4, 75 ± 12 μM, and 90 ± 18 μM after 1, 2, and 3 h of induction, respectively. The fractions of aggregated proteins were 67% ± 1 % at 1 h, 68% ± 3% at 2 h, and 72 % ± 2% at 3 h (Table A.2).

Results showed that M9A, L118V, V67A behaved similarly to WT*. They were mostly soluble (> ~82%) after 1, 2, and 3 h of induction. The levels of
chaperones and degradation enzymes of the PN are sufficient to aid their folding. By contrast, Y133S is very prone to aggregation. It aggregated ~70% starting at 1 h. Thus, the levels of the PN are insufficient to aid its folding.

2.2.2.1 Low levels of total protein for Y133S is due to the presence of Lon

The total protein concentration for Y133S is much lower than that of the WT* (Table A.2). What accounts for this? There are two possible reasons: 1) Lon degradation of unfolded or misfolded states of CRABP1(Y133S) (assuming that the synthesis rates for CRABP1(WT*) and CRABP1(Y133S) are similar), or 2) the synthesis rate for CRABP1(Y133S) is slower than that of CRABP1(WT*). The first possibility was tested. To determine the effect of Lon on the total protein, CRABP1(Y133S) and CRABP1(WT*) were expressed in Δlon E. coli K12 HMS174(DE3) cells after 0.5, 1, and 2 h. Results showed that both CRABP1(WT*) and CRABP1(Y133S) had more total protein in Δlon cells than in cells with Lon (Fig. A.2). For CRABP1(Y133S), in −Lon cells, the [total protein] was 98 ± 7 μM, 201 ± 30 μM, 415 ± 36 μM, at 0.5, 1, and 2 h, respectively. By contrast, in +Lon cells, it was 37 ± 9 μM, 67 ± 6 μM, 155 ± 14 μM at 0.5, 1 and 2 h, respectively. The concentrations were reduced by a factor of three; 2/3 of the total protein synthesized was degraded. For CRABP1(WT*), ~1/3 of the total protein synthesized was degraded. The total protein concentrations for Y133S and WT* in −Lon cells were similar.
2.2.2.2 Experimental data are consistent with computational expectations

We assess the predictions in the preceding section in the light of our experimental data. Table 2.2 shows the biophysical properties of the CRABP1 variants and their folding fates after 3 h of protein induction. According to the FoldEco results, all CRABP1 variants should be kinetically limited (LFP = \( k_f \)) due to their high \( K_f \). This is consistent with the results from experiments.

CRABP1(WT*), CRABP1(M9A), CRABP1(V67A), and CRABP1(L118V) had similar levels of total protein with 275 \( \mu \)M, 281 \( \mu \)M, 294 \( \mu \)M, and 257 \( \mu \)M, respectively, and degraded protein with 130 \( \mu \)M, 124 \( \mu \)M, 111 \( \mu \)M, and 148 \( \mu \)M, respectively (assuming, in each case, that the total amount of protein synthesized was 405 \( \mu \)M). The folding fates of these four proteins are not very different from each other because their \( k_f \) does not change much (changes only by a factor of 2.5) despite the variation in their \( K_f \) (changes by 2.5 orders of magnitude).

CRABP1(Y133S) behaves differently because its \( k_f \) is an order of magnitude (10-fold) lower than that of CRABP1(WT*). This behavior is consistent with what we expect for a thermo/kinetically limited system. Unfortunately, the LMP and LAP are not known.

Using FoldEco, the known folding properties were combined with the known \textit{in vivo} folding fates in order to semi-quantitatively localize the LMP and LAP. A graphical example of combinations of parameters for CRABP1(WT*) is shown in Fig. 2.16A. Given the fraction soluble (0.63), the fraction aggregated (0.06) and the LFP (0.77), the LMP and LAP can be estimated. Thus, the
biophysical properties of CRABP1 variants in terms of its thermo-kinetically limited parameters can be localized. The LFP, LMP, and LAP values for the CRABP1 variants are shown (Fig. 2.16B). For WT*, LFP is the $k_f$, which means that $k_f$ is limiting. The LMP (3.4) is either $k_m$ or $K_m$, and the LAP (90) is either $k_a$ or $K_a$, depending on the limiting factor. This is also true for the other variants. Using this method we can also deduce that in addition to Y133S folding the slowest among the CRABP1 variants, it also has the highest misfolding propensity (Fig. 2.16B).

2.2.3 Preliminary analyses of the effects of biophysical properties on folding fates of proteins using FoldEco

Here, I discuss my preliminary analyses on the effects of biophysical properties on protein folding fates using FoldEco.

To determine how biophysical properties of proteins affect their folding fates, I used CRABP1 variants (as discussed earlier) with known folding parameters ($k_f$ and $K_f$) from in vitro measurements (Table 2.1) but unknown misfolding ($k_m$ and $K_m$) and aggregation ($k_a$ and $K_a$) parameters. Different combinations of the six parameters ($k_f$, $K_f$, $k_m$, $K_m$, $k_a$, $K_a$) were run in FoldEco. The combinations include estimates for the upper (“high”) and lower (“low”) bounds for misfolding and aggregation parameters; the corresponding parameters for a real protein may be within the range. For each variant, the $k_m$ (low) is 10-fold lower than $k_f$ and $k_m$ (high) is 10-fold higher than $k_f$; $K_m$ (low) is 0.3 and $K_m$ (high) is 30; $k_a$ (low) is 0.3 and $k_a$ (high) is 30; $K_a$ (low) is 1 and $K_a$ (high)
is 1000. There were 64 different combinations that were run, under the basal PN condition for 1, 2, 3 h of protein expression. These combinations of parameters resulted to proteins (for all the variants) that were ~>98% soluble. These predictions were tested with wet lab experiments by expressing the variants in *E. coli* (as discussed in section 2.2.2.1). The FoldEco predictions were not consistent with the wet lab experiments, which means that none of the combinations of parameters are consistent with the behavior of the proteins.

I performed another set of FoldEco simulations, in which combinations of parameters were run (k_f and K_f from *in vitro* measurements while k_m, K_m, k_a, and K_a are changed) until they matched the *in vivo* folding fates. Determining k_f and K_f of a protein is relatively easy *in vitro*, however, determining k_m and K_m are difficult. Thus, FoldEco is very useful in estimating such parameters. Simulations were run under the basal PN for 1, 2, and 3 h protein expression time. The effective parameters are shown in Table A.3. Fig. A.5-A.9 show the images from the simulations. On the image, the middle part shows where the fluxes of proteins from one state (or chaperone) to another are going. On the left is the time of simulation and % chaperone use. On the right is the total protein, % of ongoing degradation, and ATP use. On the bottom is the list of the concentrations of each species (either bound to chaperones or free) at a given time. Based on these results, each variant has distinct properties (i.e. unique combination of k_f, K_f, k_m, K_m, k_a, and K_a). However, for each variant, the solution for each parameter is not unique (but there is only a defined range). For example, for M9A, the k_m can be 10 or 20, and still gives the same result. It is also possible that the variants can
have similar $K_m$ or $K_a$ (or probably even same $k_m$ and $k_a$). Despite the similarities, the combinations of the other folding, misfolding, and aggregation parameters make them unique. Further detailed investigation about the kinetic and thermodynamic properties of proteins should be done to have a full understanding about the influence of these properties to protein folding fates in vivo.

2.3 Discussion

Our results show that a protein’s intrinsic biophysical properties, specifically, its kinetic and thermodynamic properties, determine protein folding fates in *E. coli* under the adapted-basal PN.

The schema that we used in FoldEco to describe how biophysical properties influence protein folding fates is fairly simplified (Fig. 2.3). Using this simple schema, we are able to determine that the folding fate of a protein is under a thermo/kinetic limitation (Fig. 2.5) Thermo/kinetic limitation indicates that the fraction of soluble protein depends on either the kinetics or thermodynamics, but (for the most part) not on both at the same time (as shown by the “L-shaped” contours). Three parameters dictate protein fates. Folding is limited by the Limiting Folding Parameter (LFP), misfolding by the Limiting Misfolding Parameter (LMP), and aggregation by the Limiting Aggregation Parameter (LAP). In the simulations, degradation rate was set as constant. Changing the degradation rate would give contours that are also “L-shaped”, but the positions
of the contours are different. They either “grow” or “shrink”; they “grow” if more fraction soluble, and “shrink” if less fraction soluble. More degradation would “shrink” the contours, while less degradation would “grow” the contours. Furthermore, increasing concentrations of chaperones (for example, KJE) may also “grow” the contours. The time of expression also affects the positions of contours; at earlier times (for example 1 and 2 h), the fraction soluble is higher, which means that the contours would be bigger.

Based on FoldEco runs, proteins are categorized into three cases: 1) those that do not misfold and do not aggregate, 2) those that misfold but do not aggregate, and 3) those that misfold and aggregate. In the first case, protein fate is either folded or degraded. Folding and degradation compete over the unfolded states (U). The partitioning between folding and degradation depends on the LFP (Fig. 2.4 and 2.5). In the second case, folding still competes with degradation, but degradation is increased because misfolded states (M) are prone to be degraded (Fig. 2.6). The fraction of soluble protein depends on both LFP and LMP (Fig. 2.7). Only a few proteins (3% of all cases) are prone to aggregation and therefore fit into the third category, because aggregation requires very specific parameter combinations. Here, folding, degradation and aggregation all compete for U and M states. Native folding is decreased because more proteins are prone to both degradation and aggregation. The behavior of aggregation-prone proteins is fairly well described by just three parameters: LFP, LMP, and LAP (Fig. 2.10).
The data from wet lab experiments done using five CRABP1 variants (WT*, M9A, V67A, L118V, and Y133S) with distinct kinetic and thermodynamic properties are broadly consistent with our computational expectations (Table 2.2). At 3 h expression, the amounts of soluble protein are similar for WT*, M9A, V67A, and L118V due to their having similar $k_f$ values, despite the ~2.5 orders of magnitude differences in their $K_f$. However, further decreasing $k_f$ (by an order of magnitude relative to WT*) changes the fate of the Y133S variant from mostly soluble to mostly aggregated. This behavior reflects the thermo/kinetic limitation that we expect. The FoldEco simulations were done at 3 h expression, and since we have also experimental data for 1 and 2 h, simulations at these expression times can also be performed. We expect that thermo/kinetic limitation will also be observed at these time points.

In addition, we have used FoldEco to translate the behavior of proteins in vivo into information about their biophysical properties (Fig. 2.16). Given LFP and the in vivo folding fate data, LMP and LAP can be estimated.

Furthermore, based on the predictions, in general, aggregation of proteins under the basal PN of *E. coli* is only ~3%, which suggests that the PN is robust to most sets of protein folding and misfolding parameters and is capable of maintaining proteostasis. This result is also consistent with wet lab experiments, in which four out of five of the proteins under study are mostly soluble after 1, 2, and 3 h of induction (Fig. A.4-9). Here, the basal PN is sufficient to maintain more
soluble states and to minimize aggregation. Only Y133S aggregates significantly; the basal PN is insufficient to prevent its aggregation.

Moreover, the ranges of the values of $k_f$ ($10^{-3} \text{ to } 10^3 \text{ s}^{-1}$) and $K_f$ ($10^2 \text{ to } 10^8$) in our simulations cover the folding rates (Zou et al., 2014) and equilibrium constants (Ghosh and Dill, 2010) of most proteins in the *E. coli* proteome, which suggest selective pressure and evolution have enabled the PN to handle proteins with different biophysical properties.

Overall, the simple schema that we use to describe what is happening is adequate to recapitulate the behaviors that we see. The fits are good and the correlations that are emerging explain a lot about the behavior of proteins *in vivo*. This suggests that the contributions of other chaperones, like HtpG (the *E. coli* Hsp90) and IbpA/IbpB in FoldEco, while probably important in some cases, can often be neglected. In addition, including on-pathway intermediates (I) in FoldEco likely would not offer first-order improvements in our results in most cases, although there almost certainly are some instances in which such intermediates are important.

### 2.4 Conclusion

The use of an integrated computational and experimental approach has enabled us to investigate systematically the effects of the kinetic and thermodynamic properties of proteins on their folding fates. The folding fates of proteins are thermo/kinetically limited, based on their biophysical properties.
We were able to understand how proteins behave under one set of conditions, that is, the basal PN. Having an understanding about the intrinsic kinetic and thermodynamic properties of proteins, we can now change conditions and determine the sensitivity of different proteins under various PN conditions. This is the subject of the next chapter.

2.5 Materials and Methods

2.5.1 E. coli strains and plasmids

The E. coli strain, K12 HMS174(DE3) (Novagen) was used as the background strain in all the experiments. The Δlon K12 HMS174(DE3) (kindly provided by Kelly Lab), in which lon gene was deleted/replaced by a kanamycin-resistant (KanR) cassette, was used for control experiments. The cells (K12 HMS174(DE3)) that were used in the overexpression experiments were with a plasmid containing the DnaK/DnaJ/GrpE system (KJE) (chloramphenicol-resistant) under a pBAD promoter. The KJE system was not induced in all the experiments; its expression was suppressed by adding 0.2% w/v D-glucose, which represses the arabinose operon (Tokuriki and Tawfik, 2009). This was the “adapted-basal” condition.

The murine CRABP1 was used as a model protein. A variant with an N-terminal (His)_{10}-tag and with R131Q mutation to stabilize the protein (Clark et al., 1998). Here, it is referred to as WT*. WT* was used as the template for the other CRABP1 variants (M9A, V67A, L118V, and Y133S) (Budyak et al., 2013).
CRABP1 was inserted into pET16b vector (ampicillin-resistant) under a T7 promoter.

2.5.2 Protein expression in *E. coli*  

Bacterial transformations with the pET16b (with CRABP1) were done on Luria-Bertani (LB) plates. Bacterial cultures were grown in LB with ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) at 30 °C until OD<sub>600</sub> was ~0.4. Glucose (0.2% w/v) was added to suppress KJE expression. After 1 h, CRABP1 expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1, 2, and 3 h. Cells were harvested by centrifugation for 15 min. Cells were lysed by resuspending with Bacterial Protein Extraction Reagent (BPER-II) (ThermoFisher Scientific), lysozyme (0.05 mg/mL) and DNase I (1 μg/mL). Cell lysates were incubated at room temperature for ~15-30 min. A fraction of the lysate was transferred to one tube (this was the “T”-total lysate). An equal amount of volume of the lysate was transferred to another tube, which was then centrifuged at 13,500xg for 10 min at 4 °C to partition the supernatant from the pellet. The supernatant was collected as soluble (“S”) while the pellet as aggregated (“A”). The pellet was resuspended with the same volume of buffer. Gel loading buffer was added to each sample. Samples were boiled for 10 min. Samples were loaded on a 15% glycine gel, and were run on sodium dodecyl sulfate-polyacrylamide agarose gel electrophoresis (SDS-PAGE). Proteins were
detected by Coomassie blue staining and imaged with Odyssey Infrared Imaging System (Li-COR Biosciences).

The wild-type CRABP1 (WT*) and another variant CRABP1(Y133S) were also expressed in Δlon cells (-Lon) to determine the typical protein expression in *E. coli* (without degradation by Lon). Cells were grown in LB with ampicillin (100 μg/mL) at 30 °C until OD<sub>600</sub> was ~0.7 (same OD<sub>600</sub> when CRABP1 was induced in cells with KJE plasmid described above). CRABP1 (WT* or Y133S) expression was induced with 0.5 mM IPTG for 0.5, 1, and 2 h. Another set of experiment with +Lon cells (K12 HMS174(DE3)) was run in parallel in order to compare folding fates in –Lon and +Lon cells.

### 2.5.3 Quantitative Coomassie blue staining and Western blotting

For the quantitation, recombinant CRABP1 purified proteins with known amounts (pmol) were used as standards. These were also loaded on the same gel where the samples of unknown amounts were loaded. Each volume of sample loaded corresponded to 1x10<sup>8</sup> cells, which is equivalent to 1 mL of bacteria with OD<sub>600</sub>~0.1. After running samples and detecting them with Coomassie staining, they were imaged using Li-COR. A standard curve was generated based on the known amounts of proteins, and the value of the unknown was determined by extrapolation (Fig. A.4). The concentration of samples was calculated by dividing the amount of protein (mole) by the volume of cell cytoplasm (L). It is assumed that for *E. coli* grown in LB, the total cell volume
is 2.9 μL per mL of cells per OD\textsubscript{600} unit (Volkmer and Heinemann, 2011). Subtracting the periplasm, which is ~10% of the cell volume, the volume of cell cytoplasm is 2.6 μL per mL per OD\textsubscript{600}.

Western blotting was performed to detect and confirm the presence of CRABP1 in the samples. The primary antibody used was anti-CRABP1 (Abcam, monoclonal, mouse, 1:5000). The secondary antibody was Li-COR antibody, anti-mouse (emission at 800 nm, from Li-COR Biosciences).

The concentrations of total (T), soluble (S), and aggregated (A) protein and the fractions of soluble (f\textsubscript{s}) and fraction of aggregated (f\textsubscript{a}) protein that are existing after 1, 2, and 3 h were calculated as follows: [T] = [S] + [A], f\textsubscript{s} = [S]/[T], f\textsubscript{a} = [A]/[T]. The concentration of degraded protein was calculated by subtracting the total existing protein from the total synthesized protein after 1, 2, and 3 h, assuming that the total synthesized protein after 1 h is 135 μM, 2 h is 270 μM, and 3 h is 405 μM. Note that the calculated concentration for total protein was the sum of the concentrations of soluble and aggregated protein, and not the concentration of the total lysate (“T”) run on the “T” lane on Coomassie stained-gels or Western blots. The total lysate (“T”) lane was a control to check for mass balance; the sum of soluble and aggregated relative to the total lysate should be 1. The mass balance was good in all samples; the difference of the concentrations of total lysate on “T” lane ([T]) from the total of soluble and aggregated ([S] + [A]) was less than 20%. This difference can be due to errors in resuspending and loading the samples.
2.5.4 Measurement of relative levels of PN components

A control, in which bacteria were transformed with an empty pET16b vector (without the CRABP1 gene), was also performed in parallel to the experiments described above (section 2.5.2). The total lysates (“T”) of each sample were run together with the control. Proteins were detected with Coomassie blue stains. Using the Li-COR imaging, the levels of DnaK, GroEL, and Lon under the “adapted-basal” conditions, in which heterologous CRABP1 was expressed, were compared to those under the basal conditions, in which CRABP1 was not expressed.

Western blotting was also performed. The primary antibodies used were anti-DnaK (Enzo Life Sciences, monoclonal, mouse, 1:10000), anti-GroEL (Enzo Life Sciences, monoclonal, mouse, 1:2000), and anti-Lon (kindly provided by Prof. R. T. Sauer, 1:10000). The secondary antibodies were Li-COR antibodies (emission at 680 or 800 nm, from Li-COR Biosciences).

2.5.5 FoldEco simulations

Evan T. Powers performed this set of FoldEco simulations.

FoldEco has many parameters to consider for a simulation. Most of the parameters, however, are independent of the nature of the protein substrate (Powers et al., 2012). For the simulations, FoldEco was run under conditions in which synthesis rate, initial concentrations of PN components, and folding energetics parameters, were adjustable. Synthesis rate was set to 135 μM h⁻¹.
This was based on a typical protein expression in *E. coli*, which was determined by expressing CRABP1 in Δlon K12(DE3) cells (see section 2.5.2). The folding biophysical parameters comprised of folding rate and equilibrium constants (\(k_f\) and \(K_f\)), misfolding rate and equilibrium constants (\(k_m\) and \(K_m\)), and aggregation rate and equilibrium constants (\(k_a\) and \(K_a\)). Combinations of these parameters were generated to cover the following ranges (that are physically reasonable): 

- \(k_f\): \(10^{-3}\) to \(10^3\) s\(^{-1}\); \(K_f\): \(10^2\) to \(10^8\);
- \(k_m\): \(10^{-3}\) to \(10^3\) s\(^{-1}\); \(K_m\): \(10^{-3}\) to \(10^3\);
- \(k_a\): \(10^{-3}\) to \(10^3\) μM\(^{-1}\) s\(^{-1}\); \(K_a\): \(10^{-3}\) to \(10^3\) μM\(^{-1}\).

Each was varied in 10-fold steps to make 117,649 \((7^6)\) combinations. Degradation rate was set as constant. The model was solved for each set of parameters. The concentrations of soluble, aggregated, and degraded proteins were compared after 3 h (~405 μM of protein synthesized) as a function of their folding energetics. The concentration of soluble protein ([S]) was calculated as total protein synthesized minus aggregated ([T] – [A]). The concentration of degraded protein was total minus the sum of soluble and aggregated, [T] – ([S] + [A]).

In each case/observation, data were plotted using contour plots. The Limiting Folding Parameter (LFP) was determined by projecting points above the line of thermo-kinetic equivalence down to the \(k_f\) axis and by projecting points below over to the \(K_f\) axis and then offset. This was also done with the Limiting Misfolding Parameter (LMP) for \(k_m\) and \(K_m\), and Limiting Aggregation Parameter (LAP) for \(k_a\) and \(K_a\). The data were fit and the best fit for fraction soluble was calculated.
2.5.5.1 Preliminary FoldEco simulations

I performed this set of FoldEco simulations. Different combinations of the six parameters \( (k_f, K_f, k_m, K_m, k_a, K_a) \) were run through brute force. With known folding parameters \( (k_f \text{ and } K_f) \) from in vitro measurements (Table 2.1), estimates for the upper (“high”) and lower (“low”) bounds for misfolding and aggregation parameters were run; the corresponding parameters for a real protein may be within the range. For each variant, the \( k_m \) (low) was 10-fold lower than \( k_f \) and \( k_m \) (high) was 10-fold higher than \( k_f \); \( K_m \) (low) was 0.3 and \( K_m \) (high) was 30; \( k_a \) (low) was 0.3 and \( k_a \) (high) was 30; \( K_a \) (low) was 1 and \( K_a \) (high) was 1000. There were 64 different combinations that were run, under basal PN condition for 1, 2, 3 h of protein expression.

Another batch of FoldEco simulations were performed to translate the in vivo folding fate of proteins (from wet lab experiments) into information about their biophysical properties. Determining folding rate and folding equilibrium constants \( (k_f, K_f) \) of a protein is relatively easy in vitro, however, determining misfolding rate and misfolding equilibrium constants \( (k_m, K_m) \) are difficult. Thus, FoldEco is very useful in estimating such parameters. The \( k_f \) and \( K_f \) were kept constant for each variant while the parameters to be estimated were \( k_m, K_m, k_a, \) and \( K_a \). Simulations were run under basal PN for 1, 2, and 3 h protein expression time. FoldEco results were matched to the data from in vivo experiments through different combinations of the six parameters \( (k_f, K_f, k_m, K_m, k_a, K_a) \). The synthesis rate used was 0.001 s\(^{-1}\), which is the rate constant when the ribosomes are translationally active (Powers et al., 2012). Other parameters were assigned
values based on literature. Degradation was by Lon; for Y133S, degradation equilibrium constant was higher (by a factor of 10) than that of the other variants to match the observation \textit{in vivo} that more proteins were degraded for Y133S compared to the others.

2.5.6 Protein sequences

The sequences of CRABP1 variants are the following. The font in bold letter is the new residue from the mutation.

R131Q CRABP1 (CRABP1-WT*)

\begin{verbatim}
PNFA\textbf{GTKWMR} SSENFD\textbf{ELLK} AL\textbf{GVNAML}\textbf{RK} VAVA\textbf{AASKPH} VEIRQDGDQF YIKTSTTVRT TEINF\textbf{KVGE\textbf{G}} FEEE\textbf{TVDG\textbf{R}} KR CRSLPTWENE NKIHCTQ\textbf{T}LL EG\textbf{DGPKTYWT} RELANDELIL TFGADDVVCT Q\textbf{IYVRE}
\end{verbatim}

M9A/R131Q CRABP1 (CRABP1-M9A)

\begin{verbatim}
PNFA\textbf{GTKMAR} SSENFD\textbf{ELLK} AL\textbf{GVNAML}\textbf{RK} VAVA\textbf{AASKPH} VEIRQDGDQF YIKTSTTVRT TEINF\textbf{KVGE}\textbf{G} FEEETVDG\textbf{R} KR CRSLPTWENE NKIHCTQ\textbf{T}LL EG\textbf{DGPKTYWT} RELANDELIL TFGADDVVCT Q\textbf{IYVRE}
\end{verbatim}

V67A/R131Q CRABP1 (CRABP1-V67A)

\begin{verbatim}
PNFA\textbf{GTKWMR} SSENFD\textbf{ELLK} AL\textbf{GVNAML}\textbf{RK} VAVA\textbf{AASKPH} VEIRQDGDQF YIKTSTTVRT TEINF\textbf{KAGEG} FEEETVDG\textbf{R} KR CRSLPTWENE NKIHCTQ\textbf{T}LL EG\textbf{DGPKTYWT} RELANDE\textbf{VIL} TFGADDVVCT Q\textbf{IYVRE}
\end{verbatim}

L118V/R131Q CRABP1 (CRABP1-L118V)

\begin{verbatim}
PNFA\textbf{GTKWMR} SSENFD\textbf{ELLK} AL\textbf{GVNAML}\textbf{RK} VAVA\textbf{AASKPH} VEIRQDGDQF YIKTSTTVRT TEINF\textbf{KVGE\textbf{G}} FEEETVDG\textbf{R} KR CRSLPTWENE NKIHCTQ\textbf{T}LL EG\textbf{DGPKTYWT} RELANDE\textbf{VIL} TFGADDVVCT Q\textbf{IYVRE}
\end{verbatim}
Y133S/R131Q CRABP1 (CRABP1-Y133S):

PNFAGTWKMR SSENFDELLK ALGVANMLRK VAVAAASKPH VEIRQDGDFQF
YIKTSTTVRT TEINFKVGEF FEEETVDRGRK CRSLPTWENE NKIHTQTLLE
EDGPKTYWT RELANDELIL TFGADDVVCQI SVE

2.5.7 Protein purification

CRABP1(WT*) and CRABP1(Y133S) were purified to use as protein standards for the quantitation of absolute concentrations of CRABP1 in the in vivo experiments, as well as for in vitro refolding/unfolding experiments. The other variants (M9A, V67A, and L118V) had already been purified and available to use (from (Budyak et al., 2013). CRABP1 was expressed in E. coli BL21DE3 cells (Novagen) (carrying CRABP1 gene) by induction with 0.4 mM IPTG when OD$_{600}$ was ~0.8, at 37 ºC for 4 h. In the case of Y133S, which tends to partition to the insoluble pellet fraction, in order to improve its solubility, when OD$_{600}$ was ~0.4, the temperature was decreased to 30 ºC and L-proline (20 mM) and NaCl (0.3 M) were added. Cells were grown at 30 ºC for 4 h. After the cells were lysed, proteins were purified from the soluble fraction by using Ni-NTA (Qiagen) affinity chromatography. The molar extinction coefficient that was used to determine protein concentration was $\varepsilon_{280} = 20,970 \text{ M}^{-1}\text{cm}^{-1}$. The identity of variants was confirmed by electrospray ionization mass spectrometry of the purified proteins (Budyak et al., 2013).
2.5.8 Folding kinetics measurements

Refolding and unfolding kinetics for CRABP1(WT*) and CRABP1(Y133S) were performed as described (Budyak, et. al., 2013), with the modification of measuring kinetics at 30 °C. Kinetics traces were averaged and fit to a multiexponential equation using Origin (OriginLab).
Figure 2.1 Schematic of kinetic partitioning during protein folding *in vivo*

When an unfolded (U) polypeptide is synthesized from the ribosome (Synthesis), it can have different folding fates: It can fold into native (N), it can misfold (M), and if M self-associates, it can aggregate (A), or it can be degraded. The fate of the protein is dictated by its intrinsic biophysical properties (red text). In addition, the PN components modulate its fate *in vivo*. Kinetic partitioning affects folding. For example, in *E. coli*, GroEL and GroES (Hsp60/Hsp10) promote folding by encapsulating unfolded states and helping them fold to native. DnaK, DnaJ, and GrpE (Hsp70/Hsp40/Nucleotide exchange factor) rescue misfolded states by binding to M and helping them to become unfolded. ClpB and KJE collaborate to untangle and solubilize aggregates. (Reproduced from (Cho et al., 2015)).
Figure 2.2 Structure of murine CRABP1 (PDB 2CBR (Kleywegt et al., 1994))

Residues that are substituted are shown in spheres. The substitutions are: M9A, V67A, L118V, and Y133S.
Table 2.1 Thermodynamic and kinetic parameters of CRABP1 variants measured *in vitro*

**A.** The values were measured at 25 °C. (Table is modified from (Budyak et al., 2013)).

The $\Delta \Delta G^\circ_{U-N}$ is the energetic effect of the mutations of the unfolded state (U) with respect to the native state (N). $\Delta \Delta G^\circ_{U-N} = \Delta G^\circ_{U-N}^{WT} - \Delta G^\circ_{U-N}^{mut}$, $\Delta G^\circ_{U-N}^{WT} = 10.0 \pm 0.2$ kcal/mol.

The $k_f$ and $k_u$ values were converted from ln values determined from *in vitro* experiments at 25 °C. The $K_f$ values were converted from $\Delta G^\circ$.

*The $k_f$ for Y133S at 25 °C was extrapolated from the measured value at 30 °C.*

**B.** The $k_f$ values for WT* and Y133S measured *in vitro* at 30 °C. Y133S folds 10-fold slower than WT*. It unfolds 4-fold faster than WT*.

<table>
<thead>
<tr>
<th>CRABP1 variants</th>
<th>$\Delta \Delta G^\circ_{U-N}$ (kcal/mol)</th>
<th>$k_f$ (s$^{-1}$)</th>
<th>$k_u$ (s$^{-1}$)</th>
<th>$K_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT*</td>
<td>--</td>
<td>0.77</td>
<td>1.67 x 10$^{-5}$</td>
<td>2.17 x 10$^7$</td>
</tr>
<tr>
<td>M9A</td>
<td>-1.8 ± 0.2</td>
<td>0.31</td>
<td>1.01 x 10$^{-5}$</td>
<td>1.04 x 10$^6$</td>
</tr>
<tr>
<td>V67A</td>
<td>-3.3 ± 0.2</td>
<td>0.57</td>
<td>1.50 x 10$^{-3}$</td>
<td>8.23 x 10$^4$</td>
</tr>
<tr>
<td>L118V</td>
<td>-2.3 ± 0.4</td>
<td>0.41</td>
<td>5.02 x 10$^{-5}$</td>
<td>4.45 x 10$^5$</td>
</tr>
<tr>
<td>Y133S</td>
<td>-4.8 ± 0.2</td>
<td>0.08$^a$</td>
<td>8.96 x 10$^{-5}$</td>
<td>6.53 x 10$^3$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CRABP1 variants</th>
<th>$k_f$ (s$^{-1}$)</th>
<th>$k_u$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT*</td>
<td>3.0 ± 0</td>
<td>1.04 x 10$^{-4}$ ± 0.119 x 10$^{-4}$</td>
</tr>
<tr>
<td>Y133S</td>
<td>0.32 ± 0.04</td>
<td>4.32 x 10$^{-4}$ ± 1.79 x 10$^{-4}$</td>
</tr>
</tbody>
</table>
The synthesis rate was set to 135 μM h⁻¹ (a typical protein expression in *E. coli*). Combinations of folding, misfolding, and aggregation parameters were generated. The initial concentrations for the chaperones and degradation enzymes were their levels in the “adapted-basal” PN (Cho et al., 2015). Degradation rate was set as constant. The concentrations of soluble, aggregated, and degraded proteins were compared after 3 h as a function of their intrinsic biophysical properties. Soluble proteins include proteins that are natively folded (N) and those unfolded (U) and misfolded (M) proteins that are bound to chaperones making them soluble. Aggregated proteins (A) consist of proteins that are insoluble. Degraded proteins are from U and M states.
Figure 2.4 Fraction soluble vs. $k_f$ and $K_f$ of proteins that do not misfold and do not aggregate

**A.** Schematic of the partitioning between folding and degradation for proteins that do not misfold and do not aggregate. **B.** and **C.** L-shaped contour plots showing that the fraction of soluble protein depends on $k_f$ and $K_f$ but (most of the time) not on both parameters. Example of the dependency of fraction of soluble protein to $k_f$ or $K_f$ is shown in **C.**
Figure 2.5 Fraction of soluble protein vs. $k_f$ (at high $K_f$) and $K_f$ (at high $k_f$)

Fraction of soluble protein vs. $k_f$ or $K_f$ (A) in separate plots. When plotted in one graph (B), the equilibrium constant needs to be offset by ~4.5 to 4.6 log units to match the rate constant. C. Line of thermo-kinetic crossover on the contour plot. Above the line, fraction soluble depends on $k_f$ alone, and below the line the fraction soluble depends on $K_f$ alone. Along the line, fraction soluble depends on both $k_f$ and $K_f$. 
**Figure 2.6 Comparison of fraction soluble vs. $k_f$ and $K_f$ of proteins that do not misfold and those that misfold, but no aggregation**

**A.** Schematic for proteins that misfold but do not aggregate. **B.** The fraction of soluble protein for proteins that have higher propensity to misfold ($k_m = 100$ s$^{-1}$ and $K_m = 100$) (right) is lower compared to those that do not misfold ($k_m < 0.1$ s$^{-1}$ and $K_m < 1$) (left) at any given pair of $k_f$ and $K_f$. 
Figure 2.7 Effect of misfolding, but no aggregation

A. Misfolding penalty vs. $k_m$ and $K_m$. B. The fraction of soluble protein depends on Limiting Folding Parameter (LFP) and Limiting Misfolding Parameter (LMP).
Figure 2.8 Comparison of fraction soluble vs. $k_f$ and $K_f$ of proteins that misfold and do not aggregate and those that misfold and aggregate

A. Schematic for proteins that misfold and aggregate. B. The fraction of soluble protein for proteins that have higher propensity to aggregate is lower (right) compared to those that do not aggregate (left) at any given pair of $k_f$ and $K_f$. 
Figure 2.9 Effect of aggregation

A. Aggregation penalty vs. $k_a$ and $K_a$. B. Fraction aggregated vs. $k_a$ and $K_a$. 
Figure 2.10 The behavior of aggregation-prone proteins

The behavior of aggregation-prone proteins as described by Least Folding Parameter (LFP), Least Misfolding Parameter (LMP), and Least Aggregation Parameter (LAP). Points on the contour surfaces give the indicated fraction of soluble or aggregated protein.
Figure 2.11 Folding fate of CRABP1(WT*) upon expression in *E. coli* under adapted-basal conditions

A. Coomassie blue stain for CRABP1(WT*) induced with 0.5 mM IPTG for 1, 2 and 3 h at 30 °C. B. Cytoplasmic concentration of CRABP1(WT*). C. Fraction of soluble and aggregated proteins.
Figure 2.12 Folding fate of CRABP1(M9A) upon expression in *E. coli* under adapted-basal conditions

A. Coomassie blue stain for CRABP1(M9A) induced with 0.5 mM IPTG for 1, 2, and 3 h at 30 °C.

B. Cytoplasmic concentration of CRABP1(M9A).

C. Fraction of soluble and aggregated proteins.

T: total, S: soluble, A: aggregated
Figure 2.13 Folding fate of CRABP1(V67A) upon expression in *E. coli* under adapted-basal conditions
A. Coomassie blue stain for CRABP1(V67A) induced with 0.5 mM IPTG for 1, 2, and 3 h at 30 °C. B. Cytoplasmic concentration of CRABP1(V67A). C. Fraction of soluble and aggregated protein
Figure 2.14 Folding fate of CRABP1(L118V) upon expression in *E. coli* under adapted-basal conditions

A. Coomassie blue stain for CRABP1(L118V) induced with 0.5 mM IPTG for 1, 2, and 3 h at 30 °C. B. Cytoplasmic concentration of CRABP1(L118V). C. Fraction of soluble and aggregated proteins.

T: total, S: soluble, A: aggregated
Figure 2.15 Folding fate of CRABP1(Y133S) upon expression in *E. coli* under adapted-basal conditions
A. Coomassie blue stain for CRABP1(Y133S) induced with 0.5 mM IPTG for 1, 2, and 3 h 30 °C. B. Cytoplasmic concentration of CRABP1(Y133S). C. Fraction of soluble and aggregated proteins.
Table 2.2 Biophysical properties of CRABP1 variants and their folding fates in *E. coli* after 3 h of protein expression under the adapted-basal PN condition

*Total protein- This is the total existing protein (not total protein synthesized).

**Based on control experiments with CRABP1(WT*) expressed in Δlon *E. coli* K12 cells, in which ~1/3 of the protein that is synthesized is degraded. The total amount of protein synthesized is ~405 μM in all cases.

<table>
<thead>
<tr>
<th>CRABP1 Variant</th>
<th>k_f (s^{-1})</th>
<th>K_f</th>
<th>LFP</th>
<th>Total Protein* (μM)</th>
<th>Aggregated protein (μM)</th>
<th>Degraded protein** (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT*</td>
<td>0.77</td>
<td>2.2 × 10^7</td>
<td>0.77</td>
<td>275 μM</td>
<td>24 μM</td>
<td>130 μM</td>
</tr>
<tr>
<td>M9A</td>
<td>0.31</td>
<td>1.0 × 10^6</td>
<td>0.31</td>
<td>281 μM</td>
<td>41 μM</td>
<td>124 μM</td>
</tr>
<tr>
<td>V67A</td>
<td>0.57</td>
<td>8.2 × 10^4</td>
<td>0.57</td>
<td>294 μM</td>
<td>54 μM</td>
<td>111 μM</td>
</tr>
<tr>
<td>L118V</td>
<td>0.41</td>
<td>4.4 × 10^5</td>
<td>0.41</td>
<td>257 μM</td>
<td>45 μM</td>
<td>148 μM</td>
</tr>
<tr>
<td>Y133S</td>
<td>0.08</td>
<td>6.5 × 10^3</td>
<td>0.08</td>
<td>125 μM</td>
<td>90 μM</td>
<td>280 μM</td>
</tr>
</tbody>
</table>
Figure 2.16 Biophysical properties of CRABP1 variants in terms of thermo-kinetically limited parameters

A. Graph for CRABP1 biophysical properties. Shown is an example for WT*.

B. Table for the thermo-kinetically limited parameters of CRABP1.

<table>
<thead>
<tr>
<th>CRABP1 Variant</th>
<th>LFP</th>
<th>LMP</th>
<th>LAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT*</td>
<td>0.77</td>
<td>3.4</td>
<td>90</td>
</tr>
<tr>
<td>M9A</td>
<td>0.31</td>
<td>0.6</td>
<td>360</td>
</tr>
<tr>
<td>V67A</td>
<td>0.57</td>
<td>1.2</td>
<td>400</td>
</tr>
<tr>
<td>L118V</td>
<td>0.41</td>
<td>1.7</td>
<td>250</td>
</tr>
<tr>
<td>Y133S</td>
<td>0.08</td>
<td>10.9</td>
<td>170</td>
</tr>
</tbody>
</table>
CHAPTER 3

DISSECTING THE IMPACT OF THE PROTEOSTASIS NETWORK
ON PROTEIN FOLDING FATE


This chapter investigates the impact of different components of the PN on the folding fates of unstable and aggregation-prone variants of our model proteins: 1) cellular retinoic acid-binding protein-1 (CRABP1) expressed at different times, and 2) dihydrofolate reductase (DHFR), and retroaldolase (RA) expressed for 2 h. It also discusses the use of two quantitative models, a phenomenological model and FoldEco, to understand how PN components contribute to proteostasis and how their contributions differ depending on the substrate’s biophysical properties.

3.1 Introduction

A protein’s intrinsic biophysical properties and the cellular proteostasis network (PN) dictate the folding fate of a protein inside the cell (Cho et al., 2015) (as discussed in Chapters 1 and 2) (Fig. 1.11). In Chapter 2, we have examined about how the intrinsic biophysical properties, particularly kinetic and thermodynamic properties of proteins, influence protein folding fate under the
adapted-basal PN condition. Here, in Chapter 3, we now dissect the impact of different PN components on the folding fate of a protein.

The cellular PN, consisting of various processes (such as protein synthesis and folding, chaperone interactions, and degradation) and different components (molecular chaperones and degradation enzymes), collaborates to maintain proteostasis (Balch et al., 2008; Cho et al., 2015; Powers et al., 2009). The levels of PN components are well controlled inside the cell. For example, in *E. coli*, some components are regulated by the σ^{32} transcription factor (Zhang et al., 2014). The concentrations of PN components during normal conditions are different from those under stressed conditions. During normal conditions, for example in *E. coli*, DnaK is an abundant protein (~10000 copies) (Genevaux et al., 2007), and its monomer concentration is 30 μM (Powers et al., 2012) (Table 1.1). The concentration of DnaK is high relative to those of DnaJ (1 μM) and GrpE (15 μM) (Table 1.1). The ClpB concentration is also low relative to DnaK. GroEL and GroES are present at similar levels (Lu et al., 2007; Powers et al., 2012). During stress, such as heat, normal levels of the PN components are insufficient. The concentration of σ^{32} is increased at high temperatures, causing the up-regulation of chaperones and degradation enzymes (that are under its regulon) (Guisbert et al., 2008; Yura et al., 2007; Zhang et al., 2014). Increasing the levels of PN components enhances proteostasis capacity (Lindquist, 1986; Zhang et al., 2014) (see section 1.11, Fig. 1.10). Furthermore, increasing the levels of chaperones improves the yields of soluble proteins of overexpressed
heterologous proteins in *E. coli* (de Marco, 2007; de Marco et al., 2007; Makino et al., 2011; Zhang et al., 2014). The $\sigma^{32}$ is present at low levels and is inactive at low temperatures, as well as when FtsH (with the help of other chaperones) degrades it. It is also inactivated when free DnaK/DnaJ and GroEL/GroES bind to it (Guisbert et al., 2008). Under this condition, the concentrations of PN components are present at their normal levels.

Although extensive research has been carried out on the main functions of individual PN components, the mechanism by which these components function as a whole has not been clear (Dickson and Brooks, 2013; Kim et al., 2013; Powers et al., 2012). How does the PN, consisting of molecular chaperones and degradation enzymes, modulate the folding fate of proteins with different intrinsic biophysical properties? Which molecular chaperones and/or degradation enzymes affect the folding fate of a model protein? We hypothesized that molecular chaperones and degradation enzymes modulate the folding fate of a protein by working together to maintain proteostasis. The sensitivity of a protein to a particular chaperone or degradation enzyme depends on the protein’s intrinsic biophysical properties, and on the concentrations of the protein and chaperones and degradation enzymes at any given time. We used an integrated computational and experimental approach to address the questions presented above.

We explored the impact of different PN components on the folding fates of model proteins with low stabilities and high aggregation propensities because
these proteins challenge proteostasis capacity (Gidalevitz et al., 2006; Olzscha et al., 2011). We expressed model proteins in *E. coli* cells and determined their folding fates *in vivo* under varying PN components. We then used FoldEco to estimate the biophysical properties of the model proteins based on their *in vivo* folding fates. In our study, protein “folding fate” refers to the amount of existing total, soluble and aggregated proteins at any given synthesis rate and time. It is assumed that soluble proteins include proteins that are natively folded (N) and those unfolded (U) and misfolded (M) proteins that are free or bound to chaperones making them soluble. Aggregated proteins (A) consist of proteins that are not soluble.

Three model proteins were used. These were the unstable and aggregation-prone variants of murine cellular retinoic acid-binding protein 1 (CRABP1), *E. coli* dihydrofolate reductase (EcDHFR) and a de novo designed retroaldolase enzyme (RA114.3) (Cho et al., 2015). These three proteins differ in ancestral origins (mammalian for CRABP1, endogenous *E. coli* for DHFR, and de novo designed for RA), folds (β barrel for CRABP1, αβα sandwich for DHFR, and α/β for RA) and sequences (no significant sequence similarity) (Cho et al., 2015). By using a diverse set of proteins, we hoped to understand more about the qualities of the PN as a system, and determine the key players for the different functions of the PN (Cho et al., 2015). I will focus my discussion on results using CRABP1 as a model, because this was the protein that my studies focused on. The other two proteins were primarily investigated by our collaborators, Younhee
Cho, Xin Zhang, and Yu Liu (Cho et al., 2015). The CRABP1 variant that was used in this study was CRABP1(Y133S) because it is the most destabilized, most aggregation-prone, and slowest folder among the CRABP1 variants (Fig. 2.2, Fig. 2.15).

3.2 Results

To recap, our data from Chapter 2 showed that CRABP1(Y133S) was mostly aggregated (~>70% aggregated) after 1, 2, and 3 h of expression (Fig. 2.15) under the adapted-basal PN condition. The results suggest that the adapted-basal levels of chaperones and degradation enzymes are insufficient to prevent the aggregation of CRABP1(Y133S). Note that in our study, “adapted-basal” PN refers to the condition of PN as a result of slight perturbations in the levels of PN components due to heterologous expression of proteins in E. coli (as discussed in Chapter 2, Fig. A.3).

3.2.1 Enhancement of the whole proteostasis network (PN) through I54N σ32 overexpression reduces the fraction of aggregated protein

To assess the impact of the full PN on the folding fate of CRABP1(Y133S), the heat shock transcription factor, σ32, was co-expressed with CRABP1(Y133S) to up-regulate PN components that are under its regulon. The mutant I54N of σ32 was used because this variant is not affected by post-translational regulation unlike the wild-type (Guisbert et al., 2008; Yura et al., 2007; Zhang et al., 2014).
The I54N $\sigma^{32}$ and CRABP1(Y133S) were expressed on two different vectors and under the control of orthogonal promoters so that they could be induced independently. The I54N $\sigma^{32}$ gene was introduced into pBAD vectors (Fig. B.1). The CRABP1(Y133S) gene was inserted into a pET vector. The expression of I54N $\sigma^{32}$ for 1 h prior to expression of CRABP1(Y133S) for 2 h caused the major PN components to increase by ~3-fold compared to those in adapted-basal condition (Table B.1). The induction of the HSR through I54N $\sigma^{32}$ expression was performed to ensure sufficiently high levels of KJE, GroELS, and Lon, and other PN components. Results showed that the up-regulation of the PN components through I54N $\sigma^{32}$ overexpression virtually eliminated CRABP1(Y133S) aggregates (Fig. 3.1). The concentrations of total, soluble and aggregated protein were all decreased. The up-regulated PN through I54N $\sigma^{32}$ overexpression had a beneficial effect on CRABP1(Y133S) (Fig. 3.1).

3.2.2 Up-regulation of PN components decreases CRABP1(Y133S) aggregation to different extents

To dissect which components of the PN are responsible for the facilitation of folding and decrease in aggregation under the I54N $\sigma^{32}$ overexpression condition, chaperones and/or degradation enzymes are overexpressed individually or in combinations. For individual up-regulation, the PN components and CRABP1(Y133S) were expressed on two different vectors and under the control of orthogonal promoters. The KJE system, GroELS system, Lon, TF,
HtpG, and IbpA/B were introduced into pBAD vectors (Fig. B.1). The CRABP1(Y133S) gene was inserted into a pET vector. In our experiments, the expression of PN components was induced with arabinose, and then after 1 h, CRABP1 was induced with IPTG for 2 h. Different concentrations of arabinose were added to up-regulate PN components. Furthermore, the combinations of KJE and GroELS (KJE+GroELS), KJE and Lon (KJE+Lon), GroELS and Lon (GroELS+Lon), KJE and HtpG (KJE+HtpG), KJE and ClpB (KJE+ClpB) were inserted into vectors controlled by two promoters for each pair: one was under an arabinose promoter and the other was under the tetracycline promoter (Fig. B.1).

3.2.2.1 Up-regulation of KJE, GroELS, or Lon individually decreases CRABP1(Y133S) aggregation

To determine the effects of individual up-regulation of PN components on the folding fate of CRABP1(Y133S), individual components of KJE, GroELS, Lon, TF, HtpG, and IbpA/B were up-regulated by inducing each with appropriate concentrations of arabinose for 1 h, and then expressing CRABP1 for 2 h. The levels of PN components are controlled by the concentration of added arabinose. High concentration of arabinose is added to induce TF, HtpG and IbpA/B. High and low concentrations are added to induce KJE, GroELS, and Lon. There were two conditions for KJE (KJE-low and KJE-high), two for GroELS (GroELS-low and GroELS-high), and three for Lon (Lon-low, Lon-medium or Lon-med and Lon-high). Higher arabinose concentrations were added in KJE-high, GroELS-high
and Lon-high, lower levels in KJE-low, GroELS-low, and Lon-low, and medium levels in Lon-med. The Lon-med condition was done because total protein in the Lon-high condition decreased drastically. The resulting concentrations of PN components were increased in these conditions compared to those under adapted-basal conditions (Table B.1, Fig. B.2).

Fig. 3.2 shows the representative gels and bar graphs for the folding fates of CRABP1(Y133S) after 2 h of expression under the conditions of adapted-basal PN and when individual components of the PN were up-regulated in E. coli cells. Under adapted-basal conditions, the concentration of total CRABP1 was 106 \( \pm \) 5 \( \mu \)M. The fraction of aggregated protein was 76\% \( \pm \) 1\% aggregated (Fig. 3.2C) (Cho et al., 2015).

Up-regulation of KJE in KJE-high condition decreased the concentration of the total protein. The fraction aggregated also decreased (from 76\% \( \pm \) 1\% to 46\% \( \pm \) 5\%), but aggregates were not eliminated (Fig. 3.2). In KJE-low, the concentrations of total protein and fraction aggregated also decreased, but to a lesser extent (Cho et al., 2015).

Up-regulation of GroELS under GroELS-high condition did not eliminate aggregates and had no significant effect on the fraction of aggregates (from 76 \( \pm \) 1\% to 66 \( \pm \) 11\%) (Fig. 3.2C).

Up-regulation of Lon decreases total protein. Under Lon-low and Lon-med conditions, the concentration of aggregated protein was decreased (Fig. 3.2) while the concentration of soluble protein was not significantly affected. In Lon-
high conditions, the concentration of total, soluble, and aggregated proteins were decreased drastically. The fraction of aggregated protein was decreased while the fraction of soluble protein was increased because of the ongoing degradation (Cho et al., 2015).

Up-regulation of TF, HtpG, and IbpA/B did not have a significant effect on the fraction of aggregated protein (data not shown).

3.2.2.2 The beneficial effects of up-regulation of KJE, GroELS, and Lon in pairs is largely additive

To determine the effects of the pairwise up-regulation of PN components on the folding fate of CRABP1(Y133S), KJE+GroELS, KJE+Lon, GroELS+Lon, KJE+HtpG, KJE+ClpB were co-expressed with CRABP1 with the same method as described above except that high concentration of arabinose (0.2% (w/v) (similar concentration of arabinose for KJE-high, GroELS-high, and Lon-high) and $5 \times 10^{-7}$% (w/v) (5 ng/mL) tetracycline were added. The levels of major chaperones were ~2- to 4- fold higher than those in adapted-basal condition (Table B.1). These levels were lower than when they were up-regulated on their own (see section 3.2.2.1; Table B.1).

Up-regulation of KJE+Lon and GroELS+Lon decreased the fraction of aggregated protein significantly, from 76 ± 1% to 19 ± 3% and 10 ± 5%, respectively (Fig. 3.3). The aggregated fraction was also decreased with KJE+GroELS but to a lesser extent, from 76 ± 1% to 54 ± 5%. KJE+Lon and
GroELS+Lon were the most effective pairs in decreasing aggregation (Cho et al., 2015). The beneficial effects of KJE+GroELS, KJE+Lon, and GroELS+Lon are largely additive. For example in KJE+GroELS pair, it was mostly KJE that was helping in decreasing aggregation.

Both combinations of KJE+ClpB and KJE+HtpG decreased the fraction of aggregated protein similarly to when KJE was expressed alone even with low levels of KJE (in KJE-low) (see Fig. 3.2B-C and Fig. B.3). (Cho et al., 2015).

Together, these results indicate that KJE, GroELS, and Lon are primarily responsible for the observed effects of I54N $\sigma^{32}$ overexpression. Moreover, the up-regulation of PN components through induction of heat shock response via $\sigma^{32}$ overexpression enables the PN to prevent the aggregation of aggregation-prone protein CRABP1(Y133S) even at high expression levels (Cho et al., 2015).

3.2.3 Up-regulation of PN components decreases the aggregation of other model proteins (from work of Younhee Cho, Xin Zhang, Yu Liu)

As mentioned earlier, we wanted to understand about the qualities of the PN as a system, and determine the key players to the different functions of the PN. We then used a broad selection of proteins, including CRABP1(Y133S). With our collaborators, we studied two other proteins with low stabilities and high aggregation-propensities— an unstable variant of *E. coli* DHFR (m-EcDHFR with M42T/H114R mutation) and a de novo designed enzyme retroaldolase (m-RA114 with E10K/D120V/N124S/L225P mutation). These variants were less stable than
their wild-type counterparts (Cho et al., 2015). We performed the same sets of experiments described above. We determined the folding fates of these model proteins upon expression in *E. coli* for 2 h at 30 °C under PN conditions similar to those used for CRABP1. I will discuss the results briefly.

Under adapted-basal conditions, the total protein concentrations for m-EcDHFR and m-RA114 were 498 ± 58 μM and 385 ± 40 μM, respectively. The fractions of aggregated protein were 46 ± 3% and 86 ± 1% for m-EcDHFR and m-RA114, respectively (data not shown, but in (Cho et al., 2015)). To determine the effect of up-regulation of the whole PN, we also overexpressed I54N σ^32. Results showed that increased levels of PN components through I54N σ^32 expression virtually eliminated aggregation of m-EcDHFR and m-RA114 (similar results for CRABP1(Y133S)) (Cho et al., 2015).

To investigate which chaperones and/or degradation enzymes are responsible for the observed effects, we overexpressed individual PN components, and in pairs (same PN components used in CRABP1(Y133S). Their levels were also those presented in Table B.1 and Fig. B.2. Results show that up-regulation of KJE (in KJE-high condition) decreased the aggregated fraction of m-EcDHFR significantly from 46 ± 3% to 25 ± 4% and from 86 ± 1% to 37 ± 3% for m-RA114. The total test protein concentrations also decreased by 30-40% (Cho et al., 2015). Up-regulation of GroELS (in GroELS-high condition) decreased the fraction of aggregated protein for both m-EcDHFR and m-RA114. The aggregated fractions decreased from 46 ± 3% to 22 ± 9% for m-EcDHFR and
from 86 ± 1% to 39 ± 3% for m-RA114 (Cho et al., 2015). The results were in contrast to those of CRABP1(Y133S) (Fig. 3.2). Furthermore, up-regulation of Lon decreased the levels of both the soluble and aggregated protein of m-EcDHFR similarly. The decrease was larger for the aggregated fraction of m-RA114 (similar to CRABP1(Y133S)). Moreover, the up-regulation of TF, HtpG, and LbpA/B had no significant effect on the fraction aggregated. The pairwise up-regulation of GroELS+Lon and KJE+Lon decreased the fraction of aggregated protein for both m-EcDHFR and m-RA114, while the GroELS+KJE was less effective in suppressing aggregation (Cho et al., 2015). Furthermore, KJE+ClpB and KJE+HtpG had no significant effect on the fraction of aggregated protein. Overall, results showed that the major contributors of the effects we observed under the I54N σ32 were KJE, GroELS, and Lon.

3.2.4 Low levels of expression of CRABP1(Y133S)

In section 3.2.2 we determined the effects of varying PN conditions on the folding fate of CRABP1(Y133S) expressed at 2 h. To assess the sensitivity of lower levels of CRABP1(Y133S) on increasing levels of PN components, CRABP1(Y133S) expression was induced for 1 h under adapted-basal and varying chaperone conditions. The same sets of experiments described above were performed except that the induction time was 1 h, in which total synthesized protein was half what it was after 2 h induction.
Based on quantitative Western blotting, under adapted-basal conditions, the total protein concentration of CRABP1(Y133S) was 57 ± 4 μM (Fig. 3.4, Table B.2). The fraction of aggregated protein was 71% ± 4% aggregated (Fig. 3.4C) (Cho et al., 2015).

The expression of I54N σ^{32} for 1 h prior to expression of CRABP1(Y133S) caused the major PN components to increase by ~2- to 3-fold compared to those in adapted-basal condition (Table. B.3, Fig. B.4). Results showed that the up-regulation of the PN components through I54N σ^{32} overexpression eliminated CRABP1(Y133S) aggregates. Proteins are ~100% soluble; however, the concentrations of total, soluble and aggregated protein were all decreased (Fig. 3.4, Table B.2).

To investigate which PN components are responsible for the observed effects, we up-regulated KJE, GroELS, and Lon individually and in pairs. We chose to up-regulate these three because our results from sections 3.2.1 and 3.2.2 (Fig. 3.1-3, Fig. B.3) indicate that these were the major contributors of what we observed when I54N σ^{32} was expressed with CRABP1.

Individual up-regulation of KJE, GroELS, and Lon resulted to concentrations of major PN components (DnaK, GroEL, and Lon) to increase ~2- to 8- fold compared to those under adapted-basal condition (Table B.3).

Up-regulation of KJE under KJE-high condition decreased the concentrations of total, soluble and aggregated protein (35 ± 10 μM, 17 ± 5 μM, and 18 ± 5 μM, respectively) as well as fraction aggregated (50% ± 1%
aggregated) (Fig. 3.5, Table B.2) compared to those under adapted-basal conditions. The total protein decreased by ~40%. Under the KJE-low condition, the concentrations of total, soluble and aggregated protein (42 ± 10 μM, 16 ± 5 μM, and 26 ± 6 μM, respectively) (Fig. 3.5B), and the fraction aggregated (62% ± 2%) (Fig. 3.5C) were decreased, but proteins were mostly aggregated. The total protein decreased by ~30%.

Up-regulation of GroELS decreased the fraction of aggregated protein and increased the fraction of soluble protein compared to adapted-basal conditions. Under GroELS-high condition, the concentrations of total, soluble and aggregated proteins were 45 ± 9 μM, 35 ± 4 μM, and 10 ± 5 μM, respectively (Fig. 3.5B). The aggregated fraction decreased significantly from 71% ± 4% to 21% ± 6%, while increasing the fraction of soluble protein from 29% ± 4% to 79% ± 6% (Fig. 3.5C). The concentrations of total, soluble and aggregated proteins were 53 ± 4 μM, 26 ± 2 μM, and 27 ± 2 μM, respectively (Fig. 3.5B), and the aggregated fraction was 51% ± 1% (Fig. 3.5C) under GroELS-low condition. These results were different from those when CRABP1(Y133S) was expressed for 2 h under GroELS-low and GroELS-high conditions (compare Fig. 3.2 and Fig. 3.5). The observed increase in the levels of soluble proteins and decrease in aggregated proteins could be due to misfolded or unfolded proteins being helped by GroELS to fold to their native states before they can aggregate. Further, the slight decrease in total protein could also be due to misfolded proteins being rescued and then are degraded before they can fold.
Up-regulation of Lon decreased the levels of total protein, but increased the fraction of soluble and decreased fraction of aggregated proteins compared to those under the adapted-basal levels. Under Lon-low conditions, the concentrations of total, soluble and aggregated proteins were $25 \pm 2 \mu M$, $22 \pm 1 \mu M$, and $3 \pm 1 \mu M$, respectively (Fig. 3.5B). The aggregated fraction decreased from $71\% \pm 4\%$ to $13\% \pm 1\%$ (Fig. 3.5C). The concentration of total protein was decreased by $\sim 60\%$, the concentration of soluble protein was increased by $\sim 30\%$, and the concentration of aggregated protein was decreased by $\sim 90\%$ (Fig. 3.5B, Table B.2). Under Lon-med conditions, the concentrations of total, soluble and aggregated proteins were $12 \pm 1 \mu M$, $11 \pm 1 \mu M$, and $1 \mu M$, respectively. Aggregation was almost eliminated. The aggregated fraction decreased from $71\% \pm 4\%$ to $9\% \pm 1\%$. The total protein was decreased by $\sim 80\%$. Under Lon-high conditions, the concentrations of total, soluble and aggregated proteins were $0.6 \mu M \pm 0.2 \mu M$, $0.6 \mu M \pm 0.2 \mu M$, and $0 \mu M$, respectively. The aggregated proteins were virtually eliminated, but the total protein was very low. Apparently, the concentration of aggregated protein was greatly reduced under all Lon conditions while the concentration of soluble protein was not affected as much (except under Lon-high conditions when almost all proteins that were produced were degraded) (Fig. 3.2). The increase in the amount of soluble protein under Lon-low conditions could be due to the degradation of misfolded states, which leads to a decrease in the pool of M states that would otherwise saturate the
other PN components; thus, making the other PN components more efficient in helping the folding of other states to their native state.

Pairwise up-regulation increased the levels of major chaperones by ~2- to 6- fold compared to those in adapted-basal condition (Table B.3, Fig. B.4), but lower than their levels when they were expressed individually. Fig. 3.6 shows the representative gels and bar graphs for the quantitation of the absolute protein concentrations and the fraction of soluble and aggregated protein of CRABP1(Y133S) under the adapted-basal conditions and the pairwise up-regulation of PN components. In KJE+GroELS condition, the concentrations of total, soluble and aggregated proteins were 41 ± 20 μM, 37 ± 17 μM, and 4 ± 3 μM, respectively (Fig. 3.6B, Table B.2). The fraction of aggregated protein decreased from 71% ± 4% to 6% ± 4% (Fig. 3.6C). In the KJE+Lon condition, the concentrations of total, soluble and aggregated proteins were 43 ± 11 μM, 35 ± 7 μM, and 8 ± 4 μM, respectively (Fig. 3.6B, Table B.2). The fraction of aggregated protein decreased from 71% ± 4% to 18% ± 4% (Fig. 3.6C). In GroELS+Lon, the concentrations of total, soluble and aggregated proteins were 54 ± 3 μM, 50 ± 2 μM, and 4 ± 2 μM, respectively (Fig. 3.6B). The aggregated fraction decreased from 71% ± 4% to 8% ± 3% (Fig. 3.6C). All the pairwise conditions were efficient in suppressing aggregation of CRABP1(Y133S), and their effects were mostly additive. In all these conditions, the concentration of soluble protein increased while the levels of aggregated protein decreased (Fig. 3.6).
In sum, the sensitivity of CRABP1(Y133S) to molecular chaperones and degradation enzymes, particularly to KJE, GroELS, and Lon is concentration-dependent.

3.2.5 Analysis of CRABP1(Y133S) Protein Folding Fates

In order to understand how the *E. coli* PN maintains proteostasis for our test proteins, we performed a quantitative analysis of our data. First, we utilized a phenomenological method to extract protein-specific trends, and then secondly, we used the FoldEco program to model mechanistically how the sensitivity of a protein to the PN reports on its biophysical properties. These quantitative analyses were done by our collaborator, Evan Powers and are reported (Cho et al., 2015). I will only focus on the analysis for CRABP1(Y133S) expressed for 2 h.

3.2.5.1 Phenomenological Models

To quantify the effect of KJE, GroELS, and Lon on the test proteins, the *in vivo* folding fate data for each test protein were fit to the phenomenological models below:

\[
\text{[Agg]}_{\text{rel},X} = c_{\text{Agg},X} + a_{K,X}[\text{DnaK}]_{\text{rel}} + a_{G,X}[\text{GroEL}]_{\text{rel}} + a_{L,X}[\text{Lon}]_{\text{rel}} \quad \text{(Equation 1)}
\]

\[
\text{[Sol]}_{\text{rel},X} = c_{\text{Sol},X} + S_{K,X}[\text{DnaK}]_{\text{rel}} + S_{G,X}[\text{GroEL}]_{\text{rel}} + S_{L,X}[\text{Lon}]_{\text{rel}} \quad \text{(Equation 2)}
\]

where “X” is the test protein; “Agg” is aggregated and “Sol” is soluble; \([\text{Agg}]_{\text{rel},X}\) and \([\text{Sol}]_{\text{rel},X}\) are the concentrations of the aggregated and soluble protein normalized to the total concentration under adapted-basal conditions; \([\text{DnaK}]_{\text{rel}}\),...
[GroEL]_{rel} and [Lon]_{rel} are the concentrations DnaK, GroEL, and Lon relative to their adapted-basal concentrations; \( a_{K,X} \), \( a_{G,X} \), and \( a_{L,X} \) are the gradients of \([Agg]_{rel,X}\) for \([DnaK]_{rel}\), \([GroEL]_{rel}\), and \([Lon]_{rel}\), respectively; \( s_{K,X} \), \( s_{G,X} \), and \( s_{L,X} \) are the gradients of \([Sol]_{rel,X}\) for \([DnaK]_{rel}\), \([GroEL]_{rel}\), and \([Lon]_{rel}\), respectively; and \( c_{Agg,X} \) and \( c_{Sol,X} \) are the model intercepts. The efficacies of the PN components are quantified by the gradient parameters. For example, a large, positive value of \( s_{G,X} \) increases value of \([Sol]_{rel,X}\) as the concentration of GroEL increases, which means that the protein benefits from GroEL (Cho et al., 2015).

The in vivo folding fate data of CRABP1(Y133S) were fit to equations (1) and (2). The quality of the fit of Equation (1) to the normalized concentrations of aggregated protein is moderately good (adjusted \( R^2 = 0.68 \)) (Fig. 3.7A, red data points). The negative values for the parameters \( a_{K} \), \( a_{G} \), and \( a_{L} \) indicate that all of the PN components decrease aggregation (Fig. 3.7B, red bars). However, their effects differ. KJE is the most effective in decreasing aggregation, followed by Lon and GroELS (\( a_{K,X} < a_{L,X} < a_{G,X} \)). The fit of Equation (2) to the normalized concentrations of soluble protein is good (adjusted \( R^2 = 0.71 \)) (Fig. 3.7A, blue data points). The negative value for the parameter \( s_{L,CRABP1} \) indicates that Lon decreases the amount of soluble protein (Fig. 3.7B). The value of \( s_{L,CRABP1} \) is much smaller in magnitude than \( a_{L,CRABP1} \), which suggests that Lon up-regulation preferentially decreases the levels of aggregated protein. In addition, \( s_{K,CRABP1} \) is substantial and positive indicating that KJE is effective in increasing the concentration of soluble protein and in decreasing the concentration of
aggregated protein (Fig. 3.7B). The value of $s_{G,CRABP1}$ is much smaller than that of $s_{K,CRABP1}$ (Fig. 3.7B), which suggests that GroELS is not effective in increasing the soluble protein of CRABP1(Y133S) (Cho et al., 2015).

The phenomenological models described above were used to quantify the effects of KJE, GroELS, and Lon on the folding fate of CRABP1(Y133S); however, these models cannot inform us about the mechanism by which the protein behaves. A protein’s intrinsic biophysical properties influence folding fate as discussed in Chapter 2, and also based on FoldEco results in which a protein’s sensitivity to different chaperoning mechanisms should be influenced by a protein’s intrinsic biophysical properties (Dickson and Brooks, 2013; Powers et al., 2012). Thus, a protein’s intrinsic biophysical properties should reflect the values of the best-fit parameters for Equations (1) and (2) (Cho et al., 2015). To investigate the relationship between its biophysical properties and how CRABP1(Y133S) behaves under varying PN conditions, we used FoldEco to fit our in vivo folding fate data by using the protein’s intrinsic biophysical properties as adjustable parameters.

**3.2.5.2 Analysis of protein folding fate of CRABP1(Y133S) using FoldEco**

FoldEco models the proteostasis network in *E. coli*, in which processes such as protein synthesis and folding, chaperoning by KJE and GroELS systems, and degradation by Lon, as well as PN components are interconnected (discussed in Chapter 1, section 1.13; Fig. 1.12) (Powers et al., 2012). FoldEco
consists of ordinary differential equations that describe the kinetics of these processes \textit{in vivo} and the time-dependent concentrations of each species (e.g. concentration of unfolded U, misfolded M, or native N states). We fit FoldEco to our \textit{in vivo} folding fate data (i.e. the concentrations of soluble and aggregated protein) by changing the parameters in FoldEco until the output folding fates (i.e. concentrations of soluble and aggregated protein) optimally matched our data \textit{in vivo}. FoldEco has many parameters, but in running FoldEco (for our purpose), only the protein synthesis rate and biophysical parameters were adjustable parameters. The biophysical parameters are the folding rate and equilibrium constants \((k_f \text{ and } K_f)\), misfolding rate and equilibrium constants \((k_m \text{ and } K_m)\), and aggregation rate and equilibrium constants \((k_a \text{ and } K_a)\), which is the equilibrium constant when a misfolded monomer is added to an aggregate, i.e., the inverse of critical concentration). The other parameters are likely to be independent of, or weakly dependent on, the nature of the client protein (Powers et al., 2012). For example, the effect of one bound client on a chaperone/co-chaperone interaction is likely to be similar to that of other bound clients. In addition, chaperones interact with client proteins promiscuously (Aoki et al., 2000; Landry and Gierasch, 1991; Rudiger et al., 1997; Wang et al., 1999; Weinstock et al., 2014). which means that chaperone-client interaction parameters are also likely to be similar for most clients (Cho et al., 2015). The values for such parameters have been derived from available literature data (Powers et al., 2012). By adjusting the biophysical parameters to fit the FoldEco model to our \textit{in vivo} data, we can obtain
parameters considered to be “effective parameters” since we have applied the simple, generic folding mechanism used in FoldEco.

The quality of the FoldEco fit to our data is fairly good. The best-fit value of $[\text{Sol}]_{\text{rel,CRABP1}}$ and $[\text{Agg}]_{\text{rel,CRABP1}}$ deviate from experimental data on average by 0.12 (Fig. 3.8A). However, there are large residuals, particularly in GroELS-low and GroELS-high. The effective biophysical parameters are summarized in Table B.4. Unfortunately, the fit did not provide us definite parameter estimates. However, we were still able to define some relationships based on these parameter estimates, and extracted some general features regarding the relationship of intrinsic biophysical properties and in vivo folding fate behavior of CRABP1(Y133S) (Cho et al., 2015).

Fig. 3.8B illustrates certain aspects about the biophysical properties of CRABP1(Y133S). Under adapted-basal conditions, our modeling shows that CRABP1(Y133S) misfolds faster than it folds ($k_m > k_f$), and this causes an accumulation of misfolded states (Fig. 3.8B). These high levels of misfolded states are sensitive to KJE and Lon overexpression as shown by the increase in the utilization of KJE and Lon (Fig. 3.8B, red numbers). Here, Lon preferentially degrades misfolded states (70%) to unfolded states (30%), thereby decreasing the concentration of aggregated protein. Moreover, KJE converts misfolded states to unfolded states causing an increase in the concentration of soluble protein and a decrease in the concentration of aggregated protein (Fig. 3.8B). Unfortunately, the chaperoning function of GroELS cannot be elucidated from the
FoldEco fits of CRABP1(Y133S). According to the data acquired after 2 h of protein expression, GroELS does not have a significant effect on the fate of CRABP1(Y133S) when expressed at 2h (Fig. 3.2; Fig. 3.8A). It could be that CRABP1(Y133S) binds to GroELS differently. To determine the extent of the interaction of GroEL and CRABP1(Y133S), FoldEco was fit to the in vivo folding fate data of CRABP1(Y133S) with varying equilibrium association constant between CRABP1(Y133S) and GroEL ($K_{\text{Gro-CRABP1}}$). With a $K_{\text{Gro-CRABP1}}$ that was 1000-fold lower than the value derived from literature, the overall fit was improved by 20% (mean residuals = 0.10). However, the fit still did not provide precise estimates of the biophysical parameters. Since the fit was improved with a large decrease in $K_{\text{Gro-CRABP1}}$, this suggests that CRABP1(Y133S) binds weakly to GroEL (Cho et al., 2015).

3.3 Discussion

We determined how the PN, consisting of molecular chaperones and degradation enzymes, modulates the folding fate of three model proteins with different intrinsic biophysical properties. Using a diverse set of model proteins (CRABP1, DHFR, and RA), our results show that molecular chaperones and degradation enzymes modulate the folding fate of a protein by working together to maintain proteostasis. The up-regulation of the KJE, GroELS, and Lon decreases the aggregation of our model proteins to different extents. These are consistent with the previous studies of the effect of up-regulation of chaperones.
on the yield of overexpressed heterologous proteins (de Marco, 2007; de Marco et al., 2007; Makino et al., 2011; Zhang et al., 2014). The KJE, GroELS, and Lon pathways form an efficacious triad for maintaining proteostasis. Each pathway has a distinct role, and these roles are interconnected to each other towards maintaining proteostasis—“anti-misfolding” by KJE, “pro-folding” by GroELS, and “concentration control” by Lon. These three are especially effective when they are up-regulated via the expression of σ^{32} transcription factor.

Furthermore, our in vivo folding fate data and two quantitative models (phenomenological model and FoldEco) indicate that the contributions of KJE, GroELS, and Lon to proteostasis depend on the intrinsic biophysical properties of their substrates and on the concentrations of these PN components and substrates at any given time. For example, CRABP1(Y133S) (the model protein that I discussed thoroughly in this chapter) was ~>70% aggregated under the adapted-basal PN condition because of its kinetic and thermodynamic properties (explained in Chapter 2) and the fact that the PN is insufficient to aid its folding and minimize aggregation (Fig. 3.1). Its folding fate is changed when PN conditions are altered, such as when the KJE, GroELS, and Lon pathways are up-regulated individually, in pairs, or simultaneously, or when its concentration is varied. At low concentrations of CRABP1(Y133S) (when it is expressed for 1 h), the up-regulation of KJE, GroELS, and Lon are more effective in decreasing aggregation compared to when CRABP1(Y133S) is present at high levels (when it is expressed for 2 h) (Fig. 3.2, Fig. 3.4). The sensitivity of CRABP1(Y133S) to
PN components depends on the concentration of both CRABP1(Y133S) and the PN components. At low and high concentrations of CRABP1(Y133S), the up-regulation of KJE decreases aggregated protein but not significantly (Fig. 3.2, Fig. 3.5). It could be that DnaK, DnaJ or GrpE is limiting in this system. One possibility is DnaJ as it is shown in the FoldEco simulation (Fig. A.9), in which under the adapted-basal condition, the usage for DnaJ is ~80% at 1 h of CRABP1(Y133S) expression and ~90% at 2 h of CRABP1(Y133S). With wet lab experiments, under the KJE-low and KJE-high conditions, the levels of DnaJ at 1 h and 2 h of CRABP(Y133S) expression were increased by ~20-fold and ~23-fold, respectively (Table B.1, Table B.3). But still, it seems that it is the limiting factor (Heldens et al., 2010). Furthermore, the up-regulation of GroELS decreases aggregated protein significantly at low CRABP1(Y133S) concentration (Fig. 3.5), while it has no effect on the aggregates at high CRABP1(Y133S) concentration (Fig. 3.2). An implication of this is that the GroELS system (either GroEL or GroES or both) is limiting when concentrations of CRABP1(Y133S) are high. The system maybe saturated with high concentrations of unfolded (U) or misfolded (M) states. Moreover, up-regulation of Lon (even just by ~2-fold) increases the fraction of soluble protein and decreases the fraction of aggregated protein of CRABP1(Y133S) when it is at low concentration (Fig. 3.5). Higher levels of Lon (~8-fold higher than that of adapted-basal Lon) are needed to increase the fraction of soluble CRABP1(Y133S) when it is at high concentration (Fig. 3.2). The concentration of total protein decreases significantly when Lon is increased.
(Fig 3.2). The combinations of KJE, GroELS, and Lon are all effective in increasing fraction soluble and decreasing fraction aggregated under low concentrations of CRABP1(Y133S); however, at high concentrations of CRABP1(Y133S), only the GroELS+Lon and KJE+Lon combinations are effective (Fig. 3.3). The up-regulation of all three when expressed via $\sigma^{32}$ expression is efficient in eliminating aggregates in both low and high concentrations of CRABP1(Y133S) (Fig. 3.1, Fig. 3.4).

Under the KJE-low and KJE-high conditions (Fig. 3.2 and Fig. 3.5), the decrease in total protein may be due to the facilitation of degradation by DnaK or DnaJ by delivering substrates to proteases (Sherman and Goldberg, 1992). However, based on FoldEco, it is more likely that the decrease in the concentration of total protein are due to KJE rescuing misfolded proteins and these misfolded proteins being degraded before they can fold or re-aggregate (Fig 3.8B). Further, up-regulation of KJE (in KJE-high) and up-regulation of GroELS (in GroELS-low) have similar effects on the folding fate of CRABP1(Y133S) which suggests that KJE and GroELS can compensate for each other, depending on their levels. When the two are combined, they are effective in decreasing aggregates. It could be that KJE converts M to U and hands it off to GroELS, which in turn helps the U to fold to N.

The up-regulation of the whole HSR (particularly, KJE, GroELS, and Lon) through $\sigma^{32}$ overexpression are beneficial for all CRABP1(Y133S), m-EcDHFR, and m-RA114 (Cho et al., 2015). Like CRABP1(Y133S), m-EcDHFR and m-
RA114 are also sensitive to the up-regulation of PN components; however, their folding fates are affected differently. The extent of the effects of individual and pairwise up-regulation of KJE, GroELS, and Lon are different for the three model proteins. One particular difference is that at 2 h expression of model proteins, up-regulation of GroELS decreases the aggregates of m-EcDHFR and m-RA114, but no significant effect on CRABP1(Y133S).

The use of the phenomenological and FoldEco models enabled us to understand how KJE, GroELS, and Lon contribute to proteostasis and how their effects are dependent on the substrate’s biophysical properties. The two models have two similar features: 1) they only include KJE, GroELS, and Lon as PN components, and 2) they have no direct mechanisms by which these components collaborate (Cho et al., 2015). The good fits of the models in light of the first feature suggest that KJE, GroELS, and Lon are the major contributors to proteostasis under the conditions of our experiments. In addition, our results also indicate that KJE, GroELS, and Lon act independently under our experimental conditions, and that their effects are mostly additive. However, some of the fits were not good, and that could be due to not accounting for collaborations between KJE, GroELS, and Lon in the model. We are also aware that cooperation between chaperone systems and degradation machineries occurs and has been reported (Bershtein et al., 2013; Bissonnette et al., 2010; Gragerov et al., 1992; Huang et al., 2001; Mogk et al., 1999; Sakr et al., 2010; Thomas and Baneyx, 2000; Tomoyasu et al., 2001). Moreover, PN components may “hand off”
substrates from one to another. An example is the collaboration of DnaK with GroEL (Langer et al., 1992). In our experiments, it could be that a large amount of unfolded protein was produced such that chaperones are saturated with newly synthesized protein not giving them time to bind to other proteins from other chaperones, thus, limiting or avoiding collaboration (Cho et al., 2015).

The FoldEco model allows us to extract information about the biophysical properties of CRABP1(Y133S) (Fig. 3.8A). For example, the poor fit to the data of CRABP1(Y133S) (expressed for 2 h) with the up-regulation of GroELS implies weak binding of CRABP1 and GroELS.

3.4 Conclusion

We have examined how the PN, consisting of molecular chaperones and degradation enzymes, modulates the folding fate of proteins with different biophysical properties. The major _E. coli_ PN components, specifically, KJE, GroELS, and Lon, work together to maintain proteostasis. KJE is for “anti-misfolding”, GroELS is for “pro-folding”, and Lon is for “concentration control”. These three are especially effective when they are up-regulated via the expression of the σ^{32} transcription factor. Furthermore, the contributions of KJE, GroELS, and Lon to proteostasis depend on the intrinsic biophysical properties of their substrates and on the concentrations of these PN components and substrates at any given time. The use of an integrated experimental and computational approach has enabled us to understand how the folding fate of a
protein is determined by both its intrinsic biophysical properties and the cellular PN.

3.5 Materials and Methods

3.5.1 E. coli strains and plasmids

The E. coli strain, K12 HMS174(DE3) (Novagen) was used as the background strain in all the experiments. The cells (K12 HMS174(DE3)) that were used in the overexpression experiments were with a plasmid containing the PN components under a pBAD promoter and/or pTet promoters on low copy number plasmids (Fig. B.1). The levels of PN components were induced with different concentrations of L-arabinose (ara) (for individual up-regulation of PN components and σ^{32}) and tetracycline (tet) (for the up-regulation of combinations of PN components).

The genes for CRABP1(Y133S), m-EcDHFR, and m-RA were inserted into pET29b vectors (used for the 2 h expression). The CRABP1(Y133S) used in 1 h expression experiments had N-terminal (His)_{10}-tag and was inserted into pET16b vector. The CRABP1(Y133S) used for the 2 h expression experiments was without the His-tag. The His tag does not affect the folding fate of CRABP1. Under the adapted-basal PN condition, the folding fates of CRABP1(Y133S) (with His tag) expressed for 2 h were similar to those of CRABP1(Y133S) (without His tag) expressed for 2 h (see Fig. 2.15 and Fig. 3.1). In addition, the effects of up-
regulating KJE (in KJE-low condition) and GroELS (in GroELS-low condition) are the same for CRABP1(Y133S) with and without the His tag (data not shown).

### 3.5.2 Protein expression in *E. coli*

Bacterial cultures were grown in LB with appropriate antibiotics at 30 °C until OD$_{600}$ was ~0.4. Different concentrations of arabinose and/or tet were added to induce PN components. In KJE-high, GroELS-high, and Lon-high conditions, PN components were induced with 0.2% (w/v) arabinose. In KJE-low, GroELS-low, and Lon-low conditions, PN components were induced with 0.005%, 0.001%, or 0.0005% w/v arabinose for the KJE pathway, the GroELS pathway, and Lon, respectively. In Lon-med, Lon was induced using 0.002% (w/v) arabinose. For the combinations, 0.2% (w/v) arabinose and $5 \times 10^{-7}$% (w/v) (5 ng / mL) tetracycline were added to induce PN components (Fig. 3.9). Control experiments under adapted-basal conditions were run in parallel with the same cultures, but 0.2% (w/v) D-glucose, which represses the arabinose operon (Tokuriki and Tawfik, 2009), was added. After 1 h of adding the necessary inducers to express the PN components, CRABP1 expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1 and 2 h. Cells were harvested by centrifugation for 15 min. Cells were lysed by resuspending with Bacterial Protein Extraction Reagent (BPER-II) (ThermoFisher Scientific), lysozyme (0.05 mg/mL) and DNase I (1 μg/mL). Cell lysates were incubated at room temperature for ~15-30 min. A fraction of the lysate was transferred to one
tube (this was the “T”-total lysate). An equal amount of volume of the lysate was transferred to another tube, which was then centrifuged at 13,500xg for 10 min at 4 °C to partition the supernatant from the pellet. The supernatant was collected as soluble (“S”) while the pellet as aggregated (“A”). The pellet was resuspended with the same volume of buffer. Gel loading buffer was added to each sample. Samples were boiled for 10 min. Samples were loaded on a 12% or 15% glycine gel, and were run on sodium dodecyl sulfate-polyacrylamide agarose gel electrophoresis (SDS-PAGE). Proteins were detected by Coomassie blue staining and imaged with Odyssey Infrared Imaging System (Li-COR Biosciences). The same methods were performed for m-EcDHFR and m-RA114.3 (Cho et al., 2015).

3.5.3 Quantitative Western blotting

For the quantitation, recombinant CRABP1 purified proteins with known amounts (pmol) were used as standards. These were also loaded on the same gel where the samples of unknown amounts were loaded. Each volume of sample loaded corresponded to 1x10^8 cells, which is equivalent to 1 mL of bacteria with OD_{600} ~0.1. After running samples under SDS-PAGE, Western blotting was performed. Samples were transferred to a PVDF membrane (Millipore), then blocked with 5% milk solution in TBST (1x TBS, 0.02% tween-20) at room temperature for 1 h, and then incubated with primary antibody (anti-CRABP1, Abcam, monoclonal, mouse, 1:5000) at room temperature for 1 h, and
secondary antibody (Li-COR antibody, anti-mouse, emission at 800 nm, from Li-COR Biosciences) at room temperature for 1 h. Western blots were visualized and quantitated using the Odyssey Infrared Imaging System (Li-COR Biosciences). A standard curve was generated based on the known amounts of proteins, and the value of the unknown was determined by extrapolation. The concentration of samples was calculated by dividing the amount of protein (mole) by the volume of cell cytoplasm (L). It is assumed that for *E. coli* grown in LB, the total cell volume is 2.9 μL per mL of cells per OD$_{600}$ unit (Volkmer and Heinemann, 2011). Subtracting the periplasm, which is ~10% of the cell volume, the volume of cell cytoplasm is 2.6 μL per mL per OD$_{600}$. The calculated total protein was the sum of soluble and aggregated protein. The fractions of soluble and aggregated proteins were calculated as the ratio of soluble to the sum of soluble and aggregated protein. The total protein (“T”) lane was a control to check for mass balance; the sum of soluble and aggregated relative to the total should be 1. The mass balance was good in all samples; the difference of the total protein lane (T) from the total of soluble and aggregated (S+A) was less than 15%. The method in calculating the concentrations of total, soluble and aggregated, as well as fractions of soluble and aggregated was similar to that in Chapter 2 (see Methods section 2.5.3).
3.5.4 Measurement of relative levels of PN components

Quantitative Western blotting was performed to determine the fold changes of chaperones and Lon. The primary antibodies used were anti-DnaK (Enzo Life Sciences, monoclonal, mouse, 1:10000), anti-DnaJ (Enzo Life Sciences, polyclonal, rabbit, 1:1000); anti-GrpE (Enzo Life Sciences, polyclonal, rabbit, 1:1000), anti-GroEL (Enzo Life Sciences, monoclonal, mouse, 1:2000), anti-GroES (Enzo Life Sciences, polyclonal, rabbit, 1:5000); and anti-Lon (kindly provided by Prof. R. T. Sauer, 1:10000). The secondary antibodies were Li-COR antibodies (emission at 680 or 800 nm, from Li-COR Biosciences). These fold changes in the levels of PN components are important in calculating (or estimating) the concentrations of PN components needed as inputs for FoldEco simulations.

3.5.5 Protein sequences

The sequences of the proteins are the following. The font in bold letter is the new residue from the mutation.

R131Q/Y133S CRABP1 (CRABP1-Y133S)

<table>
<thead>
<tr>
<th>PNFAGTWKMR</th>
<th>SSENFDELLK</th>
<th>ALGVNAMLK</th>
<th>VAVAAASKPH</th>
<th>VEIRQDGDFQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>YIKSTTTVRT</td>
<td>TEINFKVEGE</td>
<td>FEETVDGRK</td>
<td>CRSLPTWENE</td>
<td>NKIHCTQTLL</td>
</tr>
<tr>
<td>EGDGPKTYWT</td>
<td>RELANDELIL</td>
<td>TFGADDVCT</td>
<td>QYVRE</td>
<td></td>
</tr>
</tbody>
</table>

M42T/H114R EcDHFR (m-EcDHFR):

<table>
<thead>
<tr>
<th>MISLIAALAV</th>
<th>DRVIGMEMN</th>
<th>PWNLPADLAW</th>
<th>FKRNTLNKPV</th>
<th>ITGRHTWESI</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRPLPGRKNI</td>
<td>ILSSQPGTDD</td>
<td>RVTWVKSDE</td>
<td>AIAACGDVPE</td>
<td>IMVGGGRVY</td>
</tr>
<tr>
<td>EQFLPQAKKQ</td>
<td>YLTRIDAEVE</td>
<td>GDTTPFDEP</td>
<td>DDWESVFSEF</td>
<td>HDADAQNSHS</td>
</tr>
<tr>
<td>YCFEILERR</td>
<td></td>
<td></td>
<td></td>
<td>YCFEILERR</td>
</tr>
</tbody>
</table>
E10K/D120V/N124S/L225P RA114 (m-RA114):

MPRYLKWLMK DVVQLSLRRP SVRASRQRP ISLNERILEF NKRNITAIIA
EYKRKDPSSL DVRDEPIEYA KFMERYAVGL FISTEEYFN GSYETLRKIA
SSVSIPILMY DFIVKESQIV DAYSGADTV ALIVKILTER ELESLEYAR
SYGMEPLIIIDNENDLDIAL RIGARFIGIA ARDWETGEIN KENQRLISM
IPSNNVKVAK EGISERNIEELRKGPGVNAF LIGSSLMRNP EKIKELIEGS
LEHHHHHH

3.5.6 Fits of Equations (1) and (2) and FoldEco to \textit{in vivo} folding fate data

Analyses of protein folding fates were done by (Cho et al., 2015). Details can be found in (Cho et al., 2015).

Briefly, equations (1) and (2) were fit by linear regression to the \textit{in vivo} folding fates of proteins. FoldEco fits were done with least squares approach. We fit FoldEco to our \textit{in vivo} folding fate data by changing the parameters in FoldEco until the output folding fates match our data \textit{in vivo} and determine the biophysical properties of proteins. FoldEco has many parameters, but in running FoldEco (for our purpose), only the protein synthesis rate and biophysical parameters were adjustable parameters (and thus, the parameters to be estimated). The biophysical parameters were the folding rate and equilibrium constants ($k_f$ and $K_f$), misfolding rate and equilibrium constants ($k_m$ and $K_m$), and aggregation rate and equilibrium constants ($k_a$ and $K_a$) were adjustable. The other parameters are likely to be independent on, or weakly dependent on the nature of the client protein (Powers et al., 2012). The values for such parameters were derived from available literature data (Powers et al., 2012). The initial concentrations for the adapted-basal PN were: trigger factor = 14 \, \mu M, DnaK = 23.4 \, \mu M, DnaJ = 1.56 \, \mu M, GrpE = 10.4 \, \mu M, GroEL tetradecamer = 2.23 \, \mu M, GroES heptamer = 3.53

\begin{verbatim}
MPRYLKWLMK DVVQLSLRRP SVRASRQRP ISLNERILEF NKRNITAIIA
EYKRKDPSSL DVRDEPIEYA KFMERYAVGL FISTEEYFN GSYETLRKIA
SSVSIPILMY DFIVKESQIV DAYSGADTV ALIVKILTER ELESLEYAR
SYGMEPLIIIDNENDLDIAL RIGARFIGIA ARDWETGEIN KENQRLISM
IPSNNVKVAK EGISERNIEELRKGPGVNAF LIGSSLMRNP EKIKELIEGS
LEHHHHHH
\end{verbatim}
μM, Lon hexamer = 0.4 μM, and ClpB hexamer = 0.22 μM. The concentrations of these PN components under various conditions were obtained by multiplying the adapted-basal concentrations by the fold changes shown in Fig. B.3. The synthesis rate was assumed to be similar under all conditions. A range of biophysical effective parameters was obtained (from upper and lower, and with the best-fit (Table B.2) (Cho et al., 2015).
A. 

Adapted-basal  
I54N σ^{32}

B. 

Concentration (μM) 

- Total 
- Soluble 
- Aggregated 

Adapted-basal  I54N σ^{32}

C. 

Relative concentration 

- $f_{soluble}$ 
- $f_{aggregated}$ 

Adapted-basal  I54N σ^{32}
Figure 3.1 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 2 h at 30 °C under adapted-basal PN conditions and up-regulation of PN components through overexpression of I54N σ32

A. Representative gels showing the Western blots for CRABP1(Y133S) when overexpressed in *E. coli* for 2 h at 30 °C under up-regulation of PN components through overexpression of I54N σ32. The lane for “T” is for total protein, which is the cell lysate before centrifugation. The lane for “S” is for soluble and “A” is for aggregated proteins from the supernatant and pellet, respectively, after centrifugation of cell lysates for 10 min at 13,500 x g. Y133S was mostly aggregated (~>70%) under adapted-basal conditions. B. Bar graph showing the cytoplasmic concentration of CRABP1(Y133S) determined by quantitative analysis of gels shown in Fig. 3.1A. White, blue and red bars represent the concentrations of total (soluble + aggregated), soluble and aggregated proteins, respectively. Error bars represent SEM. C. Bar graph showing the concentration of soluble and aggregated protein relative to the concentration of their total (soluble + aggregated) protein determined from the concentrations in Fig. 3.1B. Blue bars (“f soluble”) represent the fraction of soluble protein. Red bars (“f aggregated”) represent the fraction of aggregated protein. For the condition with the overexpression of I54N σ32, DnaK, DnaJ, and GrpE levels were increased by ~3-, 5-, and 6-fold, respectively, compared to those in adapted-basal condition; GroEL and GroES levels were increased by ~3- and 4-fold; Lon levels were increased by ~3-fold (Table B.1) (Modified from (Cho et al., 2015)).
A.

Adapted-basal

KJE up-regulation

KJE-low
KJE-high

GroELS up-regulation

GroELS-low
GroELS-high

Lon up-regulation

Lon-low
Lon-med
Lon-high

T: total  S: soluble  A: aggregated

B.

[Bar chart showing concentration (µM) for various conditions.]

KJE  GroELS  Lon

Adapted-basal  low  high  low  high  low  med  high

K: 2x  J: 10x  E: 5x  K: 4x  J: 23x  E: 10x  EL: 3x  ES: 2x  EL: 8x  ES: 5x

Total  Soluble  Aggregated
Figure 3.2 Folding fates of CRABP1(Y133S) upon expression in E. coli for 2 h at 30 °C under adapted-basal conditions and with up-regulation of individual PN components

A. As in Fig. 3.1, but with the up-regulation of individual PN components. Their fold changes relative to basal are shown (Table B.1). “K”-DnaK, “J”-DnaJ, “E”-GrpE, “EL”-GroEL, “ES”-GroES. (Modified from (Cho et al., 2015)).
A. Adapted-basal
KJE + GroELS
KJE + Lon
GroELS + Lon

B. 

C.
Figure 3.3 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 2 h at 30 °C under basal conditions and with the pairwise up-regulation of PN components

**A.** As in Fig. 3.2A, but with pairwise up-regulation of PN components. For KJE+GroELS, DnaK, DnaJ, and GrpE levels were increased by ~3-, 15-, and 10-fold, respectively while GroEL and GroES levels were increased by ~3- and 2-fold, respectively. For KJE+Lon, DnaK, DnaJ, and GrpE levels were increased by ~2-, 15-, and 5-fold, respectively, while Lon levels were increased by ~3-fold. For GroELS+Lon, GroEL and GroES levels were increased by ~3- and 8-fold, respectively, while Lon levels were increased by ~3-fold (Table B.1). (Modified from (Cho et al., 2015)).
A. Adapted-basal

B. 

C.
Figure 3.4 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 1 h at 30 °C under adapted-basal conditions and on conditions with overexpression of I54N σ^{32}

As in Fig. 3.1, but CRABP1(Y133S) was expressed for 1 h. For I54N σ^{32}, DnaK, DnaJ, and GrpE levels were increased by ~3-, 8-, and 1.1-fold, respectively; GroEL and GroES levels were increased by ~2- and 4-fold; Lon levels were increased by ~3-fold (Table B.3).
A. Adapted-basal

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>S</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>KJE up-regulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KJE-low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KJE-high</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GroELS up-regulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GroELS-low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GroELS-high</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lon up-regulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lon-low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lon-med</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lon-high</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

T: total S: soluble A: aggregated

B. Bar graph showing concentration (μM) of KJE, GroELS, and Lon in low, high, and adapted-basal conditions.

- KJE: low K: 2x, J: 9x, E: 1x; high K: 3x, J: 21x, E: 1.4x
- GroELS: low EL: 3x, ES: 3x; high EL: 8x, ES: 21x
- Lon: low Lon: 2x, med Lon: 3x, high Lon: 7x
Figure 3.5 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 1 h at 30 °C under adapted-basal conditions and with up-regulation of individual PN components

As in Fig. 3.2, but CRABP1(Y133S) was expressed for 1 h. Conditions have varying levels of KJE, GroELS, and Lon, and their fold changes compared to adapted-basal levels are shown.
A. Adapted-basal

- GroELS
- Lon

B. 

- Total
- Soluble
- Aggregated

C. 

- $f_{\text{soluble}}$
- $f_{\text{aggregated}}$

153
Figure 3.6 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 1 h at 30 °C under adapted-basal conditions and with the pairwise up-regulation of PN components

As in Fig. 3.3, but CRABP1(Y133S) was expressed for 1 h. Conditions have varying levels of KJE, GroELS, and Lon. For KJE+GroELS, DnaK, DnaJ, and GrpE levels were increased by ~2-, 15-, and 1.3-fold, respectively while GroEL and GroES levels were increased by ~6- and 8-fold, respectively. For KJE+Lon, DnaK, DnaJ, and GrpE levels were increased by ~2-, 11-, and 1.3-fold, respectively, while Lon levels were increased by ~2-fold. For GroELS+Lon, GroEL and GroES levels were increased by ~6- and 11-fold, respectively, while Lon levels were increased by ~2-fold (Table B. 3).
Figure 3.7 Dependence of CRABP1(Y133S) protein folding fates on different PN components based on phenomenological fits of *in vivo* folding fate data

A. Plot of the experimental values of \([\text{Agg}]_{\text{rel,CRABP1}}\) and \([\text{Sol}]_{\text{rel,CRABP1}}\) of CRABP1(Y133S) from Fig. 3.2 and Fig 3.3 vs. the corresponding model-derived values from the fits of Equations (1) or (2). Red data points- for \([\text{Agg}]_{\text{rel,CRABP1}}\), fit with Equation (1). Blue data points- for \([\text{Sol}]_{\text{rel,CRABP1}}\), fit with Equation (2). Dashed line- Line through the origin, slope of 1; the closer the points to the line, the better the fits.

B. Bar graph for the gradient parameters and SEM from the best-fits of Equations (1) (red bars) and (2) (blue bars) to the relative concentration. Positive values mean that increasing the concentration of a PN component increases the concentration of the aggregated (red bars) or soluble (blue bars) protein. Negative values indicate the opposite. The p-values are indicated as follows: *** for p < 0.0001; ** for p < 0.001; * for p < 0.01; and n.s. indicates p > 0.05. (Reproduced from (Cho et al., 2015)).
Figure 3.8 Results of FoldEco-derived fit of experimental (in vivo) folding fate data from the overexpression of CRABP1(Y133S) for 2 h

A. As in Fig. 3.7A, except that these are determined from the FoldEco fits to the data. The circled points are those for CRABP1 under the GroELS-low and GroELS-high conditions, which have large residuals (shown by the circled data points). (Modified from (Cho et al., 2015)).

B. Summary diagram (laid out as in Fig. 1.11) of folding fates of CRABP1 overexpressed for 2 h under adapted-basal conditions based on FoldEco simulations.

The concentrations for each state (native (N), unfolded (U), misfolded (M), and aggregated (A)) are written below the circles. The radii of the colored circles are proportional to the cube roots of the concentrations. Cube roots are used to show both the lowest and highest concentrations on one diagram. The concentrations of synthesized and degraded protein are represented as circles for “synthesis” and “degradation”, respectively. Blue text- qualitative descriptors for biophysical processes. Red numbers- percentages of misfolded states that involve in KJE recovery, aggregate, or engage the Lon degradation. Black italic numbers- percentages of degraded protein either from the unfolded or misfolded states. (Modified from (Cho et al., 2015)).
Grow E. coli K12DE3 cells (competent cells with PN components, transformed with CRABP1) at 30 °C.

\[ \text{OD}_{600} \approx 0.4 \]

Divide cultures, 10-ml each.

- 10-mL culture (1 h time point)
  - Induce PN components
  - 1 h
  - Induce CRABP1
  - 1 h

- 10-mL culture (2 h time point)
  - Induce PN components
  - 1 h
  - Induce CRABP1
  - 2 h

Harvest cells, lyse cells, quantitate total, soluble and aggregated protein

**Figure 3.9 Schematic for experimental methods**

The individual PN components were co-expressed with CRABP1 under orthogonal promoters so that they can be induced independently. First, KJE, GroELS, and Lon were induced with arabinose, and then after 1 h, CRABP1 was induced for 1 and 2 h.
CHAPTER 4

CONCLUSIONS

4.1 Summary

The overarching goal of my thesis was to determine how a protein folds inside the cell, and how the cellular environment maintains levels of functional proteins. Specifically, I addressed the question, “How does the proteostasis network (PN), consisting of molecular chaperones and degradation enzymes, modulate the folding fate of proteins with different intrinsic biophysical properties?” I utilized a holistic approach, in which I integrated computation and experiments, to understand the cooperation and/or competition among the PN components when they are present and operating simultaneously. I performed wet lab experiments in *E. coli* by expressing proteins with different biophysical properties and determining their folding fates *in vivo*. In addition, I used FoldEco and provided us insight into how the PN functions as a whole.

I have shown that a protein’s intrinsic biophysical properties and the cellular PN determine the folding fate of a protein inside the cell, through kinetic partitioning. The folding fate of a protein is under a thermo-kinetic limitation, which indicates that the fate depends on either the kinetics or thermodynamics, but (for the most part) not on both at the same time. Specifically, folding is limited by the Limiting Folding Parameter (LFP), misfolding by the Limiting Misfolding Parameter (LMP), and aggregation by the Limiting Aggregation Parameter (LAP). Different proteins behave according to their particular values of these properties.
Kinetic partitioning occurs between folding, aggregation, and degradation. In Chapter 2, we understood the influence of intrinsic biophysical properties on the fate of a protein under one set of conditions, that is, the adapted-basal PN. In Chapter 3, I have shown that the PN components, consisting of molecular chaperones and degradation enzymes, modulate the folding fate of a protein by working together to maintain proteostasis. I dissected the components of the PN to determine which chaperones and/or degradation enzymes are beneficial to proteins of low stabilities and high aggregation propensities. I was able to determine the individual and collective contributions of chaperones and degradation enzymes to proteostasis. Results show that PN components, particularly, KJE, GroELS, and Lon, work together to maintain proteostasis. These three systems form an efficacious triad for maintaining proteostasis: KJE is for “anti-misfolding”, GroELS is for “pro-folding”, and Lon is for “concentration control”. Furthermore, the contributions of KJE, GroELS, and Lon to proteostasis depend on the intrinsic biophysical properties of their substrates and on the concentrations of these PN components and substrates at any given time.

4.2 Significance of findings

The outcome of my research is a model for the interplay of a protein’s intrinsic biophysical properties and the cellular proteostasis network (PN), consisting of molecular chaperones and degradation enzymes, in determining a protein’s folding fate. This research extends our knowledge of how proteostasis is
maintained in the cell. This contribution is significant because it helps us design ways to modulate the PN, which may be therapeutically useful in the case of diseases such as Alzheimer’s disease, amyotrophic lateral sclerosis (ALS), Parkinson’s disease, and others. My work shows that one can enhance proteostasis by modulating a protein’s biophysical properties and/or regulating the proteostasis network. The findings also further advance the protein folding field. The use of integrated experimental and computational approach has enabled us to understand how a protein folds inside the cell, where proteins really fold. Using a holistic approach, I have translated a protein’s folding fate \textit{in vivo} into information about its intrinsic biophysical properties, rather than the usual reductionist approach in which protein folding is analyzed \textit{in vitro} to understand its behavior \textit{in vivo}.

4.3 Future directions

I have shown that a protein’s biophysical properties and the cellular PN determine folding fate. However, there are still unresolved questions that need to be addressed.

4.3.1 Biophysical properties

The CRABP1 variants that were used in this study had mutations that changed both the rate and equilibrium constants of the protein. While these have been useful in addressing our questions, it is not very clear how rate and
equilibrium constant contribute to the fate of the protein. We need a system that can only change either one of the parameters. The use of ligand binding can be used since ligands only bind to native states, and that only the effective equilibrium constant is changed, and not the folding rate constant, and other parameters. Furthermore, with FoldEco I have estimated effective parameters (LFP, LMP, and LAP) for the CRABP1 variants. We know the limiting parameter for folding, which is $k_f$, but we do not know yet the limiting parameters for misfolding and aggregation. Studies in vitro should be done to identify such parameters. How can we measure misfolding and aggregation parameters in vitro or in vivo? Are the effective parameters consistent with these experimental values?

Moreover, in the simulations, the degradation rate was set as constant. Based on in vivo experiments, more CRABP1(Y133S) proteins were degraded at a given time. Future research should be carried out to determine the degradation rate constant and equilibrium constant for Lon and CRABP1 measured in vitro and in vivo. These values will be input to FoldEco for simulations and test the effect of changing degradation parameters on the fate of a protein.

It would be interesting to assess kinetic partitioning between folding, aggregation, and degradation, to test the predictions from FoldEco. For example, in the case when proteins do not misfold and do not aggregate, folding and degradation only compete for unfolded states (U). How about the partitioning
between folding, aggregation, and degradation for proteins that misfold and aggregate?

### 4.3.2 Proteostasis network

Further work needs to be done to establish that PN fails due to saturation of some PN components. Our results indicate that GroEL and/or GroES are limiting when the concentrations of CRABP1(Y133S) are high. It would be interesting to perform titration experiments and determine the extent of the beneficial effect of GroELS on the folding fate of CRABP1(Y133S). Moreover, a future study could investigate the interaction between CRABP1(Y133S) and GroEL and/or GroES.

Future versions of FoldEco should be refined and expanded. As mentioned in the previous sections, the present form of FoldEco does not have direct mechanisms for the PN components to collaborate. The inclusion of direct collaborations among PN components should provide a better understanding about how the whole PN functions as a system. Moreover, although, our results suggest that HtpG and IbpA/B are not required for us to model folding fate under our conditions, it would be interesting to include them in FoldEco such that the whole *E. coli* PN will be well represented. Furthermore, FoldEco would also be improved with the addition of heat shock response, more sophisticated folding mechanisms, and when multiple proteins can be synthesized at a time.
APPENDIX A

SUPPLEMENTARY MATERIAL FOR INFLUENCE OF INTRINSIC BIOPHYSICAL PROPERTIES ON PROTEIN FOLDING FATE IN *E. COLI* UNDER THE BASAL PROTEOSTASIS NETWORK

Table A.1 Concentrations of components in an adapted-basal PN

The concentrations of PN components under the “adapted-basal” PN. The values given are the concentrations when they are active. Values taken from Cho et al. (2015).

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration (μM), active</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger factor, active as monomer</td>
<td>14</td>
</tr>
<tr>
<td>DnaK, active as monomer</td>
<td>23.4</td>
</tr>
<tr>
<td>DnaJ, active as dimer</td>
<td>1.56</td>
</tr>
<tr>
<td>GrpE, active as dimer</td>
<td>10.4</td>
</tr>
<tr>
<td>GroEL, active as 14-mer</td>
<td>2.23</td>
</tr>
<tr>
<td>GroES, active as 7-mer</td>
<td>3.53</td>
</tr>
<tr>
<td>ClpB, active as 6-mer</td>
<td>0.4</td>
</tr>
<tr>
<td>Lon, active as 6-mer</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Figure A.1 Fold changes of DnaK, GroEL, and Lon in cells expressing CRABP1 relative to cells not expressing CRABP1

The levels of DnaK, GroEL, and Lon in cells expressing CRABP1 (A. WT* B. M9A C. V67A D. L118V E. Y133S), at 1, 2, and 3 h of induction were ~10-30% lower or ~10-20% higher than those in cells not expressing CRABP1.
Figure A.2 Total protein concentration of CRABP1(WT*) and CRABP1(Y133S) upon expression in *E. coli* under -Lon and +Lon conditions

Cytoplasmic concentration for (A) CRABP1(WT*) and (B) CRABP1(Y133S) expressed under –Lon and +Lon *E. coli* cells at 30 °C for 0.5, 1, and 2 h. There are more proteins in cells without Lon compared to those with Lon in both proteins.
Figure A.3 Fraction of soluble protein when either $k_f$ or $K_f$ is at maximum value

At maximum $K_f$ ($10^8$), as $k_f$ increases from 0.001 s$^{-1}$ to 1000 s$^{-1}$, the fraction of soluble protein increases from 0.1 to 0.9. At maximum $k_f$ ($1000$ s$^{-1}$), as $K_f$ increases from $10^2$ to $10^8$, the fraction of soluble protein also increases.
Figure A4. Sample for protein concentration quantitation using standards

Increasing amounts of purified proteins (in picomoles) are loaded. Standard curves are used to calculate the absolute concentrations of test proteins. A. Coomassie stain of CRABP1. B. Standard curve. This same method is also used in quantitation of proteins detected by Western blotting.
Table A.2 Summary for the quantitation of protein folding fates of CRABP1 variants upon expression in *E. coli* under basal conditions

*T-total, S-soluble, A-aggregated, D-degraded; \( f_{\text{sol}} \)-fraction of soluble, \( f_{\text{agg}} \)-fraction of aggregated

*Total protein- This is the total existing protein (not total protein synthesized).
**Based on control experiments with CRABP1(WT*) expressed in \( \Delta \text{l}\text{o}n\) *E. coli* K12 cells, in which ~1/3 of the protein that is synthesized is degraded. The total amount of protein synthesized is ~405 \( \mu \text{M} \) in all cases.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Time (h)</th>
<th>Concentration (( \mu \text{M} ))</th>
<th>Relative concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T*</td>
<td>S</td>
</tr>
<tr>
<td>WT*</td>
<td>1</td>
<td>95±15</td>
<td>95±15</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>202±25</td>
<td>192±22</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>275±28</td>
<td>251±19</td>
</tr>
<tr>
<td>M9A</td>
<td>1</td>
<td>115±17</td>
<td>111±14</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>202±27</td>
<td>189±24</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>281±36</td>
<td>240±29</td>
</tr>
<tr>
<td>V67A</td>
<td>1</td>
<td>125±5</td>
<td>108±6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>189±11</td>
<td>159±11</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>294±32</td>
<td>240±27</td>
</tr>
<tr>
<td>L118V</td>
<td>1</td>
<td>94±11</td>
<td>82±5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>182±44</td>
<td>156±35</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>257±86</td>
<td>212±64</td>
</tr>
<tr>
<td>Y133S</td>
<td>1</td>
<td>67±5</td>
<td>22±1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>109±13</td>
<td>34±2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>125±25</td>
<td>35±7</td>
</tr>
</tbody>
</table>
Table A.3 Effective folding parameters for the different variants of CRABP1

The $k_f$ and $K_f$ values were derived from \textit{in vitro} experiments. The $k_m$, $K_m$, $k_a$, and $K_a$ values were based from FoldEco simulations.

<table>
<thead>
<tr>
<th>CRABP1 variant</th>
<th>$k_f$ (s$^{-1}$)</th>
<th>$K_f$</th>
<th>$k_m$ (s$^{-1}$)</th>
<th>$K_m$ (μM$^{-1}$ s$^{-1}$)</th>
<th>$k_a$ (μM$^{-1}$ s$^{-1}$)</th>
<th>$K_a$ (μM$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT*</td>
<td>0.77</td>
<td>2.17 x 10$^7$</td>
<td>0.077</td>
<td>0.3</td>
<td>0.3</td>
<td>1</td>
</tr>
<tr>
<td>M9A</td>
<td>0.31</td>
<td>1.04 x 10$^6$</td>
<td>3.1</td>
<td>10</td>
<td>250</td>
<td>1000</td>
</tr>
<tr>
<td>V67A</td>
<td>0.57</td>
<td>8.23 x 10$^4$</td>
<td>15</td>
<td>30</td>
<td>250</td>
<td>80</td>
</tr>
<tr>
<td>L118V</td>
<td>0.41</td>
<td>4.45 x 10$^5$</td>
<td>20</td>
<td>30</td>
<td>250</td>
<td>100</td>
</tr>
<tr>
<td>Y133S</td>
<td>0.080</td>
<td>6.53 x 10$^3$</td>
<td>8</td>
<td>30</td>
<td>100</td>
<td>80</td>
</tr>
</tbody>
</table>
Figure A.5 FoldEco simulation image for WT* at 1, 2, and 3 h

The arrows show fluxes of proteins

TF (trigger factor), KJE (DnaK/DnaJ/GrpE), ClpB+KJE (ClpB+DnaK/DnaJ/GrpE), GroELS (GroEL/GroES), Deg (degradation), N (native), U (unfolded), A (aggregate), M (misfolded)
Each circle is 1 μM.

On the image, the middle part shows where the fluxes of proteins from one state (or chaperone) to another are going. On the left is the time of simulation and % chaperone use. On the right is the total protein, % of ongoing degradation, and ATP use. On the bottom is the list of the concentrations of each species (either bound to chaperones or free) at a given time.
Figure A.6 FoldEco simulation image for M9A at 1, 2, and 3 h

The arrows show fluxes of proteins

TF (trigger factor), KJE (DnaK/DnaJ/GrpE), ClpB+KJE (ClpB+DnaK/DnaJ/GrpE), GroELS (GroEL/GroES), Deg (degradation), N (native), U (unfolded), A (aggregate), M (misfolded)

Each circle is 1 μM.

On the image, the middle part shows where the fluxes of proteins from one state (or chaperone) to another are going. On the left is the time of simulation and % chaperone use. On the right is the total protein, % of ongoing degradation, and ATP use. On the bottom is the list of the concentrations of each species (either bound to chaperones or free) at a given time.
1 h

Time: 3600 s

<table>
<thead>
<tr>
<th>Protein state</th>
<th>Total (µM)</th>
<th>Free (µM)</th>
<th>TF bound (µM)</th>
<th>In KJE cycle (µM)</th>
<th>In GroEL cycle (µM)</th>
<th>In B+KJE cycle (µM)</th>
<th>Protease bound (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>0.6529</td>
<td>0.0158</td>
<td>0.2376</td>
<td>0.3054</td>
<td>0.0338</td>
<td>N/A</td>
<td>0.0003</td>
</tr>
<tr>
<td>M</td>
<td>1.2675</td>
<td>0.0208</td>
<td>0.264</td>
<td>0.2824</td>
<td>0.7098</td>
<td>N/A</td>
<td>0.0004</td>
</tr>
<tr>
<td>N</td>
<td>76.1842</td>
<td>76.0414</td>
<td>N/A</td>
<td>N/A</td>
<td>0.1428</td>
<td>N/A</td>
<td>0.0006</td>
</tr>
<tr>
<td>Aggregates</td>
<td>17.9012</td>
<td>0.3151</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>17.5861</td>
</tr>
<tr>
<td>Total</td>
<td>96.0037</td>
<td>76.3931</td>
<td>0.4916</td>
<td>0.6478</td>
<td>0.8864</td>
<td>17.5861</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

2 h

Time: 7200 s

<table>
<thead>
<tr>
<th>Protein state</th>
<th>Total (µM)</th>
<th>Free (µM)</th>
<th>TF bound (µM)</th>
<th>In KJE cycle (µM)</th>
<th>In GroEL cycle (µM)</th>
<th>In B+KJE cycle (µM)</th>
<th>Protease bound (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>0.6313</td>
<td>0.0166</td>
<td>0.2474</td>
<td>0.3212</td>
<td>0.0349</td>
<td>N/A</td>
<td>0.0003</td>
</tr>
<tr>
<td>M</td>
<td>1.2886</td>
<td>0.0215</td>
<td>0.2619</td>
<td>0.2759</td>
<td>0.7298</td>
<td>N/A</td>
<td>0.0004</td>
</tr>
<tr>
<td>N</td>
<td>158.035</td>
<td>157.888</td>
<td>N/A</td>
<td>N/A</td>
<td>0.147</td>
<td>N/A</td>
<td>0.0006</td>
</tr>
<tr>
<td>Aggregates</td>
<td>31.224</td>
<td>0.5919</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>30.6622</td>
</tr>
<tr>
<td>Total</td>
<td>151.118</td>
<td>158.438</td>
<td>0.5093</td>
<td>0.008</td>
<td>0.9117</td>
<td>30.6622</td>
<td>0.0007</td>
</tr>
</tbody>
</table>
Figure A.7 FoldEco simulation image for V67A at 1, 2, and 3 h

The arrows show fluxes of proteins

TF (trigger factor), KJE (DnaK/DnaJ/GrpE), ClpB+KJE (ClpB+DnaK/DnaJ/GrpE), GroELS (GroEL/GroES), Deg (degradation), N (native), U (unfolded), A (aggregate), M (misfolded)
Each circle is 1 μM.

On the image, the middle part shows where the fluxes of proteins from one state (or chaperone) to another are going. On the left is the time of simulation and % chaperone use. On the right is the total protein, % of ongoing degradation, and ATP use. On the bottom is the list of the concentrations of each species (either bound to chaperones or free) at a given time.
1 h

Time: 3600 s

Chaperone use:

Protein 1

Microsome
U
M
A
N

Trigger factor

Dnak
DnaK
GrpE
GroES
GroEL
CcpA
LamH

Deg
N
M
TF
U

Color key

Synthesis

Total Protein: 98.367 μM
% degraded, protein 1: 8.55%
%

Amplicon

Amplicon

Growth

Degradation

with synthesis: 7.4 μM

Energy diagram, Protein 1

Protein state Total (μM) Free (μM) TF bound (μM) In KJE cycle (μM) In GroEL cycle (μM) In B+KJE cycle (μM) Protease bound (μM)

<table>
<thead>
<tr>
<th>Protein state</th>
<th>Total (μM)</th>
<th>Free (μM)</th>
<th>TF bound (μM)</th>
<th>In KJE cycle (μM)</th>
<th>In GroEL cycle (μM)</th>
<th>In B+KJE cycle (μM)</th>
<th>Protease bound (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>0.4858</td>
<td>0.0098</td>
<td>0.1654</td>
<td>0.2867</td>
<td>0.0237</td>
<td>N/A</td>
<td>0.0002</td>
</tr>
<tr>
<td>M</td>
<td>1.0172</td>
<td>0.0378</td>
<td>0.2134</td>
<td>0.2466</td>
<td>0.531</td>
<td>N/A</td>
<td>0.0003</td>
</tr>
<tr>
<td>N</td>
<td>85.5028</td>
<td>85.3382</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.1646</td>
<td>N/A</td>
</tr>
<tr>
<td>Aggregates</td>
<td>11.3617</td>
<td>0.1925</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>11.1691</td>
<td>N/A</td>
</tr>
<tr>
<td>Total</td>
<td>98.367</td>
<td>85.5358</td>
<td>0.3838</td>
<td>0.5363</td>
<td>0.7193</td>
<td>11.1691</td>
<td>0.0009</td>
</tr>
</tbody>
</table>

2 h

Time: 7200 s

Chaperone use:

Protein 1

Microsome
U
M
A
N

Trigger factor

Dnak
DnaK
GrpE
GroES
GroEL
CcpA
LamH

Deg
N
M
TF
U

Color key

Synthesis

Total Protein: 198.332 μM
% degraded, protein 1: 8.78%
%

Amplicon

Amplicon

Growth

Degradation

with synthesis: 7.6 μM

Energy diagram, Protein 1

Protein state Total (μM) Free (μM) TF bound (μM) In KJE cycle (μM) In GroEL cycle (μM) In B+KJE cycle (μM) Protease bound (μM)

<table>
<thead>
<tr>
<th>Protein state</th>
<th>Total (μM)</th>
<th>Free (μM)</th>
<th>TF bound (μM)</th>
<th>In KJE cycle (μM)</th>
<th>In GroEL cycle (μM)</th>
<th>In B+KJE cycle (μM)</th>
<th>Protease bound (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>0.4090</td>
<td>0.0102</td>
<td>0.1698</td>
<td>0.2051</td>
<td>0.0243</td>
<td>N/A</td>
<td>0.0002</td>
</tr>
<tr>
<td>M</td>
<td>1.3843</td>
<td>0.0383</td>
<td>0.2237</td>
<td>0.2403</td>
<td>0.5427</td>
<td>N/A</td>
<td>0.0003</td>
</tr>
<tr>
<td>N</td>
<td>176.608</td>
<td>176.44</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.1683</td>
<td>N/A</td>
</tr>
<tr>
<td>Aggregates</td>
<td>18.2204</td>
<td>0.3153</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>17.9951</td>
<td>N/A</td>
</tr>
<tr>
<td>Total</td>
<td>196.332</td>
<td>176.783</td>
<td>0.3935</td>
<td>0.5144</td>
<td>0.7353</td>
<td>17.9951</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
Figure A.8 FoldEco simulation image for L118V at 1, 2, and 3 h

The arrows show fluxes of proteins

TF (trigger factor), KJE (DnaK/DnaJ/GrpE), ClpB+KJE (ClpB+DnaK/DnaJ/GrpE), GroELS (GroEL/GroES), Deg (degradation), N (native), U (unfolded), A (aggregate), M (misfolded)

Each circle is 1 μM.

On the image, the middle part shows where the fluxes of proteins from one state (or chaperone) to another are going. On the left is the time of simulation and % chaperone use. On the right is the total protein, % of ongoing degradation, and ATP use. On the bottom is the list of the concentrations of each species (either bound to chaperones or free) at a given time.
Figure A.9 FoldEco simulation image for Y133S at 1, 2, and 3 h

The arrows show fluxes of proteins

TF (trigger factor), KJE (DnaK/DnaJ/GrpE), ClpB+KJE (ClpB+DnaK/DnaJ/GrpE), GroELS (GroEL/GroES), Deg (degradation), N (native), U (unfolded), A (aggregate), M (misfolded)

Each circle is 1 μM.

On the image, the middle part shows where the fluxes of proteins from one state (or chaperone) to another are going. On the left is the time of simulation and % chaperone use. On the right is the total protein, % of ongoing degradation, and ATP use. On the bottom is the list of the concentrations of each species (either bound to chaperones or free) at a given time.
Figure B.1 Schematic for the plasmids used for the up-regulation of PN components

Individual up-regulation of PN components were induced with L-arabinose. Up-regulation of combinations of PN components were induced with L-arabinose and tetracycline. (Reproduced from (Cho et al., 2015)).
Table B.1 Fold changes of the levels of KJE, GroELS, and Lon in cells overexpressing these proteins relative to cells under adapted-basal conditions, and when CRABP1 was expressed for 2 h

*Data from (Cho et al., 2015).

<table>
<thead>
<tr>
<th>Protein Combination</th>
<th>Fold Change (relative to adapted-basal)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>KJE-low</td>
<td>DnaK 2.1</td>
</tr>
<tr>
<td></td>
<td>DnaJ 10.1</td>
</tr>
<tr>
<td></td>
<td>GrpE 4.6</td>
</tr>
<tr>
<td>KJE-high</td>
<td>DnaK 3.7</td>
</tr>
<tr>
<td></td>
<td>DnaJ 23.4</td>
</tr>
<tr>
<td></td>
<td>GrpE 10.4</td>
</tr>
<tr>
<td>GroELS-low</td>
<td>GroEL 3.4</td>
</tr>
<tr>
<td></td>
<td>GroES 2.1</td>
</tr>
<tr>
<td>GroELS-high</td>
<td>GroEL 7.5</td>
</tr>
<tr>
<td></td>
<td>GroES 5.1</td>
</tr>
<tr>
<td>Lon-low</td>
<td>Lon 2.9</td>
</tr>
<tr>
<td>Lon-med</td>
<td>Lon 3.8</td>
</tr>
<tr>
<td>Lon-high</td>
<td>Lon 7.8</td>
</tr>
<tr>
<td>KJE+GroELS</td>
<td>DnaK 2.6</td>
</tr>
<tr>
<td></td>
<td>DnaJ 14.6</td>
</tr>
<tr>
<td></td>
<td>GrpE 10.2</td>
</tr>
<tr>
<td></td>
<td>GroEL 2.9</td>
</tr>
<tr>
<td></td>
<td>GroES 2.2</td>
</tr>
<tr>
<td>KJE+Lon</td>
<td>DnaK 2.0</td>
</tr>
<tr>
<td></td>
<td>DnaJ 15.4</td>
</tr>
<tr>
<td></td>
<td>GrpE 4.8</td>
</tr>
<tr>
<td></td>
<td>Lon 2.8</td>
</tr>
<tr>
<td>GroELS+Lon</td>
<td>GroEL 3.4</td>
</tr>
<tr>
<td></td>
<td>GroES 8.2</td>
</tr>
<tr>
<td></td>
<td>Lon 2.7</td>
</tr>
<tr>
<td>I54N σ²²</td>
<td>DnaK 3.1</td>
</tr>
<tr>
<td></td>
<td>DnaJ 4.9</td>
</tr>
<tr>
<td></td>
<td>GrpE 6.5</td>
</tr>
<tr>
<td></td>
<td>GroEL 3.0</td>
</tr>
<tr>
<td></td>
<td>GroES 4.4</td>
</tr>
<tr>
<td></td>
<td>Lon 2.6</td>
</tr>
</tbody>
</table>
Figure B.2 Fold changes of the levels of KJE, GroELS, and Lon in cells overexpressing these proteins relative to cells under adapted-basal conditions

Bar graph showing the fold changes (relative to adapted-basal) of the levels of PN components when PN components (KJE, GroELS, and Lon) were up-regulated A. individually and B. simultaneously, 1 h prior to CRABP1(Y133S) induction for 2 h. DnaK, DnaJ, GrpE, GroEL, GroES, and Lon were up-regulated to different extents.

*Data from (Cho et al., 2015) (Reproduced from (Cho et al., 2015)).
Figure B.3 Folding fates of CRABP1(Y133S) upon expression in *E. coli* for 2 h at 30 °C under adapted-basal conditions and with up-regulation of other PN components

As in Fig. 3.3, but for other PN components. (Reproduced from (Cho et al., 2015)).
Table B.2 Data on the partitioning of CRABP1(Y133S) when induced at 30 °C for 1 h under different PN conditions

The concentrations of total, soluble and aggregated proteins are shown under different conditions. The concentration of total protein was derived from soluble + aggregated. The soluble and aggregated proteins were from the supernatant and pellet, respectively, after centrifugation of cell lysates for 10 min at 13,500 x g. The fraction of soluble ("f_{sol}") and fraction of aggregated ("f_{agg}") were also shown.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Concentration (µM)</th>
<th>Relative concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Soluble</td>
</tr>
<tr>
<td>Adapted-basal</td>
<td>57 ± 4</td>
<td>17 ± 3</td>
</tr>
<tr>
<td>KJE-low</td>
<td>42 ± 10</td>
<td>16 ± 5</td>
</tr>
<tr>
<td>KJE-high</td>
<td>35 ± 10</td>
<td>17 ± 5</td>
</tr>
<tr>
<td>GroELS-low</td>
<td>53 ± 4</td>
<td>26 ± 2</td>
</tr>
<tr>
<td>GroELS-high</td>
<td>45 ± 9</td>
<td>35 ± 4</td>
</tr>
<tr>
<td>Lon-low</td>
<td>25 ± 2</td>
<td>22 ± 1</td>
</tr>
<tr>
<td>Lon-med</td>
<td>12 ± 1</td>
<td>11 ± 1</td>
</tr>
<tr>
<td>Lon-high</td>
<td>0.6 ± 0.2</td>
<td>0.6 ± 0.2</td>
</tr>
<tr>
<td>KJE+GroELS</td>
<td>41 ± 20</td>
<td>37 ± 17</td>
</tr>
<tr>
<td>KJE+Lon</td>
<td>43 ± 11</td>
<td>35 ± 7</td>
</tr>
<tr>
<td>GroELS+Lon</td>
<td>54 ± 3</td>
<td>50 ± 2</td>
</tr>
<tr>
<td>I54N σ^{2+}</td>
<td>3 ± 0</td>
<td>3 ± 0</td>
</tr>
</tbody>
</table>
Table B.3 Fold changes of the levels of KJE, GroELS, and Lon in cells overexpressing these proteins relative to cells under adapted-basal conditions, and when CRABP1 was expressed for 1 h

*Data from (Cho et al., 2015).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Protein</th>
<th>Fold Change (relative to adapted-basal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KJE-low</td>
<td>DnaK</td>
<td>1.8 ± 0.2</td>
</tr>
<tr>
<td></td>
<td>DnaJ</td>
<td>8.9 ± 1.6</td>
</tr>
<tr>
<td></td>
<td>GrpE</td>
<td>0.82 ± 0.08</td>
</tr>
<tr>
<td>KJE-high</td>
<td>DnaK</td>
<td>3.0 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>DnaJ</td>
<td>21 ± 6</td>
</tr>
<tr>
<td></td>
<td>GrpE</td>
<td>1.4 ± 0.1</td>
</tr>
<tr>
<td>GroELS-low</td>
<td>GroEL</td>
<td>3.3 ± 0.2</td>
</tr>
<tr>
<td></td>
<td>GroES</td>
<td>3.3</td>
</tr>
<tr>
<td>GroELS-high</td>
<td>GroEL</td>
<td>8.0 ± 0.8</td>
</tr>
<tr>
<td></td>
<td>GroES</td>
<td>12</td>
</tr>
<tr>
<td>Lon-low</td>
<td>Lon</td>
<td>2.1 ± 0.3</td>
</tr>
<tr>
<td>Lon-med</td>
<td>Lon</td>
<td>2.9 ± 0.6</td>
</tr>
<tr>
<td>Lon-high</td>
<td>Lon</td>
<td>7.4 ± 1.7</td>
</tr>
<tr>
<td>KJE+GroELS</td>
<td>DnaK</td>
<td>2.2 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>DnaJ</td>
<td>15 ± 9</td>
</tr>
<tr>
<td></td>
<td>GrpE</td>
<td>1.3 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>GroEL</td>
<td>6.0 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>GroES</td>
<td>8.5</td>
</tr>
<tr>
<td>KJE+Lon</td>
<td>DnaK</td>
<td>1.6 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>DnaJ</td>
<td>11 ± 7</td>
</tr>
<tr>
<td></td>
<td>GrpE</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td></td>
<td>Lon</td>
<td>2.4 ± 0.4</td>
</tr>
<tr>
<td>GroELS+Lon</td>
<td>GroEL</td>
<td>6.2 ± 0.0</td>
</tr>
<tr>
<td></td>
<td>GroES</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Lon</td>
<td>2.2 ± 0.3</td>
</tr>
<tr>
<td>I54N σ^{32}</td>
<td>DnaK</td>
<td>2.9 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>DnaJ</td>
<td>8.1 ± 4.2</td>
</tr>
<tr>
<td></td>
<td>GrpE</td>
<td>1.1 ± 0.0</td>
</tr>
<tr>
<td></td>
<td>GroEL</td>
<td>2.5 ± 0.0</td>
</tr>
<tr>
<td></td>
<td>GroES</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>Lon</td>
<td>3.4 ± 0.8</td>
</tr>
</tbody>
</table>
Figure B.4 Fold changes of the levels of KJE, GroELS, and Lon in cells overexpressing these proteins relative to cells under adapted-basal conditions

A. Table showing the fold changes (relative to adapted-basal) of the levels of PN components when PN components (KJE, GroELS, and Lon) were upregulated individually and simultaneously 1 h prior to CRABP1(Y133S) induction for 1 h. DnaK, DnaJ, GrpE, GroEL, GroES, and Lon were upregulated to different extents. B, C. Bar graph of the table in (A).
Table B.4 Effective biophysical parameters for CRABP1(Y133S)

The best fit values and the 90% confidence intervals are shown.

(Data from (Cho et al., 2015)).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>CRABP1(Y133S)</th>
<th>90% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best</td>
<td>Lower</td>
</tr>
<tr>
<td>$k_f$ (s$^{-1}$)</td>
<td>0.03</td>
<td>0.0005</td>
</tr>
<tr>
<td>$K_f$</td>
<td>N/A</td>
<td>400</td>
</tr>
<tr>
<td>$k_m$ (s$^{-1}$)</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>$K_m$</td>
<td>13</td>
<td>0.1</td>
</tr>
<tr>
<td>$k_a$ (μM$^{-1}$ s$^{-1}$)</td>
<td>250</td>
<td>8</td>
</tr>
<tr>
<td>$K_a$ (μM$^{-1}$)</td>
<td>80</td>
<td>30</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


