5-2013

The Unavoidable Threat of Aggregation: Implications for Folding and Function of a β-Rich Protein

Mylene Hazelle Anne Ferrolino
University of Massachusetts Amherst

THE UNAVOIDABLE THREAT OF AGGREGATION: IMPLICATIONS FOR FOLDING AND FUNCTION OF A β-RICH PROTEIN

A Dissertation Presented

by

MYLENE CASTELL FERROLINO

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

DOCTOR OF PHILOSPHY

May 2013

Molecular and Cellular Biology

THE UNAVOIDABLE THREAT OF AGGREGATION:
IMPLICATIONS FOR FOLDING AND FUNCTION OF A β-RICH PROTEIN

A Dissertation Presented

by

MYLENE CASTELL FERROLINO

Approved as to style and content by:

_____
Lila M. Gierasch, Chair

_____
Daniel N. Hebert, Member

_____
Richard W. Vachet, Member

_____
Peter Chien, Member

_____
Barbara A. Osborne, Director
Molecular and Cellular Biology Program

DEDICATION

I'd like to dedicate this thesis to my inspiration – my children, JAMES and JULIANNE. Always remember that your dreams have no bounds. Live your dreams with passion and you will surely accomplish them. Mommy loves you.

ACKNOWLEDGEMENTS

I am extremely grateful to the people who have been instrumental in accomplishing this dissertation. I'd like to thank my adviser, *Dr. Lila Gierasch* for mentoring me to become a good scientist while keeping my sanity. Thank you for never doubting my abilities and for generously sharing your secrets to success. Thank you for allowing me to have my own little niche - to explore and learn new things everyday over the past five years. It was indeed a great pleasure to have you as my mentor.

I'd like to thank my committee members, *Dr. Daniel Hebert, Dr. Peter Chien* and *Dr. Richard Vachet.* Thank you for helping me direct my study to the right path

I'd also like to thank my fellow Gieraschians: *Dr. Anastasia Zhuravleva* who has been there in all my critical NMR experiments. Thank you for generously sharing your expertise and for always having answers to all my questions. None would have been possible without you. *Dr. Ivan Budyak* and *Dr. Beena Krishnan* who I have worked with side by side. Thank you for allowing me to compare notes with you and for helping me develop my ideas. It was a pleasure to have worked with wonderful and hardworking scientists like you. *Dr. Zoya Ignatova* for starting out the aggregation project. Although I never got to meet you, your work was a perfect precedent for my thesis. *Dr. Eugenia Clerico* for being the first person in the lab who gave me a warm smile while answering my inane questions. You are a jewel in the lab. *Kristine Pobre* for being my constant companion and friend. It would have been such a bore not to have you in the lab all these years – (Salamat!). *Amanda Clouser* who has been a wonderful person to work with. I was so lucky to mentor someone so smart who never really needed me! *Ilene Magpiong* for all the mommy stories we share. *Karan Hingorani* and *Mandy Blackburn,*

you just don't know how it means have those laughs in my most stressful days. *Dr. Anne Gershenson,* for constantly sharing new ideas on protein aggregation. Other lab members who have been very considerate during my critical last few days in the lab: *Dr. Abhay Thakur, Dr. Weiwei Kuo, Dr. Ben Yang*. I'd like to especially thank *Ellen Kalt* for being the person to I run to with so many things!

I'd also like to thank my friends who have been my family's social support since we arrived in the US. My dear friend, *Cristina Martin*, who has always helped me in so many ways, even way back in the Philippines. *Dr. Abigail Guce, Dr. Vanessa Gill, Dr. Annie Marcelino, Dr. Risha Molato, Dr. Princess Wilson, Dr. Bernard Matute, Drs. Ron and Phey Lerum, Dr. Andrea Gomez-Escudero, Kevin* and *Dolly Dagbay, Nilima Kolli, Bay and Mahalia Serrano,* for making us feel that there is home here in Amherst. Most especially, I'd like to thank *Drs. Meme and Albert Orquiola* and family who have been our foster family in the U.S.

I'd like to also thank my family in the Philippines: My dear parents, *Mario* and *Vikki Castell,* who have been my pillar all my life. I attribute all my accomplishments to you. My brothers and my sisters-in-law, *Gerald, Cindy, Marvin* and *Angel* thanks for believing in me although we are all apart. My in-laws, *Drs. Brendan* and *Joseline Ferrolino*, *Jong, Joey* and *Brian* - although Mommy Josie is no longer with us, I am sure she is happy to see me also get a doctoral degree.

Finally, I'd like to thank my loving husband *Dr. Jose Amadeo Ferrolino* for being my source of strength throughout my graduate studies. Thank you for sharing with my frustrations and triumphs. All these would not have been possible if not for our Divine Being who has clearly paved the way for me to achieve my dreams.

ABSTRACT

THE UNAVOIDABLE THREAT OF AGGREGATION:

IMPLICATIONS ON FOLDING AND FUNCTION OF A β-RICH PROTEIN

MAY 2013

MYLENE CASTELL FERROLINO

B.S. UNIVERSITY OF THE PHILIPPINES

M.S., UNIVERSITY OF THE PHILIPPINES

PH.D., UNIVERSITY OF THE MASSACHUSETTS

Directed by: Professor Lila M. Gierasch

Protein aggregation has been implicated in several catastrophic diseases (neurodegeneration, diabetes, ALS) and its complexity has also become a major obstacle in large-scale production of protein-based therapeutics. Despite the generic behavior of proteins to aggregate, only a few globular proteins have known aggregation mechanisms. At present, there have been no clear connections between protein folding, function and aggregation. We have tackled the challenge of understanding the links between a protein's natural tendency to fold and function with its propensity to misfold and aggregate. Using a predominantly β-sheet protein whose *in vitro* folding has been explored in detail: cellular retinoic acid-binding protein I (CRABP 1), as a model, we investigated sequence determinants for folding and aggregation. In addition, we characterized the aggregation-prone intermediate under native conditions. Our studies revealed similar contiguous aggregation cores in *in vitro* and *in vivo* aggregates of CRABP 1 validating the importance of sequence information under extremely different conditions. Hydrophobic stretches that comprise the interface in aggregates include

residues surrounding the ligand binding portal and residues at the C-terminal strands of CRABP 1. Folding studies reveal that docking of the N and C terminals happen in the early stages of barrel closure of CRABP 1 emphasizing the role of folding in preventing exposure of risky aggregation-prone sequences. We further examined the intermediate that initiates aggregation under native conditions. We found that inherent structural fluctuations in the native protein, relevant to ligand binding of CRABP 1, expose aggregation-prone sequences. Binding of the ligand, retinoic acid decreases the aggregation of CRABP 1 illustrating the contribution of functional interactions in avoiding aggregation. Our study implies that because of the evolutionary requirement for proteins to fold and function, aggregation becomes an unavoidable risk.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1. Competitive Interactions in Protein Folding and Aggregation

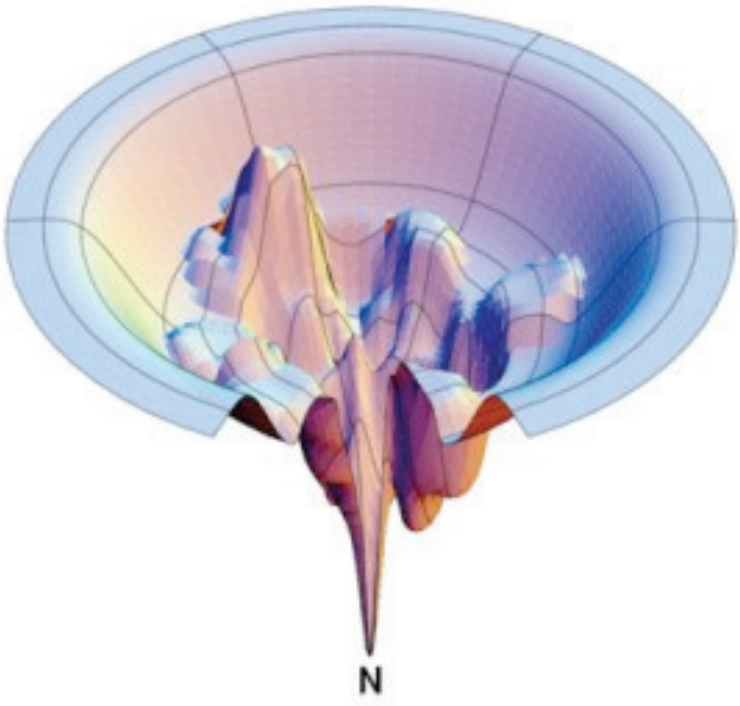
Proteins have encoded in their amino acid sequence highly feasible interactions that direct folding into their lowest energy three-dimensional functional state. The intrinsic property of proteins to achieve a shape that is thermodynamically stable has been originally demonstrated in Christian B. Anfinsen's experiment where the unfolded protein spontaneously re-acquires its three-dimensional structure upon removal of denaturant (Anfinsen, 1973). At infinite dilution, intramolecular contacts leading to natively folded proteins are highly favored. Nonetheless, above infinite dilution, folding competes with stable intermolecular contacts leading to oligomers and aggregates.

Proteins are able to fold efficiently in a biologically suitable timescale despite the enormous conformational space that a protein can sample under a given condition (Bartlett & Radford, 2009). The paradox between the time required for random unbiased conformational search to achieve the native state and the folding time of proteins can be explained by the existence of well-defined pathways (Karplus, 1997, Bartlett & Radford, 2009). Dill and Chan have described the protein folding framework using folding not only as two dimensional pathways but as multi-dimensional funnels and free energy landscapes (Figure 1) (Dill & Chan, 1997). Folding free energy landscapes describe the narrowing of conformational space of the unfolded state in a downward direction to the lowest free energy native state (Dill & Chan, 1997). Robust folders containing naturally selected amino acid sequences are viewed as smooth funnels (Dill & Chan, 1997). On the other hand, proteins with more complex folding pathways are represented with

rugged energy landscapes containing kinetically trapped intermediates. The hierarchy of local minima represents bottleneck conformations of either on-pathway or off-pathway intermediates (Dill & Chan, 1997, Milanesi*, et al.*, 2012). The ability to surmount the energy barriers associated with kinetic traps dictates the rate of folding of a protein.

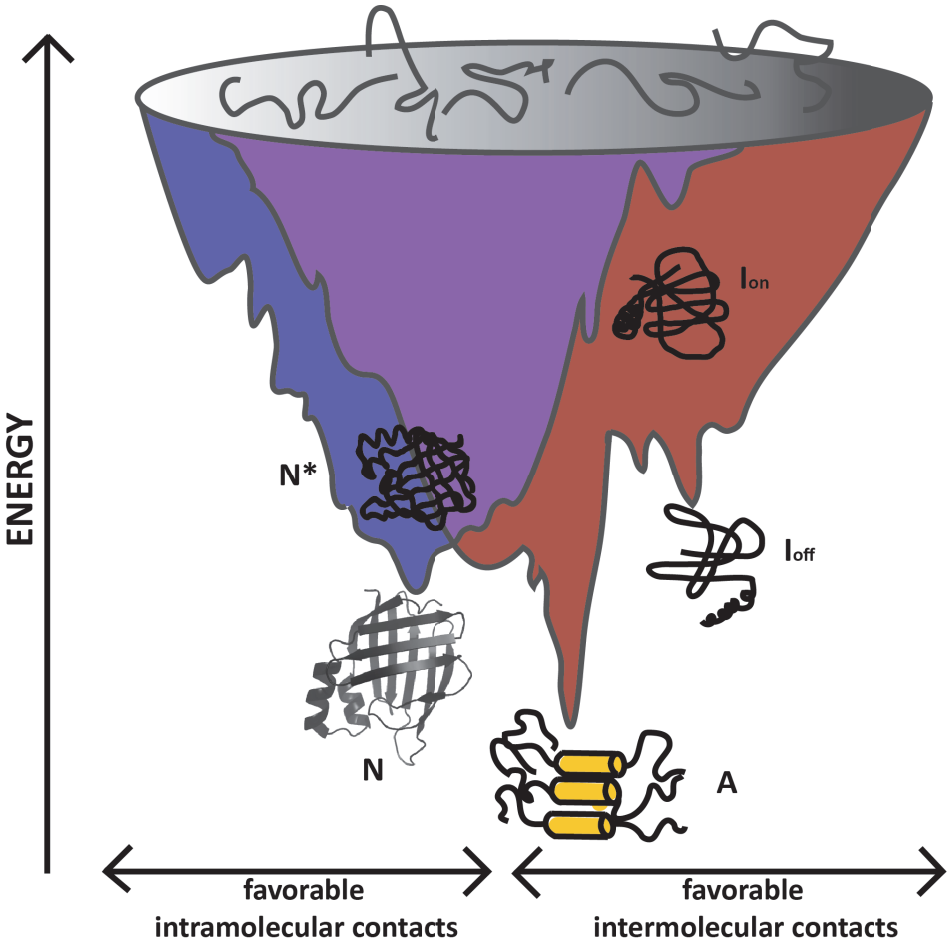
Energy landscapes are not only used to describe protein folding but protein motions as well. Although the native state is illustrated as the single lowest energy minima in a folding landscape, a closer examination reveals multiple low-barrier energy wells representing populations in the native state ensemble (Frauenfelder*, et al.*, 1991, Frauenfelder*, et al.*, 2007). These conformations are thermodynamically distinct but structurally similar (Chiti & Dobson, 2009). Structural fluctuations have been associated with biological functions of proteins and thus are very important aspects in the study of proteins. “Breathing” motions of proteins have recently been linked to aggregation. In Chapter 3 of this thesis, we will describe protein dynamics in more detail including its coupled contributions to function and aggregation.

Figure 1.1. Example of free energy landscape for folding of proteins. Folding of proteins can be illustrated from physical point of view using energy funnels. Multi-state folders are represented with rugged energy landscapes containing several energy minima portraying kinetically trapped intermediates. Energy barriers govern the rate of protein folding. This figure is re-printed from Ken Dill's website: http://dillgroup.stonybrook.edu/energy-landscapes.


N

The primary sequence of a protein defines its energy landscape and its path to the native state. However, the amino acid sequence also encodes the protein's ability to form intermolecular contacts required for functional oligomerization or dysfunctional aggregation. The burial of hydrophobic segments in a native protein and in surfaces stabilizing protein aggregates suggests common physico-chemical constraints governing both folding and aggregation (Linding*, et al.*, 2004, Routledge*, et al.*, 2009). Hence, an integrated energy landscape for folding and aggregation is necessary to predict the fate of a polypeptide chain. Under finite dilutions, intramolecular associations leading to stably folded states compete with intermolecular contacts that lead to functional or non-functional oligomers or highly stable insoluble aggregates. This can be envisioned as overlapping energy landscapes for folding and aggregation (Figure 1. 2) (Clark, 2004, Hartl & Hayer-Hartl, 2009). Energy wells in the folding landscape represent intermediates where destabilizing conditions can overcome an energy barrier and access the aggregation landscape (Clark, 2004, Hartl & Hayer-Hartl, 2009). As a consequence the equilibrium is pulled into the extremely stable aggregated state. Energy landscapes suggest correlation between thermodynamic stabilities and aggregation propensities and their dependence on the kinetic barriers that separate the states (Agostini*, et al.*, 2012). Thus, changes in the folding energy landscape from destabilizing mutations or alterations in solution conditions (pH, temperature, ionic strength, pressure) may increase the aggregation propensity (Linding*, et al.*, 2004). There is however growing evidence, suggesting aggregation of proteins may occur under native conditions which do not require significant destabilization of proteins (Chiti & Dobson, 2009). Structural fluctuations result in an ensemble of native-like species described in multiple low energy barriers. Low populations of slightly higher energy states can derail folding and lead to aggregation (Neudecker*, et al.*, 2012).

Figure 1.2. Schematic integrated folding and aggregation energy landscape of proteins. Similar physico-chemical properties that govern intramolecular and intermolecular contacts result in an overlap (purple) in energy landscapes for folding (blue) and aggregation (red). Wells on energy landscapes represent kinetically trapped intermediates - on-pathway (I_{on}), off-pathway (I_{off}) or near native (N*) intermediates, that can lead either to natively folded proteins or aggregates.


ENERGY
N*
Ion
Ioff
N
A
favorable
intramolecular contacts
favorable
intermolecular contacts

1.2. Characteristics of Protein Aggregates

Aggregation is a concentration-dependent self-assembly of misfolded proteins yielding either highly ordered or amorphous insoluble deposits (Sugiyama*, et al.*, 2010). Protein aggregation typically results in loss-of-function or a gain-of-function resulting to cellular toxicity. For decades, huge interest in understanding protein aggregation has been attributed to its direct implications in serious diseases. Cellular protein aggregates are highly abundant in several neurodegenerative diseases (e.g. Alzheimer's diseases, Parkinson's disease and Huntington's disease), prion-related diseases (Creutzfeldt-Jacob disease and kuru), type 2 diabetes, hemodialysis related diseases, inherited cataracts and amyloidosis (Chiti & Dobson, 2006). From a cellular point of view, toxicity derived from aggregation is linked to obstruction of the protein quality control machineries, which include molecular chaperones and proteases, depletion of essential proteins and build-up of toxic polypeptides (Sanchez de Groot*, et al.*, 2012). However, aggregation is not only associated with pathological conditions but can be beneficial as well. Under highly regulated conditions several organisms are able to exploit aggregation for function. Functional aggregates are exemplified in the case of an *E. coli* protein, curlin, found to be essential in colonizing inert surfaces, forming biofilms and mediating host protein binding (Chapman*, et al.*, 2002). Functional aggregates are also found in higher organisms such as aggregates formed by the protein Pmel17 involved in melanin production (Berson*, et al.*, 2003). Peptide and protein hormones are also stored in cells in the form of protein aggregates (Maji*, et al.*, 2009). In addition, aggregates also function as inheritable non-chromosomal genetic elements which is exhibited by prion protein aggregates (Wickner*, et al.*, 2013).

The most characterized forms of protein aggregates are the highly structured amyloid fibrils. Amyloids consist of repeated β-strands that run perpendicular to the fiber axis assembling into stacked cross-β sheets (Tycko, 2006). These very stable amyloid "cores" are highly solvent protected and protease resistant (Hoshino*, et al.*, 2002, Frare*, et al.*, 2006). Interestingly these stable aggregation cores are only comprised of very short (5-15 residues) hydrophobic and low charged segments (Chiti*, et al.*, 2003, Fernandez-Escamilla*, et al.*, 2004, Beerten*, et al.*, 2013). Aggregation cores are typically made up of amino acids with aliphatic side chains (valine, leucine and isoleucine) and aromatic side chains (Sanchez de Groot*, et al.*, 2012). Amyloid fibrils contain rope-like structures of twisted protofilaments (Sunde*, et al.*, 1997, Serpell*, et al.*, 2000). The regular cross-β cores of amyloids specifically bind to congo red and thioflavin T (LeVine, 1999, Westermark*, et al.*, 1999, Biancalana & Koide, 2010). Amyloids are formed via an initial nucleation lag phase involving the population of a rare misfolded intermediate, followed by a rapid exponential growth phase, where monomers or oligomers associate with the nucleus (Chiti & Dobson, 2006).

In contrast to amyloids, the mechanism of amorphous aggregation is less explored. Amorphous aggregates have been of significant interest in the microbiology and biotechnology fields as these are commonly found in bacterial inclusion bodies during over-expression of heterologous proteins (de Groot*, et al.*, 2008). Despite the non-fibrillar, amorphous appearance, several studies revealed presence of ordered amyloid-like core structures in bacterial inclusion body deposits (Carrio*, et al.*, 2005, Wang*, et al.*, 2008, de Groot*, et al.*, 2009). Interestingly, globular proteins in these bacterial aggregates are still able to retain their functional native-like conformations (Garcia-Fruitos*, et al.*, 2005, Ventura & Villaverde, 2006). Recently, amorphous aggregates have also been implicated in diseases such as in aggregated lenticular αB-

crystallins in cataracts (Sugiyama*, et al.*, 2010). Furthermore, prefibrillar aggregates that are currently argued to cause toxicity in amyloid diseases appeared amorphous (Dobson, 2003, Meredith, 2005). The kinetics of amorphous aggregation is proposed to be similar to a glass transition (Yoshimura*, et al.*, 2012). Glass-like properties in amorphous aggregates observed swap domains and restructure into stable species during intense salting-out (Yoshimura*, et al.*, 2012). This suggested the possible role of amorphous aggregation as a preliminary step towards amyloid formation (Yoshimura*, et al.*, 2012). It has been shown that under defined conditions proteins either form amyloids and amorphous aggregates or both (Morshedi*, et al.*, 2010). Although the morphologies of aggregates may vary within the same or across different proteins, stretches engaging in stabilization of aggregates have similar properties. Experimentally determined properties of aggregation prone segments can be used to train algorithms for forecasting aggregation tendencies. Thus, sequence-based predictions for aggregation have been useful tools in screening aggregation-prone segments and evaluating aggregation propensities for a large number of proteins.

1.3. Prediction of Aggregation-Prone Sequences Proteins

A protein's thermodynamic stability, structural class and amino acid sequence of proteins all govern aggregation propensities. (Conchillo-Sole*, et al.*, 2007, Niwa*, et al.*, 2009, Tartaglia & Vendruscolo, 2010, Agostini*, et al.*, 2012). However, there is not a huge amount of information available on stabilities of proteins. We also do not know much about the aggregation propensities of different structural classes. Hence, correlating aggregation with amino acid sequences is the only convenient way to provide initial clues on the aggregation tendencies of proteins. Furthermore, we can apply these to find relationships for natural selection and to optimize conditions in controlling self-

assembly. Several aggregation prediction algorithms have been developed over the years that exploit the plethora of studies on experimentally identified residues of proteins participating in aggregation. These algorithms aimed to distinguish short specific amino acid stretches engaging in aggregates and predict overall aggregation propensities of proteins. At present these have been used to forecast aggregation propensities for an enormous number of proteins both in bacterial and eukaryotic proteomes (de Groot & Ventura, 2010, Tartaglia & Vendruscolo, 2010, Grana-Montes*, et al.*, 2012). Here we will briefly describe experimentally validated aggregation predictors namely, AGGRESCAN, PASTA, TANGO and Zyggregator that we used in the study.

AGGRESCAN predicts “aggregation hotspots” based on hydrophobicity, β-sheet propensity and charge of amino acid side chains (de Groot*, et al.*, 2006, Conchillo-Sole*, et al.*, 2007). Aggregation propensities are calculated from the analysis of single residue substitutions in an aggregation-prone sequence (the hydrophobic cluster of Aβ) in the context of the full-length polypeptide (GFP fusion acting as an aggregation reporter) and not as an isolated short peptide (Conchillo-Sole*, et al.*, 2007). The calculations are based on previously determined aggregation-propensity values per amino acid (a^3v) averaged over a sliding window of a given length (a^4v) (Conchillo-Sole*, et al.*, 2007). A region in the polypeptide sequence is considered an aggregation "hot spot" (HS) if there are five or more contiguous residues (none is a proline) with an a^4v larger than the HST ("hot spot" threshold defined as the average of the a^3v of the 20 amino acids weighted by their frequencies in the SwissProt)(Conchillo-Sole*, et al.*, 2007). AGGRESCAN has been used in analyzing the aggregation propensities of protein kinase sequences in human and other organisms to identify evolutionary pressures in this family of proteins (Grana-Montes*, et al.*, 2012). This method has also been exploited to find associations between aggregation propensity with length, conformation, location, function, and

abundance in cells (de Groot & Ventura, 2010). AGGRESCAN has been found to be consistent with other aggregation predictors (de Groot & Ventura, 2010).

TANGO predicts β-aggregation in peptides and proteins based on a phase-space covering major conformational states, namely the folded state, β-sheet, β-turn, α-helix and β-aggregate (Fernandez-Escamilla*, et al.*, 2004). Segments of a peptide can populate each of these states following the Boltzmann distribution. Prediction of aggregation segments by TANGO is simply calculating the partition function of the phase-space (Fernandez-Escamilla*, et al.*, 2004). TANGO has been previously used in studying structural properties and mutational effects linked to aggregation (Linding*, et al.*, 2004). TANGO has also been used analyze several proteomes to find evolutionary pressures that minimize aggregation (Rousseau*, et al.*, 2006, Chen & Dokholyan, 2008).

PASTA calculates aggregation propensities based on similar parameters such as β-propensity, hydrophobicity and charge of amino acid side chains. However PASTA can also predicts the registry of the intermolecular hydrogen bonds formed between amyloid sequence stretches (Trovato*, et al.*, 2006). This is useful in envisioning the orientation of β-strands in the aggregation cores with respect to each other.

Zyggregator, similar to all other aggregation prediction algorithms, takes into account the hydrophobicity and secondary structure propensities of the amino acid side chains to predict differential aggregation behaviors between polypeptides (Tartaglia & Vendruscolo, 2008). In addition to this, Zyggregator also considers the intrinsic aggregation propensity and the stability of a polypeptide to have a significant tendency to form intermolecular interactions (Tartaglia & Vendruscolo, 2008). Aggregation

propensities are reported as Z_{agg}. Zyggregator has been employed to identify key regions in proteins that affect aggregation rates (Routledge*, et al.*, 2009). Also, it has also been used to predict relationships of aggregation propensity with evolutionary constraints at a proteome level (Tartaglia & Vendruscolo, 2010).

1.4. Aggregation Mechanisms of Globular Proteins

Despite the generality of aggregation in proteins and the abundance of aggregation-prone proteins *in vivo*, only a handful of globular proteins have well-described aggregation mechanisms. Originally, it was recognized that aggregation requires population of alternative protein conformations characterized by partially or globally unfolded intermediates that transitions the native state over a high-energy barrier of unfolding (Kelly, 1996, Chiti & Dobson, 2009). At conditions that favor aggregation such as low pH, high temperature, high pressure, presence of co-solvents or destabilizing mutations, the free energy of native state is increased relative to the unfolded and intermediate states hence increasing the populations of these states (Chiti & Dobson, 2009). When there is none or very minimal change in the free energy of the transition state, then the native state will have increased kinetic accessibility to the unfolded and intermediate states (Chiti & Dobson, 2009). Thus, population of these high energy misfolded species promotes aggregation (Chiti & Dobson, 2009). Several proteins have demonstrated tendencies to aggregate from a disordered state for obvious reasons that aggregation-prone segments are exposed when unfolded. These have added interest in the study of intrinsically disordered proteins (IDPs) in aggregation.

Some examples of proteins with partially unfolded aggregation-prone intermediates identified include β2-microglobulin (β2m), lysozyme, transthyretin and

copper-zinc superoxide dismutase I (SOD I). β2-microglobulin, a 99 residue β-sandwich protein deposited in the joints of patients suffering from dialysis related amyloidosis was shown to aggregate in an acidic solvent (Platt & Radford, 2009). A 10-residue aromatic rich amino acid stretch which controls the rate of aggregation of β2m is found to be exposed in the a partially unfolded intermediate at low pH (Platt & Radford, 2009). Human wild-type lysozyme at pH 5 and its mutants form molten globule-like intermediates during aggregation (Booth*, et al.*, 1997). In case of transthyretin, aggregation accedes via a partially denatured monomer at conditions that mimic the pH of the lysosome (Colon & Kelly, 1992). Several mutations in copper-zinc SOD I that lead to amyotrophic lateral sclerosis significantly destabilize the β-barrel and dimerization interfaces leading to aggregation (Perry*, et al.*, 2010).

As mentioned earlier, there is a growing number of proteins demonstrating aggregation without largely changing their conformation and surpassing high energy barriers (Kelly, 1996, Kelly, 1998, Jahn*, et al.*, 2006, Chiti & Dobson, 2009, Neudecker*, et al.*, 2012). Conformations of aggregation-prone states can be thermodynamically distinct but almost structurally identical to the native state. These native-like or N* states are accessed from the native state through thermal fluctuations (Chiti & Dobson, 2009). N* states are at higher energy than the native proteins, but are only separated by a relatively low energy barrier (Chiti & Dobson, 2009). Interestingly, globular proteins under extremely different conditions (destabilizing or native conditions) contain the same driving hydrophobic sequences to produce similar types of aggregates (Calamai*, et al.*, 2005). This is consistent with proteins possessing rugged aggregation energy landscapes consisting of structurally diverse precursors.

The proteins mentioned earlier alternatively aggregate at near physiological conditions. A rarely populated native-like intermediate of β2-microglobulin containing a non-native *trans*-proline isomer is found to favor aggregation (Eakin*, et al.*, 2006, Jahn*, et al.*, 2006). Edge strands of β2m that normally protect β-sandwich proteins from self-association are perturbed in the aggregation precursor (Jahn*, et al.*, 2006). Inherent fluctuations associated with the native state also lead to aggregation of the mutant lysozyme proteins (Chiti & Dobson, 2009). The aggregation-prone intermediate is a partially unfolded state formed through a locally cooperative process and a conformation thermodynamically distinct from the native state, but still on the native side of the free energy barrier for unfolding (Chiti & Dobson, 2009). In the case of copper-zinc SOD I, a metal and non-metal bound mutant S134N populates a near native intermediate that permits transient soluble oligomers through an unfolded loop VII or electrostatic loop leading to aggregation (Banci*, et al.*, 2005). Transthyretin aggregates under native conditions when the monomer undergoes local unfolding of peripheral strands yielding exposed aggregation-prone β-strands (Olofsson*, et al.*, 2004). Proteins like insulin and acylphosphatase (AcP) from *Sulfolobus solfataricus* form native-like α-helical oligomeric precursors that later undergo global structural reorganization to form amyloids (Bouchard*, et al.*, 2000, Chiti & Dobson, 2009). The growing evidence suggesting that proteins apparently access aggregation-prone conformations that are near-native indicates that the folded state remains at risk of aggregation.

1.5. Protein Folding and Aggregation in the Cell

It is imperative to note that events observed *in vitro* do not fully recapitulate folding and aggregation in the complex environment of the cell. Thus, what we learn *in vitro* from a few model proteins is only seeing the tip of the iceberg. *In vivo* conditions pose several challenges in predicting the fate of a protein. In the cell, newly synthesized

polypeptides vectorially emerge from the ribosome permitting the N-terminal segments of the protein to encounter an extremely complex and crowded environment during folding that can exacerbate aggregation (Ellis & Minton, 2006, Gershenson & Gierasch, 2011, Zhang & Ignatova, 2011). In addition, the protein load in cells that needs to fold is relatively large. It is estimated that 10-20% of residues in proteomes from prokaryotic and eukaryotic organisms are aggregation-prone (Rousseau*, et al.*, 2006). Although macromolecular crowding is biased towards more compact states like the native fold, it also favors specific and nonspecific intermolecular associations (Gierasch & Gershenson, 2009). Inside the cell protein mobilities are also hindered, increasing the chances for productive and non-productive encounters. In addition, macromolecular crowding affects the viscosity of the protein environment thus affecting their folding rates and mechanism (Gershenson & Gierasch, 2011). Increased risk of aggregation caused by macromolecular crowding in the cell emphasizes the critical role of folding assistants or molecular chaperones and efficient degradation machineries (Ellis & Minton, 2006, Gershenson & Gierasch, 2011, Vendruscolo, 2012). This evolutionarily conserved network of chaperones and proteases provides immediate cellular counter measures to environmental stress in order to maintain protein homeostasis also known as "proteostasis" (Hartl*, et al.*, 2011, Powers & Balch, 2013). However, their roles in assisting de novo folding and refolding have also been recognized under non-stressed conditions (Richter*, et al.*, 2010). In fact, there are a remarkable number of proteins highly dependent on chaperones (Kerner*, et al.*, 2005, Niwa*, et al.*, 2009, Fujiwara*, et al.*, 2010, Calloni*, et al.*, 2012). Molecular chaperones remodel the energy landscape of proteins to favor productive folding over aggregation (Hartl*, et al.*, 2011). There are five major classes of molecular chaperones, which include: Hsp100s, Hsp90s, Hsp70s, Hsp60s (chaperonins), and small heat shock proteins. These proteins either recognize unfolded proteins to assist folding, bind native proteins to facilitate functions or disrupt

insoluble aggregates (Richter*, et al.*, 2010). Molecular chaperones either enclose unfolded proteins in a chamber to actively or passively assist folding as in the case of Hsp60 or protect hydrophobic stretches from exposure as in the case of Hsp 70 (Zhu*, et al.*, 1996, Hartl*, et al.*, 2011, Clare*, et al.*, 2012). Hsp 90 assists in the late stage in protein folding and is required for regulation of conformational states of proteins (Saibil, 2008). Hsp 100s are involved in disaggregation by pulling apart proteins stuck in together in aggregates (Mayer, 2010). Lastly small heat shock proteins bind and protect unfolded proteins from aggregation (Hilton*, et al.*, 2013). Molecular chaperones work simultaneously in an intricate network with proteases that degrade misfolded adds to the increasing complexity of competing interactions with the polypeptide chain (Powers*, et al.*, 2012). Thus, it remains an enormous challenge to understand protein folding in the cell.

1.6. Statement of Dissertation

Despite the vast amount of information that is available, there is a lack of a holistic view of the entwined folding and aggregation energy landscapes. Fundamental questions continue to confound us: What are the sequence features of a polypeptide that link folding and aggregation? Are mechanisms of aggregation observed *in vitro* maintained *in vivo*? In addition, there has been very scarce evidence linking aggregation propensity and evolutionary requirement to function. Hence we also ask: What can be learned from our model protein CRABP 1 which has utilized several mechanisms to avoid aggregation? How do inherent protein motions associated with function endanger proteins towards aggregation? How do functional interactions offset aggregation propensity? What are the cellular mechanisms that modulate aggregation without compromising functions?

In Chapter 2 of this thesis, sequence determinants that drive *in vitro* and *in vivo* aggregation of CRABP 1 were explored. Here, CRABP 1 will be presented as an excellent model to study aggregation revealing conserved residue stretches that drive aggregation *in vitro* and *in vivo*. Furthermore, our findings will illustrate how critical amino acid segments expose mechanistic linkages between folding and aggregation of CRABP 1. In Chapter 3, we will present extensive work done to characterize the molecular basis of aggregation of CRABP 1 under native conditions. This chapter will describe findings supporting intimate connections between a low-population intermediate closely resembling the functional conformation and the tendency to aggregate. We will also describe existing intrinsic and extrinsic protection mechanisms against protein aggregation. The final chapter will provide a summary of what we have learned about the aggregation mechanism of CRABP 1 under native conditions, its implications in the cellular context, and further discuss directions currently being undertaken to explore the entire aggregation energy landscape of CRABP 1.

CHAPTER 2

AGGREGATION SEQUENCE DETERMINANTS: IMPLICATIONS FOR FOLDING OF A β-RICH PROTEIN

This chapter describes results of collaborations with Anastasia Zhuravleva, Ivan Budyak, Beena Krishnan, Annie Marcelino and Lila Gierasch [Publication: Budyak I.L., Krishnan B., Marcelino-Cruz A.M., Ferrolino M.C., Zhuravleva A. & Gierasch L.M. (2013) Early Folding Events Protect Aggregation-Prone Regions of a β-Rich Protein. *Structure* **21**: 476-485.

2.1. Introduction

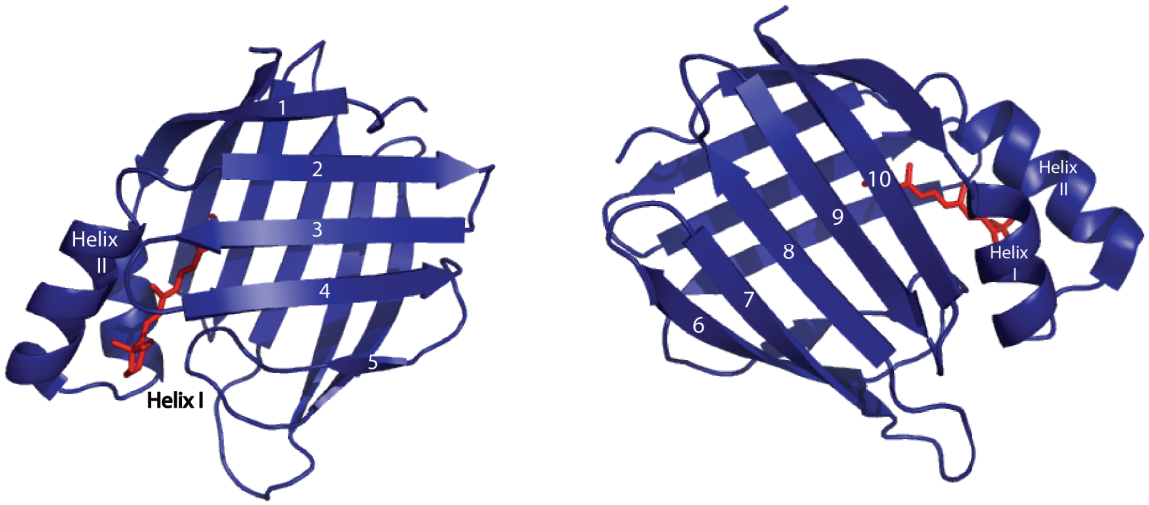
Competition between folding and aggregation is imminent in proteins with rugged energy landscapes such as β-sheet rich proteins. β-sheet rich proteins have complex topologies and exhibit high degrees of energetic frustration. Intracellular lipid binding proteins (ILBPs) are a family of proteins with frustrated energy landscapes yet have not been implicated in any misfolding diseases (Budyak*, et al.*, 2013). Over the past decades cellular retinoic acid binding protein (CRABP 1), a member of the ILBP family has been used as an excellent model to study the folding and aggregation of β sheet-rich proteins (Liu*, et al.*, 1994, Clark*, et al.*, 1996, Clark*, et al.*, 1997, Clark*, et al.*, 1998, Eyles*, et al.*, 1999, Rotondi & Gierasch, 2003, Ignatova & Gierasch, 2004, Ignatova & Gierasch, 2005, Ignatova & Gierasch, 2006, Marcelino & Gierasch, 2008, Ignatova & Gierasch, 2009, Budyak*, et al.*, 2013). However, similar to other proteins, a link between these two distinct overlapping processes is not yet established. This chapter will present findings from the experiments we carried out with an aim to fill the gap in knowledge between folding and aggregation of CRABP 1. We will show common sequences in our model protein that are critical for both folding and aggregation. Furthermore, we will demonstrate how in the case of CRABP 1 physico-chemical properties required for aggregation *in vitro* are maintained *in vivo* and validate predictions offered using several

aggregation algorithms. Finally, this chapter will create an integrated picture describing how key sequence features in CRABP 1 that direct aggregation are protected early by folding.

2.1.1. Structure, Function and Dynamics of CRABP 1

CRABP 1 is a 136-residue β-sheet-rich protein and a member of the ILBP family. Its proposed role in the cell is to control the nuclear uptake of retinoic acid at different stages of development (Donovan*, et al.*, 1995). Like other ILBPs, CRABP 1 has a β-barrel structure made up of two orthogonal five stranded β-sheets with an open angle between and two helices that connect strands 1 and 2 (Kleywegt*, et al.*, 1994, Thompson*, et al.*, 1995) (Figure 2.1). Except for strands 5 and 6, which are connected by a variable Ω-loop, other β-strands are linked by short reverse turns (Kleywegt, Bergfors et al. 1994). A hydrogen bond ladder stabilizes the β-strands in the barrel and is broken between strands 4 and 5 (Thompson, Bratt et al. 1995). Like other ILBPs, it forms a ligand binding solvent-shielded cavity filled with ordered water (Kleywegt*, et al.*, 1994, Thompson*, et al.*, 1995). Retinoic acid is sequestered in the barrel by hydrophobic interactions with helix 2, turn 2 (between strands 3 and 4), turn 4 (between strands 5 and 6) and polar interactions at the innermost cavity making the barrel almost inaccessible (Kleywegt*, et al.*, 1994, Li, 1999). The helix-turn-helix structural motif acts as lid that opens up a dynamic portal between turns 2 and 4, allowing ligand entry and egress (Thompson*, et al.*, 1995, Krishnan*, et al.*, 2000, Xiao & Kaltashov, 2005). Binding of retinoic acid reduces the dynamics of this portal and locks the ligand inside the protein. In chapter 3, we will discuss the structurally dynamic behavior of CRABP 1 and other ILBPs that engages in ligand binding.

Figure 2.1. Two projections of the structure of cellular retinoic acid binding protein 1 (CRABP 1) bound to retinoic acid. CRABP 1 consists of two anti-parallel five stranded β-sheets forming a large ligand binding cavity and connected by a helix-turn-helix motif. Turn II connecting strands 3 and 4 and turn 4 which links strand 5 and 6 interact with the helical region and form the ligand portal site. (PDB ID: 1CBR)
(Kleywegt, G. J., T. Bergfors, et al. (1994). "Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid." Structure **2**(12): 1241-1258)


1
2
3
4
5
Helix
II
Helix I
10
9
8
7
6
Helix
II
Helix
I

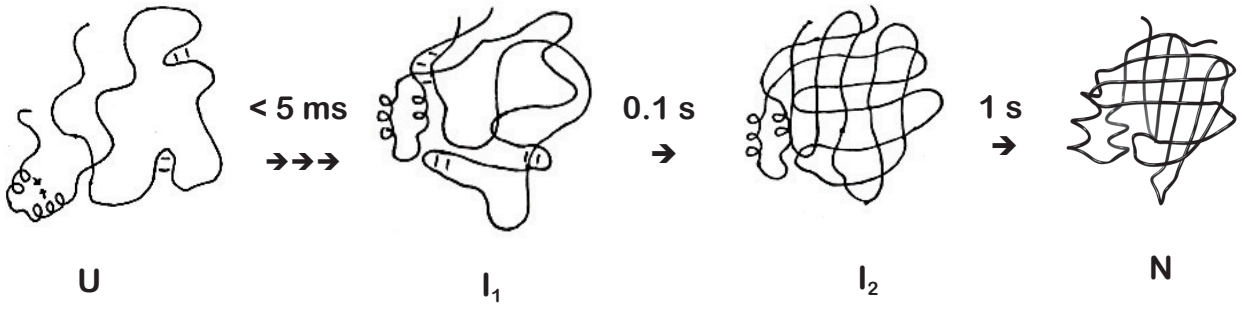
2.1.2. Folding and Aggregation of CRABP 1

CRABP 1 has been used as a suitable model protein for understanding the folding and aggregation landscapes of β-sheet proteins (Clark*, et al.*, 1998, Ignatova & Gierasch, 2005, Ignatova*, et al.*, 2007). CRABP 1 like other ILBPs has a rugged folding landscape, which has been described over the years using a wide array of biophysical techniques such as intrinsic tryptophan fluorescence, ANS binding and circular dichroism methods (Clark*, et al.*, 1996, Clark*, et al.*, 1997, Clark*, et al.*, 1998). CRABP 1 folds via three well-defined kinetic phases. The initial phase is characterized by a rapid hydrophobic collapse (τ=250 μs) where local segments of chain adopt significant secondary structure (Clark*, et al.*, 1997). This is followed by the establishment of specific interactions that restrict the arrangement of the chains into a native topology (τ=100 ms)(Clark*, et al.*, 1998). The final phase (1s) consists of specific packing of the β-sheet side chains and formation of stable inter-strand hydrogen bonding (Clark*, et al.*, 1998) (Figure 2.2).

As mentioned earlier despite the high beta sheet content and the ruggedness of folding landscapes, CRABP 1 and other ILBPs have not been associated with any misfolding diseases. This observation implies that this class of proteins may have evolved this robust folding mechanism with built-in aggregation protection (Budyak*, et al.*, 2013). Although CRABP 1 is found to be a successful folder, we have the capacity to explore its aggregation landscape by introduction of single residue substitutions on CRABP 1 or by modest alterations in solution conditions (addition of low concentrations of denaturant or at slightly acidic pH). Previously, it has been found that *in vitro* and upon overexpression in *E. coli*, CRABP 1 mutants form amorphous aggregates containing β-lamellar structures with 10.03Å spacing between adjacent beta-sheets as revealed by wide-angle X-ray scattering (WAXS) (Ignatova & Gierasch, 2004, Ignatova*, et al.*, 2007).

CRABP 1 folding and aggregation in *E. coli* have been monitored using a real-time fluorescence labeling method (Ignatova & Gierasch, 2004). This strategy involves incorporation of structurally non-perturbing, specific tetra-Cys binding site for a bis-arsenical fluorescein-based dye FlAsH that reports the folding and solubility of the protein (Ignatova & Gierasch, 2004). Using this approach, aggregation of various CRABP 1 mutants was monitored (Z. Ignatova & L. Gierasch, unpublished). Aggregation of a slow-folding CRABP 1 mutant (Ignatova & Gierasch, 2005) showed similarities to amyloids, following a nucleation-dependent polymerization requiring the formation of an energetically unfavorable nucleus. (Ignatova & Gierasch, 2005, Ignatova*, et al.*, 2007). However, we have not yet fully understood the molecular basis of aggregation propensity of CRABP 1. With the growing number of aggregation-prone mutants of CRABP 1 being identified in the lab, we have been equipped to explore sequences that determine the tendency of CRABP 1 to aggregate (Z. Ignatova and L. Gierasch, unpublished). In the following section, we will describe hydrogen exchange experiment we have utilized to identify regions sequestered in CRABP 1 aggregates.

Figure 2.2. Folding pathway of CRABP 1. The proposed folding pathway for CRABP 1 proceeds through an early hydrophobically collapsed to form intermediate, I_1 (τ= ~250 μs); followed by formation of I_2 which has the native topology and ligand binding cavity (τ= ~100 ms); and then by establishment of a hydrogen-bonding network and specific tertiary interactions to form the native state, N (τ= ~1 s).


< 5 ms
0.1 s
1 s
U
I_1
I_2
N

2.1.3. Hydrogen-Deuterium Exchange NMR to Map Aggregation Cores

Short segments of a protein of at least five amino acid residues are sufficient to form ordered aggregates (Chiti*, et al.*, 2003, Fernandez-Escamilla*, et al.*, 2004). These segments have strong propensities to exhibit cross-β structures in aggregates and are highly occluded from the environment (Hoshino*, et al.*, 2002, Frare*, et al.*, 2006). Aggregation core sequences have been identified experimentally using methods that probe for solvent accessibility as well as protease resistance (Hoshino*, et al.*, 2002, Frare*, et al.*, 2006, Wang*, et al.*, 2008, Vilar*, et al.*, 2012). Amide hydrogen deuterium exchange (HX) has been used to map core segments of β2m (Hoshino*, et al.*, 2002), transthyretin (Olofsson*, et al.*, 2004), Aβ peptide (Kheterpal*, et al.*, 2006) amyloids as well as of bacterial inclusion bodies (Hoshino*, et al.*, 2002, Olofsson*, et al.*, 2004, Kheterpal*, et al.*, 2006, Vilar*, et al.*, 2012). This method sensitively probes for highly ordered regions of the protein molecule forming amyloid fibrils at amino acid residue resolution. In this thesis we have successfully employed hydrogen exchange in resolving aggregation core sequences of CRABP 1.

Hydrogen deuterium exchange measures the rate of exchange between the amide hydrogen in the polypeptide backbone and deuterium in the solvent. Exchange rates can be monitored using available methods that can distinguish between proton and deuterium such as nuclear magnetic resonance (HX-NMR) and mass spectrometry (HX-MS) (Scholtz & Robertson, 1995, Hoofnagle*, et al.*, 2004). Measurements of exchange rates can unveil the presence or absence of protecting structure thus are valuable in determining thermodynamic stability, studying protein dynamics and following protein folding (Eyles & Kaltashov, 2004, Krishna*, et al.*, 2004). Amide hydrogens in an unfolded polypeptide are labile and exchange readily with the solvent. On the other hand, in

folded proteins where secondary and tertiary structures prevail, exchange is retarded by H-bonding networks or by burial of hydrophobic residues in the core of a protein. The extent of protection from exchange in the folded protein correlates with the degree of ordered structures.

Hydrogen exchange consists of two steps: First is the exposure of backbone amides to the solvent by reversible unfolding of the protein. Second is exchange between deuterium from the solvent and the exposed amide protons. The overall exchange kinetics are governed by both reactions.

$$NH_c \underset{k_c}{\overset{k_o}{\rightleftharpoons}} NH_o \xrightarrow{k_{int}} ND$$

NH_c and NH_o represent the amide groups in the “closed” (folded) and “open” (unfolded) ND is the exchanged amides; The measured exchange rate k_{ex} is determined by the opening k_o, closing k_c rate constants and the expected rate of exchange assuming the site is completely exposed, k_{int}. The relationship is shown as:

$$k_{ex} = \frac{k_o k_{int}}{k_o + k_c + k_{int}}$$

Under conditions where HX is generally performed, the $k_c >> k_o$ to approximate the previous equation as:

$$k_{ex} = \frac{k_o k_{int}}{k_c + k_{int}}$$

There are two extreme limits of exchange EX1 and EX2. The first case is when $k_c >> k_{int}$, an behavior expected at low pH and temperature, the exchange rate becomes:

$$k_{ex} = \frac{k_o k_{int}}{k_c} = K_o k_{int}$$

where $K_o = k_o/k_c$. This is an EX2 or bimolecular exchange limit. A heterogeneous pattern of exchange results from incomplete exchange at all sites. The second case is when $k_c << k_{int}$, all protons in the unfolded segment will exchange, in this case, unfolding of the protein is a rate-limiting step thus,

$$k_{ex} = k_o$$

This limit of exchange is referred to as EX1 or monomolecular exchange. EX1 appears as a bimodal pattern with all amides either exchanged or unexchanged. This is usually observed at high pH and temperature the exchange reaction rate is limited by the opening event of the amide site.

Hydrogen exchange combined with high-resolution NMR spectroscopy provides detailed information on the structure and thermodynamic stability of proteins at residue-specific resolution. This method has been widely applied in probing various conformational states of proteins, including denatured states, and equilibrium or kinetic folding intermediates with complex three-dimensional structure (Hoshino*, et al.*, 2007). A typical HX-NMR experiment requires high concentrations of proteins in solution and an appropriate NMR measurement is recorded. Cross-peak intensities used to detect amide protons such as ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) directly correlates with the occupancy of amide protons. Exchange rates are determined by monitoring the decrease in peak intensity in the spectrum. A single exponential decay plot of peak intensity as a function of exchange time is generated as described in the equation:

$$I(t) = I(0)\ exp(-k_{ex}t)$$

where t is the exchange time, I(0) and I(t) are peak intensity at time zero and time t, respectively, and kex is the exchange rate constant.

In a hydrogen exchange experiment protein samples are incubated in a deuterated solvent for a period of time followed by a quenching reaction at low pH and temperature. At low temperatures and pH, the rates of exchange are reduced preventing back exchange in deuterated amides. HX detection using solution NMR poses inherent limitation on the apparent size of the target protein in which signal is approximately inversely proportional to the molecular weight of the protein (Hoshino*, et al.*, 2007). Line width broadening of NMR signal results from increase in rotational correlation time as molecular weight of the protein increases. Since the line widths are infinitely broadened for supramolecular complexes, direct measurement of NMR spectrum is impossible. Nevertheless, a modified hydrogen exchange NMR method has been developed to study the structural and dynamic properties of β-microglobulin and cold shock protein A (CspA) aggregates (Alexandrescu, 2001, Hoshino*, et al.*, 2002). Despite the fact that this does not offer information about the geometric constraints of the aggregates, it provides details on structural reorganizations in the native protein accompanying aggregation (Vilar*, et al.*, 2012). In addition it also allows identification of peptide segments involved in the formation of secondary structures that stabilize aggregates (Vilar*, et al.*, 2012). An indirect HX technique uses an organic aprotic solvent dimethyl sulfoxide (DMSO) to quench exchange and dissolve aggregates in solution to permit analysis by solution NMR (Hoshino*, et al.*, 2007). At very high DMSO concentrations, aggregates are dissolved into monomers within the dead time of NMR measurements (Hirota-Nakaoka*, et al.*, 2003). The exact oligomeric state of the protein in DMSO is not known however no difference in line widths in the NMR spectra is evident between the monomer and dissolved aggregates (Hoshino*, et al.*, 2007).

The rate of exchange in DMSO is 100-fold lower in 90-95% DMSO than in water (Zhang*, et al.*, 1995). Furthermore, pH dependent exchange is different in water and in

DMSO shifting the minimum from pH 3 to pH 5 (Zhang*, et al.*, 1995). In DMSO, formation of deprotonated amide necessary in the base-catalyzed exchange is unfavorable resulting in this pH minimum shift (Zhang*, et al.*, 1995). This also shifts the pKas of acids to a great extent whereby strong acids in aqueous solutions such as dichloroacetic acid (DCA) and trifluoroacetic acids may act as buffers in DMSO (Zhang*, et al.*, 1995). DMSO not only dissolves aggregates but also offers as a solvent without labile hydrogens that may cause back-exchange. At high DMSO concentrations, solubilized proteins are denatured thus amide proton exchange for each residue is susceptible to exchange with trace water in the solvent. However, the rates of back-exchange is reduced in high DMSO solvents thus permits acquisition of multiple HSQC spectra for analysis (Hoshino*, et al.*, 2007).

The steps for DMSO-quenched hydrogen exchange are as follows: 1) Preparation of aggregates in aqueous solution; 2) incubation of aggregates in deuterated buffer; 3) collection of aggregates at various exchange times; 4) exchange quenching and dissolution of aggregates at high (>95%) concentrations of DMSO-d6 at an appropriately adjusted pH value; 5) acquisition of two-dimensional NMR spectra successively to detect resonance signals from remaining protonated residues (Hoshino*, et al.*, 2007).

Several factors may interfere with this experiment. First is the fact that DMSO is an aprotic and hygroscopic solvent it is difficult to maintain pH and to prevent water contamination (Hoshino*, et al.*, 2007). In addition even if water is completely eliminated, internal exchange between amide protons of residues may complicate the experiment (Hoshino*, et al.*, 2007). To minimize the effect of water contamination it is recommended that 95% (v/v) DMSO-d6 and 5% (v/v) D_2O at pD 5 adjusted by DCA-d_2 be used as a

quenching buffer for the exchange reaction. The presence of residual exchange in 5% (v/v) of D_2O is not expected to complicate the experiment as long as the rate of exchange and water contamination is very minimal (Hoshino*, et al.*, 2007).

2.2. Results

2.2.1. Aggregation Propensity of CRABP 1

In our study, we have used a stabilized variant of CRABP 1, WT* containing R131Q mutation. Mutation of Arg 131 breaks the charged interactions between retinoic acid leading to less affinity to the ligand (Zhang*, et al.*, 1992). This variant has been previously used in understanding the folding landscape of CRABP 1 (Clark*, et al.*, 1996, Clark*, et al.*, 1998, Eyles*, et al.*, 1999, Eyles & Gierasch, 2000, Marcelino & Gierasch, 2008). While this mutant has a lower ligand binding affinity, it is more thermostable and highly soluble when over-expressed in *E. coli*. Thus, WT* CRABP 1 serves as a suitable background to monitor effects of residue substitutions or solution conditions on aggregation propensity.

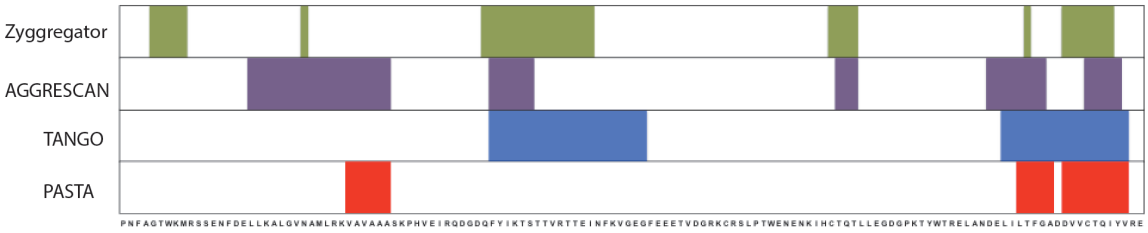
Using several available aggregation prediction algorithms namely, AGGRESCAN (Conchillo-Sole*, et al.*, 2007, de Groot*, et al.*, 2012), TANGO (Fernandez-Escamilla*, et al.*, 2004), PASTA (Trovato*, et al.*, 2006) and Zyggregator (Tartaglia & Vendruscolo, 2008) aggregation hotspots in CRABP 1 were predicted (Appendix A). Predictions varied only slightly from one algorithm to another, as aggregation prone regions are usually predicted by at least two algorithms. Aggregation propensity scores have been consistently high in strands 3, 4, 9 and 10 and helix II (Figure 2.2a).

Figure 2.3 Predicted aggregation-prone regions of CRABP 1.

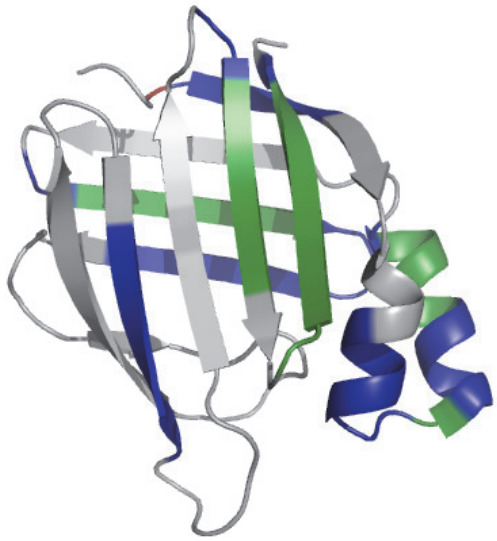
A) Predicted aggregation-prone regions using Zyggregator, AGGRESCAN, TANGO and PASTA mapped on to the sequence of CRABP 1.

B) Aggregation-prone regions projected on to the structure of CRABP 1. Sequences predicted by at least two algorithms are painted in green. Sequences predicted by only one algorithm are painted in blue. All predictors identified the C-terminal strands to be aggregation prone.

A)


Zyggregator
AGGRESCAN
TANGO
PASTA
PNFAGTWKMRSSENFDELLKALGVNAMLRKVAVAAASKPHVEIRQDGDQFYIKTSTTVRTTEINFKVGEGFEEETVDGRKCRSLPTWENENKIHCTQTLLEGDGPKTYWTRELANDELILTFGADDVVCTQIYVRE

B)



WT* CRABP 1 can be overexpressed as soluble protein in *E. coli*, but we found a number of single residue substitutions that result in different extents to aggregate (Table 2.1. The aggregation propensities of CRABP 1 variants were determined using cell fractionation method. This was accomplished by overexpressing mutant protein in BL21(DE3) *E. coli* strains at 37^{o}C for 3 hours and then lysing the cells using a commercially available mild non-ionic detergent. Initial experiments to test the reproducibility of cell disruption techniques revealed that optimal lysis can be achieved without breaking the aggregates using this mild detergent as compared to sonication and freeze thaw methods. Following cell disruption, lysates were separated into supernatant and pellet fractions and run on an SDS-PAGE. The distributions between the soluble and pellet fractions of the protein were quantified by measuring the optical densities of protein bands corresponding to CRABP 1 and taking the ratio of the optical densities of the pellet and soluble fractions with respect to the lysate fraction (Table 2.1).

To test whether aggregation propensities of the CRABP 1 mutants can be predicted, we again employed the aggregation predictors. We have found that except for a couple of mutants, aggregation predictors did not distinguish aggregation propensities between CRABP 1 variants with very different experimental aggregation propensities. An example is the case of F71A, which is a mutation in strand 5 and involved the hydrophobic core network of CRABP 1. We experimentally find this mutant to be highly aggregation-prone since it can only be expressed as inclusion bodies. However, all the aggregation predictors did not show any increase in aggregation propensity for this mutant. We also found a couple of mutations, which emerged to affect the predicted aggregation propensity of CRABP 1. These include mutations I52A and L118V. isoleucine 52 is found in strand 3 and leucine 118 is found in strand 9 (Appendix B). I52A substitution had a lower overall predicted aggregation propensity resulting from a

decrease in CRABP 1 β-propensity and hydrophobicity attributed to the substitution of isoleucine to alanine (Appendix C). However, in contrast to predictions, compared to the WT protein, I52A mutant was completely insoluble upon overexpression in *E. coli* (Table 2.1). In the case of L118V the substitution increased the hydrophobicity of the protein thus a higher aggregation propensity was predicted (Appendix D) Experimentally, we find L118V is moderately (~50% soluble after 3 hours of protein overexpression) aggregation-prone in comparison to WT CRABP 1 (Table 2.1). Examination of aggregation propensities of CRABP 1 mutants suggests that effects of single amino acid substitutions on globular proteins cannot be sensitively predicted based solely on protein sequence. Other factors such as thermodynamic stability and structural fluctuations inherent for different proteins, which dictate the exposure of aggregation-prone sequences must be integrated in models to forecast aggregation propensities.

Table 2.1. Aggregation propensities and thermodynamic stabilities of CRABP 1 variants

$\Delta\Delta G_{U\text{-}N}$ (energetic effect of mutations on the free energy of the unfolded state (U) with respect to the native state (N)) = $\Delta G_{U\text{-}N}^{WT} - \Delta G_{U\text{-}N}^{Mutant}$
Aggregation propensities were determined based on partitioning of the protein into soluble and insoluble fractions after overexpression in *E. coli* for 3 hours at 37^{o}C.

Mutant	Aggregation Propensity	ΔΔG (kcal/mol)
WT*	0	0
R29A	0	0.8 ± 0.3
E69A	10	1.3 ± 0.2
T75A	20	2.4 ± 0.3
R79A	30	3.1 ± 0.2
G68A	50	2.6 ± 0.2
G78A	50	2.5 ± 0.4
L118V	50	2.3 ± 0.4
F50M	50	2.0 ± 0.2
F65M	50	3.3 ± 0.2
F71M	50	2.7 ± 0.2
Y133S	90	4.8 ± 0.2
R135G	90	3.9 ± 0.3

2.2.2. Sequestered sequences in CRABP 1 *in vitro* and *in vivo* aggregates

In order to find significance in our experiments, we first explored how aggregation *in vitro* compared *in vivo*. Thus we performed experiments to compare morphologies and molecular interactions stabilizing *in vitro* and *in vivo* aggregates. To inspect the morphologies of the *in vitro* and *in vivo* aggregates, transmission electron microscopy (TEM) was performed. Inclusion bodies were purified from washing the cell pellet with non-ionic detergent and treating with lysozyme and DNase. To yield *in vitro* aggregates, aggregation-prone F71A CRABP 1 mutant proteins isolated from inclusion bodies were refolded from high urea concentrations into buffer without urea at pH 7.0 containing 150 mM NaCl and 5 mM DTT at 37^{o}C. Insoluble aggregates were formed during refolding of aggregation-prone mutants at high concentrations (above 50 μM) under these conditions. Aggregates generated under these different conditions appeared amorphous and non-fibrillar consistent with previous observations (Ignatova & Gierasch, 2005). Although both were amorphous, the morphologies were different. *In vivo*, CRABP 1 forms uniform globular aggregates while aggregate formed *in vitro* did not (Appendix E).

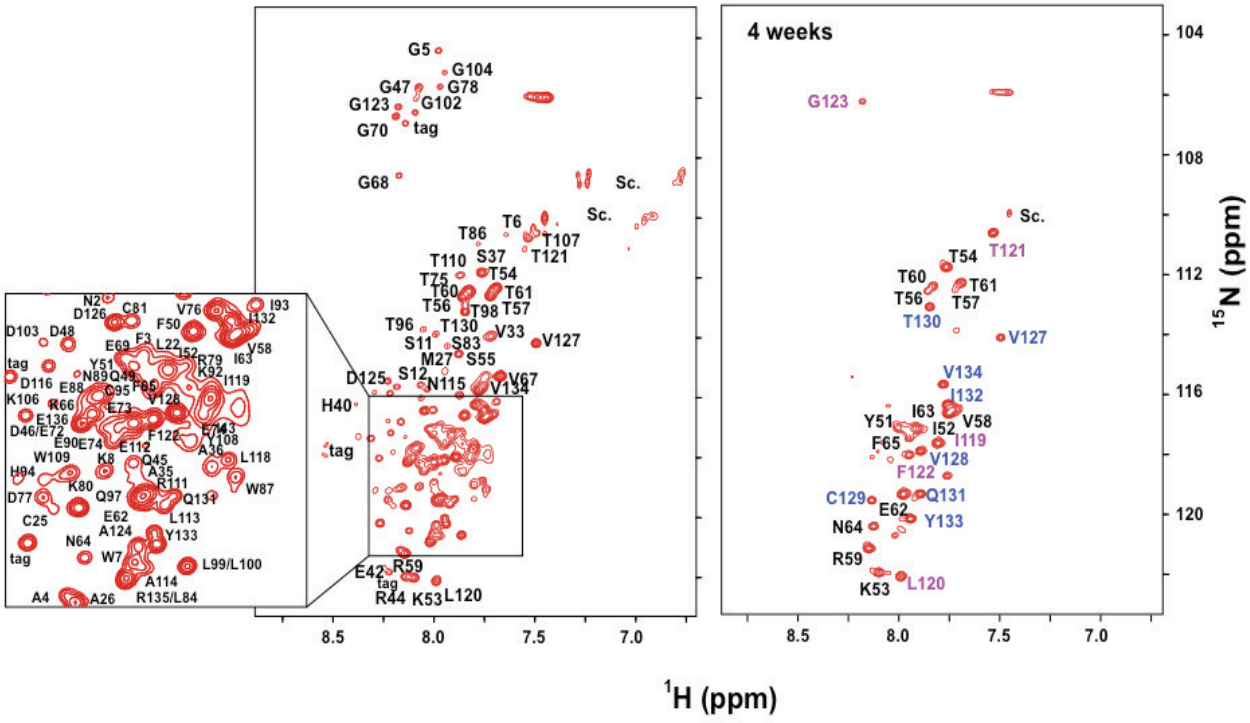
To interrogate whether there are contiguous sequences sequestered in these amorphous aggregates and whether they are consistent with predicted aggregation hotspots, aggregate cores of both *in vitro* and *in vivo* aggregates were mapped. Using DMSO-quenched hydrogen deuterium exchange NMR as an approach, aggregation cores of *in vitro* aggregates and inclusion bodies of CRABP 1 mutants were determined. ^{15}N-labeled *in vivo* aggregates of F71A mutant were isolated using a non-ionic detergent. We estimated the purity of the inclusion bodies to be around 80-90% (Appendix F). Previous reports on the quality of purified inclusion bodies revealed that although inclusion bodies generally consist of overexpressed recombinant proteins, a small fraction of contaminating proteins are present (Ventura & Villaverde, 2006). These

include proteolytic fragments of the recombinant protein, molecular chaperones such as small heat-shock proteins (IbpA and IbpB), DnaK and GroEL (Ventura & Villaverde, 2006, Dasari*, et al.*, 2011). Purified CRABP 1 inclusion bodies were tested for the presence of molecular chaperones DnaK and small heat shock protein, IbpA by western blot. Purified inclusion bodies from several aggregation-prone CRABP 1 mutants did not show presence of DnaK but were positive for IbpA (Appendix G).

Isotopically-labeled proteins in inclusion bodies and in *in vitro* aggregates for NMR studies were prepared as previously mentioned. Aggregates were exchanged in D_2O for four weeks to determine the highly protected cores. Incubation was performed for four weeks to ensure exhaustive exchange. Reactions implemented at different time intervals showed that no significant exchange occurs after four weeks. Aggregates were collected, lyophilized and resuspended in DMSO-d_6 containing 0.1% (v/v) trifluoroacetic acid (TFA), 50 mM DTT and 5% (*v*/*v*) D_2O (pD 3.0-3.5) to a final protein concentration of 200 µM. Quench reactions were done at pD=3.0-3.5 ^{1}H-^{15}N HSQC NMR experiments were performed to monitor amide protons. We performed our experiments in buffers with pD=3.0 to 3.5 instead of pD=5.0, the condition suggested by Hoshino and Goto that minimizes back-exchange, since we find less pH-dependent chemical shifts under this condition. We performed backbone residue assignments using purified $^{13}C/^{15}N$-labeled WT* CRABP 1 dissolved in quench buffer. We observed a narrow and crowded 2D spectrum with a number of overlapping peaks. Hence, we added another dimension by performing several triple-resonance NMR experiments to resolve individual residues. Backbone assignments were transferred to the spectra of the mutant protein and adjusted accordingly for any shifts associated with the mutations. Figure 2.4 shows a typical HSQC spectra of mutant CRABP 1 aggregates dissolved in DMSO. In total 128 out of 158 amino acid residues (around 80%) have peaks assigned. NMR spectra for

both *in vitro* and *in vivo* aggregates were compared before and after exchange. For controls, the unexchanged samples were incubated in water for the same amount of time and quenched under the same conditions as in the exchanged samples. It is worth mentioning that aggregates remained stable even after long periods of incubation in either water or D_2O. We have also performed initial control experiments to monitor back-exchanges to determine the duration for data acquisition by taking HSQC measurements of unexchanged aggregates every hour over a 24-hour period. We found little back-exchange if measurements were performed within two hours after dissolution of aggregates in DMSO. In our experiments, our data acquisitions were completed within 30 minutes after dissolving the aggregates.

Figure 2.4. $^{1}H^{15}N$ HSQC Spectra of F71A CRABP 1 before and after hydrogen-deuterium exchange. The NMR spectrum before exchange (left), shows several intense peaks with their residue assignments. On the right is the NMR spectrum after exchange showing less peaks. These remaining peaks correspond to the aggregation cores.


4 weeks
G5
G104
G47
G78
G123
G102
G70
tag
G68
Sc.
T6
T86
T107
T110
S37
T121
T54
T75
T60
T61
T56
T98
T57
T96
T130
V33
V127
S11
S83
M27
S55
D125
S12
N115
V67
V134
H40
tag
E42
R59
tag
R44
K53
L120
N2
D126
C81
V76
I93
I132
D103
D48
F50
F3
L22
E69
I52
V58
tag
Y51
R79
I63
N89
Q49
K92
I119
D116
E88
C95
F85
V128
K106
R66
E73
E136
D46/E72
E90
E74
E112
F122
E71
Y108
A36
L118
W109
K8
Q45
H94
A35
K80
R111
W87
D77
Q97
Q131
C25
E62
L113
A124
Y133
N64
tag
W7
L99/L100
A114
A4
A26
R135/L84
G123
Sc.
T121
T54
T60
T61
T56
T57
T130
V127
V134
I132
Y51
I63
I52
V58
F65
I119
V128
F122
C129
Q131
E62
Y133
N64
R59
K53
L120
8.5
8.0
7.5
7.0
104
108
112
116
120
^{15}N (ppm)
^{1}H (ppm)

A comparison of the acquired HSQC NMR spectra from *in vivo* and *in vitro* aggregates revealed that *in vivo* aggregates contained additional peaks not originating from CRABP 1 (Figure 2.5). These peaks may be attributed to the presence of contaminating proteins in the inclusion bodies. As we have mentioned earlier, we found a small percentage of other proteins in our samples. Additional peaks have also been observed in HSQC spectra of inclusion bodies from other proteins (Dasari*, et al.*, 2011).

As we expected, both *in vitro* and *in vivo* aggregates had only few but strong peaks remaining after exchange. Individual residues corresponding to intense peaks after exchange were identified. For *in vitro* aggregates, residues 51-65 with the exception of serine 55 (core 1), 119-123 (core 2) and 127-134 (core 3), were extremely protected (Figure 2.5). A comparison of aggregation core sequences with predictions correlated very well with predicted aggregation hotspots (Figure 2.6). Except for residues along the helical domain, we were able to observe high solvent protections in β-strands predicted by the different algorithms. Core 1 is located in the strand 3, turn II, and strand 4 while cores 3 and 4 are situated in strands 9 and 10 respectively (Figure 2.7). Surprisingly the exact same regions of CRABP 1 were found for the *in vivo* aggregates. Backbone amides of residues 52-64, 120-123 and 127-134 remain protonated.

To check whether the same sequences form the aggregation cores for other mutants, hydrogen exchange NMR experiments were also performed on L118V and I52A CRABP 1 inclusion bodies. All mutants contained highly protected residues and were consistent with those identified for F71A mutant (Figures 2.5). It is interesting that although both mutants are found close to two different aggregation cores, there was no preference for aggregation cores close to the mutation.

Figure 2.5. HSQC spectra of CRABP 1 inclusion bodies before and after hydrogen exchange. Inclusion bodies from moderately aggregation-prone mutant L118V (left) and highly aggregation-prone mutants F71A and I52A (center and right) were allowed to exchange in D_2O containing 0.025% NaN_3 for four weeks. Inclusion bodies were dissolved in DMSO-d6 containing 0.1% TFA and 5 mM DTT. Mutants of CRABP 1 contain similar aggregation core segments as revealed by similar peaks remaining after exchange.

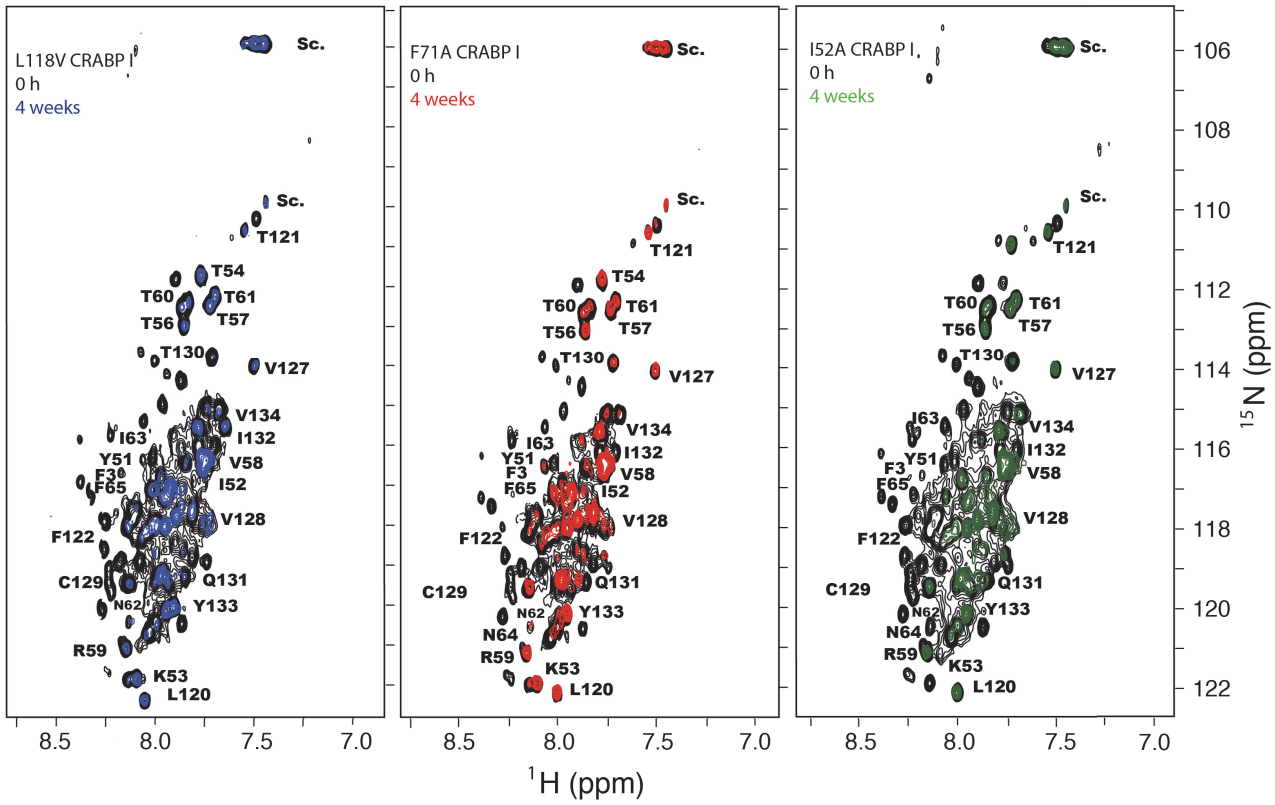

L118V CRABP I
0 h
4 weeks
F71A CRABP I
0 h
4 weeks
I52A CRABP I,
0 h
4 weeks
Sc.
T121
T54
T60
T61
T56
T57
T130
V127
V134
I63
I132
Y51
V58
F3
F65
I52
V128
F122
C129
Q131
N62
Y133
N64
R59
K53
L120
106
108
110
112
114
116
118
120
122
15N (ppm)
8.5
8.0
7.5
7.0
1H (ppm)

Figure 2.6. Experimentally determined and predicted aggregation-prone regions of CRABP 1. Aggregation cores of CRABP 1 determined by hydrogen exchange mapped against the amino acid sequence. Also aligned with the sequences are the predicted aggregation segments. These cores were consistent to predictions

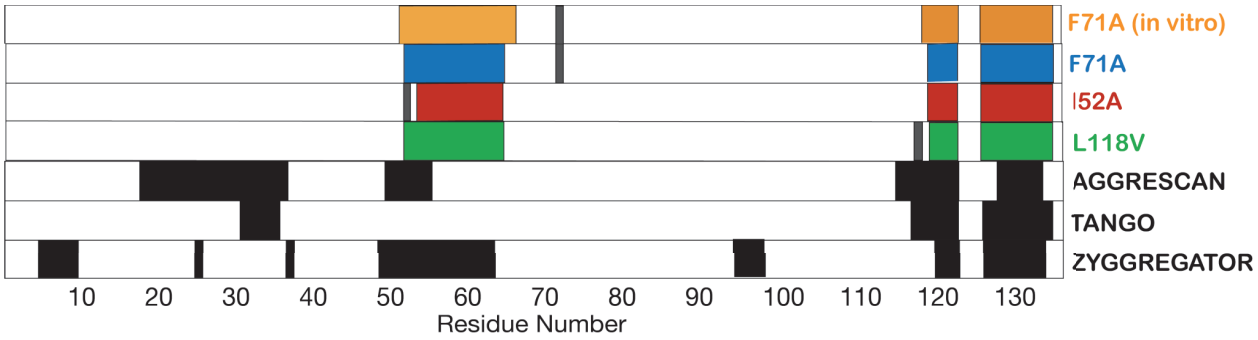
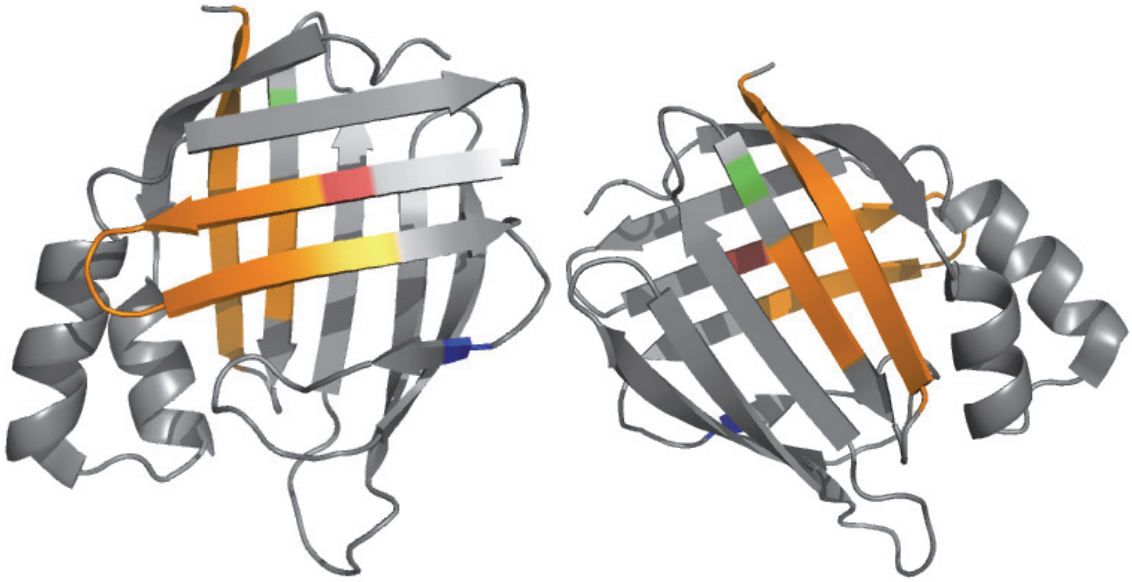

F71A (in vitro)
F71A
I52A
L118V
AGGRESCAN
TANGO
ZYGGREGATOR
10
20
30
40
50
60
70
80
90
100
110
120
130
Residue Number

Figure 2.7. Aggregation cores of CRABP 1.
Aggregation-prone segments mapped on the structure of CRABP 1 (orange highlights). Aggregation cores are found at strands 3, 4, 9 and 10 and in turn 2 of CRABP 1. Indicated in the structure are the locations of the mutations: Phe71 (blue); Ile52 (red) and Leu118 (green).



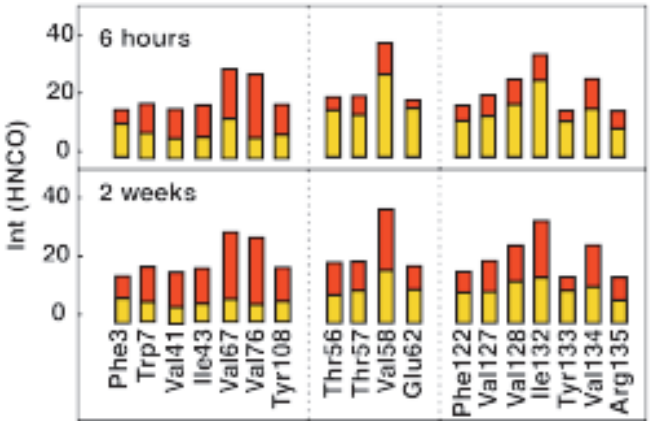
To verify whether the very slow exchange behavior of the aggregation cores is distinct from the other residues, residual peak intensities from triple resonance HNCO experiments were compared at different time points. In order to minimize acquisition time, sparse HNCO experiments (Orekhov*, et al.*, 2003, Jaravine*, et al.*, 2008). Sparse 3D experiments applies non-Fourier transform methods for analysis and non-uniform sampling to minimize data acquisition time while maintaining good resolution. HNCO experiments were performed on I52A CRABP 1 inclusion bodies exchanged at varying time intervals. Intensities of individual, non-overlapping peaks were determined and plotted with time (Figure 2.8). We did not extract thermodynamic parameters from our quantitation since the peak intensities had low signal to noise ratios. However, we were able to classify residues based on two types of behaviors; fast and slow exchanging residues. In fast exchanging residues, we observed a rapid decrease in intensity within a few hours of exchange and after two weeks, signals have deceased at least five-fold (Figure 2.8). However, for slow exchanging residues, signal intensities decrease slower and still contain at least 30% of their intensities after two weeks of exchange.

Figure 2.8. HNCO peak intensities of CRABP 1 residues from aggregates exchanged over time

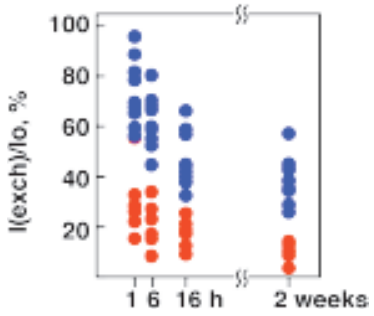
A) Peak intensities obtained from sparse HNCO spectra (each spectra were recorded for less than 2 hours to minimize H/D exchange during experiments of highly aggregation-prone CRABP1 mutant I52A aggregates in 95% d6-DMSO/5% D2O for the unexchanged sample (red) and samples that were H/D-exchanged (yellow) for 6 hours (top) and 2 weeks (bottom). All assigned residues those peaks have signal-to-noise ratio more than 10 (and errors in peak intensities more than 10%) for the unexchanged sample were used for analysis. Resonances for residues from the aggregation core(s) (middle and right) retained significant (30-60% from intensities for the unexchanged sample) intensities after 2 weeks of exchange, while intensities for the rest resonances (with exception of Phe3) decreases in more than 5 times (left).

B) Relative peak intensities as a function of H/D-exchange time, $I(exch)/I_o \times 100\%$, where I(exch) and I_o are peak intensities for residues shown on (A) [with exception of Phe3, the quality of data for this residue is too noisy for analysis] obtained from H/D-exchanged and unexchanged samples, respectively. Residues from the aggregation cores (shown in blue) have a significantly slower decay for relative peak intensities as comparing with the rest protein (red); while behavior of the (different) aggregation core(s) are similar within experimental errors (10-20%).

A)


6 hours
2 weeks
Int (HNCO)
40
20
0
Phe3
Trp7
Val41
Ile43
Val67
Val76
Tyr108
Thr56
Thr57
Val58
Glu62
Phe122
Val127
Val128
Ile132
Tyr133
Val134
Arg135

B)


I(exch)/Io, %
100
80
60
40
20
1
6
16 h
2 weeks

2.3. Discussion

2.3.1. Common Sequences Direct Aggregation of CRABP 1

Protein folding and aggregation are driven by similar amino acid properties such as hydrophobicity, secondary structure propensity and charge. Thus, competition between intramolecular and intermolecular interactions, which direct folding and aggregation respectively, is inevitable. However, the balance between folding and aggregation can be shifted under defined conditions allowing a plethora of independent mechanistic studies. This study aims to make a connection between folding and aggregation to provide a comprehensive picture that will illustrate the fate of a polypeptide chain. Here, sequence determinants that link folding and aggregation of a β-protein were described. In addition, we demonstrated the specificity of aggregation by comparing aggregate structures of various mutants generated *in vitro* and *in vivo*.

CRABP 1, a β-clam protein whose folding mechanism has been studied in detail was used as a model. Owing to their rugged folding energy landscapes with high degree of frustration, multiple intermediate states and increased propensity to aggregate, predominantly β-sheet proteins are an apt structural class for examining molecular details of the balance between folding and aggregation. CRABP 1 folds via two intermediate states an early collapsed state (I_1) and a later β-molten globule-like state with native topology (I_2) – and thus exemplifies β-barrel frustration (Clark*, et al.*, 1996, Clark*, et al.*, 1997, Clark*, et al.*, 1998). Despite this, CRABP 1 and members of its class fold and avoid aggregation efficiently. To study how CRABP 1 protects itself from aggregation, we mapped sequences that are sequestered in aggregates. As we discover the role of these sequences in folding of CRABP 1 from previous studies we were able to find clear links to aggregation.

First we interrogated whether aggregation is specific by comparing the molecular characteristics of *in vitro* and *in vivo* aggregates of CRABP 1. We initially developed protocols to generate *in vitro* and *in vivo* aggregates. For *in vitro* aggregates we aimed to produce aggregates that will simulate how proteins aggregate in cells but in the absence of other complexities such as a crowded environment and other cellular machineries such as molecular chaperones and proteases. An aggregation-prone mutant F71A was overexpressed as inclusion bodies and purified in high urea. With the protein denatured in urea, it mimics an unfolded nascent chain. From a highly denaturing environment, the protein was refolded into native-like buffer conditions. During this process, depending on the initial protein concentration, a significant amount of aggregates are formed permitting us to analyze its structure. To generated *in vivo* aggregates, the mutant protein was overexpressed as inclusion bodies and purified using a mild non-ionic detergent. Using this method, cell lysis was efficient yet inclusion bodies remained intact. As mentioned in the previous section, several other proteins co-aggregated with CRABP 1. We have noted the presence of small heat shock protein IbpA. Small heat shock proteins IbpA and IbpB bind to aggregated proteins and change their physical properties to mediate disaggregation and refolding by chaperones (Ratajczak*, et al.*, 2009). Aggregates derived *in vitro* and *in vivo* were strikingly different in morphologies. *In vitro* aggregates appeared heterogeneous with no distinct shape while inclusion bodies had the typically observed round shaped structures (Ignatova & Gierasch, 2005, Morell*, et al.*, 2008). It appears that in a crowded cellular environment, aggregation of proteins is confined in space giving rise to compact inclusions. An interesting aspect to further explore is the potential role of bound small heat shock proteins in modulating aggregate morphologies and molecular structures in cells.

To characterize the interactions that stabilize aggregates, DMSO-quenched hydrogen exchange was performed and monitored using solution NMR. The use of this protocol allowed us to resolve regions of protein sequestered from the solvent in aggregates. DMSO dissolved *in vitro* aggregates almost instantly while inclusion bodies were only partly soluble. A small percent of the inclusion bodies, which may be non-protein components, was insoluble in DMSO. However, these did not interfere with the analysis. Dissolution of aggregates in DMSO provided high sensitivity allowing us to study consecutive regions of CRABP 1 extremely resistant to solvent exchange by NMR. Analysis of the aggregates derived under extremely diverse conditions revealed common sequence motifs that drive aggregation. These sequences correlated to a high extent with that of predicted aggregation hotspots. These suggests that the physico-chemical properties of protein sequences as evaluated by several aggregation predictors dictate the potential to be sequestered in aggregates. The identical sequences in aggregation cores *in vitro* and *in vivo* supports the idea that although aggregation involves interaction between hydrophobic patches it is specific even under extremely variant conditions. In separate reports, it has been shown that inclusion bodies, although amorphous, have high specificity similar to highly ordered amyloid structures (de Groot*, et al.*, 2008, Morell*, et al.*, 2008, Wang*, et al.*, 2008). Thus inclusion bodies have been proposed to be useful models for intermediates in amyloid formation (Dasari*, et al.*, 2011).

To further investigate whether other aggregation-prone mutants contain the same core structures, L118V and I52A CRABP 1 mutants were also tested. L118V has a high predicted and moderate experimental aggregation propensity. However, I52A has a lower predicted aggregation propensity but was highly aggregation-prone. Leu118 is close to aggregation core 2 and Ile52 is in core 1 based on the core sequences identified

on F71A CRABP 1 mutant. Using the same hydrogen exchange method, the aggregation cores determined in these mutants corresponded to the same amino acid stretches determined for F71A mutant. These suggest that the population of common aggregation-prone intermediate/s that expose the same consensus sequences dictates the extent of aggregation of CRABP 1. In the following chapter, we will describe experiments performed to resolve the aggregation-prone intermediate of CRABP 1 under native conditions.

2.3.2. Common Sequence Motifs Drive Aggregation and Folding of CRABP 1

An extensive study on the early events in folding of CRABP 1 was done in parallel with this study (Budyak*, et al.*, 2013). In this study single residue substitutions at 33 sites were introduced in the CRABP 1 sequence with comprehensive coverage of structural elements including the minor hydrophobic core near the barrel closure region (Budyak*, et al.*, 2013). The observed impact of these mutations on CRABP 1 stability and unfolding kinetics revealed that its rate-determining transition state (TS) is highly polarized (Budyak*, et al.*, 2013). It showed that the barrel closure interactions formed before the TS, and interactions in the major hydrophobic core developing only after the TS (Budyak*, et al.*, 2013). Barrel closure of CRABP 1 involves partner interactions between strand 10 and strand 1. It is speculated that the early docking and specific interactions between these terminal strands help simultaneously assemble the front and back sheets of the barrel assisting the rapid folding of CRABP 1 (Budyak*, et al.*, 2013).

Consistent with the results on the folding studies, strand 10 was invariably predicted to be an aggregation hotspot. Experimentally, we have also resolved these regions of the proteins to be sequestered in the cores of aggregates. Strand 10 is a typical edge strand, an interface that can transiently be exposed leading to aggregation

(Richardson & Richardson, 2002, Soldi*, et al.*, 2008, Chiti & Dobson, 2009). Early barrel closure in the folding mechanism of CRABP 1 would protect strand 10 by providing a set of partner interactions, thus mitigating its vulnerability as an unpartnered edge strand (Budyak*, et al.*, 2013). The aggregation core 2, involving strand 9 is also intrinsically labile (Krishnan*, et al.*, 2000, Xiao & Kaltashov, 2005) also becomes protected upon closure of the barrel. Thus, taken together, it is hypothesized that protection of aggregation-prone regions in CRABP 1 by structural features and barrel closure occurs early in folding and significantly reduces its risk of aggregation. Our findings on a common sequence motif in CRABP 1 that initiate folding and aggregation demonstrate a clear connection between the two processes involving very distinct interactions. It is argued in this case that folding involves protective mechanisms that may have evolved to prevent aggregation. However, we also show that there is a very delicate balance that can be shifted in either direction. The following chapter will demonstrate how despite several protective mechanisms, local fluctuations in the native state can trigger aggregation.

2.4. Experimental Procedures

Cell partitioning experiments

Variants of CRABP 1 were overexpressed until OD_{600}=0.8 and protein expression were induced for three hours at 37^{o}C. Cells were lysed using bacterial protein extraction reagent BPER II (Thermo Scientific) following manufacturer's protocol. Lysates were spun down at 20,000 for five minutes. After centrifugation, soluble and pellet fractions were separated. Pellet fractions were dissolved 8M urea using equal volumes with the soluble fractions. Soluble fraction and dissolved pellet were mixed SDS-loading dye and boiled for five minutes then run on 12% Tricine SDS-PAGE. Protein bands were stained with Coomassie blue and observed using a GelDoc system (Biorad). Distributions of

CRABP 1 mutant proteins in each fraction were determined by taking the ratio of the optical density of the band corresponding to CRABP 1 in both pellet and soluble fractions with respect to the lysate fraction.

Expression and purification of proteins from inclusion bodies

CRABP 1 mutants were generated using a stabilized variant of murine CRABP 1 (WT* CRABP 1) (Clark*, et al.*, 1998) and N-terminal $(His)_{10}$-tag as background template incorporated into a PET16b vector. Single residue substitutions were introduced into the WT* CRABP 1 sequence by site-directed mutagenesis using a QuikChange protocol (Stratagene). The WT* and mutant proteins were expressed in BL21(DE3) *E. coli* strain (Novagen) either in Luria-Bertani media (non-isotopically labeled proteins) or M9 minimum media containing $^{15}NH_4Cl$ and/or ^{13}C-glucose (for isotopically labeled proteins). *E. coli* cells containing plasmids for WT* CRABP 1 were grown until OD_{600}=0.7-0.8 and protein expression was induced using 0.4mM IPTG for four hours at 30^{o}C. To produce CRABP 1 mutants as inclusion bodies, cells were grown at OD_{600}=0.8-1.0 and overexpressed at 37^{o}C for four hours for non-isotopically labeled proteins and six hours for isotopically labeled proteins. Cells were resuspended in lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl) and disrupted using Microfluidizer M-110L processor (Microfluidics). To purify soluble WT* CRABP 1, we collected the supernatant and ran on a Ni-NTA agarose affinity column (Qiagen) at over an increasing imidazole gradient. Fractions containing pure proteins were dialyzed in 10 mM ammonium bicarbonate and lyophilized overnight. For proteins isolated from inclusion bodies, pellet fractions were collected and resuspended in lysis buffer containing 8M urea. Protein solutions were eluted on a Ni-NTA affinity chromatography column at room temperature using buffers containing 8M urea with increasing imidazole concentrations. Absorbance at 280 nm

was taken to determine the concentration of purified proteins and using the extinction coefficient 21,294 M^{-1} cm^{-1}.

Preparation of *in vitro* aggregates

To generate in vitro aggregates, fractions containing purified F71A CRABP 1 mutant in urea were dialyzed in 10 mM phosphate buffer pH 7.0 containing 150 mM NaCl and 5 mM DTT at 37^{o}C. Aggregates were collected from the dialysate by centrifugation. To determine the concentration of proteins in aggregates, aggregates were dissolved in 8M urea and concentrations were determined by taking the absorbance at 280nm.

Purification of bacterial inclusion bodies

Inclusion bodies for these aggregation-prone mutant proteins were purified using a non-ionic detergent, BPER II combined with lysozyme treatment. Purification of inclusion bodies was performed following manufacturer's instructions with slight variations. The following procedures to purify inclusion bodies were performed in small aliquots (15mL cell culture) for efficient separation of contaminants. Cells were collected after centrifugation at 4,000 rpm for 10 min and resuspended in BPER II bacterial protein extraction reagent (Thermo Scientific) at 1:10 (BPER II: bacterial growth culture) ratio. The cell suspension was vortexed and centrifuged to collect the pellet. The pellet was resuspended in the same volume of BPER II and treated with lysozyme (0.4mg/mL). Pellet was washed with twenty-fold diluted BPER reagent. Wash steps were repeated trice. Pellet was washed with 10 mM Tris-HCl pH 8.0 and then with distilled deionized water. To check the purity of the inclusion bodies, these were resuspended in 8M urea and ran on a 12% SDS-PAGE. In order to check the concentration of CRABP 1 mutants in the aggregates, inclusion bodies were ran on an SDS-PAGE along with protein

standards of known concentrations. Band intensities of standard proteins were taken to generate a standard curve and from this, the concentrations of inclusion bodies were determined.

Hydrogen exchange NMR of aggregates

We identified the aggregation core residues of CRABP 1 inclusion bodies using hydrogen-deuterium (H/D) exchange and monitored by solution NMR spectroscopy as described previously (Hoshino*, et al.*, 2002). Briefly, $^{15}N/^{13}C$-labeled CRABP 1 inclusion bodies of F71A, L118V and I52A CRABP 1 mutants were incubated in D_2O for four weeks at 4 °C. Aggregates were collected, lyophilized and then resuspended in d_6-DMSO containing 0.1% (*v*/*v*) TFA, 50 mM DTT and 5% (*v*/*v*) D_2O (pD 3.0-3.5) to a final protein concentration of at least 200 μM. The solution was immediately transferred to an NMR tube, and the heteronuclear single quantum coherence (HSQC) spectrum was recorded at 26 °C on a 600-MHz Bruker Avance spectrometer using a TXI cryoprobe. For unexchanged samples, inclusion bodies were resuspended in water containing 0.025% (*w*/*v*) NaN_3 and incubated at 4°C for four week. NMR data were processed using NMRPipe (Delaglio*, et al.*, 1995). Backbone assignments of CRABP 1 aggregates dissolved in DMSO were obtained using a standard set of triple resonance experiments, including HNCACB, CBCA(CO)NH, HBHANH, HNCO, and HNCACO.

Sparse 3D HNCO experiments were performed on DMSO dissolved inclusion bodies initially incubated in D_2O at different time points, using a 700-MHz NMR Varian spectrometer. Relative peak intensities as a function of H/D-exchange time, $I(exch)/I_o \times 100\%$, where I(exch) and I_o are peak intensities for assigned residues were determined.

CHAPTER 3

PROTEIN AGGREGATION: BALANCE BETWEEN INHERENT DYNAMICS AND FUNCTIONAL INTERACTIONS OF A β-PROTEIN

This chapter describes results of collaborations with Anastasia Zhuravleva, Ivan Budyak and Lila Gierasch

3.1. Introduction

In order to function proteins have evolved to correctly self-assemble into their three-dimensional native fold. In a protein's energy landscape it is inevitable that the multiple intermediate species populated prior to the native state may tip the balance between folding and aggregation (Brockwell & Radford, 2007, Jahn & Radford, 2008). In addition to this complexity, local fluctuations driving partial exposure of hydrophobic stretches in the native state may precede aggregation (Bemporad*, et al.*, 2012, Neudecker*, et al.*, 2012, Sabate*, et al.*, 2012). Recently, it has been proposed that functional interactions rather than folding primarily compete with aggregation (Masino*, et al.*, 2011, Pastore & Temussi, 2012). Over the past years, mechanisms of aggregation have been studied on a few small globular proteins but no clear link between folding, function and aggregation has been elucidated (Booth*, et al.*, 1997, Villanueva*, et al.*, 2004, Yamaguchi*, et al.*, 2004, Hamada*, et al.*, 2009, Pagano*, et al.*, 2010). It remains very confounding how a protein is fine-tuned to favor protein stability and necessity to function over aggregation. In this chapter, we interrogate the following: How does a natively folded protein become susceptible to aggregation? How do functional interactions counteract the risk of aggregation? What intrinsic and extrinsic protective mechanisms help attenuate aggregation?

These questions were addressed using our model protein CRABP 1 to examine the complex balance between folding, aggregation and functional interactions. As discussed in Chapter 2, CRABP 1 possesses a robust folding mechanism that protects it from aggregation. However, the remaining question is how despite protective mechanisms, CRABP 1 is drawn to aggregate. WT* CRABP 1 under native-like conditions *in vitro*, aggregates to a minimal extent. However, this propensity is further aggravated by addition of small amounts of denaturant or introduction of single residue substitutions that do not affect the protein's overall stability. In this chapter an aggregation-prone intermediate populated under near native conditions was identified. Using several biophysical tests, we show that the inherent "breathing" of CRABP 1 exposes the highly aggregation-prone segments identified in Chapter 2. This dynamic motion is associated with the innate movement of the helical domain away from the β-barrel body and the loosening of interactions in the ligand portal region. However, previous studies show that interaction with the ligand halts this motion and stabilizes the complex (Krishnan*, et al.*, 2000, Xiao & Kaltashov, 2005). Here we will demonstrate inhibition of aggregation by retinoic acid suggests the role of ligand interactions in offsetting aggregation. Our findings on the aggregation of CRABP 1 demonstrate the complex intimate overlap between folding, functional roles and aggregation of proteins.

3.1.1. Native State Dynamics and Aggregation Propensity

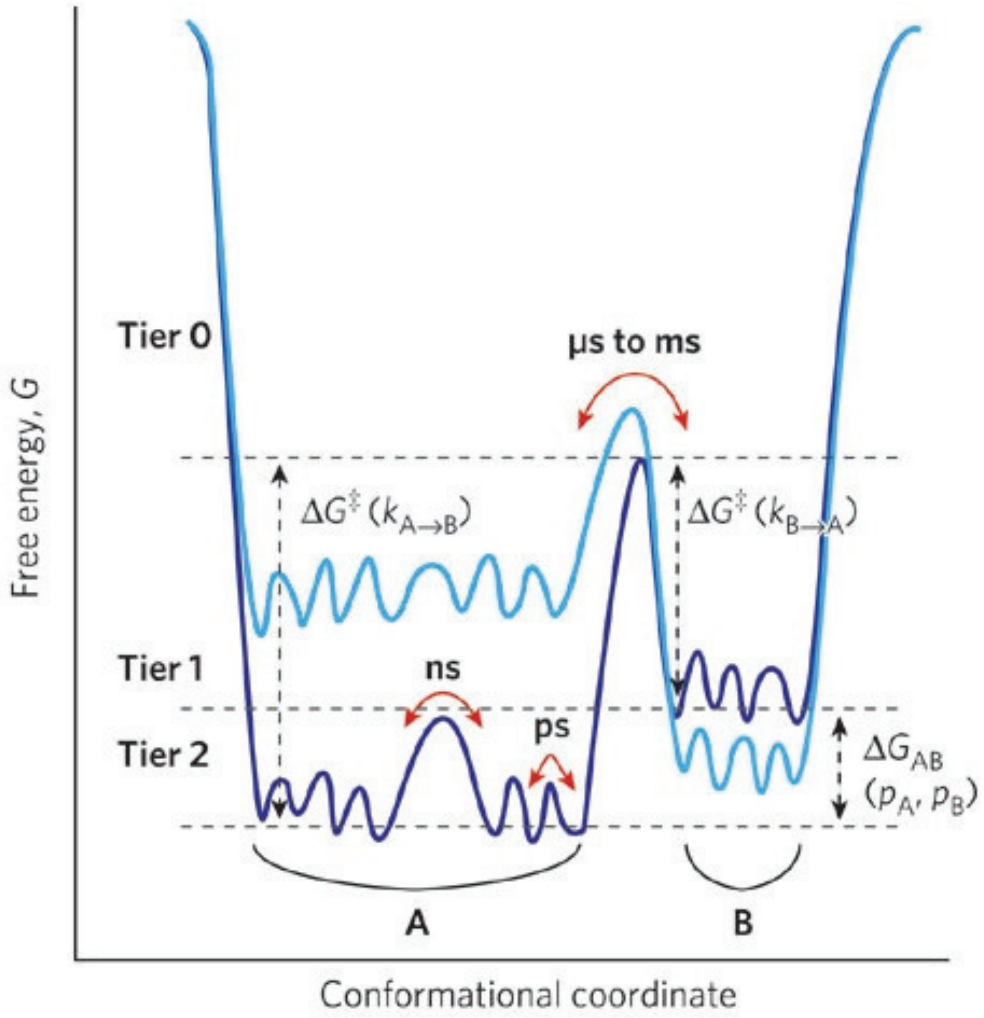
In this section, we will recollect how the behavior of native proteins are currently viewed and from here highlight circumstances that endanger native proteins towards aggregation. By looking at crystal structures one might picture folded proteins as having rigidly fixed conformations. But in reality, protein fluctuations around the native structure may be attributed to its function to either form interactions with other proteins or binding specific ligands. Descriptions of how proteins form specific interactions have developed

through the years. The classical mechanism of a “lock and key” model initially proposed by Emil Fischer where an enzyme interacts with its substrate via exact geometric complementarity has evolved into a more mechanistically dynamic interaction referred to as the “induced fit” model (Koshland, 1958, Csermely*, et al.*, 2010). In the induced fit model, binding events between a protein and a rigid ligand promotes conformational changes (Koshland, 1958). However, more recently it has been proposed that the native state is not only one distinct conformation but an ensemble of states for which protein binding follows a “conformational selection” process. This paradigm suggests that proteins fluctuate forming a conformational ensemble and a ligand selects one form compatible to binding and shifts the equilibrium towards this state (Csermely*, et al.*, 2010). This is described as “conformational selection and population shift phenomena”. In Chapter 1 native state ensembles have been described through multiple shallow wells in the energy landscape. The heights of energy barriers associated with protein motions can be described based on their timescales (Henzler-Wildman & Kern, 2007) (Figure 3.1). Fast protein dynamics involves local flexibility through bond vibrations, methyl rotations, loop motions and side-chain rotamers, which are in femtoseconds to nanoseconds timescales (Henzler-Wildman & Kern, 2007). Slow protein dynamics are in the microseconds to millisecond timescales demonstrated by large domain movements and collective motions (Henzler-Wildman & Kern, 2007). The rapid fluctuations have very low energy barriers compared to slower motions (Henzler-Wildman & Kern, 2007). Of the two types of dynamics, ensembles in slow timescales are more relevant to protein functions.

Selection for functional conformations of proteins have already been described experimentally (Csermely*, et al.*, 2010). Nevertheless, although the initial steps for ligand recognition of conformational selection and induced fit paradigms are different, the

distinction is not very clear since several interactions appear to require both. The model was then further refined into an “extended conformational selection” process (Wlodarski & Zagrovic, 2009). This extension describes that the binding process not only changes the energies of the interacting partners but also alters the shape of the entire energy landscape (Csermely*, et al.*, 2010). This was clearly illustrated in an NMR study on structural ensembles of ubiquitin during binding revealing equal contributions from both conformational selection and conformational rearrangements (Wlodarski & Zagrovic, 2009). The ability of proteins to form interactions by sampling several conformations implies the possibility of visiting a high-energy poorly populated state prone to form undesired self-interaction (Neudecker*, et al.*, 2012). In the following section, we will show a few examples of proteins that demonstrate the progressive view of an interconnection between function and aggregation.

Figure 3.1. One-dimensional cross-section through the high-dimensional energy landscape of a protein showing the hierarchy of protein dynamics and the energy barriers. $\Delta G^{\pm}$ corresponds to the free energy barrier between the populated states. Tiers are divided according to the timescales of protein motions. Tier 0 represents slow dynamics associated with large domain motions important for protein functions.


Free energy, G
Tier 0
μs to ms
$\Delta G^{\ddagger}(k_{A\rightarrow B})$
$\Delta G^{\ddagger}(k_{B\rightarrow A})$
Tier 1
ns
ps
Tier 2
ΔG_{AB}
(p_A, p_B)
A
B
Conformational coordinate

3.1.2. Linkage between Functional Interactions and Aggregation

In Chapter 1 we have mentioned several proteins that aggregate under physiological conditions suggesting a connection between functional dynamics and aggregation. From the view of conformational selection, aggregation may be an outcome of high energy less populated states of the native state ensemble (Csermely*, et al.*, 2010, Neudecker*, et al.*, 2012). A systematic study conducted on protein-protein complexes from structures deposited in the Protein Data Bank (PDB) showed evidence on shared physicochemical properties between functional interfaces and aggregation-prone regions (Pechmann*, et al.*, 2009). In this study, the interface regions correlated with aggregation hotspots projected on protein surfaces (surface aggregation propensity) of homodimers, heterodimers and homotrimers (Pechmann*, et al.*, 2009). In all cases, hydrophobic interactions were the best indicator for functional and aberrant protein-protein associations (Pechmann*, et al.*, 2009). There have been a small number of studies directly showing functional interfaces in proteins forming aggregates. We will briefly discuss these few examples, which include the cases of ataxin-3, p53, β-lactoglobulin and SUMO proteins.

3.1.2.1. Ataxin-3

Ataxin-3 (atx3) is a conserved and ubiquitously expressed enzyme whose function is not yet clearly understood (Matos*, et al.*, 2011). However, there have been several evidence pointing to its important role in protein turnover by the Ubiquitin-Proteasome Pathway because of its ability to recognize polyubiquitin chains and polyUb-editing proteolytic activity. Atx3 interacts with a number of proteins involved in protein homeostasis maintenance, transcription regulation, cytoskeleton organization and myogenesis. The structure of ataxin-3 is composed of a structured globular N-terminal domain referred to as Josephin domain (JD), which is responsible for the protease

activity followed by a flexible C-terminal tail consisting of ubiquitin binding motifs and a polyglutamine (polyQ) region. Increased length of the polyQ expansion has been associated with neurodegeneration called Machado-Joseph disease (MJD), the most common form of spinocerebellar ataxia. Atx3 presents as inclusions in neurons and axons in brains of patients suffering from MJD. In addition, quality control proteins including ubiquitin and molecular chaperones co-aggregate with atx3 in diseased brains. It has been shown that in the absence of the polyQ tail, atx3 aggregates under native conditions (Gales*, et al.*, 2005). It is shown that the folded Josephin domain modulates the stability and regulates self-association (Masino*, et al.*, 2004). Under these native conditions, Josephin domain nucleates atx3 fibrillation suggesting it role in the early stages of aggregation (Masino*, et al.*, 2004, Gales*, et al.*, 2005). A detailed study to identify aggregation-prone regions of Josephin domain uncovered two solvent exposed patches (Masino*, et al.*, 2011). These solvent exposed interfaces have been identified to be important in ubiquitin binding (Masino, Nicastro et al. 2011). To further demonstrate competition between self-association and functional interactions, they showed that addition of ubiquitin slowed down aggregation and changed the morphology of aggregates (Masino, Nicastro et al. 2011).

3.1.2.2. p53 protein

p53 is a tumor suppressor protein that is central to an extremely complicated cellular signaling network. p53 mediates a wide range of responses by recognizing signals for oncogene activation, DNA damage and oxidative stress and to initiate repair mechanisms, cell cycle arrest or apoptosis (Aylon & Oren, 2007). Approximately half of known cancers have mutations in the p53 gene. Majority of these cancers have impairment in the positive and negative transcriptional regulation of function of p53. p53 is active as a tetramer. The N-terminus consists of disordered transactivation domain

(TAD) and a proline-rich region (Joerger & Fersht, 2008). The central region consists of a folded DNA-binding core domain linked to a tetramerization domain (Joerger & Fersht, 2008). At the C-terminus of p53 is a disordered basic regulatory domain which binds DNA non-specifically (Joerger & Fersht, 2008). Post-transcriptional modification of the p53 by phosphorylation, acetylation, methylation and modifications with ubiquitin-like proteins affects its function and stability (Vousden & Lane, 2007). Aside from the effects of mutations in p53 for its multiple functions, p53 is also shown to aggregate. Large aggregates of wild-type p53 protein have been observed in the cytoplasm and nucleus of neuroblastomas, carcinomas and myelomas (Moll*, et al.*, 1996, Ostermeyer*, et al.*, 1996). Succeeding experiments to reveal the nature of the aggregates revealed the tendency of its functional domains to self-associate. The central DNA binding domain, the tetramerization domain and the disordered TAD have all been found to be involved in the cores of aggregates (Ishimaru*, et al.*, 2003, Higashimoto*, et al.*, 2006, Rigacci*, et al.*, 2008). It has been reported that kinetic partitioning between folded and misfolded proteins lead to a decrease in DNA binding for p53. Consistent with a common DNA binding interface and aggregation cores, interaction of p53 with a consensus DNA sequence inhibits aggregation.

3.1.2.3. β-lactoglobulin

β-lactoglobulin is a non-obligate homodimeric, β-barrel protein consisting of eight anti-parallel strands (A-H) with an extra stand, I, which is involved in dimerization and a helix at the C-terminal end. At low pH, the dimer dissociates into folded monomers (Uhrinova*, et al.*, 2000, Hamada & Dobson, 2002). β-lactoglobulin is a member of the lipocalin family and has a large structural plasticity (Gutierrez-Magdaleno*, et al.*, 2013). Its large hydrophobic cavity enables it to bind to diverse hydrophobic ligands (Gutierrez-Magdaleno*, et al.*, 2013). Based on calorimetric studies, ligand binding promotes dimer

dissociation (Gutierrez-Magdaleno*, et al.*, 2013). The protein also contains two disulfide bonds which links strands D and I and G and H. Under denaturing conditions, β-lactoglobulin forms aggregates but only after prolonged incubation. However, aggregation can be triggered by reduction of disulfide bonds. Under oxidizing environments, strand I is stabilized by disulfide bonds and favors dimerization. However, under reducing conditions, strand I frays away from the β-barrel and leading to aggregation (Hamada*, et al.*, 2009). β-lactoglobulin exhibits another case where functional dimerization interfaces coincide with aggregation-prone regions. In addition, the ability of ligand binding to compete with dimerization may also provide some hint on the possible role of natural ligands in preventing aggregation.

3.1.2.4. SUMO Proteins

Small ubiquitin-related modifier (SUMO) proteins are highly conserved proteins covalently bound to target proteins. Unlike ubiquitin, SUMO proteins do not target proteins to degradation rather they modulate protein-protein interactions. SUMO post-translationally attaches to proteins via a isopeptide bond and participate in transcriptional regulation, nuclear transport, maintenance of genome integrity, and signal transduction (Johnson, 2004). The downstream effect of sumoylation varies for every substrate. SUMO can either promote complex formation of substrates but can also prevent interactions such as blocking ubiquitination. Similar to ubiquitin, SUMO proteins have a ββαββαβ fold. The two C-terminal Gly residues required for isopeptide bond formation is conserved between ubiquitin and SUMO. SUMO-1 contains a long and flexible N terminus, which projecting from the core of the protein not found in ubiquitin. There are four isoforms of SUMO (1 to 4). SUMO 2 and 3 are highly homologous. SUMO conjugation has been found to increase the solubility of proteins leading to exploration of its possible role in regulating aggregation-prone proteins (Krumova*, et al.*, 2011). In

addition, SUMO proteins have been popularly used as solubility tags for recombinant protein expression (Wang*, et al.*, 2010, Peroutka Iii*, et al.*, 2011). However, despite the very high stability of SUMO proteins, slight perturbation of this protein leads to aggregation. Detailed analysis of the aggregate structure of SUMO proteins reveals that strand 2, is both involved in the binding interface with its substrates as well as in the aggregation cores (Sabate*, et al.*, 2012). Furthermore regions of SUMO that have high aggregation propensities are also found in its functional interface (Sabate*, et al.*, 2012). Aggregation of SUMO provides direct evidence that the need for productive interactions also provides risks towards dysfunctional interactions.

3.1.3. Protective Intrinsic Mechanisms From Aggregation

Although surfaces important for functional and dysfunctional interactions are the same, proteins have been designed to prevent aggregation. Proteins have evolved utilizing several intrinsic strategies to disfavor aggregation. Even during aggregation at physiological conditions, the nature of energy barriers in the native state prevents the population of aggregation-prone species (De Simone*, et al.*, 2011). It is through these barriers that are dictated by the amino acid sequence by which proteins remain soluble *in vitro* or in cells (De Simone*, et al.*, 2011). Here we will discuss concrete specific interactions that create energy barriers to disfavor proteins form populating the aggregation-prone state. These include presence of charged residues close to interfaces and/or specific interactions involving disulfide linkages and salt bridges (Pechmann*, et al.*, 2009).

3.1.3.1. Charges and Aggregation Gatekeepers in Preventing Aggregation

Charged residues specify binding by either opposing or promoting interactions thus playing a role in favoring functional protein-protein interactions over aggregation.

(Vendruscolo & Dobson, 2007, Zhang*, et al.*, 2011). Analysis of regions around protein-protein interfaces revealed charged residues in proximity to these hydrophobic patches (Pechmann*, et al.*, 2009). Interface rims formed by the residues that bury <25 $Å^2$ upon binding are mostly composed of charged amino acids (Pechmann*, et al.*, 2009). The importance of charges on the solubility and binding of proteins has been emphasized in a study of a charge-depleted S6 ribosomal protein (Kurnik*, et al.*, 2012). The absence of charges did not affect native interactions but triggered aggregation (Kurnik*, et al.*, 2012). The presence of charged amino acids commonly found close to aggregation-prone regions also referred as aggregation "gatekeepers" inhibit aggregation (Rousseau*, et al.*, 2006). The existence of these structural gatekeeper residues flanking highly aggregating segments suggests selective pressure against aggregation (Rousseau*, et al.*, 2006, Beerten*, et al.*, 2012). Obstruction of interactions between aggregation-prone sequences is achieved in several ways: 1) repulsive effect of the charged side chains (arginine, lysine, aspartate, glutamate; 2) the entropic penalty on aggregate formation by large side chains such as arginine and lysine and 3) incompatibility of residues with the β-aggregate structure as in the case of proline residues (Reumers*, et al.*, 2009). It has been reported that mutation of gatekeeper residues result in a number of misfolding diseases and other aggregation unrelated diseases such as van der Woude syndrome (VWS), Fabry disease (FD), and limb-girdle muscular dystrophy (Reumers*, et al.*, 2009). In addition, aggregation-prone segments of critical proteins are the most gate-kept (Reumers*, et al.*, 2009). Intrinsically disordered proteins (IDPs) have also been shown to flank aggregation-prone sequences (Abeln & Frenkel, 2008). IDPs are natively unfolded proteins consisting of only few hydrophobic residues, more hydrophilic and charged amino acids and plenty of repeats in their sequence (Rezaei-Ghaleh*, et al.*, 2012). IDPs have been proposed obstruct or sterically hinder interactions between aggregation-prone

regions (Abeln & Frenkel, 2008). In addition to their safeguarding properties electrostatic patches also facilitate chaperone interactions for proper folding (Rousseau*, et al.*, 2006, Lawrence*, et al.*, 2007, Kurnik*, et al.*, 2012).

3.1.3.2. Specific Interactions and Structural Adaptations to Prevent Aggregation

Largely hydrophobic regions in globular proteins are deeply buried in the core of proteins to prevent aggregation. Specific interactions ensure that exposure of these segments are avoided. Disulfide bonds and salt bridges stabilizing native interactions are preferentially found in proximity to interfaces in protein-protein complexes (Pechmann, Levy et al. 2009). In the case of β-lactoglobulin, reduction of disulfide bonds lead to aggregation (Hamada*, et al.*, 2009). In β-lactoglobulin, the strand involved in dimerization is an edge strand. An edge strand is defined as a β-strand found at the end of β-sheets or β-sandwiches, that are inherently aggregation-prone by readily forming intermolecular H-bonding with a neighboring β-strand (Richardson & Richardson, 2002). Edge strands are commonly exposed in proteins that natively form β hydrogen bonds to form dimers or oligomers strands (Richardson & Richardson, 2002). In β-lactoglobulin, the edge strand is structurally restrained by the presence of disulfide bonds (Hamada, Tanaka et al. 2009). In a survey by Richardson and Richardson in 2002 of β-rich proteins they identified several “negative designs” or structural adaptation strategies to prevent aggregation. A number of specific interactions are employed by β-proteins to screen unwanted edge strand interactions. β-barrels prevent edges through continuous H-bonding around the barrel cylinder. Parallel β-helices protect β-sheet ends by interaction with flexible loops. β-propeller, β-sandwich and single-sheet proteins combine β-bulges, prolines, addition of charged side chains in the middle of the strand, shortened edge strands and coverage by loop structures to protect edge strands (Richardson &

Richardson, 2002). Free edges also contain bends and twists as protective mechanisms strands (Richardson & Richardson, 2002).

3.1.3.3. Cellular Strategies to Favor Functional Interactions and Folding over Aggregation

Although there have been several examples showing that protein aggregation can be beneficial, it is generally a deleterious outcome in cells. However despite the negative implications of aggregation, evolution has not successfully eliminated aggregation-prone segments in proteins. This can be attributed to their important roles in forming compact folded structures as well as in facilitating functional interactions. As discussed in the previous section, proteins bear gatekeeping sequences and form specific interactions to disfavor aggregation. However, these protection mechanisms may be impeded in the cell. During synthesis of proteins, the nascent chains may have not yet formed the necessary interactions to protect aggregation-prone sequences. Moreover, in addition to the natural motions of native proteins that expose aggregation hotspots the extremely crowded cellular environment makes it highly desirable to form aggregates. Thus, evolution has accounted several ways to prevent these dreadful outcomes. The most obvious cellular strategy is the existence of robust quality control machineries that either assist proper folding or eliminate aberrant proteins. The complex network of molecular chaperones consists of facilitators in *de novo* folding, refolding of stress-denatured proteins, disaggregation, oligomeric assembly, protein trafficking and assistance in proteolytic degradation (Hartl*, et al.*, 2011). Interplay between molecular chaperones and proteases are crucial in maintaining a functional proteome (Powers*, et al.*, 2012). However, since protein production can overwhelm the proteostasis network, protein expression is also regulated. There is a very strong anti-correlation between *in vivo* expression levels and rates of aggregation in proteins (Tartaglia*, et al.*, 2007). In

addition, since aggregation is concentration dependent, abundant proteins were found to be highly soluble in cells (Castillo*, et al.*, 2011). However, these proteins although highly soluble can become susceptible to aggregation at high concentrations therefore these are tightly regulated at the gene-expression level (Castillo*, et al.*, 2011). Another protective mechanism is cellular organization. It is tempting to think that the cellular environment is a random mesh of non-specific interactions. In reality the cellular environment is highly organized and designed for specific protein-protein interactions (Pastore & Temussi, 2012). Furthermore, cellular machineries are spatially organized and compartmentalization allows efficient folding *in vivo* (Gershenson & Gierasch, 2011). In addition to this, proteins in cells are designed to self-organize into large multiprotein complexes to prevent exposure of sticky surfaces (Gierasch & Gershenson, 2009). There is also a high correlation between aggregation propensity and turnover rate of proteins (De Baets*, et al.*, 2011). Short-living proteins have been found to be more aggregation-prone and have less chaperone interactions than proteins that live longer (De Baets*, et al.*, 2011).

3.1.4. Functional Dynamics of CRABP 1

The functional dynamics and interactions with the ligand of members of ILBP family including CRABP 1, our model protein is well characterized. Here, I will review what is known about the shared dynamic behavior of this family and CRABP 1 to clearly demonstrate further on their structural susceptibility of towards aggregation.

3.1.4.1. Function of CRABP 1

The intracellular lipid binding protein family is involved in sequestering and transport of fatty acids in the cytoplasm. These proteins have been shown to be important cellular processes such as signal transduction, gene expression, growth and

differentiation (Zimmerman & Veerkamp, 2002). Cellular retinoic acid binding proteins in particular have two isoforms, CRABP 1 and CRABP 2. CRABPs bind to all-trans retinoic acids, to regulate gene transcription, cell proliferation and cell di*ff*erentiation. CRABP 1 and CRABP 2 are highly homologous with 74% sequence identity and are extremely conserved among species (Noy, 2000, Marcelino*, et al.*, 2006). Despite the high similarity CRABPs have different patterns of expression across cell types and at various stages of development suggesting functional diversity (Noy, 2000). Both isoforms are found in the cytoplasm while CRABP 2 is also localized in the nucleus. CRABP 1 and CRABP 2 similarly bind retinoic acid but have distinct modes of transport (Noy, 2000, Stachurska*, et al.*, 2011). CRABP 1 is ubiquitous and is involved in enhancing retinoic acid catabolism and buffering retinoic acid concentrations in cells (Stachurska*, et al.*, 2011). CRABP 1 modulates catalysis of retinoic acid metabolism by associating with cytochromes P450 and discriminates trans-retinoic acid from cis-retinoic acid (Fiorella & Napoli, 1991, Noy, 2000). However the exact mechanism of how CRABP 1 regulates retinoic acid in cells still remains unclear. On the other hand, CRABP 2 activates nuclear retinoic acid receptors (RAR) and regulates retinoic-acid-dependent genes (Stachurska*, et al.*, 2011). The affinity of CRABP 1 (Kd<0.4nM) for retinoic acid is higher compared to CRABP 2 (Kd=2nM) (Norris*, et al.*, 1994).

3.1.4.2. Dynamic Properties of CRABP 1 and Implications for Ligand Binding

Attributing to similar functions to transport large hydrophobic ligands ILBPs have strong structural homology across species, cell types and ligands for this protein family (Marcelino, Smock et al. 2006). The overall structure of ILBPs consists of an internal cavity filled with ordered water molecules and a helix-turn-helix cap. A small opening enclosed between the helical domain and turns 2 and 4 acts as a portal region of the protein. Based on several crystallographic studies on ILBPs, ligands are bound within

the barrel in a central internal water-filled cavity (Jamison*, et al.*, 1994, Donovan*, et al.*, 1995, Jamison*, et al.*, 1998, Zhu*, et al.*, 1999, Krishnan*, et al.*, 2000). The side chains of buried hydrophobic amino acids govern the specificity and volume of the internal cavity (Noy, 2000). In addition, water molecules in the cavity contribute to the protein's stability and proposed to assist in the displacement of ligands. The volume of cavities varies for members of the family and correlates with the size of its ligand. Common ligand-binding residues located at the helical motif, strands 2, 3, 4, 8, 9, 10 and turn 2 have been identified in most ILBPs (Marcelino*, et al.*, 2006). However the nature of the cavity in this family is not well conserved. It is proposed that ILBPs require favorable hydrophobic interactions to encapsulate the ligand but maintain large solvent-filled cavities to accommodate diverse lipids (Massolini & Calleri, 2003, Marcelino*, et al.*, 2006).

In particular the crystal structure of holo-CRABP 1 shows retinoic acid buried in the binding cavity where the carboxy group interacts with two arginine residues and one tyrosine residue at the end pocket. The β-ionone ring is twisted into a cis-like configuration relative to the isoprene tail, and fits snugly at the portal region leaving only one end accessible to the solvent (Kleywegt*, et al.*, 1994, Thompson*, et al.*, 1995). Thus in this closed ligand-bound conformation access to the entrance of the ligand-binding pocket of CRABP 1 is restricted, implying that significant structural changes are required to release the ligand.

It is currently recognized that the conformational flexibility of CRABP 1 like other members of the iLBP family is associated with its function to recognize and bind ligand. The highly flexible and conserved helical subdomain identifies the correct ligand to bind. It is proposed that electrostatic interactions at the edge of the portal attract ligand into the interior of the cavity (Zimmerman & Veerkamp, 2002). A conserved gap is found

between strands 4 and 5 with no inter-strand hydrogen bonds is instead occupied by ordered water molecules.

There have been three proposed mechanisms of ligand entry in iLBPs. One mechanism based on well-ordered and superimposable crystal structures of apo and holo forms of rat intestinal fatty acid binding protein, states that entry of the ligand through the small portal requires minimal changes in the backbone conformation (Sacchettini*, et al.*, 1992, Sacchettini & Gordon, 1993). It is hypothesized that entry of the ligand into the portal pushes the water molecules through the conserved gap similar to a "water pump" (Sacchettini*, et al.*, 1992). However this was challenged by another hypothesis where ligand binding was shown to promote significant backbone conformational changes. Limited proteolysis on cellular retinol binding proteins (CRBP I and CRBP II), CRABP 1 and heart FABP revealed differential proteolytic patterns between the apo and holo forms of the proteins (Jamison*, et al.*, 1994). Increased susceptibility for digestion at the entry site of the apo form compared to the holo-form suggests an open conformation in the absence of a ligand. This was also supported by prominent differences in crystal structures of apo and holo forms of CRABP 1, showing a ligand accessible form where turn II moves away from the β-barrel (Thompson*, et al.*, 1995). Upon ligand binding the protein is locked in a closed state (Thompson*, et al.*, 1995). In an NMR study the apo-form of CRABP 1 shows a more flexible C-terminus of helix II permitting increased access to the cavity (Rizo*, et al.*, 1994).

However, a refined and detailed mechanism based on NMR studies suggests an extended conformational selection process by which ligand binds. Hodsdon and Cistola in 1997 proposed that existing locally disordered and ordered states of the protein in equilibrium in the apo state. Ligand binding consequently stabilizes cooperative

interactions shifting the equilibrium towards the ordered conformation. This is also referred to as the 'dynamic portal hypothesis". Hudsdon and Cistola described four key features in the dynamic portal hypothesis. First, there are several locally disordered regions in the unbound. These include the C-terminal helix II, the linker between helix II and strand 2, turn II and turn IV. These regions are flexible and capable of large backbone structural fluctuations opening the portal region. This conformational change is a transition between order and disorder where helix II and turn II interactions dissociate and the C-terminal helix II is unraveled. Second, binding of ligand shifts the equilibrium into an ordered or closed state by stabilizing interactions in the C-termini of helices I and II. The C-terminus of helix I is stabilized by a Schellman motif interaction. On the other hand the C-terminus of helix II is held by ligand-induced interactions with turn II. Third, hydrogen bond interactions between a polar residue in turn II and helix II regulate the order-disorder transition, hence the entry and exit of the ligand as well. Lastly, ligands are released from the cavity by destabilization of turn II-helix II interactions to shift the equilibrium into the disordered state. This may be driven by collision of the holo-protein with target intracellular organelles. One proposal regarding the delivery of ligands into the membranes requires the portal region (helix II, turns II and IV) to dock into the membrane while the other side (strands 2, 3, 6, 7, turn 7 and the N-terminus) is exposed to the solvent (Mihajlovic & Lazaridis, 2007).

Conformational exchanges along the portal regions have been consistently observed in several ILBPs (Zhu*, et al.*, 1999, Krishnan*, et al.*, 2000, Lu*, et al.*, 2000, Xiao & Kaltashov, 2005, Li & Frieden, 2006). In addition, molecular dynamics simulations have revealed backbone flexibility in apo-forms of iLBPs (Woolf*, et al.*, 2000, Bakowies & Van Gunsteren, 2002, Levin*, et al.*, 2010). Examination of motions and flexibility of CRABP 1 showed consistent dynamic behavior at the portal region. Using rapid

hydrogen-amide exchange NMR, motions of the apo and holo-forms of CRABP 1 were examined (Krishnan*, et al.*, 2000). Significant distinct millisecond timescale motions were observed in the portal region of the apo-form of the protein consisting of the C-terminus of helix II, helix II-strand 2 linker, turn 2 and turn 4 supporting the dynamic portal hypothesis (Krishnan*, et al.*, 2000). In a separate experiment using slow hydrogen exchange mass spectrometry, transient structural disorder was observed in the same regions of the unbound protein (Xiao & Kaltashov, 2005). When bound to retinoic acid, these fast motions are impeded and the flexibility in the portal is significantly reduced (Krishnan*, et al.*, 2000, Xiao & Kaltashov, 2005). Retinoic acid also increased the conformational stability of strand 10, which contains residues that form two salt bridges with the carboxylate group of retinoic acid (Xiao & Kaltashov, 2005). Studies on CRABP 1, consistent with other ILBPs reveal the crucial role of localized dynamics along the portal region in dictating ligand recognition and subsequent conformational stabilization. In the next section we will show our results suggesting that despite several protective mechanisms against aggregation, this dynamic behavior predisposes the protein to aggregation.

3.2. Results

3.2.1. Aggregation Propensities and Themodynamic Stabilities are Coarsely Correlated

In Chapter 2 we have identified aggregation cores in CRABP 1, which were consistent under extremely different conditions and across mutants with different propensities to aggregate. Our first instinct was to check whether global unfolding exposes these segments of the protein favoring intermolecular contacts. To do this we looked for correlation between thermodynamic stabilities and aggregation propensities. In order to do this, we checked the solubilities of several CRABP 1 mutants by cell

fractionation after overexpression in *E. coli* with $\Delta\Delta G_{U\text{-}N}$ (energetic effect of mutations on the free energy of the unfolded state (U) with respect to the native state (N) = $\Delta G_{U\text{-}N}^{WT}$ - $\Delta G_{U\text{-}N}^{Mutant}$) values determined by previous members of the lab. While we have found high correlation between aggregation propensities and thermodynamic stabilities based on $\Delta\Delta G_{U\text{-}N}$ for weakly and highly aggregation-prone variants, the range of stabilities of moderately aggregating CRABP 1 variants were relatively broad (Table 1). This coarse correlation between stability and aggregation propensities may suggest other mechanisms other than global destabilization are involved. The comparable tendencies of moderately aggregating mutants to both fold and misfold may suggest that aggregation may emerge from a folded intermediate. These mutants are able to acquire the native fold based on their high solubility early on during overexpression or during expression at low temperatures. The natively folded mutants have been confirmed by CD spectroscopy (Appendix H) and intrinsic tryptophan fluorescence spectroscopy experiments (Appendix I).

3.2.2. Aggregation of CRABP 1 Proceeds Under Native Conditions

To test whether a folded intermediate precedes aggregation, we examined whether aggregation of CRABP 1 takes place under native conditions. We tested *in vitro* conditions that lead to aggregation of WT* CRABP 1. Addition of low urea concentrations or adjusting the pH below 8.0 leads increases aggregation propensity. Here we followed aggregation of CRABP 1 at pH 7.0 in the presence of low concentrations of urea ranging from 0.5 to 3M (Figure 3.2). We have found an increasing trend towards aggregation up to 2.5M urea. Above 3M, aggregation is no longer observed. At this range of urea concentrations, CRABP 1 has not yet globally unfolded as shown in the equilibrium unfolding of CRABP 1 monitored by intrinsic tryptophan

fluorescence (Figure 3.2). At higher urea concentrations (>4M), CRABP 1 begins to unfold globally, however we do not observe aggregation under these conditions. To observe the effects of low urea concentrations on the aggregation kinetics of CRABP 1, we monitored the linear light scattering of CRABP 1 with and without urea. We observe a faster rate of aggregation in the protein solution containing 1 M urea (Figure 3.3). Our results reveal the existence of a near-native intermediate populated at low urea concentrations. In addition, we show that this intermediate cannot be sensitively detected using intrinsic tryptophan fluorescence, which has been used to monitor unfolding of CRABP 1 (Clark*, et al.*, 1996, Clark*, et al.*, 1998). We also examined the pH dependence of CRABP 1 aggregation by monitoring the aggregation propensity of CRABP 1 at pH conditions below 8.0. We have found an increasing propensity to aggregate as pH is adjusted away from pH 8.0 (Figure 3.4)

Figure 3.2. Urea-dependent aggregation and unfolding of CRABP 1.

In red, aggregation of CRABP 1 plotted with urea concentration. Aggregation was performed by incubation of 200 uM WT* CRABP 1 in buffer (10 mM phosphate buffer, 5 mM DTT pH 7.0) with increasing urea concentrations for 12 hours at 37^{o}C. Aggregated proteins were quantified by measuring amount of remaining soluble protein after aggregation.

In blue, is a representative plot of fraction unfolded protein (blue) as a function of urea concentration. Fraction unfolded is measured at different urea concentrations using intrinsic tryptophan fluorescence spectroscopy.

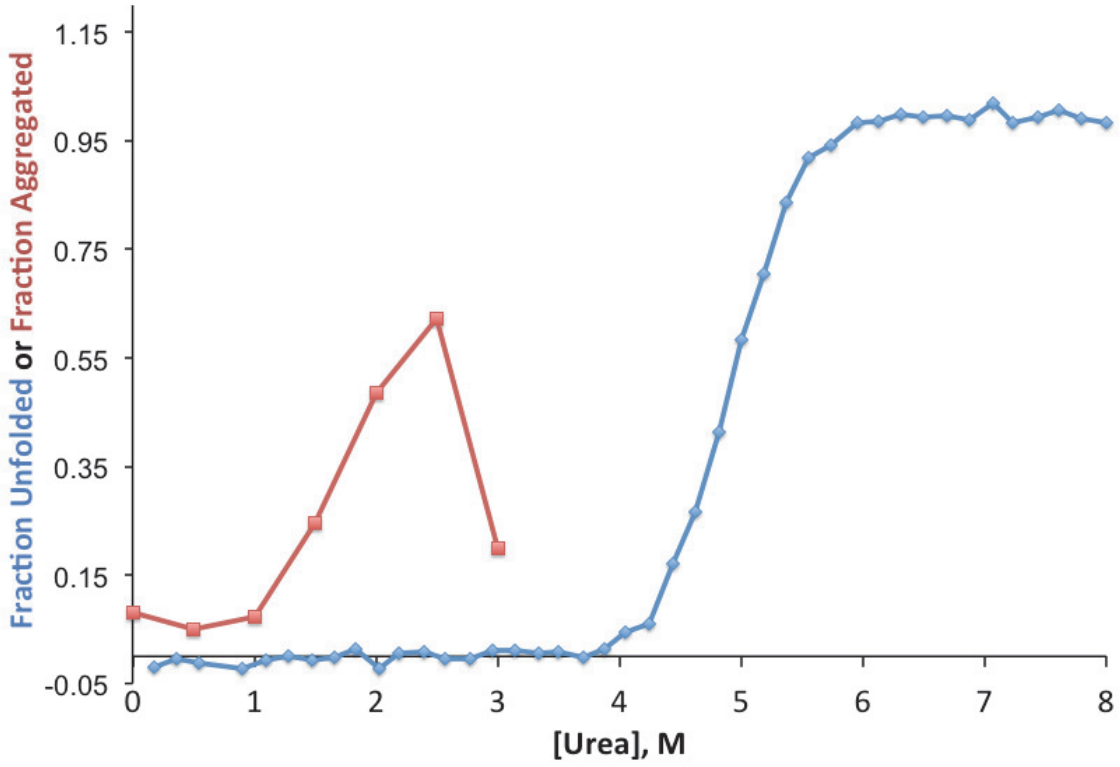

Fraction Unfolded or Fraction Aggregated
1.15
0.95
0.75
0.55
0.35
0.15
-0.05
0
1
2
3
4
5
6
7
8
[Urea], M

Figure 3.3. Aggregation kinetics of CRABP 1 with and without urea
Aggregation of 300 μM WT* CRABP 1 in 10 mM phosphate buffer, 5 mM DTT pH 7.0 monitored by linear light scattering without (filled circles) and with (open circles) 1.5M urea.

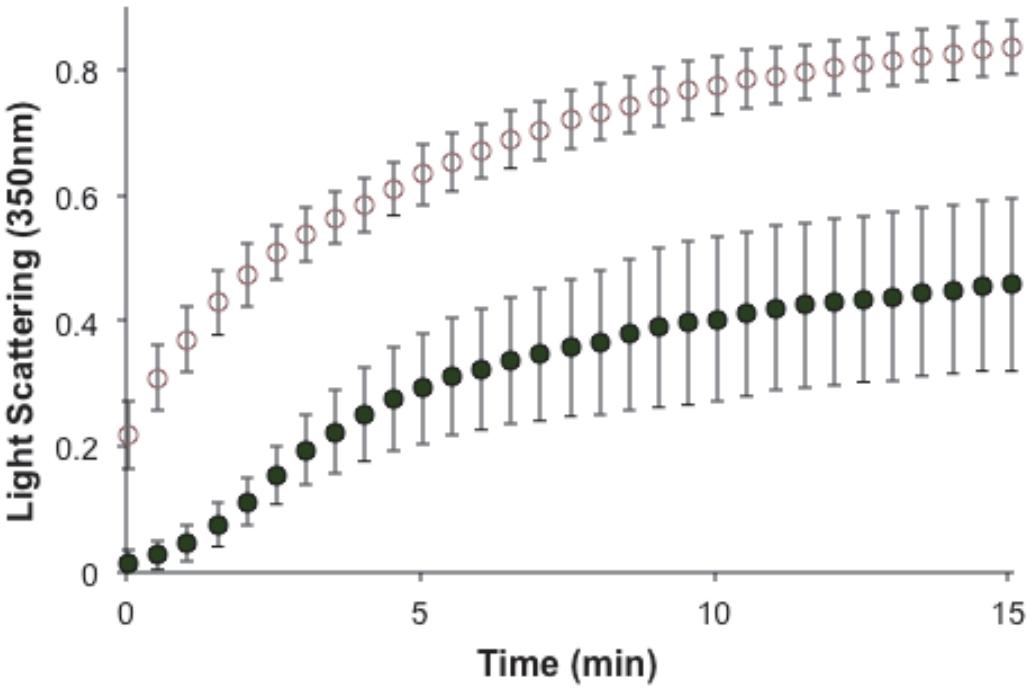
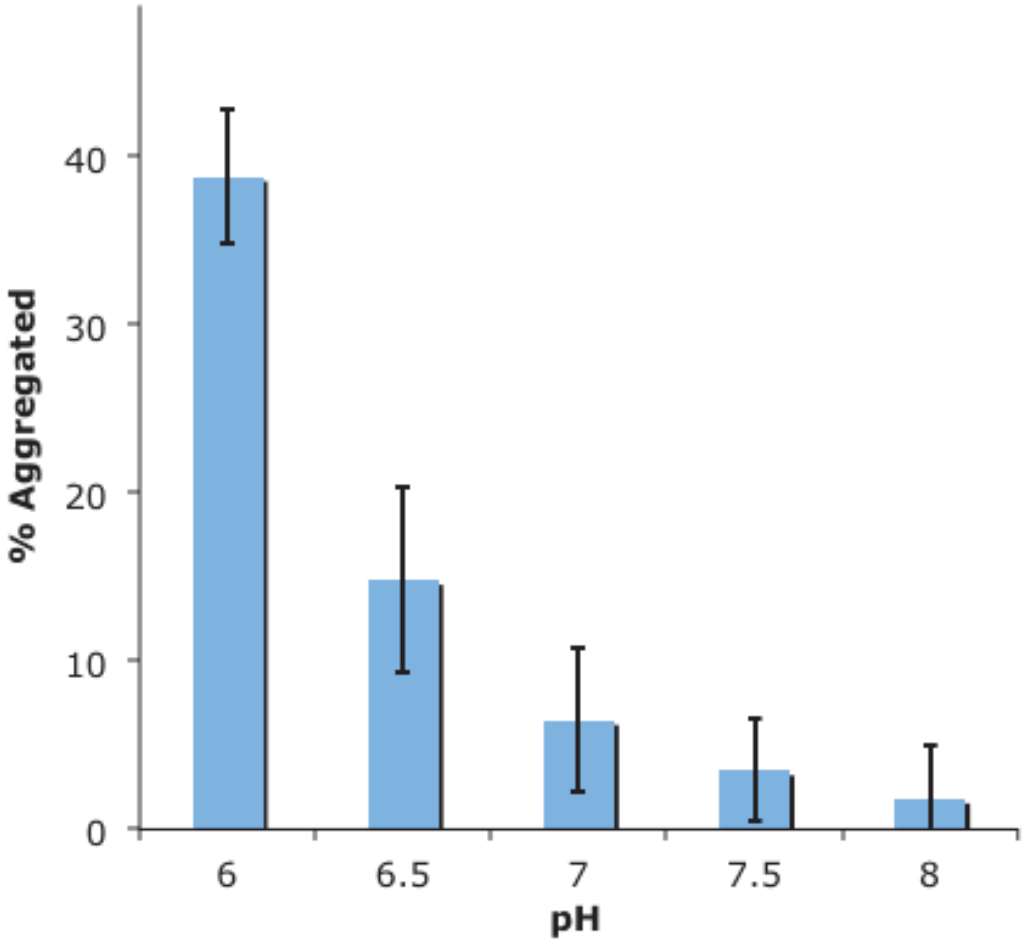

Light Scattering (350nm)
0
0.2
0.4
0.6
0.8
0
5
10
15
Time (min)

Figure 3.4. pH-dependence of CRABP 1 aggregation.
Aggregation reactions were performed on 200 μM WT* CRABP 1 at 37^{o}C at different pH conditions.


% Aggregated
40
30
20
10
0
6
6.5
7
7.5
8
pH

3.2.3. Near-Native Dynamic Intermediate of CRABP 1 Leads to Aggregation

To provide detailed structural information regarding the species responsible for the low urea dependent aggregation, we conducted equilibrium unfolding experiments and monitored chemical shift perturbations by NMR. By monitoring changes in chemical shifts of amino acid residues upon urea titration, we are able to identify regions of the proteins that are perturbed by urea. Consistent with our model we observed the population of a native-like intermediate prior to unfolding of CRABP 1. We describe our findings on the urea unfolding of CRABP 1 as a two-step process. The first step takes place at low urea concentrations manifesting a two-state transition, which is fast in the NMR timescale (us-ms timescale) (Figure 3.5). A quick glance at the overall spectra shows that only a few peaks are affected by urea. Most of the residual peaks at low urea concentrations did not deviate from the chemical shifts corresponding to the native protein (0M urea). We have looked closely at the few peaks perturbed by low urea and found a two state equilibrium exists under these conditions (Figure 3.6). This suggests a dynamic interconversion between states. This suggests that the intermediate being populated as urea concentration is increased resembles that of the native state. The interchange between the two species is fast as evidenced by single broadened peaks. Broadening of peaks means that the rate of exchange between is faster compared to the chemical shift differences of the two resonance lines. This suggests a dynamic interconversion exists between the two structurally similar species. We will refer to this intermediate as the N* state. The second step involves global unfolding depicted by the slow transition from the N* state into the unfolded protein (Figure 3.7). In the second step, as shown in concentrations above 3.5M, two species are in equilibrium, the N* state and the unfolded state (U). This can be observed in the HSQC spectrum of CRABP 1 in 4M urea where peaks corresponding to both the N* state and U states are visible (Figures 3.7 and 3.8). At this urea concentration, the interconversion is said to be slow

because the rates of exchange is smaller compared to the chemical shift differences allowing the two resonance lines to be resolved.

Figure 3.5. Overlay of ^{1}H-^{15}N HSQC spectra of CRABP 1 at different urea concentrations showing the first (fast) step of urea equilibrium unfolding of CRABP 1 (N→N*). Colors represent increasing urea concentrations from 0.5 M (yellow) to 3.5 M (red). Unperturbed chemical shifts in the absence of urea, in black.

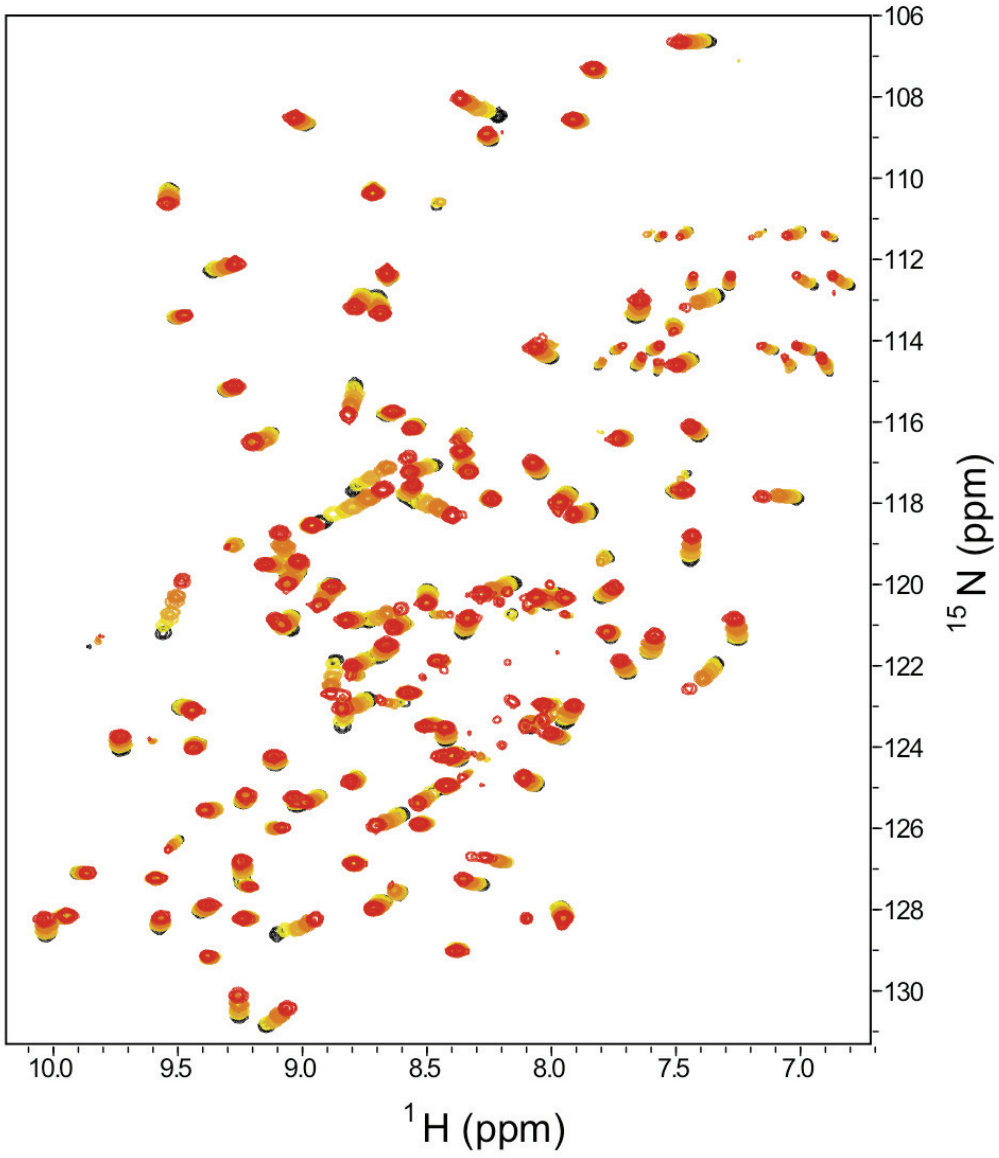

106
108
110
112
114
116
118
120
122
124
126
128
130
15N (ppm)
10.0
9.5
9.0
8.5
8.0
7.5
7.0
1H (ppm)

Figure 3.6. ^{1}H and ^{15}N chemical shifts monitored as a function of urea concentration for highly perturbed residues. CRABP 1 undergoes a transition from N-state → N*-state. at low urea concentrations

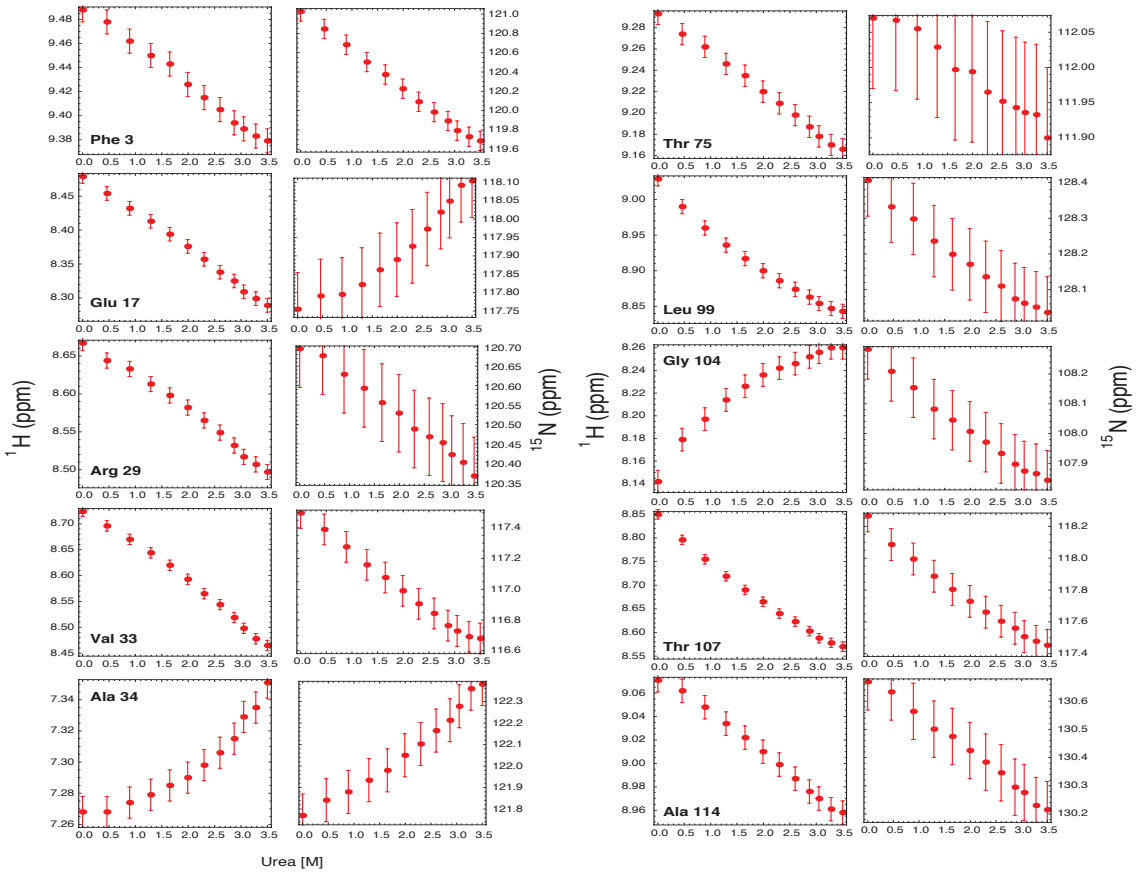

Phe 3
Glu 17
Arg 29
Val 33
Ala 34
Thr 75
Leu 99
Gly 104
Thr 107
Ala 114
^{1}H (ppm)
^{15}N (ppm)
Urea [M]

Figure 3.7. Overlay of ^{1}H-^{15}N HSQC spectra of CRABP 1 at different urea concentrations showing the second (slow) step of urea equilibrium unfolding of CRABP 1 (N*→U). Colors represent urea concentrations 4M (blue) and 8M (red). Unperturbed chemical shifts in the absence of urea, in black. In 4M urea, CRABP 1 exists in equilibrium as N, N* and U states.

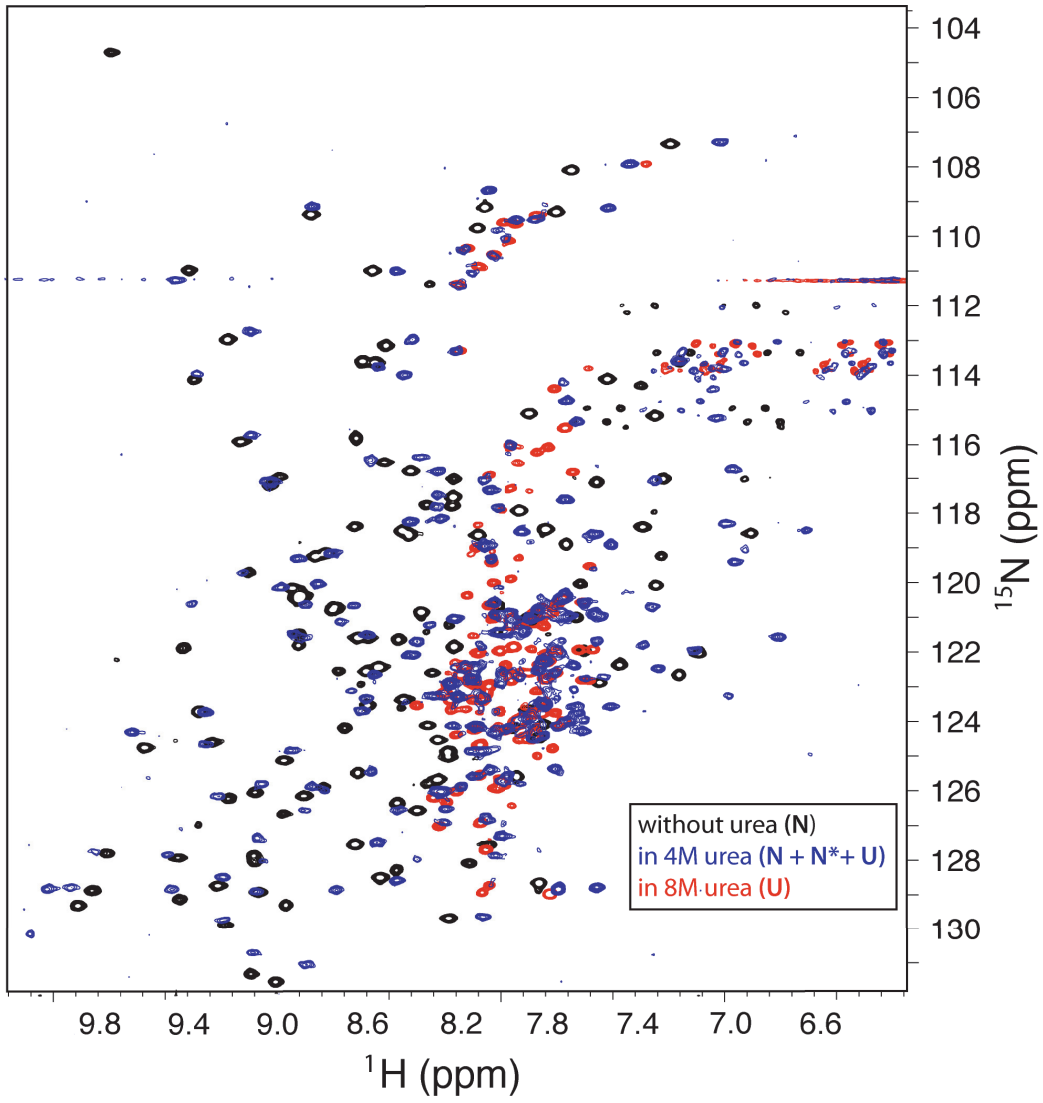
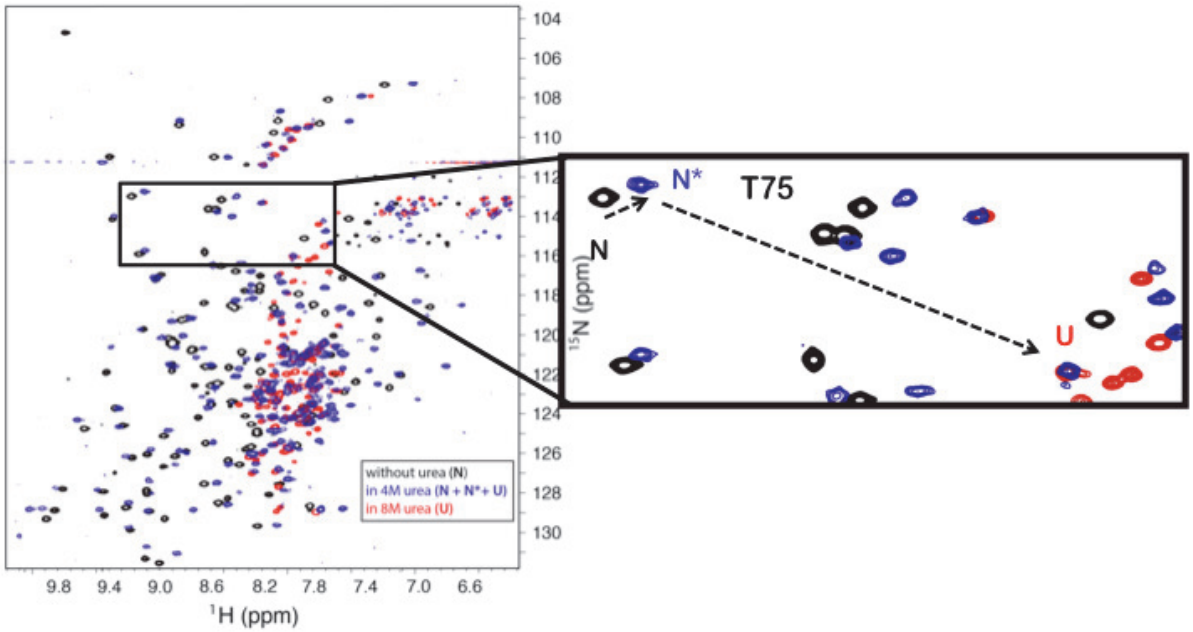

104
106
108
110
112
114
116
118
120
122
124
126
128
130
^{15}N (ppm)
9.8
9.4
9.0
8.6
8.2
7.8
7.4
7.0
6.6
^{1}H (ppm)
without urea (N)
in 4M urea (N + N*+ U)
in 8M urea (U)

Figure 3.8. Selected regions of the overlapped ^{1}H-^{15}N HSQC spectra of CRABP 1 at different urea concentrations showing the second (slow) step of urea equilibrium unfolding of CRABP 1 (N*→U). Colors represent urea concentrations 4M (blue) and 8M (red). Unperturbed chemical shifts in the absence of urea, in black. Shown is residue T75 where peaks appear in both the N* and U in states in 4M urea.


104
106
108
110
112
114
116
118
120
122
124
126
128
130
9.8
9.4
9.0
8.6
8.2
7.8
7.4
7.0
6.6
^{1}H (ppm)
without urea (N)
in 4M urea (N + N* + U)
in 8M urea (U)
N*
T75
N
U
^{15}N (ppm)

To gain insight on this intermediate, we identified residues perturbed by urea and projected them on to the structure of CRABP 1 (Figure 3.9). These residues include the helical subdomain and several loop regions of the protein. Inspection of the highly perturbed regions supports a dynamic movement of the helical region away from the β-barrel structure. Furthermore, the intermediate is characterized by moderate perturbations in the beta strands suggesting an increased breathing in the beta barrel subdomain with large motions along the loop regions.

We further examined this intermediate using native state hydrogen deuterium exchange NMR. We performed hydrogen exchange on WT* CRABP 1 with and without urea. Our experiments showed fast amide backbone proton exchange along the helical subdomain, loops and turns of the protein (Figure 3.10). Addition of urea increased the rates of exchange in the same regions of the protein (Figure 3.11A). Figure 3.11B shows that urea only increased the rate of exchange as seen with the faster decrease in intensity with urea for representative residues. Our observations on our hydrogen exchange experiments cannot completely dissociate the possibility of unfolding. However, we found very good correlation between the perturbed residues from our chemical shift perturbation analysis and highly solvent accessible residues identified by hydrogen exchange (Figure 3.12). In addition residues that have been identified using both methods corresponded to dynamic regions in the portal of the protein reported previously by fast hydrogen exchange NMR and slow hydrogen exchange MS experiments on CRABP 1 (Krishnan*, et al.*, 2000, Xiao & Kaltashov, 2005). Previous rapid water-amide hydrogen exchange and our chemical shift perturbation analysis showed consistent dynamic regions in apo-CRABP 1 (Appendix J). In addition to this, we performed partial trypsin digestion of CRABP 1. We observed faster digestion of trypsin with upon addition of urea (Appendix K and Appendix L). The two major masses that

were generated from the mass spectra of all the tryptic peptide fragments have values of 11079.2 and 7127.5 (average masses). The analyses yielded other peptide masses that are within the range of 10000-12000 for the larger fragments and 6000-8000 for the smaller ones. The predominant fragment (MW=6393) corresponded to residues along turn 4 (regions around R78, K79 and R81), which is part of the ligand portal. Increased protease accessibility was also previously reported corresponding to increased dynamics of the apo CRABP 1 (Jamison*, et al.*, 1994). We are currently working on experiments to identify other regions of the CRABP 1 that become more accessible to proteases in the N* state. Our data suggests that enhancement of the movement of the helical subdomain and increased breathing of the beta barrel region of CRABP 1 which is an intrinsic behavior of the protein to accommodate its ligand also relaxes stable interactions between the beta stands leading to exposure of the aggregation-prone sequences.

Figure 3.9. Residues perturbed by urea mapped on to the structure of CRABP 1. Residues in CRABP 1 that showed high (green spheres) and moderate (green highlight) chemical shift perturbation upon addition of low amounts of urea.

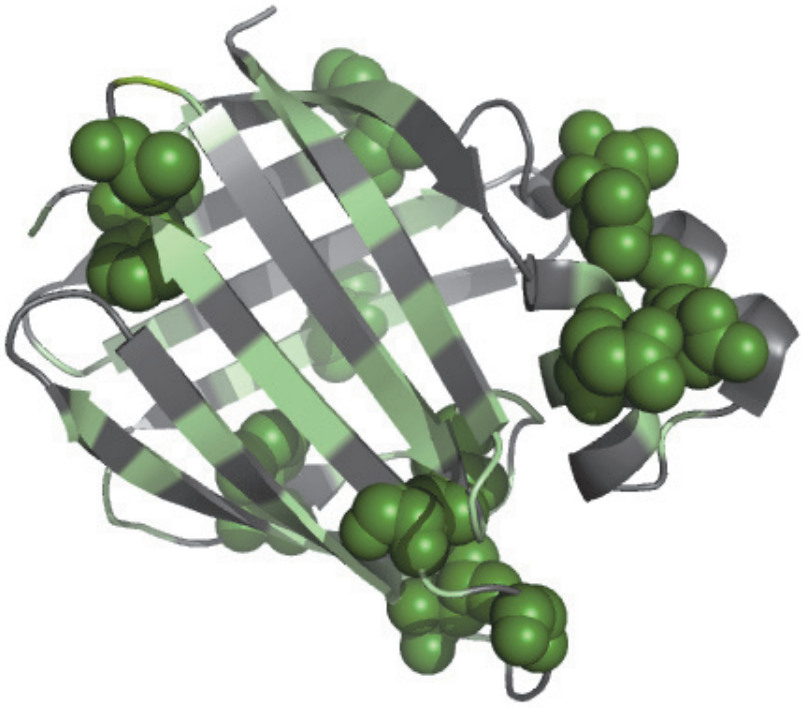


Figure 3.10. Backbone solvent accessibility of WT* CRABP 1. Fast exchanging (yellow) and slow exchanging residues (blue) monitored by H/D exchange were consistent in with and without urea. Residues not analyzed are in black.

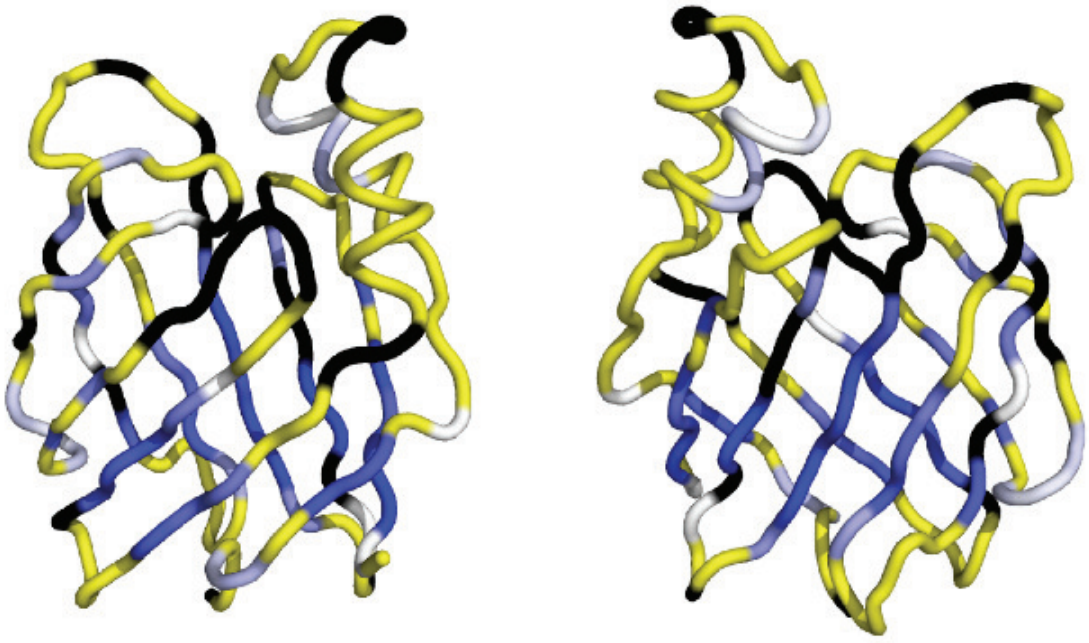
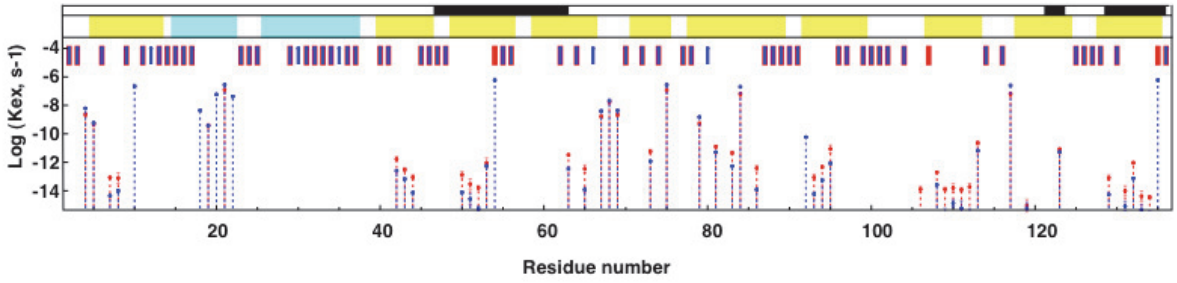


Figure 3.11. Exchange rates of CRABP 1 residues with and without urea

A) Rates of backbone amide proton exchange of amino acid residues of WT* CRABP 1 with (red dotted lines) and without (blue dotted lines) of urea. β-strands (highlighted in yellow) and α helices (highlighted in teal) in the sequence are shown. Aggregation core sequences are highlighted in black.

B) Ratio of peak intensities with time of representative residues in WT* CRABP 1 with (red) and without (blue) urea. I_0: peak intensity before exchange; I_{ex}: peak intensity after exchange.

A)


Log (Kex, s-1)
-4
-6
-8
-10
-12
-14
20
40
60
80
100
120
Residue number

B)

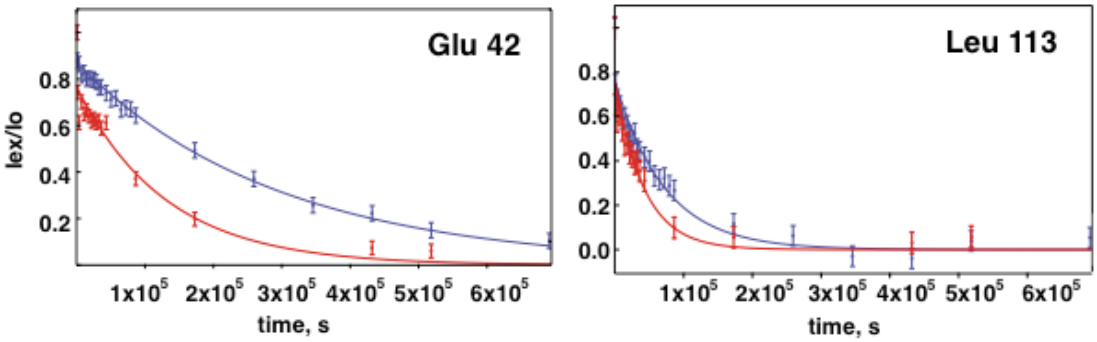
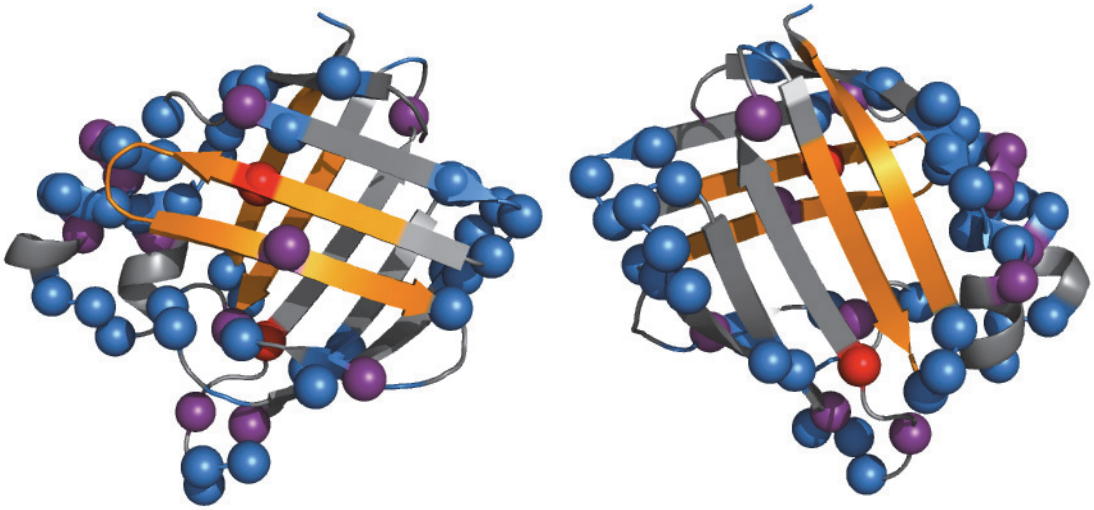

Glu 42
Iex/Io
0.8
0.6
0.4
0.2
1x10^5
2x10^5
3x10^5
4x10^5
5x10^5
6x10^5
time, s
Leu 113
0.8
0.6
0.4
0.2
0.0
1x10^5
2x10^5
3x10^5
4x10^5
5x10^5
6x10^5
time, s

Figure 3.12. Highly dynamic near native intermediate exposes aggregation-prone sequences.

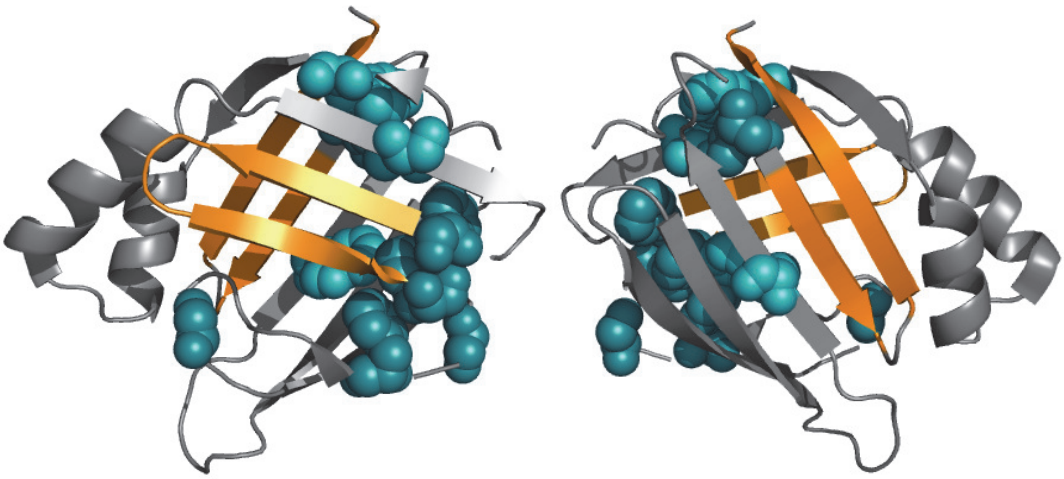
Regions of CRABP 1 with high chemical shift perturbation and fast exchanging residues monitored by HX-NMR (purple spheres), high chemical shift perturbation only (blue spheres), fast exchanging residues only monitored H/D exchange NMR (red spheres) and aggregation cores (orange highlight) mapped on to the CRABP 1 structure. Residues sensitive to urea cluster along regions of the protein that allows exposure of aggregation cores.



3.2.4. Effect of single residue substitutions on CRABP 1 aggregation

To examine the effects of mutations that leads to aggregation, we monitored chemical shift perturbations brought about by single residue substitutions around the aggregation cores F50M, F65M, G68A, F71M, G78A (core residues 51-65) and L118V (C-terminal core) (Figure 3.13). We have only observed local perturbations near the mutation sites in these variants suggesting that global unfolding of the protein is not required in these mutants. Our data suggests that local perturbations around the core sequences in the folded structure may be increasing the dynamics of the protein and enhances the exposure of these aggregation-prone segments. We are currently validating our observations by monitoring the dynamics of aggregation-prone mutants using hydrogen exchange.

Figure 3.13. Locations of single residue substitutions that favor aggregation mapped with aggregation cores. Aggregation-prone mutants CRABP 1 only lead to local perturbations in the structure. Amino acid residues that have been mutated and shown increased propensities to aggregate (teal spheres) are clustered close to the aggregation cores. Chemical shift analysis of these mutant proteins showed only local perturbations near the mutation sites.

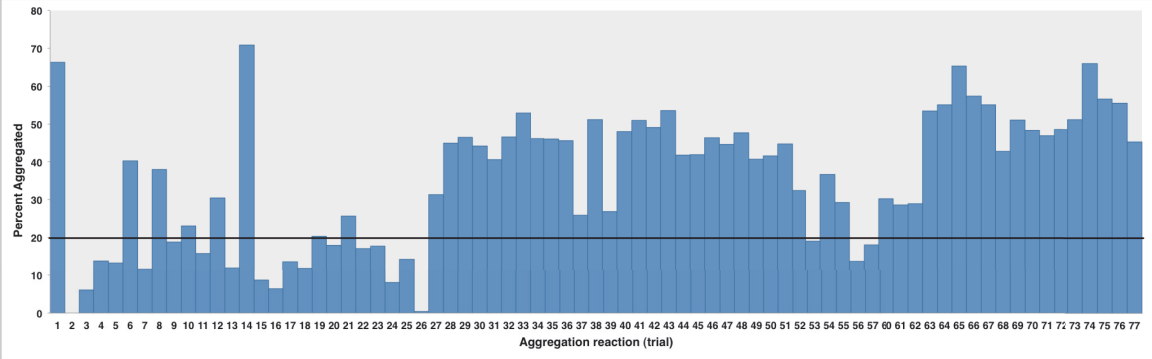


3.2.5. Retinoic acid binding decreases aggregation of WT* CRABP 1

It has been previously reported that binding of retinoic acid reduces the dynamics of CRABP 1 (Krishnan*, et al.*, 2000, Xiao & Kaltashov, 2005). To test whether stabilization of CRABP 1 by retinoic acid can decrease the aggregation of CRABP 1 we performed an *in vitro* aggregation assay and monitored formation of aggregates with and without retinoic acid. Since aggregation is a stochastic process, we performed multiple trials to test the effect of retinoic acid on the aggregation propensity of CRABP 1. We performed our assay under conditions that favor aggregation of CRABP 1 - high concentrations (200 uM and above) in the presence of 1.5 M urea at pH 7.0 (Figure 3.14A). We found that the propensity to aggregate was greatly reduced by retinoic acid (Figure 3.14). Out of 69 independent aggregation reactions, only 15 reactions showed at least 20% of monomers forming aggregates in the presence of retinoic acid (Figure 3.14B) in contrast to 54 out of 77 independent aggregation reactions with at least 20% monomers forming aggregates without retinoic acid (Figure 3.14A). Our findings support the model that ligand binding prevents CRABP 1 from accessing this open dynamic state by locking this helical domain close to the barrel inhibiting exposure of aggregation cores preventing aggregation (Figure 3.15).

Figure 3.14. Inhibition of CRABP 1 aggregation by retinoic acid. 54 out of 77 independent aggregation reactions showed at least 20% of CRABP 1 monomers forming aggregates (A) while in the presence of retinoic acid, 15 out of 69 independent aggregation reactions showed at least 20% of CRABP 1 monomers forming aggregates (B). Aggregation reactions were performed by incubating 250-300 μM of soluble CRABP 1 in phosphate buffer pH 7.0 containing 1.5M urea at 37^{o}C for 2 hours. Amount of aggregated proteins were determined by measuring the amount of remaining soluble monomers after incubation.

A)


Percent Aggregated
Aggregation reaction (trial)

B)

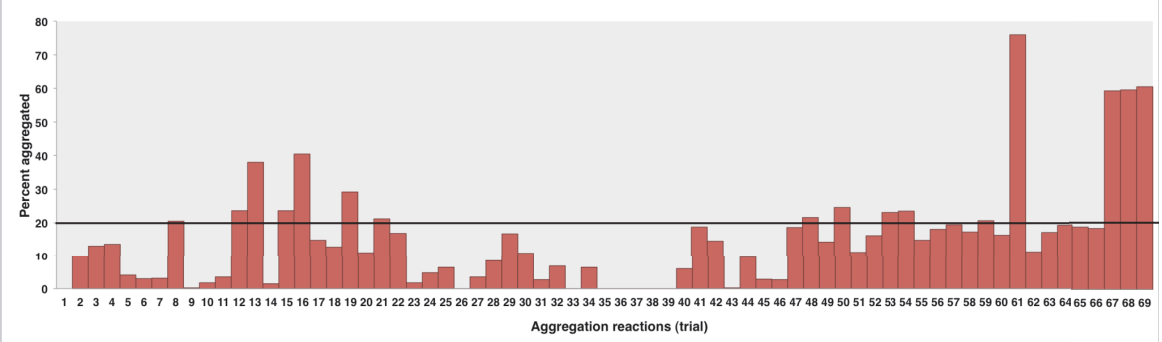
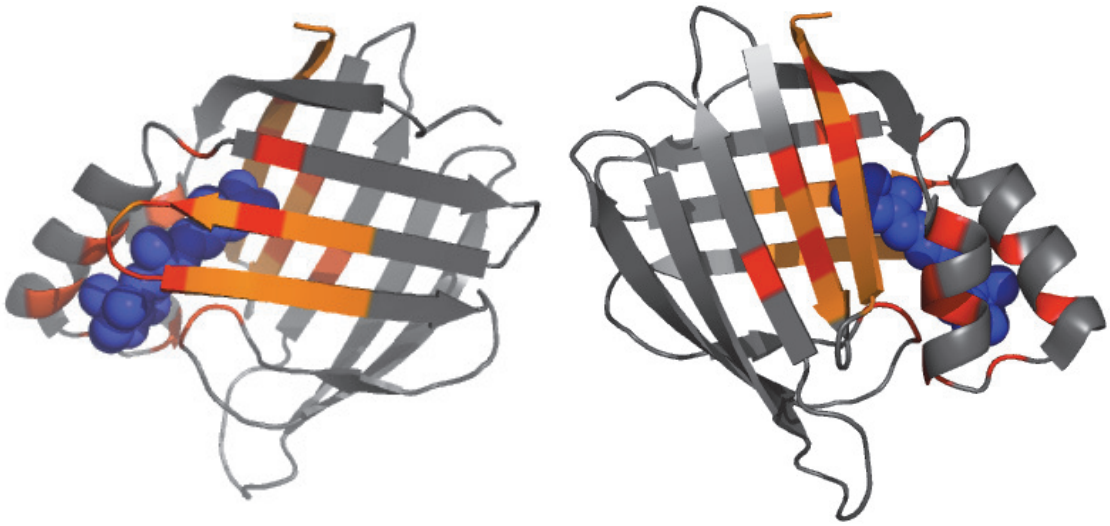

Percent aggregated
Aggregation reactions (trial)

Figure 3.15. Conserved sequences in CRABP 1 important for ligand binding and aggregation. Conserved ligand binding residues in the iLBP family mapped on to CRABP 1 (in red) are critical in not only in ligand interactions but in stabilization of aggregation cores in the complex.



3.3. Discussion

Despite several decades of trying to understand protein aggregation, there have only been a handful of globular proteins with a well-explored aggregation energy landscapes. The idea that native proteins are prone to aggregate has been progressively explored recently. However, there is lack of a complete picture describing aggregation risks and protective mechanisms from the early steps of folding to the fully functional protein. For a long time we have been interested in questions on β-rich proteins: What signals in the protein sequence favor folding over aggregation of these highly aggregation-prone motifs? How do these proteins perform their functions despite the risk of aggregation? We have used knowledge gained over the years on a β-rich protein CRABP 1 to address these questions and provide clear connections between folding, function and aggregation. In Chapter 2, we demonstrated how a protein's sequence can dictate both folding and aggregation. We have learned that folding assures early protection of hydrophobic sequences to avoid aggregation. However, the question remains why in spite of this folding driven protection of aggregation-prone segments CRABP 1 is susceptible to aggregation under native conditions? We therefore hypothesize that similar to other globular proteins, local fluctuations in CRABP 1 may be leading to undesirable interactions. In order to provide connections between aggregation and motions of the native protein, we identified conditions that favor aggregation and examined the species populated. From here, we learned that indeed the protein's dynamic behavior directs exposure of these aggregation-prone segments.

We have found that addition of small amounts of denaturants increases the aggregation propensity of CRABP 1. We were able to establish a strong correlation between urea concentration and aggregation at concentrations prior to global unfolding

of CRABP 1. Over the years, we have used intrinsic tryptophan fluorescence and circular dichroism to monitor global unfolding. However, these methods have not been sensitive in detecting species that become populated at low urea concentrations. Our inability to detect these species using fluorescence methods and CD provided us more hints that species leading to aggregation may not be largely destabilized. We therefore used NMR and chemical shift perturbation analysis to sensitively monitor populations. In addition, HX-NMR and partial proteolysis to monitor solvent and protease accessibility were used to support our findings from the chemical shift perturbation analysis. Our data from all three methods illustrate a dynamic breathing of CRABP 1 under native conditions. Our NMR data on aggregation-prone variants with substitutions around the aggregation cores displaying only local perturbations. In addition, we also found that aggregation is sensitive to pH. NMR analysis of CRABP 1 in at pH's from 7 to 8 showed perturbation in two histidine residues, His40 and His94. His40 is slightly conserved and interacts with Lys8 and Ser 55 (residue in strand 3). Protonation of His40 may break interactions with strand 3 (part of aggregation core 1) thus exposing this core to aggregation. The pH dependence of CRABP 1 aggregation has to be explored more in detail especially at lower pH's when acidic residues functioning as gatekeepers may be perturbed. All together our observations supporting aggregation from mutations that only locally perturb CRABP 1 further support an aggregation-prone intermediate with a native-like structure.

The conformation we have identified been previously described as a natural tendency of the molecule in the absence of retinoic acid. Fluctuations that occur along the portal region of CRABP 1 bare the entry site for ligand to access and bind inside the cavity. Although this dynamic portal model favoring ligand binding has already been out for quite sometime, our data extends brand new insights regarding its role in preceding

aggregation. In this dynamic open state, the helical domain dissociates from the β-barrel body and breaks interactions with loop II, thus exposing the highly fluctuating aggregation-prone strands 3 and 4. In addition, this movement partly unmasks strand 10, transforming it into an edge strand exposing it to unfavorable intermolecular interactions.

In previous reports, it has been emphasized that binding to retinoic halts CRABP 1 dynamics and stabilizes the protein (Krishnan, Sukumar et al. 2000(Xiao & Kaltashov, 2005). To support further our claim that similar motions predispose the protein to aggregation, we tested whether a stabilized, rigid ligand-bound CRABP 1 is less prone to aggregate. Consistent with our hypothesis, we show that retinoic acid is capable of inhibiting aggregation. Protective strategies using functional interactions have already been observed in proteins as discussed in the previous sections. CRABP 1 aggregation presents an excellent case where a hydrophobic ligand protects the same hydrophobic regions that drive aggregation. A similar competition between intermolecular interactions and ligand binding has also been observed in another lipid binding protein β-lactoglobulin. This protein has a similar cavity that binds hydrophobic ligands. As discussed earlier, binding of ligand dissociates the dimer, which is stabilized by the same regions in aggregates (Gutierrez-Magdaleno, Bello et al. 2013; (Hamada*, et al.*, 2009). The case of acylphosphatase also shows also a similar behavior. NMR experiments combined with molecular dynamic simulations show that phosphate ions acts as an inhibitor for protein aggregation of acylphosphatase (De Simone*, et al.*, 2011). It has been demonstrated that the binding interaction perturbs the energy landscape and the dynamics of the protein to prevent aggregation even under conditions where it is otherwise favored (De Simone*, et al.*, 2011). This observation is proposed to be a

general behavior in protein-ligand interactions to avoid aggregation (De Simone*, et al.*, 2011).

The functional relevance of the dynamic movement of CRABP 1 is shared with other members of the ILBP family. Our findings on the aggregation mechanism of CRABP 1 may have similar implications with other ILBPs. This family is not known for any aggregation-related diseases. Hence we propose that this may be ascribed to the ability of these proteins to shield aggregation via several mechanisms such as rapid folding and protective ligand interactions. Ligand specificities of different ILBPs are different, however there are conserved binding sites for these proteins. We find several conserved binding sites at the helical sub-domain and aggregation-prone segments (Figure 3.16). Our results imply that the interaction between the ligand with the helical lid, loop 2, strands 3, 4, 9 and 10 have multiple functions: 1) to functionally lock the ligand into the cavity with the helical lid; 2) to halt the motions along the portal region to stabilize the complex; 3) to stabilize fluctuations in loop 2, strands 3 and 4 to prevent exposure to aggregation; 4) to stabilize strand 10 (which forms a salt bridge with the ligand) and subsequently strand 9 to avoid aggregation. Ironically, in all our folding studies we have used the stabilized variant of CRABP 1, R131Q. As can be recalled, Arg131 (which is part of the C-terminal aggregation-prone strand) forms a salt bridge with the carboxyl group of retinoic acid to stabilize the complex. Mutation of Arg131 into glutamine destabilizes the complex however increases the solubility and stability of CRABP 1. This further suggests Arg131 interaction with retinoic acid offsets the lowered solubility and stability of CRABP 1.

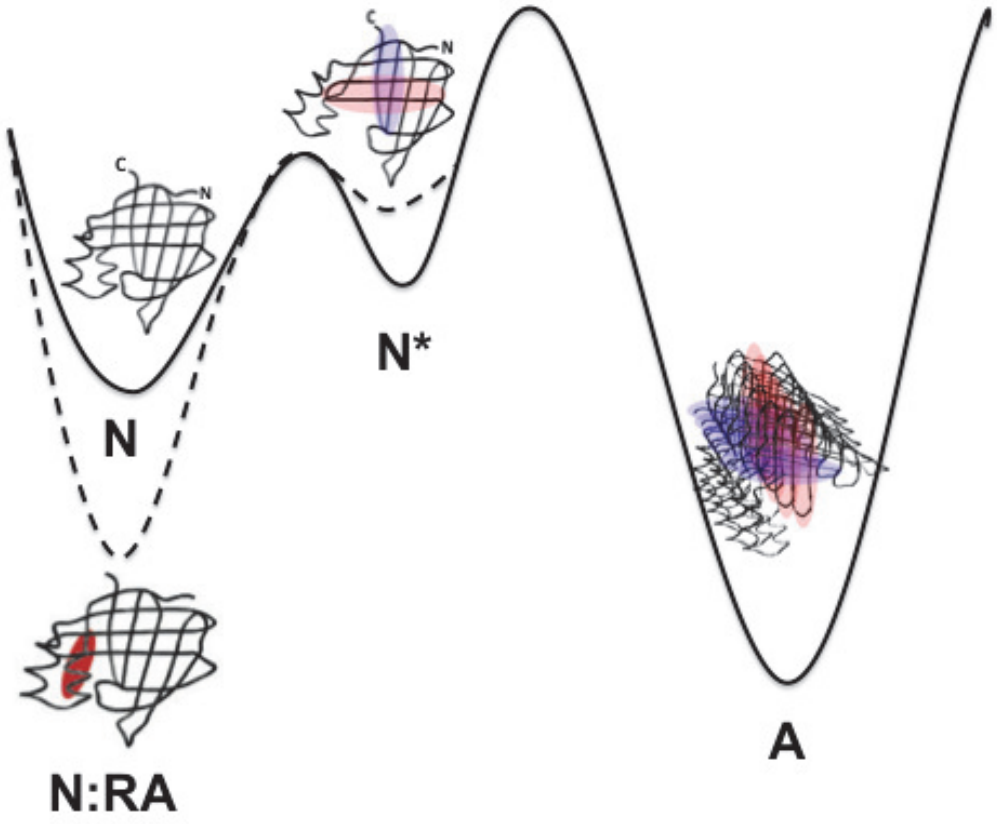
Hence we propose a model where under native conditions CRABP 1 is an ensemble of low-energy states with slightly different energies separated by defined

barriers. One of these states is a relatively higher energy open conformation, N* state capable of both ligand binding and intermolecular association. In the presence of a ligand, the equilibrium is shifted towards the lowest energy ligand-bound state. However, in the absence of the ligand at high CRABP 1 concentrations, this high-energy state becomes sufficiently populated permitting intermolecular associations shifting the equilibrium towards the very low energy aggregated state (Figure 3.16).

In the absence of the ligand, how is the native protein protected from aggregation? Analysis of the CRABP 1 sequence reveals presence of aggregation gatekeepers (Clouser, 2011). These include D16, E17, K20, K30, D125 and D126. It is predicted that helix II and partly helix I is highly aggregation-prone however we have not found these in the aggregation cores. Thus, the presence of these charged residues (D16, E17, K20 and K30) possibly safeguards this region against aggregation. We also observed that residues D125 and D126 disrupt the aggregation core in the C-terminal end of the protein resulting in a small gap between the aggregation cores. Thus, these charged residues identified function as aggregation gatekeepers of CRABP 1. In addition cellular conditions also protect CRABP 1 aggregation.

As we have discussed previously, aside from the general housekeeping machineries that prevent aggregation in cells, protein concentrations of aggregation-prone species are largely regulated. In human cells, CRABP 1 concentration is very low (1-10 μM) and has relatively low saturation for retinoic acid (40nm) around 5% saturated (Donovan*, et al.*, 1995). Despite not having enough ligand in to bind retinoic acid, the concentration of CRABP 1 is regulated to prevent aggregation. The cellular maintenance of low CRABP 1 concentrations may be a mechanistic control to limit the amount of N* state enough to trigger aggregation.

Figure 3.16. Proposed aggregation free energy landscape of CRABP 1 under native conditions. Native CRABP 1 samples a near-native conformation (N*), which is linked to both ligand binding and aggregation. N* state can bind to retinoic acid to form a stable complex or self-associate to form highly stable amorphous aggregates. This illustrates the role of retinoic acid interaction in restructuring the energy landscape of CRABP 1.


C
N
C
N
N*
N
N:RA
A

In summary our work not only provides another case of a globular protein aggregating from the native state but also describes natural features of the protein that present risks to aggregation. As described earlier in this chapter, there have been only a very few cases reported wherein oligomerization and binding interfaces also predispose proteins to aggregation. The case of CRABP 1 presents a clear illustration of functional interactions as protective mechanisms against aggregation. The overlap between functional ligand binding interfaces and aggregation-prone regions highlights the importance of natural ligands in designing small molecules against aggregation. In addition, our contribution to the growing number of cases where globular proteins aggregate from the native state supports the idea that not only folding dictates aggregation propensity but the inherent motions of proteins as well. Although this implies specific approaches in studying aggregation of globular proteins, shared characteristics within protein families can also provide hints on their behaviors. Lastly, we have shown that evolutionarily fine-tuned folding of a protein does not guarantee protection from aggregation therefore this emphasizes the critical role cellular housekeeping machineries to regulate protein concentrations in their soluble and functional states.

3.4. Experimental Procedures

In Vitro Aggregation Assays

WT* CRABP 1 proteins were generated using the same protocol described in Chapter 2. Protein solution aliquots with known concentrations were lyophilized. Lyophilized proteins were resuspended in buffer (10 mM phosphate buffer, 5 mM DTT pH 7.0) containing different urea concentrations (0 to 3.0M) to a total reaction volume of 100uL. Each tube contains 200 μM of pure CRABP 1 protein. Aggregation reactions were incubated and agitated at 1000rpm using a thermo shaker at 37^{o}C for 12 hours.

The amounts of soluble protein before and after aggregation were measured from absorbance at 280nm (A_{280}) of protein diluted in 8M urea. The same protocol was also used to determine aggregates formed at different pH conditions. Briefly, aggregation of CRABP 1 was monitored using phosphate buffers with pH's from 6.0 to 8.0. Percent of aggregated protein was determined by taking the ratio of the difference in A_{280} before and after aggregation with respect to the initial concentration. $A_{280}(t=0)$ and $A_{280}(t=12)$ are absorbance readings at 280nm before and after aggregation.

$$\% \text{ Agg} = \frac{A_{280}(t=0) - A_{280}(t=12)}{A_{280}(t=0)} \text{ X } 100$$

To monitor the effect of retinoic acid on WT* CRABP 1 aggregation, we compared the amounts of aggregates formed after two hours of aggregation at 37^{o}C using the same method described above in the presence and absence of 100 μM retinoic acid.

The effect of urea on the rate of aggregation was determined by measuring the light scattering at 350nm every 30 seconds of 300 μM WT* CRABP 1 in buffer (10 mM phosphate buffer pH 7.0 containing 5 mM DTT) with and without 1.5M urea using Varian Cary 50 UV-Vis Spectrophotometer.

Equilibrium Unfolding Experiments

Equilibrium unfolding of WT* CRABP 1 in the presence of 0 to 8M urea were monitored using solution NMR spectroscopy. ^{15}N-labeled WT* CRABP 1 was dissolved in buffer (20 mM Tris buffer, 5 mM DTT, 5% *(v/v)* D_2O pH=8.0) with or without urea. HSQC spectra of WT* CRABP 1 for each urea concentration were recorded at 26 °C on a 600-MHz Bruker Avance spectrometer using a TXI cryoprobe. Data were processed

using NMRPipe (Delaglio*, et al.*, 1995) and analyzed using Cara software (Keller, 2004). ^{15}N and proton chemical shifts for amino acid residues were assigned de novo using triple resonance experiments HNCO, HNCACB, HNcoCACB.

Proteolysis of WT* CRABP 1

To determine the solvent accessibility of WT* CRABP 1 urea, we performed limited trypsin digestion on WT* CRABP 1 with and without urea. WT* CRABP 1 was dissolved in proteolysis buffer (10 mM Tris-HCl pH 7.6, 50 mM NaCl, 2mM $CaCl_2$) and treated with Trypsin Gold (Promega) to a trypsin:protein (w/w) concentration of 1:25 at different time points at 30^oC. Tryptic digests were ran in 15% Tricine SDS-PAGE for analysis. MS experiments were done on a QStar XL Quadrupole TOF Mass Spectrometer using TurboIon electrospray ionization. LC-MS/MS was done after separation of the tryptic peptides on a C18 RP column using a gradient elution of 0-95% acetonitrile/0.1% formic acid at a flow rate of 200 μL/min over 30 mins. MS peaks were further fragmented for sequencing by collision-induced dissociation (CID). The mass spectra were analyzed using the Bayesian Protein Reconstruct tool in AnalystQS software while peptide sequences were determined by *de novo* sequencing.

Hydrogen Exchange of NMR

Hydrogen exchange was performed on WT* CRABP 1 in the presence and absence of 1.5M urea. Lyophilized ^{15}N-labeled WT* CRABP 1 was dissolved in buffer (20 mM Tris-d11, 5 mM DTT pH 8.0) with or without 1.5M urea and lyophilized overnight. Lyophilized protein was re-dissolved in D_2O. H/D exchange was followed by measurement of peak intensities in HSQC spectra acquired at different time times after transfer of the protein in D2O.

HSQCs were recorded at 26^{o}C using a 700 MHz Varian NMR system equipped with a cryogenically cooled triple-resonance probe. Total acquisition time for each 2D HSQC spectrum was about 15 min. Data were processed using NMRPipe (Delaglio*, et al.*, 1995). Peak intensities for assigned residues as a function of time were analyzed Peak intensities for assigned residues as a function of time were analyzed using homemade Mathematica scripts.

CHAPTER 4

CONCLUSIONS

4.1. Summary

A polypeptide's amino acid sequence has evolved to contain all necessary information for proper folding and function. However, these interactions likewise govern aggregation, which is associated in a number of catastrophic diseases. Thus, inherent risks are also encoded in the protein sequence. In this thesis, we investigated the existence of sequence determinants for folding and misfolding. To explore the molecular basis of protein aggregation and to find connections with folding, we employed CRABP 1, a model β-rich protein with a complex topology and rugged energy landscape. Taking advantage of several aggregation-prone mutants of CRABP 1, we have identified aggregation-driving sequences in *in vitro* and *in vivo* aggregates. These short contiguous amino acid segments encompass the ligand-binding portal and strands at the C-terminal end of the protein consisting the aggregation interfaces. These aggregation cores constitute aggregates derived *in vitro* and *in vivo*. Moreover, the experimentally determined aggregation cores validated several sequence-based aggregation predictions. However, we find that sequence-based predictions cannot fully recapitulate the overall aggregation propensities of single residue mutants. Hence, considerations for additional parameters are necessary to predict aggregation propensities such as protein thermodynamic stabilities and dynamics. Our analysis of the aggregation core sequences of CRABP 1 revealed clear connections between one of the cores, the C-terminal strand and folding. Folding studies indicate the critical role of N and C-terminal docking during the early stages of barrel closure. Thus, we have learned that fundamental steps in folding are directed by the protein's natural tendency to avoid aggregation.

Our awareness on the role of folding as a protective mechanism from aggregation led us to more questions on how proteins evade this. We further examined the behavior of CRABP 1 under native conditions and resolved intermediates that precede aggregation. Under non-denaturing conditions, we found that the natural motions of CRABP 1 associated with its role to bind ligand triggers aggregation. The dynamic behavior of CRABP 1 to open its ligand-binding portal to permit ligand access exposes the aggregation cores. We further found the role of the ligand, retinoic acid in stabilizing the complex, essential in preventing aggregation. We also established connections with members of intracellular binding proteins that contain conserved binding residues for ligands as these are located along the aggregation cores, further justifying the importance of ligand interaction as a protection from aggregation. Finally, we emphasized the relevance of low expression levels *in vivo* of CRABP 1 in facilitating folding and function over aggregation.

4.2. Implications for Future Studies

This thesis has provided a holistic examination on three important intrinsic behaviors of proteins: folding, functional dynamics and aggregation. It has provided the much-needed linkage between the evolutionary requirement for proper folding for function, as well as the inability to completely eliminate the risk of aggregation. The foundations contributed on understanding the energy landscape of CRABP 1 opens opportunities to explore other areas of the landscape including aggregation under non-native conditions. In addition, our knowledge of aggregation-prone sequences in the protein will assist further studies on conserved residues (aggregation gatekeepers) that restrict aggregation. A more detailed analysis of the aggregation kinetics of CRABP 1 and its mutants may also provide additional insights in guiding predictions for aggregation propensities. Likewise, CRABP 1 and its mutants may be used for

interrogating the roles of cellular quality control machineries (chaperones, degradation). Currently, there is still a modest amount of information showing direct links between protein sequence and interactions with the different quality control machineries. Interplay and competition for hydrophobic sequences in folding, aggregation, function and that of chaperones and degradation enzymes still remains to be understood in detail. CRABP 1 may be an excellent model for these studies to extract generic concepts in explaining protein folding in the cell.

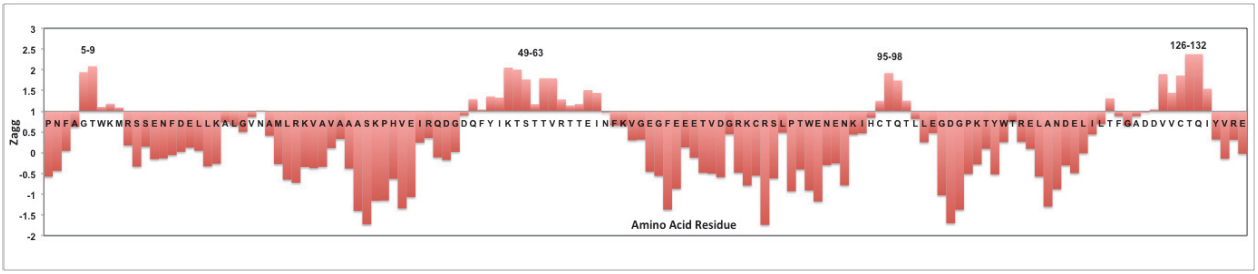
From an evolutionary point of view, it remains intriguing how proteins have not been efficiently selected against sequences that trigger aggregation. It is comprehensible that aggregation-prone sequences and dynamic motions of proteins are critical for function thus cannot be completely selected against, giving rise to intrinsic and cellular protection mechanisms against aggregation. However, it remains confounding why despite redundant protective mechanisms, there are still a growing number of misfolding diseases. Hence, it brings also to the table the evolutionary constraint for proteins to be multifunctional in the form of intrinsically disordered proteins. The existence of disorder in proteins permits multiple functions but also consequently lead to aggregation-related disease states. Therefore, understanding the balance between evolutionary selection for optimal protein function and threat for misfolding and aggregation remains as a huge area of study.

APPENDIX A

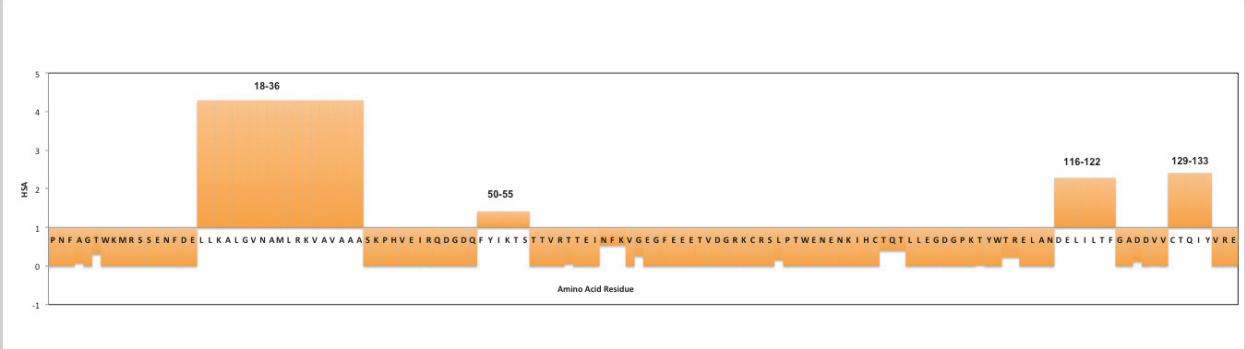
PREDICTED AGGREGATION PROPENSITY OF CRABP 1

Aggregation propensities were predicted using Zyggregator (a), AGGRESCAN (b) and TANGO (c).

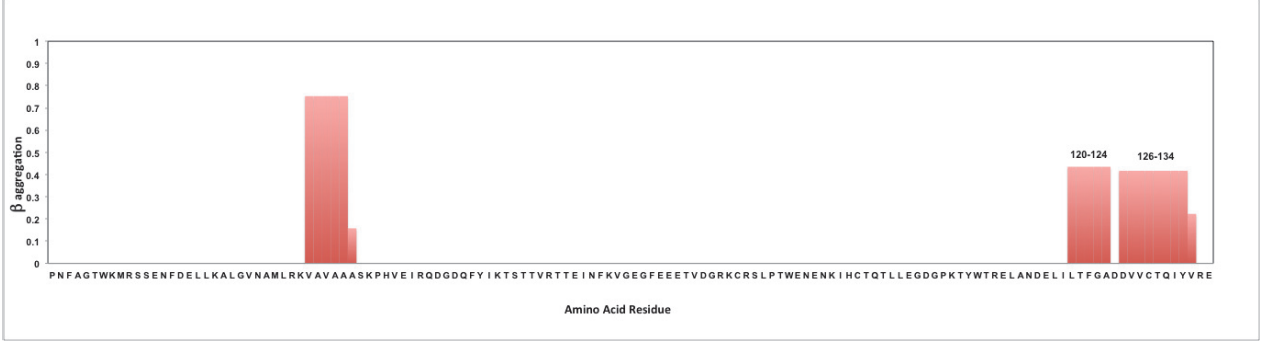
A



B



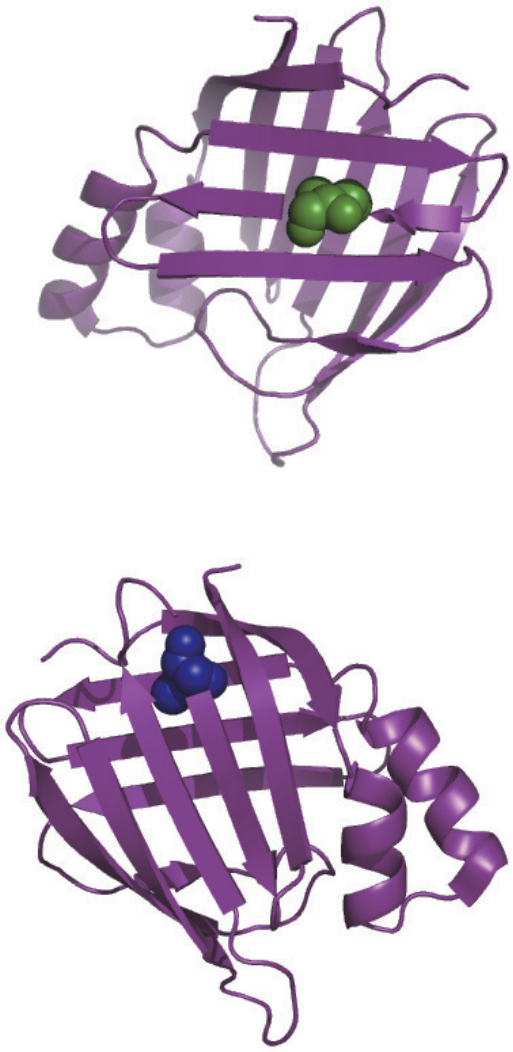
C



APPENDIX B

LOCATION OF MUTATIONS AFFECTING THE PREDICTED AGGREGATION PROPENSITY OF CRABP 1

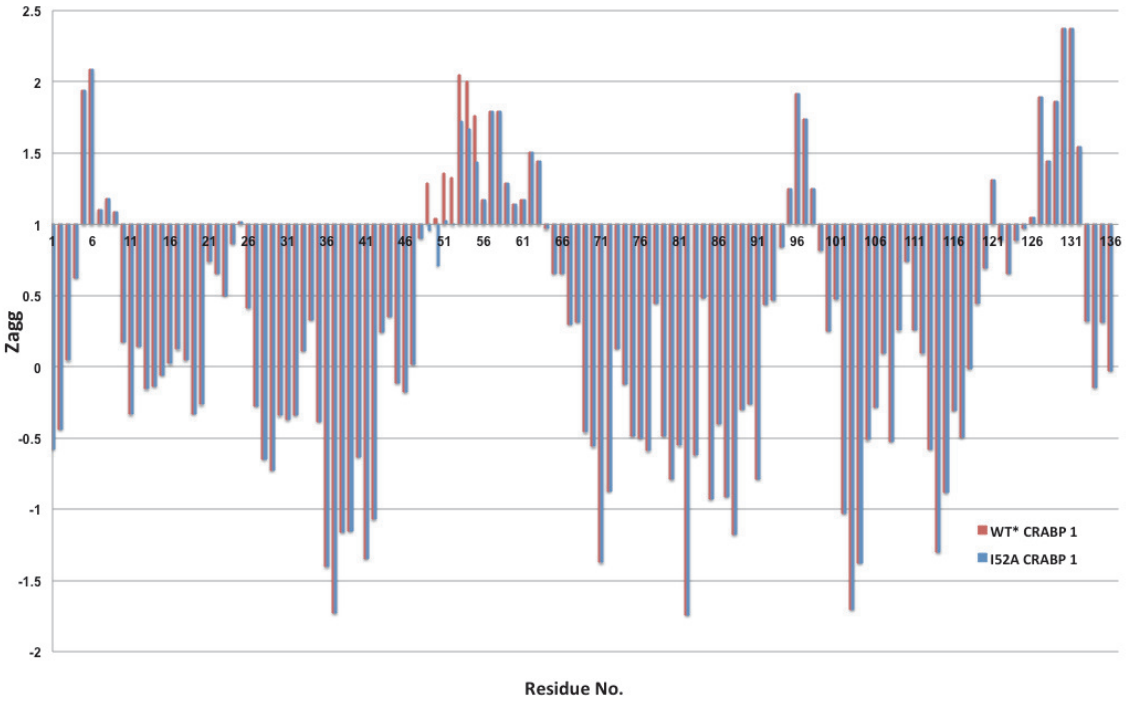
Isoleucine 52 is located in strand 2 (green sphere) and leucine 118 (blue sphere) is in strand 9 of CRABP 1.



APPENDIX C

PREDICTED AGGREGATION PROPENSITIES OF WT* AND I52A CRABP 1

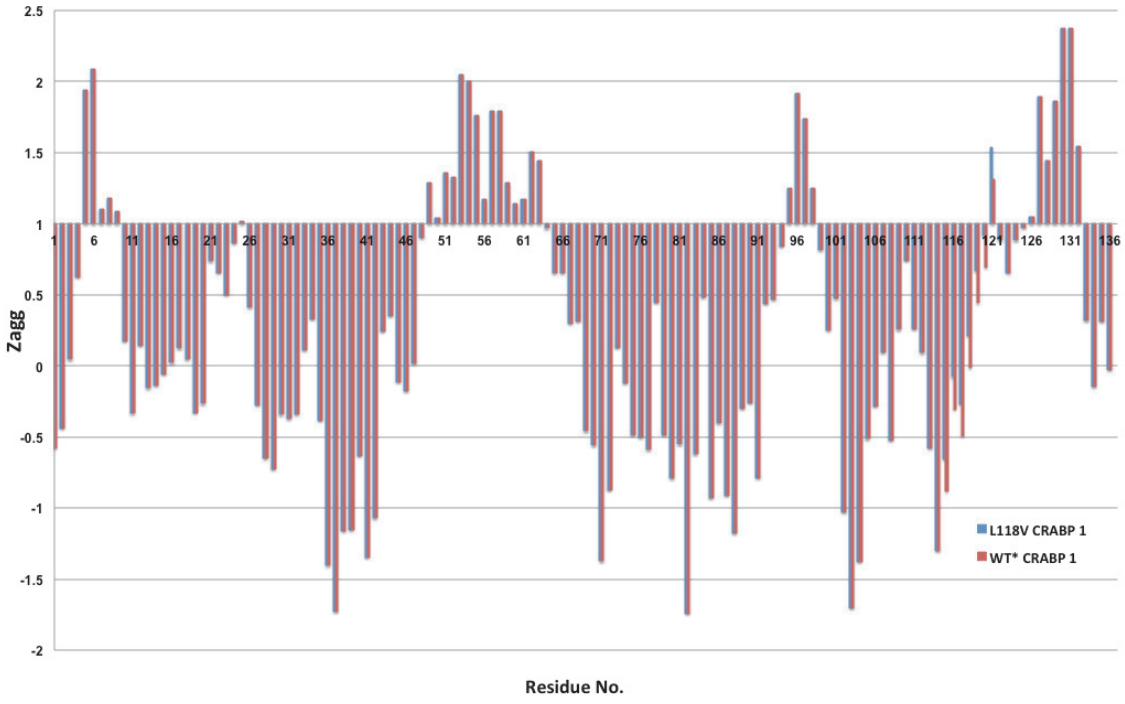
Decreased aggregation propensities were observed around the mutation site predicted by Zyggregator.



APPENDIX D

PREDICTED AGGREGATION PROPENSITIES OF WT* AND L118V CRABP 1

Increased aggregation propensities were observed around the mutation site predicted by Zyggregator.

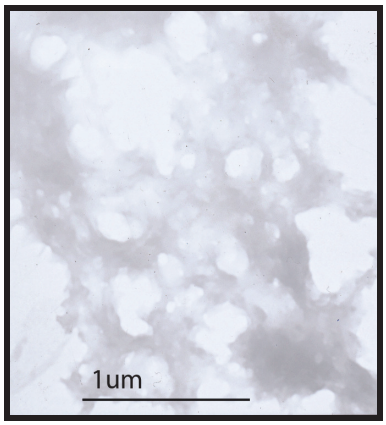


APPENDIX E

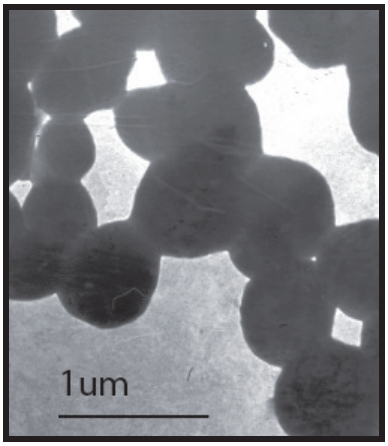
TRANSMISSION ELECTRON MICROSCOPY IMAGES OF CRABP 1 AGGREGATES

F71A *in vitro* and *in vivo* aggregates were resuspended in water. Ten microliters of the aggregate resuspension was dropped on to carbon-coated copper discs. Concentrations of samples were optimized so that only a very thin layer of aggregates forms at the surface of the disc. Samples were dried out for a few minutes. Ten microliters of 1% phosphotungstic acid was dropped on to the disc and allowed to dry for a few minutes. After this, excess stain was washed off with distilled water. Discs were dries overnight prior to visualization. *In vivo* aggregates (A) appeared as uniform globular structures in contrast to the very amorphous *in vitro* aggregates (B).

A)



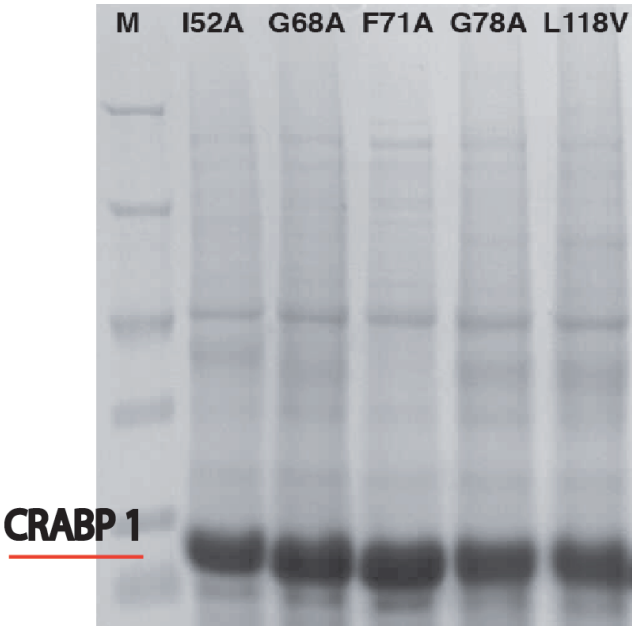
B)



APPENDIX F

SDS-PAGE OF PURIFIED INCLUSION BODIES FROM SEVERAL AGGREGATION-PRONE MUTANTS

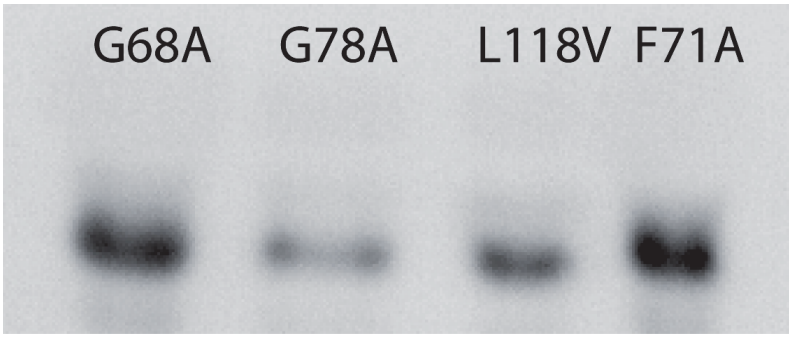
Several aggregation-prone mutants were purified using a mild non-ionic detergent. Mutant CRABP 1 proteins predominantly consist the purified inclusion bodies.



APPENDIX G

WESTERN BLOT OF CRABP 1 INCLUSION BODIES TO DETECT SMALL HEAT SHOCK PROTEIN IBPA.

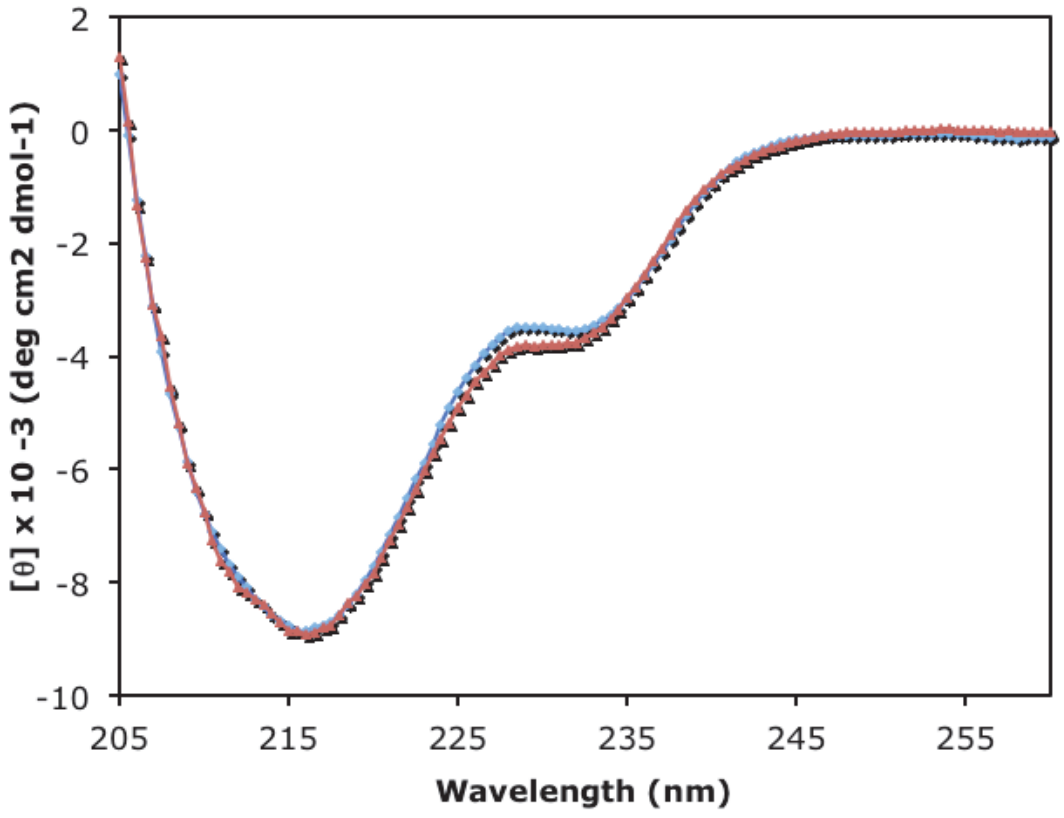
Inclusion bodies were probed against IbpA using rabbit anti-IbpA serum (from Dr. Tania Baker, MIT) and anti- rabbit IgG AP conjugate as secondary antibody. Blots were developed using Lumiphos reagent (Thermo Scientific).



APPENDIX H

CD SPECTRA OF WT* AND L118V MUTANT CRABP 1

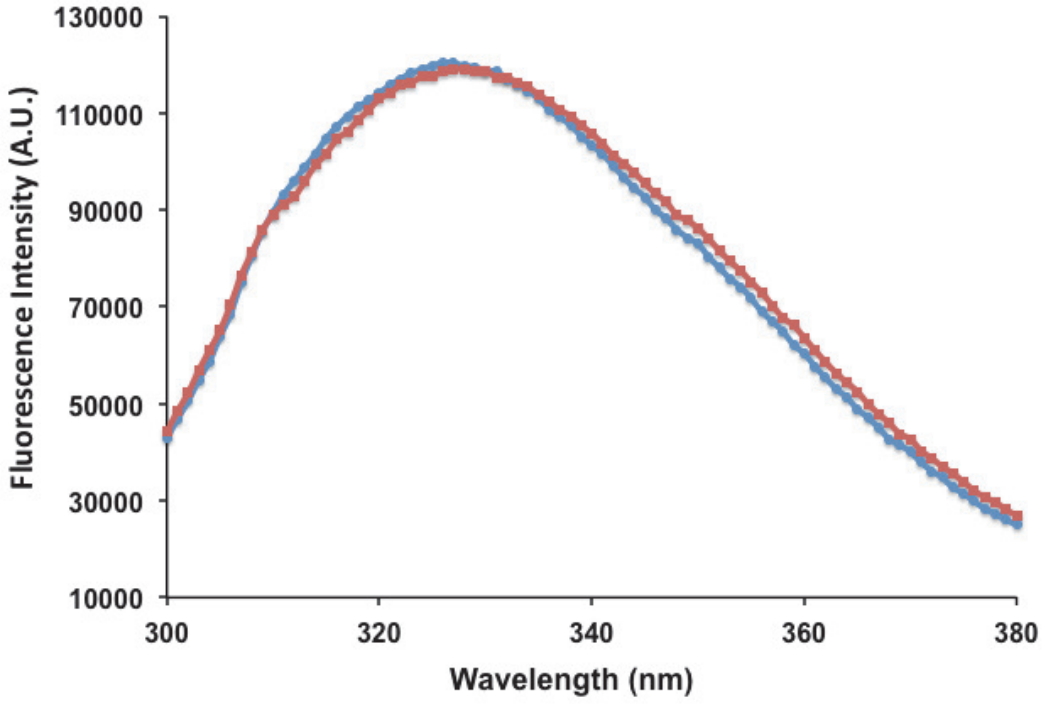
WT* (triangles) and L118V (squares) mutant CRABP I proteins (10 μM protein in 10 mM Tris-HCl, 1mM TCEP pH 8.0) show similar CD spectra. CD signals were converted to molar ellipticity (θ). CD spectra collected was an average of 10 scans. Mutant CRABP 1 contains the CD inflection at 228nm, which is characteristic of tertiary packing involving aromatic clustering around Trp 87 and the aromatic-charge interaction between Arg 111 and Trp 109 in the native protein (Clark*, et al.*, 1996)



APPENDIX I

FLUORESCENCE SPECTRA OF WT* AND L118V MUTANT CRABP 1

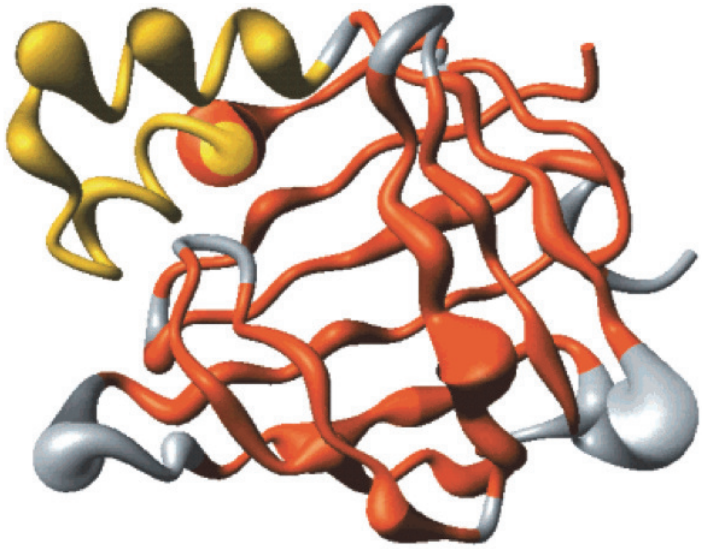
Intrinsic tryptophan fluorescence spectroscopy was performed on 4uM WT* and L118V mutant proteins in 10 mM Tris-HCl pH 8.0, 1mM DTT. Spectrum of L118V CRABP 1 (squares) was very similar to that of the WT* CRABP 1 spectrum (triangles). Maximum fluorescence is observed at 330nm is observed in the native protein and is quenched relative to the denatured state. Fluorescence quenching is attributed to the proximity of Trp 109 to Cys95 in the native state (Clark*, et al.*, 1996).



APPENDIX J

DYNAMIC REGIONS OF APO-CRABP 1 FROM FAST HYDROGEN EXCHANGE EXPERIMENTS

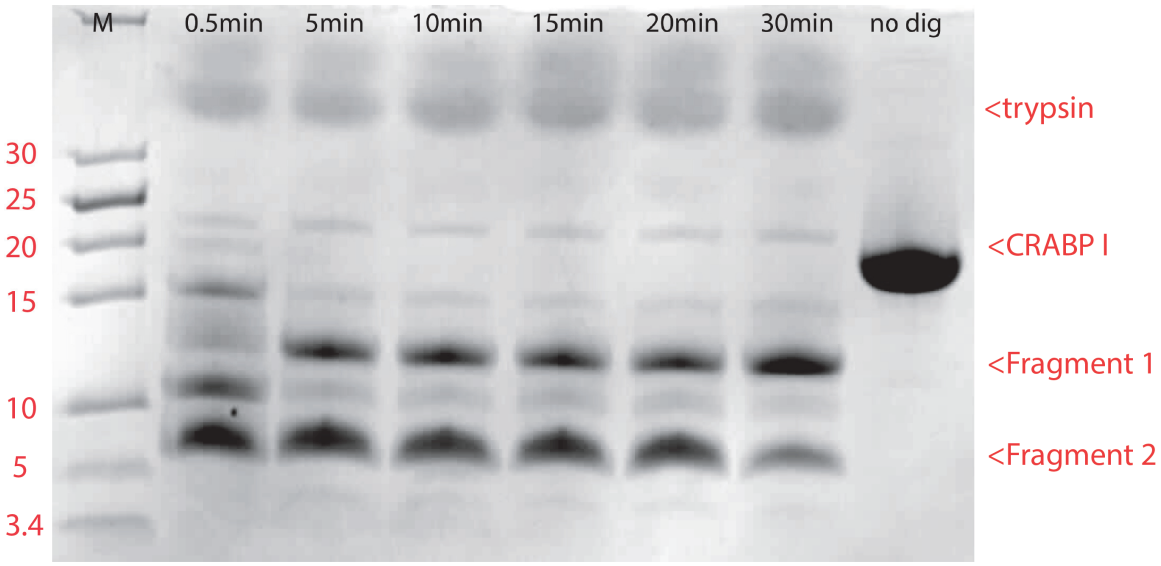
Shown in thick tubes are residues of CRABP 1 undergoing fast fluctuations in the absence of CRABP 1. This figure is re-printed from (Krishnan*, et al.*, 2000)



APPENDIX K

PARTIAL PROTEOLYSIS OF CRABP 1

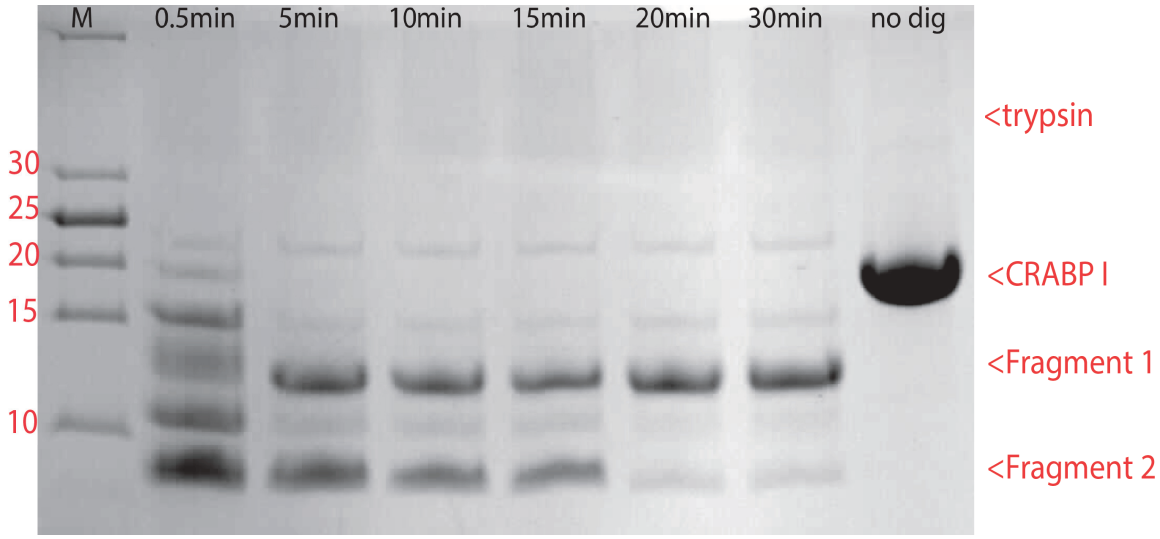
Partial trypsin digestion of CRABP 1 in 10 mM Tris-HCl pH 7.6, 50 mM NaCl, 2mM $CaCl_2$ showed appearance of two major fragments. Positions of fragments 1 (12KDa) and 2 (6KDa) are indicated.



APPENDIX L

PARTIAL PROTEOLYSIS OF CRABP 1 IN THE PRESENCE OF UREA

Partial trypsin digestion of CRABP 1 in 10 mM Tris-HCl pH 7.6, 50 mM NaCl, 2mM $CaCl_2$,1M urea showed appearance of two major fragments. The rate of proteolysis is faster with 1M urea than without (Appendix K). Positions of fragments 1 (12KDa) and 2 (6KDa) are indicated.



BIBLIOGRAPHY

Abeln S & Frenkel D (2008) Disordered flanks prevent peptide aggregation. *PLoS computational biology* **4**: e1000241.

Agostini F, Vendruscolo M & Tartaglia GG (2012) Sequence-based prediction of protein solubility. *Journal of molecular biology* **421**: 237-241.

Alexandrescu AT (2001) An NMR-based quenched hydrogen exchange investigation of model amyloid fibrils formed by cold shock protein A. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 67-78.

Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* **181**: 223-230.

Aylon Y & Oren M (2007) Living with p53, dying of p53. *Cell* **130**: 597-600.

Bakowies D & Van Gunsteren WF (2002) Water in protein cavities: A procedure to identify internal water and exchange pathways and application to fatty acid-binding protein. *Proteins* **47**: 534-545.

Banci L, Bertini I, D'Amelio N*, et al.* (2005) Fully metallated S134N Cu,Zn-superoxide dismutase displays abnormal mobility and intermolecular contacts in solution. *The Journal of biological chemistry* **280**: 35815-35821.

Bartlett AI & Radford SE (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nature structural & molecular biology* **16**: 582-588.

Beerten J, Schymkowitz J & Rousseau F (2013) Aggregation prone regions and gatekeeping residues in protein sequences. *Current topics in medicinal chemistry*.

Beerten J, Jonckheere W, Rudyak S*, et al.* (2012) Aggregation gatekeepers modulate protein homeostasis of aggregating sequences and affect bacterial fitness. *Protein engineering, design & selection : PEDS* **25**: 357-366.

Bemporad F, De Simone A, Chiti F & Dobson CM (2012) Characterizing intermolecular interactions that initiate native-like protein aggregation. *Biophysical journal* **102**: 2595-2604.

Berson JF, Theos AC, Harper DC, Tenza D, Raposo G & Marks MS (2003) Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. *The Journal of cell biology* **161**: 521-533.

Biancalana M & Koide S (2010) Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochimica et biophysica acta* **1804**: 1405-1412.

Booth DR, Sunde M, Bellotti V*, et al.* (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**: 787-793.

Bouchard M, Zurdo J, Nettleton EJ, Dobson CM & Robinson CV (2000) Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein science : a publication of the Protein Society* **9**: 1960-1967.

Brockwell DJ & Radford SE (2007) Intermediates: ubiquitous species on folding energy landscapes? *Current opinion in structural biology* **17**: 30-37.

Budyak IL, Krishnan B, Marcelino-Cruz AM, Ferrolino MC, Zhuravleva A & Gierasch LM (2013) Early Folding Events Protect Aggregation-Prone Regions of a beta-Rich Protein. *Structure*.

Calamai M, Chiti F & Dobson CM (2005) Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophysical journal* **89**: 4201-4210.

Calloni G, Chen T, Schermann SM*, et al.* (2012) DnaK functions as a central hub in the E. coli chaperone network. *Cell reports* **1**: 251-264.

Carrio M, Gonzalez-Montalban N, Vera A, Villaverde A & Ventura S (2005) Amyloid-like properties of bacterial inclusion bodies. *Journal of molecular biology* **347**: 1025-1037.

Castillo V, Grana-Montes R & Ventura S (2011) The aggregation properties of Escherichia coli proteins associated with their cellular abundance. *Biotechnology journal* **6**: 752-760.

Chapman MR, Robinson LS, Pinkner JS*, et al.* (2002) Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science* **295**: 851-855.

Chen Y & Dokholyan NV (2008) Natural selection against protein aggregation on self-interacting and essential proteins in yeast, fly, and worm. *Molecular biology and evolution* **25**: 1530-1533.

Chiti F & Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annual review of biochemistry* **75**: 333-366.

Chiti F & Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nature chemical biology* **5**: 15-22.

Chiti F, Stefani M, Taddei N, Ramponi G & Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**: 805-808.

Clare DK, Vasishtan D, Stagg S*, et al.* (2012) ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell* **149**: 113-123.

Clark PL (2004) Protein folding in the cell: reshaping the folding funnel. *Trends in biochemical sciences* **29**: 527-534.

Clark PL, Weston BF & Gierasch LM (1998) Probing the folding pathway of a beta-clam protein with single-tryptophan constructs. *Fold Des* **3**: 401-412.

Clark PL, Liu ZP, Zhang J & Gierasch LM (1996) Intrinsic tryptophans of CRABPI as probes of structure and folding. *Protein Sci* **5**: 1108-1117.

Clark PL, Liu ZP, Rizo J & Gierasch LM (1997) Cavity formation before stable hydrogen bonding in the folding of a beta-clam protein. *Nat Struct Biol* **4**: 883-886.

Colon W & Kelly JW (1992) Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* **31**: 8654-8660.

Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X & Ventura S (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC bioinformatics* **8**: 65.

Csermely P, Palotai R & Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in biochemical sciences* **35**: 539-546.

Dasari M, Espargaro A, Sabate R*, et al.* (2011) Bacterial Inclusion Bodies of Alzheimer's Disease beta-Amyloid Peptides Can Be Employed To Study Native-Like Aggregation Intermediate States. *Chembiochem : a European journal of chemical biology*.

Dasari M, Espargaro A, Sabate R*, et al.* (2011) Bacterial inclusion bodies of Alzheimer's disease beta-amyloid peptides can be employed to study native-like aggregation intermediate states. *Chembiochem : a European journal of chemical biology* **12**: 407-423.

De Baets G, Reumers J, Delgado Blanco J, Dopazo J, Schymkowitz J & Rousseau F (2011) An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. *PLoS computational biology* **7**: e1002090.

de Groot NS & Ventura S (2010) Protein aggregation profile of the bacterial cytosol. *PloS one* **5**: e9383.

de Groot NS, Sabate R & Ventura S (2009) Amyloids in bacterial inclusion bodies. *Trends in biochemical sciences* **34**: 408-416.

de Groot NS, Aviles FX, Vendrell J & Ventura S (2006) Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. *The FEBS journal* **273**: 658-668.

de Groot NS, Espargaro A, Morell M & Ventura S (2008) Studies on bacterial inclusion bodies. *Future microbiology* **3**: 423-435.

de Groot NS, Castillo V, Grana-Montes R & Ventura S (2012) AGGRESCAN: method, application, and perspectives for drug design. *Methods in molecular biology* **819**: 199-220.

De Simone A, Dhulesia A, Soldi G, Vendruscolo M, Hsu ST, Chiti F & Dobson CM (2011) Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 21057-21062.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**: 277-293.

Dill KA & Chan HS (1997) From Levinthal to pathways to funnels. *Nature structural biology* **4**: 10-19.

Dobson CM (2003) Protein folding and misfolding. *Nature* **426**: 884-890.

Donovan M, Olofsson B, Gustafson AL, Dencker L & Eriksson U (1995) The cellular retinoic acid binding proteins. *The Journal of steroid biochemistry and molecular biology* **53**: 459-465.

Eakin CM, Berman AJ & Miranker AD (2006) A native to amyloidogenic transition regulated by a backbone trigger. *Nature structural & molecular biology* **13**: 202-208.

Ellis RJ & Minton AP (2006) Protein aggregation in crowded environments. *Biological chemistry* **387**: 485-497.

Eyles SJ & Gierasch LM (2000) Multiple roles of prolyl residues in structure and folding. *J Mol Biol* **301**: 737-747.

Eyles SJ & Kaltashov IA (2004) Methods to study protein dynamics and folding by mass spectrometry. *Methods* **34**: 88-99.

Eyles SJ, Dresch T, Gierasch LM & Kaltashov IA (1999) Unfolding dynamics of a beta-sheet protein studied by mass spectrometry. *J Mass Spectrom* **34**: 1289-1295.

Fernandez-Escamilla AM, Rousseau F, Schymkowitz J & Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature biotechnology* **22**: 1302-1306.

Fiorella PD & Napoli JL (1991) Expression of cellular retinoic acid binding protein (CRABP) in Escherichia coli. Characterization and evidence that holo-CRABP is a substrate in retinoic acid metabolism. *The Journal of biological chemistry* **266**: 16572-16579.

Frare E, Mossuto MF, Polverino de Laureto P, Dumoulin M, Dobson CM & Fontana A (2006) Identification of the core structure of lysozyme amyloid fibrils by proteolysis. *Journal of molecular biology* **361**: 551-561.

Frauenfelder H, Sligar SG & Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* **254**: 1598-1603.

Frauenfelder H, Fenimore PW & Young RD (2007) Protein dynamics and function: insights from the energy landscape and solvent slaving. *IUBMB life* **59**: 506-512.

Fujiwara K, Ishihama Y, Nakahigashi K, Soga T & Taguchi H (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. *The EMBO journal* **29**: 1552-1564.

Gales L, Cortes L, Almeida C*, et al.* (2005) Towards a structural understanding of the fibrillization pathway in Machado-Joseph's disease: trapping early oligomers of non-expanded ataxin-3. *Journal of molecular biology* **353**: 642-654.

Garcia-Fruitos E, Gonzalez-Montalban N, Morell M*, et al.* (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial cell factories* **4**: 27.

Gershenson A & Gierasch LM (2011) Protein folding in the cell: challenges and progress. *Curr Opin Struct Biol* **21**: 32-41.

Gierasch LM & Gershenson A (2009) Post-reductionist protein science, or putting Humpty Dumpty back together again. *Nature chemical biology* **5**: 774-777.

Grana-Montes R, Sant'anna de Oliveira R & Ventura S (2012) Protein aggregation profile of the human kinome. *Frontiers in physiology* **3**: 438.

Gutierrez-Magdaleno G, Bello M, Portillo-Tellez MC, Rodriguez-Romero A & Garcia-Hernandez E (2013) Ligand binding and self-association cooperativity of beta-lactoglobulin. *Journal of molecular recognition : JMR* **26**: 67-75.

Hamada D & Dobson CM (2002) A kinetic study of beta-lactoglobulin amyloid fibril formation promoted by urea. *Protein science : a publication of the Protein Society* **11**: 2417-2426.

Hamada D, Tanaka T, Tartaglia GG*, et al.* (2009) Competition between folding, native-state dimerisation and amyloid aggregation in beta-lactoglobulin. *Journal of molecular biology* **386**: 878-890.

Hartl FU & Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. *Nature structural & molecular biology* **16**: 574-581.

Hartl FU, Bracher A & Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. *Nature* **475**: 324-332.

Henzler-Wildman K & Kern D (2007) Dynamic personalities of proteins. *Nature* **450**: 964-972.

Higashimoto Y, Asanomi Y, Takakusagi S*, et al.* (2006) Unfolding, aggregation, and amyloid formation by the tetramerization domain from mutant p53 associated with lung cancer. *Biochemistry* **45**: 1608-1619.

Hilton GR, Lioe H, Stengel F, Baldwin AJ & Benesch JL (2013) Small heat-shock proteins: paramedics of the cell. *Topics in current chemistry* **328**: 69-98.

Hirota-Nakaoka N, Hasegawa K, Naiki H & Goto Y (2003) Dissolution of beta2-microglobulin amyloid fibrils by dimethylsulfoxide. *Journal of biochemistry* **134**: 159-164.

Hoofnagle AN, Resing KA & Ahn NG (2004) Practical methods for deuterium exchange/mass spectrometry. *Methods in molecular biology* **250**: 283-298.

Hoshino M, Katou H, Yamaguchi K & Goto Y (2007) Dimethylsulfoxide-quenched hydrogen/deuterium exchange method to study amyloid fibril structure. *Biochimica et biophysica acta* **1768**: 1886-1899.

Hoshino M, Katou H, Hagihara Y, Hasegawa K, Naiki H & Goto Y (2002) Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange. *Nature structural biology* **9**: 332-336.

Hoshino M, Katou H, Hagihara Y, Hasegawa K, Naiki H & Goto Y (2002) Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange. *Nat Struct Biol* **9**: 332-336.

Ignatova Z & Gierasch LM (2004) Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc Natl Acad Sci U S A* **101**: 523-528.

Ignatova Z & Gierasch LM (2005) Aggregation of a slow-folding mutant of a beta-clam protein proceeds through a monomeric nucleus. *Biochemistry* **44**: 7266-7274.

Ignatova Z & Gierasch LM (2006) Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proc Natl Acad Sci U S A* **103**: 13357-13361.

Ignatova Z & Gierasch LM (2009) A method for direct measurement of protein stability in vivo. *Methods Mol Biol* **490**: 165-178.

Ignatova Z, Krishnan B, Bombardier JP, Marcelino AM, Hong J & Gierasch LM (2007) From the test tube to the cell: exploring the folding and aggregation of a beta-clam protein. *Biopolymers* **88**: 157-163.

Ishimaru D, Andrade LR, Teixeira LS*, et al.* (2003) Fibrillar aggregates of the tumor suppressor p53 core domain. *Biochemistry* **42**: 9022-9027.

Jahn TR & Radford SE (2008) Folding versus aggregation: polypeptide conformations on competing pathways. *Archives of biochemistry and biophysics* **469**: 100-117.

Jahn TR, Parker MJ, Homans SW & Radford SE (2006) Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nature structural & molecular biology* **13**: 195-201.

Jamison RS, Newcomer ME & Ong DE (1994) Cellular retinoid-binding proteins: limited proteolysis reveals a conformational change upon ligand binding. *Biochemistry* **33**: 2873-2879.

Jamison RS, Newcomer ME & Ong DE (1998) Detection of conformational changes in cellular retinoid-binding proteins by limited proteolysis. *Methods in molecular biology* **89**: 165-176.

Jaravine VA, Zhuravleva AV, Permi P, Ibraghimov I & Orekhov VY (2008) Hyperdimensional NMR spectroscopy with nonlinear sampling. *Journal of the American Chemical Society* **130**: 3927-3936.

Joerger AC & Fersht AR (2008) Structural biology of the tumor suppressor p53. *Annual review of biochemistry* **77**: 557-582.

Johnson ES (2004) Protein modification by SUMO. *Annual review of biochemistry* **73**: 355-382.

Karplus M (1997) The Levinthal paradox: yesterday and today. *Folding & design* **2**: S69-75.

Kelly JW (1996) Alternative conformations of amyloidogenic proteins govern their behavior. *Current opinion in structural biology* **6**: 11-17.

Kelly JW (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Current opinion in structural biology* **8**: 101-106.

Kerner MJ, Naylor DJ, Ishihama Y*, et al.* (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **122**: 209-220.

Kheterpal I, Cook KD & Wetzel R (2006) Hydrogen/deuterium exchange mass spectrometry analysis of protein aggregates. *Methods in enzymology* **413**: 140-166.

Kleywegt GJ, Bergfors T, Senn H, Le Motte P, Gsell B, Shudo K & Jones TA (1994) Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. *Structure* **2**: 1241-1258.

Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **44**: 98-104.

Krishna MM, Hoang L, Lin Y & Englander SW (2004) Hydrogen exchange methods to study protein folding. *Methods* **34**: 51-64.

Krishnan VV, Sukumar M, Gierasch LM & Cosman M (2000) Dynamics of cellular retinoic acid binding protein I on multiple time scales with implications for ligand binding. *Biochemistry* **39**: 9119-9129.

Krumova P, Meulmeester E, Garrido M*, et al.* (2011) Sumoylation inhibits alpha-synuclein aggregation and toxicity. *The Journal of cell biology* **194**: 49-60.

Kurnik M, Hedberg L, Danielsson J & Oliveberg M (2012) Folding without charges. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 5705-5710.

Lawrence MS, Phillips KJ & Liu DR (2007) Supercharging proteins can impart unusual resilience. *Journal of the American Chemical Society* **129**: 10110-10112.

Levin LB, Ganoth A, Amram S, Nachliel E, Gutman M & Tsfadia Y (2010) Insight into the interaction sites between fatty acid binding proteins and their ligands. *Journal of molecular modeling* **16**: 929-938.

LeVine H, 3rd (1999) Quantification of beta-sheet amyloid fibril structures with thioflavin T. *Methods in enzymology* **309**: 274-284.

Li E (1999) Structure and function of cytoplasmic retinoid binding proteins. *Molecular and cellular biochemistry* **192**: 105-108.

Li H & Frieden C (2006) Fluorine-19 NMR studies on the acid state of the intestinal fatty acid binding protein. *Biochemistry* **45**: 6272-6278.

Linding R, Schymkowitz J, Rousseau F, Diella F & Serrano L (2004) A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *Journal of molecular biology* **342**: 345-353.

Liu ZP, Rizo J & Gierasch LM (1994) Equilibrium folding studies of cellular retinoic acid binding protein, a predominantly beta-sheet protein. *Biochemistry* **33**: 134-142.

Lu J, Lin CL, Tang C, Ponder JW, Kao JL, Cistola DP & Li E (2000) Binding of retinol induces changes in rat cellular retinol-binding protein II conformation and backbone dynamics. *Journal of molecular biology* **300**: 619-632.

Maji SK, Perrin MH, Sawaya MR*, et al.* (2009) Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**: 328-332.

Marcelino AM & Gierasch LM (2008) Roles of beta-turns in protein folding: from peptide models to protein engineering. *Biopolymers* **89**: 380-391.

Marcelino AM, Smock RG & Gierasch LM (2006) Evolutionary coupling of structural and functional sequence information in the intracellular lipid-binding protein family. *Proteins* **63**: 373-384.

Masino L, Nicastro G, Calder L, Vendruscolo M & Pastore A (2011) Functional interactions as a survival strategy against abnormal aggregation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **25**: 45-54.

Masino L, Nicastro G, Menon RP, Dal Piaz F, Calder L & Pastore A (2004) Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3. *Journal of molecular biology* **344**: 1021-1035.

Massolini G & Calleri E (2003) Survey of binding properties of fatty acid-binding proteins. Chromatographic methods. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **797**: 255-268.

Matos CA, de Macedo-Ribeiro S & Carvalho AL (2011) Polyglutamine diseases: the special case of ataxin-3 and Machado-Joseph disease. *Progress in neurobiology* **95**: 26-48.

Mayer MP (2010) Gymnastics of molecular chaperones. *Molecular cell* **39**: 321-331.

Meredith SC (2005) Protein denaturation and aggregation: Cellular responses to denatured and aggregated proteins. *Annals of the New York Academy of Sciences* **1066**: 181-221.

Mihajlovic M & Lazaridis T (2007) Modeling fatty acid delivery from intestinal fatty acid binding protein to a membrane. *Protein science : a publication of the Protein Society* **16**: 2042-2055.

Milanesi L, Waltho JP, Hunter CA*, et al.* (2012) Measurement of energy landscape roughness of folded and unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 19563-19568.

Moll UM, Ostermeyer AG, Haladay R, Winkfield B, Frazier M & Zambetti G (1996) Cytoplasmic sequestration of wild-type p53 protein impairs the G1 checkpoint after DNA damage. *Molecular and cellular biology* **16**: 1126-1137.

Morell M, Bravo R, Espargaro A, Sisquella X, Aviles FX, Fernandez-Busquets X & Ventura S (2008) Inclusion bodies: specificity in their aggregation process and amyloid-like structure. *Biochimica et biophysica acta* **1783**: 1815-1825.

Morshedi D, Ebrahim-Habibi A, Moosavi-Movahedi AA & Nemat-Gorgani M (2010) Chemical modification of lysine residues in lysozyme may dramatically influence its amyloid fibrillation. *Biochimica et biophysica acta* **1804**: 714-722.

Neudecker P, Robustelli P, Cavalli A*, et al.* (2012) Structure of an intermediate state in protein folding and aggregation. *Science* **336**: 362-366.

Niwa T, Ying BW, Saito K, Jin W, Takada S, Ueda T & Taguchi H (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 4201-4206.

Norris AW, Cheng L, Giguere V, Rosenberger M & Li E (1994) Measurement of subnanomolar retinoic acid binding affinities for cellular retinoic acid binding proteins by fluorometric titration. *Biochimica et biophysica acta* **1209**: 10-18.

Noy N (2000) Retinoid-binding proteins: mediators of retinoid action. *The Biochemical journal* **348 Pt 3**: 481-495.

Olofsson A, Ippel JH, Wijmenga SS, Lundgren E & Ohman A (2004) Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy. *The Journal of biological chemistry* **279**: 5699-5707.

Orekhov VY, Ibraghimov I & Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. *Journal of biomolecular NMR* **27**: 165-173.

Ostermeyer AG, Runko E, Winkfield B, Ahn B & Moll UM (1996) Cytoplasmically sequestered wild-type p53 protein in neuroblastoma is relocated to the nucleus by a C-terminal peptide. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 15190-15194.

Pagano K, Bemporad F, Fogolari F, Esposito G, Viglino P, Chiti F & Corazza A (2010) Structural and dynamics characteristics of acylphosphatase from Sulfolobus solfataricus in the monomeric state and in the initial native-like aggregates. *The Journal of biological chemistry* **285**: 14689-14700.

Pastore A & Temussi PA (2012) The two faces of Janus: functional interactions and protein aggregation. *Current opinion in structural biology* **22**: 30-37.

Pechmann S, Levy ED, Tartaglia GG & Vendruscolo M (2009) Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 10159-10164.

Peroutka Iii RJ, Orcutt SJ, Strickler JE & Butt TR (2011) SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes. *Methods in molecular biology* **705**: 15-30.

Perry JJ, Shin DS, Getzoff ED & Tainer JA (2010) The structural biochemistry of the superoxide dismutases. *Biochimica et biophysica acta* **1804**: 245-262.

Platt GW & Radford SE (2009) Glimpses of the molecular mechanisms of beta2-microglobulin fibril formation in vitro: aggregation on a complex energy landscape. *FEBS letters* **583**: 2623-2629.

Powers ET & Balch WE (2013) Diversity in the origins of proteostasis networks - a driver for protein function in evolution. *Nature reviews. Molecular cell biology*.

Powers ET, Powers DL & Gierasch LM (2012) FoldEco: A Model for Proteostasis in E. coli. *Cell Rep* **1**: 265-276.

Powers ET, Powers DL & Gierasch LM (2012) FoldEco: a model for proteostasis in E. coli. *Cell reports* **1**: 265-276.

Ratajczak E, Zietkiewicz S & Liberek K (2009) Distinct activities of Escherichia coli small heat shock proteins IbpA and IbpB promote efficient protein disaggregation. *Journal of molecular biology* **386**: 178-189.

Reumers J, Maurer-Stroh S, Schymkowitz J & Rousseau F (2009) Protein sequences encode safeguards against aggregation. *Human mutation* **30**: 431-437.

Rezaei-Ghaleh N, Blackledge M & Zweckstetter M (2012) Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *Chembiochem : a European journal of chemical biology* **13**: 930-950.

Richardson JS & Richardson DC (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2754-2759.

Richter K, Haslbeck M & Buchner J (2010) The heat shock response: life on the verge of death. *Molecular cell* **40**: 253-266.

Rigacci S, Bucciantini M, Relini A, Pesce A, Gliozzi A, Berti A & Stefani M (2008) The (1-63) region of the p53 transactivation domain aggregates in vitro into cytotoxic amyloid assemblies. *Biophysical journal* **94**: 3635-3646.

Rizo J, Liu ZP & Gierasch LM (1994) 1H and 15N resonance assignments and secondary structure of cellular retinoic acid-binding protein with and without bound ligand. *Journal of biomolecular NMR* **4**: 741-760.

Rotondi KS & Gierasch LM (2003) Role of local sequence in the folding of cellular retinoic abinding protein I: structural propensities of reverse turns. *Biochemistry* **42**: 7976-7985.

Rousseau F, Serrano L & Schymkowitz JW (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. *Journal of molecular biology* **355**: 1037-1047.

Routledge KE, Tartaglia GG, Platt GW, Vendruscolo M & Radford SE (2009) Competition between intramolecular and intermolecular interactions in an amyloid-forming protein. *Journal of molecular biology* **389**: 776-786.

Sabate R, Espargaro A, Grana-Montes R, Reverter D & Ventura S (2012) Native structure protects SUMO proteins from aggregation into amyloid fibrils. *Biomacromolecules* **13**: 1916-1926.

Sacchettini JC & Gordon JI (1993) Rat intestinal fatty acid binding protein. A model system for analyzing the forces that can bind fatty acids to proteins. *The Journal of biological chemistry* **268**: 18399-18402.

Sacchettini JC, Scapin G, Gopaul D & Gordon JI (1992) Refinement of the structure of Escherichia coli-derived rat intestinal fatty acid binding protein with bound oleate to 1.75-A resolution. Correlation with the structures of the apoprotein and the protein with bound palmitate. *The Journal of biological chemistry* **267**: 23534-23545.

Saibil HR (2008) Chaperone machines in action. *Current opinion in structural biology* **18**: 35-42.

Sanchez de Groot N, Torrent M, Villar-Pique A, Lang B, Ventura S, Gsponer J & Babu MM (2012) Evolutionary selection for protein aggregation. *Biochemical Society transactions* **40**: 1032-1037.

Scholtz JM & Robertson AD (1995) Hydrogen exchange techniques. *Methods in molecular biology* **40**: 291-311.

Serpell LC, Blake CC & Fraser PE (2000) Molecular structure of a fibrillar Alzheimer's A beta fragment. *Biochemistry* **39**: 13269-13275.

Soldi G, Bemporad F & Chiti F (2008) The degree of structural protection at the edge beta-strands determines the pathway of amyloid formation in globular proteins. *Journal of the American Chemical Society* **130**: 4295-4302.

Stachurska E, Loboda A, Niderla-Bielinska J*, et al.* (2011) Expression of cellular retinoic acid-binding protein I and II (CRABP I and II) in embryonic mouse hearts treated with retinoic acid. *Acta biochimica Polonica* **58**: 19-29.

Sugiyama M, Fujii N, Morimoto Y, Itoh K, Mori K & Fukunaga T (2010) SAXS and SANS observations of abnormal aggregation of human alpha-crystallin. *Chemistry & biodiversity* **7**: 1380-1388.

Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB & Blake CC (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of molecular biology* **273**: 729-739.

Tartaglia GG & Vendruscolo M (2008) The Zyggregator method for predicting protein aggregation propensities. *Chemical Society reviews* **37**: 1395-1401.

Tartaglia GG & Vendruscolo M (2010) Proteome-level interplay between folding and aggregation propensities of proteins. *Journal of molecular biology* **402**: 919-928.

Tartaglia GG, Pechmann S, Dobson CM & Vendruscolo M (2007) Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends in biochemical sciences* **32**: 204-206.

Thompson JR, Bratt JM & Banaszak LJ (1995) Crystal structure of cellular retinoic acid binding protein I shows increased access to the binding cavity due to formation of an intermolecular beta-sheet. *Journal of molecular biology* **252**: 433-446.

Trovato A, Chiti F, Maritan A & Seno F (2006) Insight into the structure of amyloid fibrils from the analysis of globular proteins. *PLoS computational biology* **2**: e170.

Tycko R (2006) Molecular structure of amyloid fibrils: insights from solid-state NMR. *Quarterly reviews of biophysics* **39**: 1-55.

Uhrinova S, Smith MH, Jameson GB, Uhrin D, Sawyer L & Barlow PN (2000) Structural changes accompanying pH-induced dissociation of the beta-lactoglobulin dimer. *Biochemistry* **39**: 3565-3574.

Vendruscolo M (2012) Proteome folding and aggregation. *Current opinion in structural biology* **22**: 138-143.

Vendruscolo M & Dobson CM (2007) Chemical biology: More charges against aggregation. *Nature* **449**: 555.

Ventura S & Villaverde A (2006) Protein quality in bacterial inclusion bodies. *Trends in biotechnology* **24**: 179-185.

Vilar M, Wang L & Riek R (2012) Structural studies of amyloids by quenched hydrogen-deuterium exchange by NMR. *Methods in molecular biology* **849**: 185-198.

Villanueva J, Hoshino M, Katou H, Kardos J, Hasegawa K, Naiki H & Goto Y (2004) Increase in the conformational flexibility of beta 2-microglobulin upon copper binding: a possible role for copper in dialysis-related amyloidosis. *Protein science : a publication of the Protein Society* **13**: 797-809.

Vousden KH & Lane DP (2007) p53 in health and disease. *Nature reviews. Molecular cell biology* **8**: 275-283.

Wang L, Maji SK, Sawaya MR, Eisenberg D & Riek R (2008) Bacterial inclusion bodies contain amyloid-like structure. *PLoS biology* **6**: e195.

Wang Z, Li H, Guan W, Ling H, Mu T, Shuler FD & Fang X (2010) Human SUMO fusion systems enhance protein expression and solubility. *Protein expression and purification* **73**: 203-208.

Westermark GT, Johnson KH & Westermark P (1999) Staining methods for identification of amyloid in tissue. *Methods in enzymology* **309**: 3-25.

Wickner RB, Edskes HK, Bateman DA, Kelly AC, Gorkovskiy A, Dayani Y & Zhou A (2013) Amyloids and yeast prion biology. *Biochemistry* **52**: 1514-1527.

Wlodarski T & Zagrovic B (2009) Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 19346-19351.

Woolf TB, Grossfield A & Tychko M (2000) Differences between apo and three holo forms of the intestinal fatty acid binding protein seen by molecular dynamics computer calculations. *Biophysical journal* **78**: 608-625.

Xiao H & Kaltashov IA (2005) Transient structural disorder as a facilitator of protein-ligand binding: native H/D exchange-mass spectrometry study of cellular retinoic acid binding protein I. *J Am Soc Mass Spectrom* **16**: 869-879.

Xiao H & Kaltashov IA (2005) Transient structural disorder as a facilitator of protein-ligand binding: native H/D exchange-mass spectrometry study of cellular retinoic acid binding protein I. *Journal of the American Society for Mass Spectrometry* **16**: 869-879.

Yamaguchi K, Katou H, Hoshino M, Hasegawa K, Naiki H & Goto Y (2004) Core and heterogeneity of beta2-microglobulin amyloid fibrils as revealed by H/D exchange. *Journal of molecular biology* **338**: 559-571.

Yoshimura Y, Lin Y, Yagi H*, et al.* (2012) Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 14446-14451.

Zhang G & Ignatova Z (2011) Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Current opinion in structural biology* **21**: 25-31.

Zhang J, Liu ZP, Jones TA, Gierasch LM & Sambrook JF (1992) Mutating the charged residues in the binding pocket of cellular retinoic acid-binding protein simultaneously reduces its binding affinity to retinoic acid and increases its thermostability. *Proteins* **13**: 87-99.

Zhang YZ, Paterson Y & Roder H (1995) Rapid amide proton exchange rates in peptides and proteins measured by solvent quenching and two-dimensional NMR. *Protein science : a publication of the Protein Society* **4**: 804-814.

Zhang Z, Witham S & Alexov E (2011) On the role of electrostatics in protein-protein interactions. *Physical biology* **8**: 035001.

Zhu L, Kurian E, Prendergast FG & Kemple MD (1999) Dynamics of palmitic acid complexed with rat intestinal fatty acid binding protein. *Biochemistry* **38**: 1554-1561.

Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME & Hendrickson WA (1996) Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**: 1606-1614.

Zimmerman AW & Veerkamp JH (2002) New insights into the structure and function of fatty acid-binding proteins. *Cellular and molecular life sciences : CMLS* **59**: 1096-1116.