MECHANISMS OF PROTEIN AGGREGATION AND REFOLDING

by

JEFFREY LYNN CLELAND

B.S. Chemical Engineering
University of California, Davis
(June 1986)

Submitted to the Department of Chemical Engineering in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

April, 1991

Signiture of Author_____
Department of Chemical Engineering
April 2, 1991

Certified by_____
Professor Daniel I.C. Wang
Thesis Supervisor, Department of Chemical Engineering

Accepted by_____
Professor William Deen
Chairman, Departmental Committee for Graduate Students

MECHANISMS OF PROTEIN AGGREGATION AND REFOLDING

BY
JEFFREY LYNN CLELAND

Submitted to the Department of Chemical Engineering
Massachusetts Institute of Technology
on April 2, 1991
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Chemical Engineering

ABSTRACT

The mechanisms of protein aggregation during refolding were determined for a model protein, bovine carbonic anhydrase B (CAB). Both kinetic and thermodynamic studies were used to develop an understanding of the aggregation process. The kinetics of refolding and aggregation were elucidated for four different regimes which were dependent on the final protein and denaturant concentrations. During these studies, the first intermediate in the refolding pathway of CAB was observed to aggregate and form both dimers and trimers prior to precipitation. The first intermediate also formed dimers and trimers at equilibrium in 2 M guanidine hydrochloride (GuHCl) at high protein concentrations (>10 μM). From kinetic and thermodynamic studies, model pathways were developed and confirmed for the refolding of CAB at different final GuHCl and protein concentrations.

To improve the recovery of active protein, polyethylene glycol (PEG) was successfully used to prevent aggregation during refolding. PEG was observed to enhance refolding at stoichiometric concentrations (< 3 g/l) which were dependent on the polymer molecular weight. Equilibrium binding studies showed that PEG bound to a single site on the first intermediate in the refolding pathway of CAB. In addition, equilibrium protein association studies in 2.0 M GuHCl showed that low concentrations of PEG (< 30 g/l) reduced aggregation of the first intermediate. The rate of refolding in PEG solutions was observed to be equivalent to rate of refolding measured at nonassociating conditions. A model was successfully developed which included the formation of a nonassociating PEG - first intermediate complex.

The general application of PEG enhanced refolding was assessed with three recombinant human proteins: deoxyribonuclease (rhDNAse), tissue plasminogen activator (rtPA), and gamma interferon (rIFN-γ). Refolding of rhDNAse and rtPA with PEG (3350 MW) resulted in prevention of aggregation and increased recovery of activity. A relationship was also observed between the PEG concentration required for enhancement and the hydrophobicity of the protein.

Thesis supervisor: Dr. Daniel I.C. Wang
Title: Chevron Professor of Chemical Engineering

TABLE OF CONTENTS

LITERATURE REVIEW (cont.)

MATERIALS AND METHODS (cont.)

RESULTS AND DISCUSSION (Cont.)

ACKNOWLEDGEMENTS

Several people contributed to the successful completion of this thesis as well as the development of my skills in biotechnology. I must start by thanking the person that comforted and assisted me through both the good and bad times. My wife, Ann, has helped me make it through every day and I would like to thank her for her love and support. This thesis is dedicated to her and our new baby boy, Derrick Anthony. Derrick will never know how much he catalyzed the completion of this thesis. Next, I would like to thank my parents, Mary and Lynn Cleland, for supporting me and providing me with the skills necessary to succeed. I also appreciate the assistance that my wife's parents, Sharon and Tom Barry, have provided for the past seven years. My other relatives and friends from U.C. Davis also played an essential role over the past five years.

The initial development of my skills in biotechnology must be attributed to the people that assisted me at Genentech. In particular, I would like to thank Bob Konopacz for providing me the opportunity to experience the biotechnology industry. I am also very grateful for all the guidance and support that both Jim Swartz and Armin Ramel have provided me. Their input in the initial stages of my graduate studies was essential to my development. In addition, Genentech contributed both recombinant deoxyribonuclease and tissue plasminogen activator for this work. I would also like to thank Stuart Builder and Marge Winkler for their assistance in the refolding work performed with these proteins.

Additional support was provided by Biogen. I appreciate the help and guidance of Parrish Gallaher at Biogen. I would like to thank Parrish Gallaher and Biogen for the donation of recombinant gamma interferon and the performance of several assays.

My survival in the early stages of graduate school can be attributed to a few students whom I would like to sincerely thank for their efforts. Jeff Butterbaugh, Paul Northey, Gino Grampp and Shin-gyoo Khan made the first and subsequent years more enjoyable. Matt Croughan, John Aunins, Brian Kell, Steve Lee, Jim McMillan, Max Kennedy, Neal Gordon and Noubar Afeyan taught me about life at BPEC. Anna Hagen provided me with initial insights into protein refolding. Ed Osawa and Mike Thien gave me the opportunity to become "Mr. BPEC". Ed Osawa also provided me with much needed frienship through the Mr. BPEC years and beyond.

The newer students have also taught me a great deal. Many thanks to Brian Kelley and Bruce Woodson for taking over the role of Mr. BPEC. Thanks also to Cliff Rutt and Eric Scharin for their friendship and intellectual conversations. Also, I must thank Jean-Francois Hamel, Bob Murray, Gary Husted, and Sergio Sanchez for teaching me about large scale operations. I would like to thank Peggy Foster, Luvenia Evelyn, and Rita Caissie for making the early mornings bearable. Many parts of this research must be attributed to several undergraduates. In particular, I would like to thank Ben Urmaza, Mark Hall, and Chester Hedgepeth for making my life a little easier. Finally, I would

like to thank all the other people that have been a part of my life at M.I.T.: Jack Prior, Craig Zupke, Mark Applegate, Wen Chiou, Dawn Orton, Christine Moore, Raju Gokuraju, Greg O'Connor, Rahul Singvi, Dave Chang, Dan Lasko, Per Lindell, and David Robbins.

In addition, several BPEC staff people made the task of purchasing new equipment and supplies enjoyable. I would especially like to thank Ruth Ayers, Sonia Foster, Audrey Childs, and Lynne Lenker for all their help.

Last, but certainly not least, my thesis committee members were absolutely essential in developing this thesis. First of all, I would like to thank my thesis supervisor, Professor Daniel I.C. Wang, for his patience and guidance through this work. Professor Wang taught me how to put the pieces of the puzzle together or in his words: "exchange your different types of currency for Swiss Francs". The thoughtful discussions and suggestions provided by Professor Peter Kim were essential to many parts of this thesis. In addition, the guidance provided by both Professors Charles Cooney and Robert Sauer was instrumental in designing the focus of this work at the outset. The late stages of this research could not have been accomplished without the support and direction of Professors Clark Colton and Alan Hatton.

Beyond my thesis committee, several other faculty were vital to this thesis. Professor Robert Langer provided financial support for a new QLS machine which significantly hastened this research. I would also like to thank Professor Langer for his helpful insights into formulations and our discussions ultimately led me to pursue a career in this area. I would like to thank Professor Ted Randolph at Yale University for his assistance with the electron spin resonance studies. In addition, I must thank Professor Jonathan King and Dr. Anna Mitraki for teaching me a great deal about protein folding. I look forward to continued research and collaborations with each faculty member as well as many of the students.

Finally, the Biotechnology Process Engineering Center has been an excellent facility for the development of students and provided me with support. This work was supported by National Science Foundation Grant CDR-88-03014. I hope that the National Science Foundation will continue to provide students with the opportunities which I have had throughout this thesis.

In conclusion, the following anonymous poem is a statement which every new student should use as a guide for thesis work.

Do not follow where
the path may lead
Go, instead, where
there is no path
And leave a trail.

LIST OF FIGURES

Literature Review

Results and Discussion

LIST OF TABLES

INTRODUCTION

The pharmaceutical industry has evolved over the past several years by applying the new tool of genetic engineering. Recombinant DNA techniques have allowed pharmaceutical companies to develop novel therapeutic proteins for the treatment of many diseases. Currently, only a few protein drugs have been approved by the Food and Drug Administration (FDA) and these include tissue plasminogen activator for the treatment of acute myocardial infarction (heart attack) and gamma interferon for the treatment of chronic granulomatous disease (white blood cell defect). Both of these therapeutic proteins have been studied in this work. In addition, another protein, deoxyribonuclease, has been used in this work and clinical trials are currently underway to test the ability of this drug to break down mucous in the lungs of cystic fibrosis patients. All of the above currently approved protein therapeutics are produced by recombinant methodologies.

Commercial Protein Production Methods

When applying genetic engineering to the production of proteins, the researcher has the option to use a variety of different host organisms. The most common host organisms are *Escherichia coli*, *Saccharomyces cerevisiae*, and Chinese Hamster Ovary (CHO) cells. Since the majority of proteins expressed by recombinant techniques are originally derived from eukaryotic organisms, CHO cells provide the most natural production method. In most cases, CHO cells will produce the desired protein with the same composition and structure as the natural protein. Unfortunately, the use of CHO cells for protein production is often cost intensive since these cells require complex media

for growth, have low productivities, and often result in the use of expensive purification techniques. As an alternative, yeast can be used as an expression system. Yeast will usually synthesize and process the protein in the correct conformation, but it will not provide the proper carbohydrate structures on the protein. If the protein does not require glycosylation for activity, it could be expressed and produced in *E. coli*. This host expression system provides the least expensive production method for many proteins. However, *E. coli* often does not produce the protein in its native state and, therefore, additional processing is required to refold the protein to regain its biological activity. If these additional processing costs can be reduced by increasing the yield of active protein during refolding, the production of therapeutic proteins in *E.coli* could be a very cost effective method. Therefore, one of the major goals of this work is to develop refolding methodologies which result in an increase in the recovery of active protein.

Before developing improved refolding methods, the general process of protein production in *E. coli* must be understood. The production of recombinant proteins in *E. coli* often results in the formation of insoluble protein aggregates called inclusion bodies. These inclusion bodies must be isolated from the cell through homogenization, centrifugation, and washing procedures which have been the focus of several patents (cf. Builder & Ogez, 1985). The inactive protein aggregates are then solubilized using denaturants such as urea or guanidine hydrochloride (GuHCl). During this solubilization procedure, the protein usually becomes completely unfolded to a random coil state. Therefore, the protein must be refolded by using techniques such as dilution or dialysis to remove the denaturant. This procedure often results in the reformation of insoluble

protein aggregates. The formation of protein aggregates during refolding can significantly reduce the overall yield of active protein. Thus, the major goal of this work is to develop a fundamental understanding of the phenomenon of protein aggregation during refolding.

Problem Description and Thesis Objectives

The use of bacterial systems such as *E. coli* for protein production often results in the formation of aggregates which must be denatured and refolded. During refolding of the protein from the denatured state, protein aggregates are formed. Protein aggregation during refolding reduces the yield of active protein and introduces other problems in purification. To improve recovery of active protein, it is necessary to develop an understanding of the mechanisms of protein aggregation during refolding. With a knowledge of these mechanisms, a model can be devised to assess the kinetics of aggregation during refolding. A valid model can then be used to design general methodologies which prevent aggregation during refolding. The methodology developed to prevent aggregation can then be studied to elucidate a relationship between the improved refolding methodology and the properties of the protein.

To achieve these goals, the thermodynamics and kinetics of protein aggregation have been extensively studied. The underlying mechanism of protein aggregation during refolding has been elucidated from these studies. In addition, a model of the refolding and aggregation process has been developed and used as a basis for understanding the interaction of solvent components with the protein. This model was also used to assist in the development of a general method to enhance refolding and reduce aggregation.

Methods of Refolding

Several methods of protein refolding have been developed in an attempt to reduce aggregation while recovering higher yields of active protein. Since inclusion body proteins are usually solubilized with a denaturant such as urea, guanidine hydrochloride (GuHCl), or sodium dodecyl sulfate (SDS), the denaturant must be removed to allow the protein to regain its native state. Denaturant removal has typically been performed by dilution or dialysis with a stabilizing buffer solution. However, both of these processes must be performed at low protein concentrations (μg/ml) and high denaturant concentrations to reduce the degree of aggregation. To avoid aggregation during refolding, several methods of refolding have been developed. Most of these methods operate on the principle of isolating each individual protein molecule and allowing it to refold before it can associate. For example, reversed micelles have been used to assist refolding by enclosing a single protein within each micelle and, then, removing the denaturant (Hagen et.al., 1990). Unfortunately, this system could not be used with hydrophobic proteins which interact with the surfactant. Another method for isolating denatured protein involved ion exchange chromatography. In this case, the denatured protein was bound to the column resin followed by denaturant removal (Creighton, 1985). Creighton has also suggested binding of the denatured protein to a solid support at a specific surface residue which would not interfere with refolding (Creighton, 1985). As an alternative to isolating the protein, attempts have been made to increase the rate of refolding. If the rate of refolding can be significantly increased, the aggregation of intermediates could be reduced. Refolding studies have therefore been performed with

ligands and antibodies which are specific to the native structure. In particular, creatine kinase was refolded in the presence of antibodies to the native structure. The antibody was observed to sterically interfere with the interaction between the protein domains which was required for refolding (Morris et.al., 1987). Other studies of the effect of antibodies on refolding have shown some success in stabilizing intermediate structures and increasing the refolding (Carlson et.al., 1990). In addition to antibodies, ligands and cofactors have been used to enhance refolding. For example, the rate of refolding of ferricytochrome c was enhanced by the extrinsic ligand for the axial position of the heme iron (Brems & Stellwagen, 1983). Each technique used for refolding has been limited to the specific protein studied and general protein refolding methods have not been developed to improve the recovery of active protein and reduce aggregation.

LITERATURE REVIEW

Review of Protein Folding *In vivo* and *In vitro*

To develop general methods of protein refolding, it is necessary to understand the mechanisms which govern folding both *in vivo* and *in vitro*. The pathway of folding is dependent on the host cell processing as well as the protein composition. In addition, *in vitro* folding is affected by the solvent environment and physical conditions such as temperature. To understand protein folding, one must have a rudimentary knowledge of a few specific principles which relate the final protein structure to its processing and environment.

Folding in Eukaryotic and Prokaryotic Hosts

Since most proteins can be expressed in their native form in eukaryotic hosts such as CHO cells, a comparison between eukaryotic and prokaryotic protein expression could provide insight into folding and inclusion body formation in prokaryotic systems. Protein expression in eukaryotic hosts involves several processing steps which have been described previously (Darnell et. al., 1986; Robinson & Austin, 1987). The cellular machinery involved in the production and export of proteins in eukaryotes diverges significantly from that used by *E. coli*. Several cellular compartments, post-translational modifications and folding aides such as molecular chaperones (discussed below) are present in eukaryotes. In particular, a majority of the protein folding reactions in animal cells occur in the endoplasmic reticulum (ER). Proteins designated for secretion in animal cells are initially synthesized with precursors which are subsequently removed in the ER.

The leader sequence or precursor segment usually consists of a small hydrophilic section followed by hydrophobic core segment (< 11 residues) (Sabatini et.al., 1982). This leader sequence is removed in the ER through a series of binding interactions with proteins in the ER as described previously (Walter et. al., 1981; Meyer, 1982). This procedure of protein processing by the ER usually results in the secretion of native protein. However, protein aggregation in the ER of mammalian cells has been observed to occur as the result of overexpression (Lodish, 1988). Overexpression could strain the cellular machinery required to process the protein and, thus, partially folded protein structures would accumulate. High levels of protein expression in yeast has also resulted in the formation of intracellular aggregates (Cousens et.al., 1987). In contrast to eukaryotic systems, protein expression in prokaryotes such as *E. coli* occurs by significantly different mechanisms as described in reviews by Randall (Randall & Hardy, 1984; Randall et. al., 1987). Recent studies have shown that the export of folded protein from *E. coli* must proceed by the cellular recognition of a leader sequence of the protein (Park et.al., 1988). However, attempts to produce secreted and properly folded protein from *E. coli* have had only limited success. In addition, it has been postulated that the coexpression of eukaryotic folding aides could increase folding yields in bacterial systems (Schein, 1989; Goloubinoff et. al., 1989). Unfortunately, this method may have limited utility and does not address the other environmental differences between folding in prokaryotes and eukaryotes (eg. differences in redox potential). As observed in some cases for eukaryotes, overexpression of proteins in *E. coli* can result in the formation of inclusion bodies irregardless of protein composition or origin, prokaryotic or eukaryotic (Marston, 1986).

Inclusion Body Formation

Since prokaryotic organisms do not have the necessary machinery for protein processing, production of eukaryotic proteins in prokaryotes such as *E. coli* often results in the formation of inactive or aggregated protein. Proteins expressed in *E. coli* can result in three different final forms (Kane & Hartley, 1988). The protein can be expressed as either a stable species which forms soluble native protein or an unstable species. The unstable intermediate species can then be degraded or accumulate in the form of inclusion bodies (Kane & Hartley, 1988). A recent review of inclusion body formation suggested that proteins with low proline content and long sequences of acidic amino acids do not form insoluble aggregates (Schein, 1989). However, this observation may not be generally valid as noted in reviews by Mitraki and coworkers (Mitraki & King, 1989; Mitraki et.al., 1991). Mitraki and King have studied the mechanisms of inclusion body formation and concluded that these aggregates are the result of partially folded intermediates (Mitraki & King, 1989). The aggregation of these intermediates was observed to be a function of the environment of the protein. In particular, increased temperatures were observed to enhance the formation of aggregates as discussed below (Mitraki et.al., 1991). Since the environment of the intermediate affects its aggregation properties, manipulation of the growth conditions for *E. coli* could result in modifications of inclusion body formation. For example, β-lactamase inclusion body formation which occurred in the periplasmic space of *E. coli* was reduced by the addition of nonmetabolized sugars such as sucrose and raffinose to the culture media (Georgiou et.al., 1986; Bowden & Georgiou, 1988). Therefore, it should be possible to control the extent

of inclusion body formation by modifications in the growth conditions. These modifications could include growth at low temperatures and addition of aggregation inhibitors such as sugars to the media which should result in higher concentrations of soluble protein.

Effect of Folding Mutations

In vivo folding and inclusion body formation have been altered by folding mutations which stabilize a folding intermediate. Reviews of the methodology for applying protein engineering to folding studies are useful in providing conceptual background (King, 1986; Goldenberg, 1988). An excellent genetic study of protein refolding *in vivo* has been published by King and coworkers. Initially, the trimeric tailspike protein from phage p22 was mutated such that the protein folding was temperature sensitive (Goldenberg et.al., 1983). Refolding at 40°C prevented the formation of the native trimer and resulted in the accumulation of partially folded intermediates and aggregates (King et.al., 1987). The temperature sensitive mutant was shown to form the native trimer and did not result in a thermolabile native state (Sturtevant et.al., 1989). Therefore, this mutant provided a mechanism for studying the *in vivo* folding pathway and the formation of inclusion bodies. In particular, *in vivo* refolding at thermolabile conditions resulted in the stabilization of an intermediate which aggregated to form inclusion bodies (Haase-Pettingell & King, 1988). These studies revealed that aggregation occurred through a hydrophobic intermediate species which was kinetically trapped by the temperature sensitive mutation. Recent work with this protein has shown that the temperature sensitive mutations can be suppressed by additional

modifications in the protein sequence (Mitraki et.al., 1991; Fane et.al., 1991). Therefore, it may be possible to make single amino acid substitutions in other proteins which will allow them to fold in the absence of inclusion body formation.

Genetic engineering approaches to provide refolding analysis have also been applied to *in vitro* protein refolding. As in the *in vivo* studies, the approach primarily involved the stabilization of an intermediate on the refolding pathway. For example, an intermediate of bovine growth hormone was stabilized by a point mutation. This mutation was postulated to have caused enhanced hydrophobic attraction and increased aggregation (Brems et.al., 1988a). In addition, the structure of an intermediate in the refolding of barnase, ribonuclease from *Bacillus amyloliquefaciens*, was elucidated by nuclear magnetic resonance (NMR) techniques and mutants which resulted in a stable intermediate that folded to the native state (Matoushek et.al., 1990). The comparison of mutant protein refolding results to the wild-type refolding pathway could be valid for many proteins since it has been shown that substitutions can be made which do not alter the native structure (Lim & Sauer, 1989; Bowie et.al., 1990). On the other hand, compact proteins such as dihydrofolate reductase which have hydrophobic cores can be destabilized by substitution of hydrophilic residues for core hydrophobic residues (Garvey & Matthews, 1989). As pointed out by Shortle, single site mutations of proteins can be used to probe folding pathways, but the effect of the mutation on the thermodynamics of the protein must also be studied to validate the analysis (Shortle, 1989). In general, protein engineering can be used to assist in the determination of refolding pathways and the structure of folding intermediates.

Molecular Chaperones

To avoid the aggregation of partially folded intermediates, eukaryotic organisms utilize proteins which bind to these intermediates. The binding proteins or chaperonins which have been observed in the *in vivo* folding pathway were initially believed to occur only when the cells were exposed to external stresses such as heat. Heat shock proteins such as the Hsp70 class of proteins have been shown to act under normal conditions by facilitating folding and these proteins may require energy in the form of ATP for activity (Beckman et.al., 1990). Chaperonins have been postulated to have many roles in cellular processes. These roles include catalysis of folding, prevention of aggregation, and removal of denatured non-host proteins. In addition, the onset of several diseases could be the result of a failure of these proteins to recognize foreign proteins (Horwich et.al., 1990). Recent reviews of chaperonins as mediators of folding have provided an overview of the mechanisms of the binding process (Randall et.al., 1987; Rothman, 1989). For example, Rothman has concluded that chaperonins operate primarily by binding to partially folded proteins in a stepwise manner which allows folding to occur while preventing aggregation (Rothman, 1989). Relationships between eukaryotic and prokaryotic binding proteins has also been proposed from a study of mitochondrial protein processing (Hartl & Neuport, 1990). *E. coli* has at least one protein which will bind partially folded proteins. The *E. coli* binding protein, SecB, will bind to partially folded proteins with high affinity, but it does not bind selectively (Hardy & Randall, 1991). Overall, chaperonins are likely to have the primary function of binding to partially folded protein structures with high affinity and low selectivity. Many of these proteins block the

formation of aggregates by binding to hydrophobic regions of the intermediate. Therefore, it should be possible to apply these proteins to *in vitro* refolding and design refolding aides which mimic the action of chaperonins.

In addition to simple binding, proteins which catalyze specific folding reactions have also been observed in eukaryotes. Two catalytic proteins have been studied in detail. Protein disulfide isomerase (PDI) has been observed to catalyze the formation of correct disulfide bonds during folding (Freedman, 1989). PDI has also been shown to act independently of the second observed catalytic protein, peptidyl-prolyl *cis-trans* isomerase (PPI) (Lang & Schmid, 1988). PPI catalyzed the slow proline isomerization step in the refolding of the S protein of bovine RNAse A and can be effective in catalyzing the folding of other proteins (Lang et.al., 1987). The slow step in folding of several proteins has been postulated to be the result of proline isomerization or disulfide bond formation (Stellwagen, 1979: Semisotnov et.al., 1990; Ptitsyn et.al., 1990; Kim & Baldwin, 1990). Therefore, PDI and PPI could be utilized to enhance refolding. Unfortunately, both proteins must be able to access the substrate residues and, therefore, the application of these proteins *in vitro* may be limited (Lang et.al., 1987). However, a better understanding of the mechanisms of these catalysts would assist in the design of improved folding processes both *in vivo* by protein engineering of the catalysts and *in vitro* with the development of methods to circumvent accessibility problems.

Models of In vitro Protein Refolding

The *in vitro* refolding of denatured proteins has resulted in the formulation of several different models of refolding (Karplus & Weaver, 1976; Kim & Baldwin, 1982;

King, 1989). These models include biased random search, nucleation-growth, diffusion-collision-adhesion, and sequential folding. From an analysis of several reviews of potential models for refolding, a general theme can be derived. First of all, the formation of secondary structure in the denatured protein occurs very rapidly in most cases and can be described by the diffusion-collision-adhesion model where microdomains are formed from a few residues and, then, collide to form a stable structure. According to the diffusion-collision theory of protein folding developed by Karplus and Weaver, microdomains are segments of unstable native secondary structure and consist of only a few amino acid residues (8-10) such that all of the conformational alternatives can be searched in a short time period (< 1 msec) (Karplus & Weaver, 1979). More recent development of this model has suggested that the diffusion-collision process occurs prior to the formation of an initial compact intermediate (Bashford et.al., 1988). In addition, this model has been partially confirmed by NMR studies of the initiation of protein folding (Wright et.al., 1988). The diffusion-collision model also complements the observations of a molten globule state which rapidly forms in the initial stages of refolding for many proteins (Ptitsyn, 1987; Kuwajima, 1989; Ptitsyn et. al., 1990). The compact globular structure has several general features which include a native-like secondary structure and exposed hydrophobic surfaces (Kuwajima, 1989). In addition, this compact state has been observed to be general to several proteins independent of class or composition (Ptitsyn et.al., 1990). The driving force for the diffusion-collision process to form the compact molten globule was postulated to be hydrophobic interactions (Bashford et.al., 1988; Kuwajima, 1989). This hypothesis is supported by the analysis of

protein structural interactions in a recent review (Dill, 1990). Dill concluded that hydrophobic interactions are the strongest driving forces for the formation of native protein structure. Another recent review by Kim and Baldwin led to the conclusion that the initial structural intermediates in refolding form quickly and that the rate limiting step in folding then becomes the correct "adhesion" of different segments of structure (Kim & Baldwin, 1990). In accordance with the diffusion-collision-adhesion model, the next reaction would then proceed with a pairing of the domains formed from microdomains to give a stable compact native structure (Karplus & Weaver, 1979). Domain pairing does not occur in small proteins such as ribonuclease or lysozyme since they do not contain multiple domains (Karplus & Weaver, 1976). However, the slow step in folding of single domain proteins may be the result of inherent structural limitations such as proline isomerization or correct disulfide bond formation (Kim & Baldwin, 1990). Overall, refolding may be described by the rapid formation of a hydrophobic compact structure followed by the slow shuffling of the protein to its native conformation.

In vitro Folding Dependence on Denaturant Concentration

The stabilizing interactions in the formation of the native protein structure are all affected by the environmental conditions during the refolding process. Therefore, any model or analysis must include these environmental parameters in the refolding process. These physical parameters include effects such as denaturant concentration, viscosity, temperature, ionic strength, and solvent properties. The denaturant concentration has been observed to be one of the critical factors in protein refolding. Denaturants such as urea and GuHCl have been shown to bind to proteins (Prakash et.al., 1981; Arakawa &

Timasheff, 1984). The removal of these denaturants from the protein surface is required to refold the protein. The denaturant bound to the protein is in equilibrium with the bulk denaturant. Therefore, as the bulk concentration of the denaturant is reduced, the equilibrium concentration of denaturant bound to the protein will also decrease allowing the protein to refold. During denaturant removal, the unfolded protein will form stable intermediates and their rates of formation are dependent on the structural properties of the protein, the denaturant removal rate, and the final denaturant concentration. Protein folding intermediates have been studied by incrementally reducing the denaturant concentration and measuring the equilibrium formation of protein structure (cf. Rodionova et.al., 1989; Brems, 1988b). Additional analysis of the effect of denaturant on refolding has been detailed by Creighton (Creighton, 1984). For example, the rate of refolding of bovine serum albumin was enhanced by decreasing the urea concentration below 2.0 M (Damodaran, 1987). The rate of refolding was also dependent on the final denaturant concentration for refolding of lactate dehydrogenase (LDH) (Rudolph et.al., 1979). In general, the optimum denaturant concentration and removal rate would be a balance between the formation of native structure and the reduction in misfolded intermediates or aggregates.

Temperature Effects on Refolding

In addition to the solvent environment, temperature can effect the rate and extent of refolding. In particular, two proteins, phosphoglycerate kinase (PGK) and aspartokinase-homoserine dehydrogenase (AK-HDH), have been shown to be dependent on temperature for refolding (Porter & Cardenas, 1980; Vaucheret et.al., 1987). The

extent of reactivation of PGK decreases with temperature above 32°C and remains constant from 0 to 25°C. The decrease in refolding of PGK at higher temperatures is due to protein aggregation and the formation of incorrectly folded monomers (Porter & Cardenas, 1980). The slow reaction step in the folding of AK-HDH monomers is also strongly dependent on temperature from 0 to 50°C (Vaucheret et.al., 1987). The exact role of temperature in the kinetics of refolding and aggregation was not determined for these proteins. However, temperature directly affects the molecular interactions within the protein. Hydrophobic interactions which have been postulated as the major driving force for folding are endothermic (Dill, 1990). Therefore, increases in temperature should result in a faster rate of folding. A balance must be achieved, however, between the enhancement in the hydrophobic interactions and the thermal denaturation of the protein. The stabilization of denatured states at higher temperatures can also result in protein aggregation. Since protein aggregation may be driven by hydrophobic interactions between proteins, higher temperatures should also stabilize aggregates. In general, an optimum temperature probably exists for each set of solvent conditions such that the rate of refolding can be maximized for a given protein.

Aggregation During Refolding

As observed in the *in vivo* studies of inclusion body formation, protein aggregation of intermediates can also occur during *in vitro* folding. The similarity between *in vivo* and *in vitro* aggregation during folding has been accessed for the phage p22 tailspike protein (Seckler et.al., 1989). For this protein, the *in vitro* folding was observed to occur with different kinetics than in the *in vivo* studies. The difference in folding kinetics was

postulated as the lack of molecular chaperonins for the *in vitro* systems (Seckler et.al., 1989). Early studies on protein aggregation revealed that the kinetics of aggregation competed with the folding kinetics (Zettlmeissl et.al., 1979). The rate and extent of aggregation was also shown to be dependent on the final protein and denaturant concentrations for LDH (Zettmeissl et.al., 1979; Rudolph et.al., 1979). In addition, refolding of PGK at a "critical" concentration of denaturant resulted in the formation of aggregates (Mitraki et.al., 1987). The phenomena of a critical denaturant concentration for irreversible aggregation has been observed in the refolding of several other proteins (Mitraki et.al., 1987). At the critical concentration of denaturant, the rate of aggregation for a folding intermediate may exceed the rate of refolding to a nonassociating intermediate resulting in a low recovery of active protein. Folding intermediate association has also been observed at equilibrium by stabilizing the intermediate in high denaturant and protein concentrations. For example, Brems and coworkers have shown that a folding intermediate of bovine growth hormone can be stabilized at 3.7 M GuHCl or 8.5 M urea and will associate at high protein concentrations (Brems et.al., 1986; Havel et.al., 1986). The association was reduced by the addition of a peptide fragment (37 amino acids) from the native protein (Brems et.al., 1986). These studies suggest that aggregation between specific structural units may occur as hypothesized previously (Mitraki & King, 1989). The intramolecular and intermolecular interactions which occur during refolding result in the formation of native protein as well as aggregates. Furthermore, the rate and extent of this aggregate formation is dependent on the final protein and denaturant concentrations.

Noncovalent interactions are the most likely driving force for aggregation as postulated by several researchers (Ikai et.al., 1978; Fish et.al., 1985; Fuke et.al., 1985; Tandon & Horwitz, 1987; Mitraki & King, 1989). Hydrophobic interactions are the strongest nonpolar forces and have also been postulated as the driving force for refolding (Dill, 1990). The evidence that hydrophobic interactions are responsible for aggregation include the endothermic kinetics characteristic of hydrophobic reactions (Mitraki et.al., 1987; Brems et.al., 1986; Stellwagen & Wilgus, 1978), concomitant loss of entropy upon aggregation (Ghelis & Yon, 1982; Mitraki et.al., 1987), and exposure of hydrophobic regions by chaotropic agents or solvents (Marston, 1986; Mitraki et.al., 1987; Ikai et.al., 1978). A specific example of hydrophobic effects is the removal of the hydrophobic end group of a viral glycoprotein which resulted in higher protein solubility (Marston, 1986). In addition, rhodanese aggregates with concomitant exposure of hydrophobic surfaces. The domain interfaces of rhodanese are hydrophobic and cause protein-protein interactions leading to aggregation (Tandon & Horwitz, 1986; Tandon & Horwitz, 1987). Another example is the partial denaturation of bovine growth hormone which results in the exposure of a hydrophobic face on an alpha helix. The hydrophobic surfaces interact between molecules resulting in aggregation (Brems et.al., 1986). Mitraki and coworkers have provided additional evidence to support the hypothesis that hydrophobic interactions are the driving force for aggregation (Mitraki et.al., 1991). In conclusion, aggregation during refolding can occur when hydrophobic intermediates exist on the folding pathway and the extent of aggregation will depend on the final denaturant and protein concentrations both of which impact the kinetics of refolding and aggregation.

Analytical Techniques

To measure the refolding and aggregation of proteins, several different analytical techniques have been utilized. These techniques provided several independent methods to study the phenomena of aggregation during refolding. In addition, many of these techniques were useful in elucidating the protein-solvent interactions. A discussion of each technique and its applications is included below to provide a basis for data interpretation in the subsequent Results and Discussion sections.

Quasi-elastic light scattering (QLS) analysis

Light Scattering Theory: Quasi-elastic light scattering provides information on the diffusivity of particles in solution based on their Brownian motion in solution. The rapid decay of light as a function of time is measured from the scattering of the incident light. The light scattering of a sample is collected by a photomultiplier which counts the frequency of the photons and the photomultiplier then sends a signal to a series of autocorrelation channels which receives the light signals at short segments or sample times on the order of microseconds. The autocorrelation function, $G(\tau)$, is the time averaged scatter intensity between time, t, and a later time, $t+\tau$. The autocorrelation time is then plotted as a function of time, τ, as shown in Figure 1. For small particles (< 5 ηm), the total time required to achieve a complete decay, $\tau_{baseline}$, is a few microseconds due to the rapid Brownian motion of these particles. On the other hand, larger particles (>100 ηm) require on the order of 100 microseconds to completely decay. The relationship between the autocorrelation function and the Brownian motion of the particles in solution can be described as a first approximation by a summation of the exponential

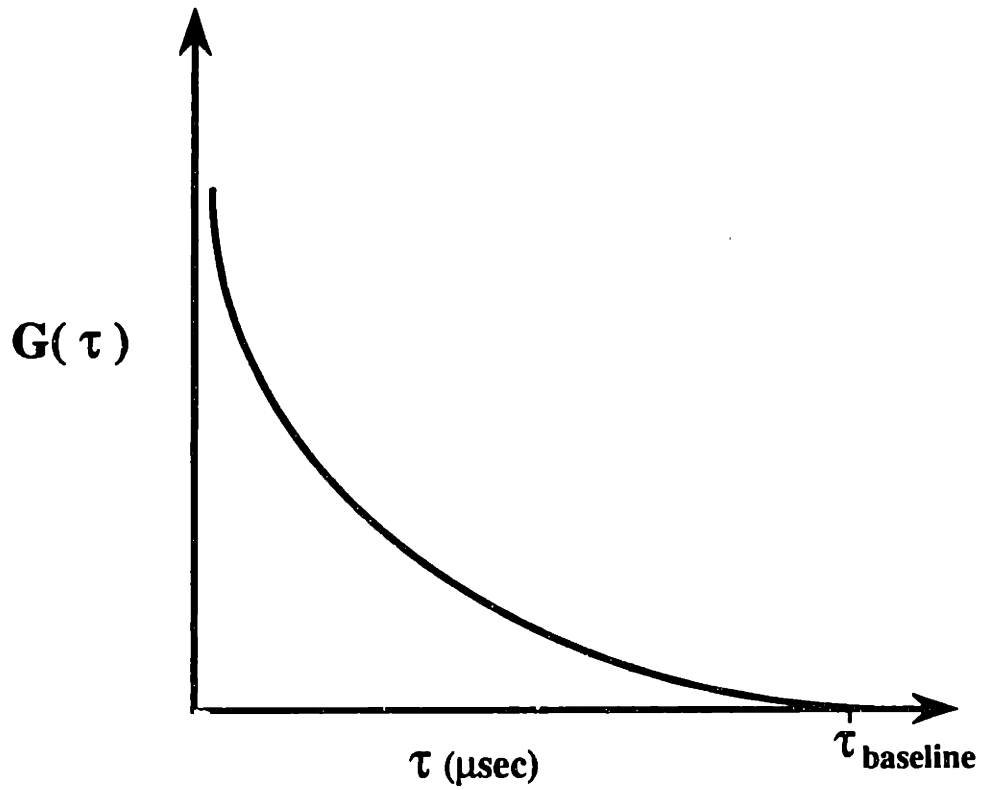


FIGURE 1: Autocorrelation function, G(τ), plotted with respect to the sample time, τ. Single exponential decay represents a monodisperse population of particles. The overall time required to achieve the baseline where the decay reaches a minimum is marked as $\tau_{baseline}$.

decay of each particle, i, in a solution of n particles.

$$G(\tau) = \sum_{i=1}^{n} G_{i_0} \exp(-\Gamma_i \tau) \quad (1)$$

The exponential term, Γ_i, is a function of the particle diffusivity, D_{s_i}, the solution conditions, and the instrument settings.

$$\Gamma_i = D_{s_i} K^2 \quad (2)$$

$$K = \frac{4\pi\eta}{\lambda} \sin(\theta/2) \quad (3)$$

The instrument and solution conditions are represented in the constant, K, to include the solution viscosity, η, the wavelength of the incident light, λ, and the angle of the light scattering measurement with respect to the incident light, Θ. The derivation of these formula and the physical basis for QLS has been described previously (Bloomfield & Lim, 1978). In addition, the summation method as shown in Equation 1 is the simplest deconvolution method and several more precise and complex data analysis methods have been detailed by Stock and Ray (Stock & Ray, 1985). The particle diffusivities derived from these analyses can used to determine the hydrodynamic diameter, D_h, of the particles by the Stokes-Einstein equation:

$$D_h = \frac{k_B T}{3 \pi D_s \eta} \quad (4)$$

where k_B is Boltzman's constant, T is the temperature of the sample, D_s is the diffusivity of the particles in solution (Brownian motion), and η is the viscosity of the solution.

The hydrodynamic diameter, D_h, of a particle may lie in one of three basic light scattering regimes: Rayleigh scattering, Mie scattering, or the Extinction Paradox. In the Rayleigh regime, the particle hydrodynamic diameter is much smaller than the wavelength of the incident light beam ($D_h << \lambda$). For this regime, QLS analysis will result in the prediction of a bimodal particle size distribution for separate particles of diameter, d, and 3d and a unimodal particle size distribution for any combinations of intermediate particle diameters ($d < D_h < 3d$). The Mie scattering regime applies to the range of particle diameters which are on the order of the wavelength of the incident light ($D_h \sim \lambda$). Mie scattering is dependent on the angle at which the scattering is measured and the resolution is defined as a bimodal distribution for discrete particles with diameters, d and 2d. Finally, the Extinction Paradox becomes relevant for hydrodynamic radii which are much greater than the wavelength of the incident light ($D_h >> \lambda$). The Extinction Paradox states that smaller particles may become effectively shadowed or eclipsed by the larger particles resulting in failure to detect the smaller species. The major difficulty with the QLS technique, therefore, lies in the Mie and Extinction regimes. However, the Mie and Extinction regimes are usually not important for QLS studies which involve measurement of gross conformational changes or the formation of submicron protein aggregates since

these cases usually result in particle diameters which are much smaller than the wavelength of the incident light (Rayleigh scattering).

Application of QLS to the Study of Proteins: QLS has been used extensively to study different aspects of proteins in solution. Initial studies involved the characterization of the hydrodynamic properties and molecular weight of proteins in solution (McDonnell & Jamieson, 1976). In addition, early studies focussed on the change in conformational state of several globular proteins and the thermal unfolding of these proteins (Nicoli & Benedek, 1976). Protein aggregation has also been studied using QLS. These studies primarily involved the analysis of large protein aggregates which formed as the result of a change in the solvent environment (Jedziniak et.al, 1978; Donovan et.al., 1987). Specific protein-protein interactions have been studied extensively in regards to the formation of antigen-antibody complexes (Yarmush et.al., 1987; Murphy et.al., 1988; Yarmush et.al., 1988; Murphy et.al., 1990). These investigators provided insight into the fundamental mechanisms of antigen-antibody complex formation. Since QLS is a relatively rapid analysis technique (minimum duration time of 30 seconds), the early stages of protein association where submicron particle sizes predominate can be measured with this technique as demonstrated for antigen-antibody complex formation (Yarmush et.al., 1988). The measurement of initial protein crystallization has also been studied utilizing QLS techniques (Mikol et.al.,1989; Kadima et.al., 1990). The inherent advantages of this technique include noninvasive analysis and rapid measurement which facilitates the study of the kinetics of protein association or aggregation during refolding.

Size Exclusion High Performance Liquid Chromatography (HPLC)

In addition to QLS, size exclusion HPLC will also provide information on the hydrodynamic properties of particles in solution. The basic principle of size exclusion HPLC involves the diffusion of each particle into and out of a porous column matrix. The extent of diffusion of each particle is dependent upon the particle size, pore diameter of the matrix and hydrodynamic conditions used for the separation (ie. flowrate, viscosity, etc.). As a result of the difference in diffusivity of each particle, the particles are separated as they pass through the column. The large particles which can not diffuse into the interior of the column matrix will be eluted in the void fraction of the column and very small particles and solvent species which readily diffuse into the matrix will elute at a volume equivalent to the total column volume. As the particle size decreases, the elution time for the particle will increase until the aforementioned limit is achieved.

Since size exclusion HPLC can also discriminate between different particle sizes in solution, it has been applied to the study of protein structures, protein association, and protein refolding. Initial size exclusion HPLC studies of proteins primarily involved measurement of conformational changes which resulted from solvent denaturation of the protein (Martenson, 1978; Endo et.al., 1983; Corbett & Roche, 1984; Brems et.al., 1985). Recently, the refolding of several proteins including thioredoxin, ribonuclease, and trypsinogen has been studied by size exclusion HPLC techniques (Shalango et.al., 1987; Al-Obeidi & Light, 1988; Shalango et.al., 1989). Protein association of unfolding intermediates has also been studied with this technique (Havel et.al., 1986). In general, size exclusion HPLC can be applied to the study of protein association or protein

conformation. However, the analysis must be performed more rapidly than the equilibrium between the each of the species in solution to obtain a representative measurement of the population. Therefore, most studies of rapid equilibria must be analyzed at high flowrates (≥ 1 ml/min) to avoid artifacts resulting from the analysis time.

Spectroscopic Techniques

Several spectroscopic techniques have been used to study protein structure and refolding. These techniques include ultraviolet absorbance, fluorescence, circular dichroism (CD), and infrared (IR) spectroscopy. In this work, ultraviolet absorbance and fluorescence have both been applied to the study of protein refolding, protein association, and protein-solvent interactions.

Absorbance: Changes in the absorbance spectra of proteins has been extensively used to measure changes in the protein structure. In general, the aromatic residues in the protein structure primarily contribute to the absorbance spectra between 230 and 300 ηm. As the protein changes in structure, the exposure of these residues is altered such that their ability to absorb high energy light (ultraviolet) changes. The interaction between adjacent aromatic groups allows for a distribution of light energy and a higher absorption of the incident light. Since most aromatic residues are clustered in hydrophobic regions of the native protein structure, the absorbance of ultraviolet light will decrease as the protein unfolds to expose and disrupt these regions. The rate and extent of unfolding or refolding can then be assessed by the change in absorbance relative to the native state. Absorption at both 280 ηm and 292 ηm has been used to study changes in the protein structure during refolding (ie. bovine growth hormone, Brems et.al., 1987; carbonic

anhydrase B (CAB), Semisotnov et.al., 1990). This technique therefore results in a method for analyzing the gross conformational changes which occur during protein refolding.

Fluorescence: Fluorescence spectroscopy can be used to measure changes in the protein structure, protein-protein interactions, and protein-solvent interactions. To apply this technique, the basic principles of fluorescence should be understood. In general, fluorescence can be described as a loss of energy from an excited state molecule resulting in the emission of light. For proteins, the aromatic residues, in particular, tryptophan and tyrosine, can be excited by high energy light (ultraviolet) and the return to the ground state of the electrons in the π orbitals of aromatic rings will result in the emission of light in the visible range, fluorescence. The rate with which the light is emitted and the emission spectra are dependent upon the environment around the excited molecule. Therefore, the fluorescence of aromatic residues in proteins provides a method of probing their local environment. The intrinsic fluorescence of a protein can then be measured by exciting the protein with ultraviolet light and measuring the emission in the visible range. Typically, an excitation wavelength of 280 ηm will excite both tyrosine and tryptophan residues within the protein. To avoid the additive fluorescence of both residues, an excitation wavelength of 296 ηm can be used to excite primarily the tryptophan residues since tyrosines absorb less energy than tryptophans at this wavelength. Measurement of the emission spectra for both these excitation wavelengths will yield insight into the conformation of the protein (see Pesce et.al., 1971 for detailed discussion). The conformational states of several proteins have been studied by fluorescence (ie. von

Willebrand protein, Loscalzo & Handin, 1984; bovine growth hormone, Holzman et.al., 1990). In addition, fluorescence has been used to study the refolding of proteins (ie. several studies on CAB refolding, Stein & Henkens, 1978; Henkens et.al., 1982; Semisotnov et.al., 1990).

Further measurements of protein structure during refolding can be performed by fluorescence quenching which probes the solvent accessibility of the aromatic residues. Quenching agents of various sizes and types can be used to investigate the environment around tryptophans and tyrosines. Large size quenching agents such as iodide can not readily penetrate into the interior of the protein and, thus, they can only reveal the surface properties of proteins. Small molecules such as acrylamide can diffuse through the protein matrix and quench internal residues. The ability of these agents to quench the protein fluorescence is related to the frequency and number of collisions between the excited molecules and the quencher. The Stern-Volmer equation relates the relative fluorescence quenching (F_o/F) to the collisional quenching constant, K_{SV}, and the concentration of the quencher, Q.

$$F_o/F = 1 + K_{SV}[Q] \qquad (5)$$

Eftink and Ghiron provide an excellent review of fluorescence quenching analysis methods (Eftink & Ghiron, 1981). Fluorescence quenching analysis has been applied to the study of changes in protein structure (ie. staphylococcal nuclease mutants, Wright & Freedman, 1989). In addition to conformational studies, protein-protein association has been measured with fluorescence quenching (ie. bovine growth hormone, Havel et.al.,

1988). Fluorescence quenching analysis has proven to be a useful method for measuring changes in the protein structure and surface properties.

In addition to fluorescence quenching, fluorescence probes can be used to study the surface properties of proteins. For example, hydrophobic probes such as 8-anilino-1-naphthalene sulfonate (ANS) which fluorescence when bound to the protein have been used to characterize hydrophobic intermediates during refolding (Rodionova et.al., 1989; Semisotnov et.al., 1987; Ptitsyn et.al., 1990). The binding of fluorescence probes or ligands to the protein surface yields insight into the properties of the protein and its solvent environment. The changes in the intrinsic protein fluorescence can provide information on the conformational or surface modifications which occur after ligand binding. An excellent review of the application of fluorescence to analyze ligand binding is provided by Pesce (Pesce at.al., 1971).

Electron Spin Resonance (ESR)

Electron spin resonance (ESR) or electron paramagnetic resonance (EPR) provides a method for the measurement of specific changes in protein structure, protein-protein interactions, and protein-solvent interactions. Therefore, this technique is complementary to other structural probing techniques such as fluorescence, circular dichroism and nuclear magnetic resonance.

Fundamental Theory: To produce an electron spin resonance, a molecule must have a free electron which can resonant between different energy levels within a magnetic field (Figure 2). Molecules which possess this ability include paramagnetic metals such as Mn^{2+} and Fe^{3+} and stable radical species such as nitroxide radicals. Since proteins

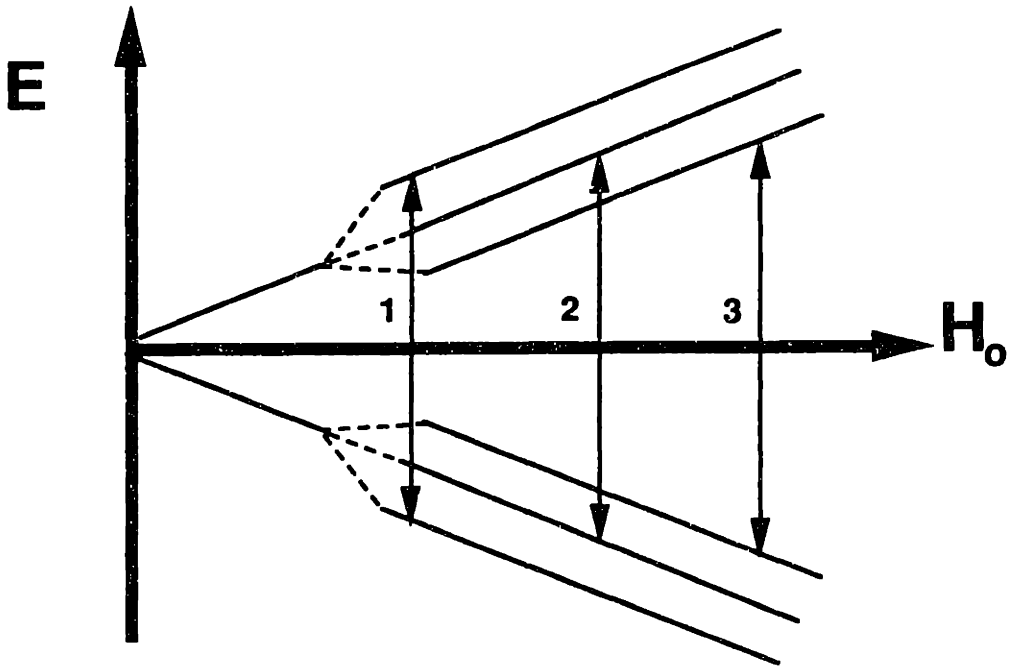


FIGURE 2: Illustration of the electron spin resonance between energy levels as a function of the field strength. An electron will resonate between energy levels, E, as a function of the applied magnetic field, H_o. The triplet splitting shown above is representative of a nitroxide radicals which have a nucleus of spin 1 (Nordio, 1976).

typically do not contain paramagnetic metals, a spin probe or spin label which contains a nitroxide radical group must be attached to the protein (Figure 3). The rotation of the spin label about the bonds which attach it to the protein and the rotation of the protein molecule dictate the relaxation rate of the spin label. To measure this rotation, a modulating magnetic field strength is applied perpendicularly to the detector current resulting in a first derivative of the sample absorption spectrum. The shape of this first derivative absorption is related to the molecular relaxation of the spin probe and the rate of molecular motion of the probe within the sample (see Kosman & Bereman (1981) for detailed discussion). The rotation rate or the rotational diffusivity, D_R, of the spin label in solution is then related to the rotational correlation time, τ_c.

$$\tau_c = (6\ D_R)^{-1} \tag{6}$$

From either of these two parameters, an effective radius of gyration for the spin label, R_g, can be calculated from the Debye diffusion model:

$$D_R = \frac{k_B\ T}{8\ \pi\ \eta\ R_g^3} \tag{7}$$

where k_B is Boltzman's constant, T is the temperature of the solution, and η is the kinematic viscosity of the solution. As the spin label becomes constrained in its motion, the first derivative absorption spectra changes significantly as shown in Figure 4 (Likhtenstein, 1976). For example, unattached spin label in solution has a rotational

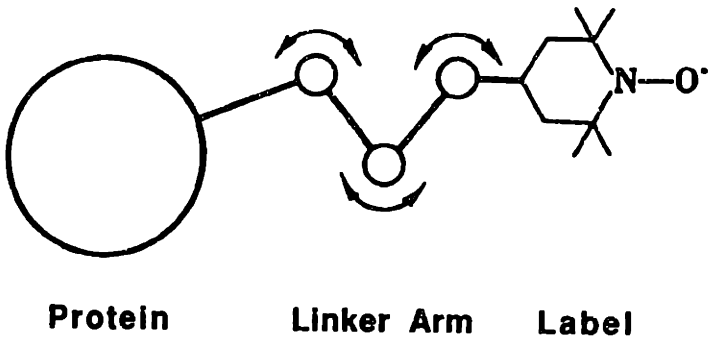


FIGURE 3: Spin labelled protein for electron spin resonance (ESR) studies (not to scale). The nitroxide radical group is attached by a linker arm to residues on the protein. The motion about the bonds of the linker arm as well as the overall motion of the protein can be measured by ESR analysis of the spin labelled protein (Likhtenstein, 1976).

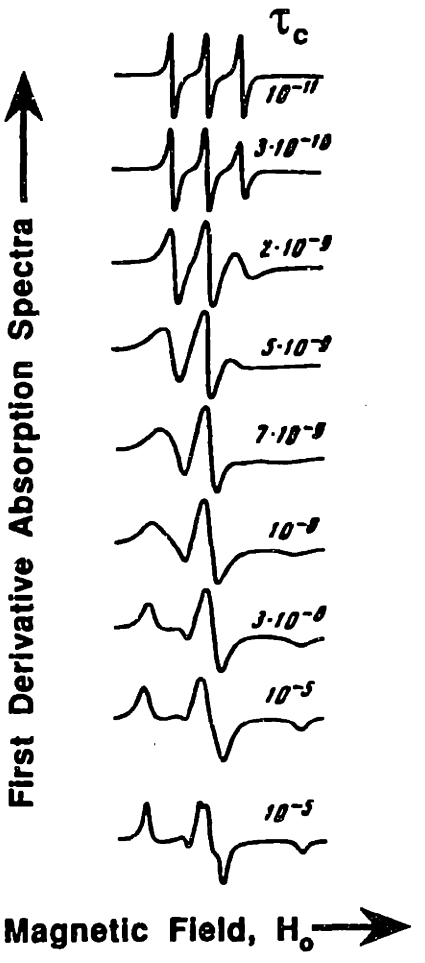


FIGURE 4: Examples of first derivative absorption spectra dependence upon the rotational correlation time, τ_c (sec), of the nitroxide spin label taken from Likhtenstein (1976).

correlation time on the order of 10^{-10} sec whereas the label attached to a random coil such as an unfolded protein or polymer has a rotational correlation time of 10^{-9} sec. In addition, if the spin label becomes incorporated into the interior of a protein or becomes constrained in its rotation as the result of the solvent environment (ie. reduction in temperature), the rotational correlation time will increase (10^{-8} sec) and the first derivative absorption spectra will be significantly altered. Therefore, the spin probe provides sensitive information on the local protein environment. This information includes the motion of the protein in the solution, the orientation of the label in relationship to other labels, the polarity of the spin environment, and the magnetic environment of the spin label. The spin label then acts a probe to measure the environment around the location of attachment.

Application of ESR to the Study of Proteins: Spin label probes such as paramagnetic metals and nitroxide radicals have been used extensively to measure changes in protein conformation, active site configurations, association phenomena, and protein refolding. Excellent reviews of several applications for probing active sites and studying conformational changes are presented by Likhtenstein and Berliner (Likhtenstein, 1976; Berliner, 1981). In particular, the active site of bovine carbonic anhydrase has been studied in detail by using spin labelled inhibitors (Haffner & Coleman, 1975; Hower et.al., 1971; Wee et.al., 1976; Mushak & Coleman, 1972). Recombinant DNA techniques combined with ESR have been utilized to determine the structure of the transmembrane protein, bacteriorhodopsin, and its interaction with membranes (Altenbach et.al, 1990). In addition, the combining sites of antibodies has also been analyzed by spin labelling

methods (Anglister, 1989). ESR techniques have also been used to study the aggregation of several different proteins (Laurie & Oakes, 1976; Thiyagarajan & Johnson, 1983; Napier et.al., 1987). In addition to protein-protein interactions, protein-solvent interactions have also been studied by electron spin resonance (Oakes & Cafe, 1973). Recent studies have also applied ESR techniques to the analysis of protein refolding and the kinetics of intermediate species formation (Ebert et.al., 1990; Semisotnov et.al., 1987). The high sensitivity of the spin label to the local environment provides a powerful tool for probing the microenvironments within a protein structure and at the surface of the protein.

Protein Systems

Several different proteins were used in this work to develop a fundamental understanding of the phenomena of protein aggregation during refolding. The primary focus of the initial refolding and aggregation studies was bovine carbonic anhydrase B (CAB). After the intensive study of the refolding process with this protein, a methodology for preventing aggregation was developed. To test the generality of this method and further characterize aggregation, three recombinant human proteins, deoxyribonuclease, tissue plasminogen activator, and gamma interferon, were studied. Each of these proteins presents a unique and difficult refolding problem and few published refolding studies have been conducted on these proteins. Each protein and its known refolding properties are described below.

Bovine Carbonic Anhydrase B (CAB)

To study the phenomena of aggregation, bovine carbonic anhydrase B (CAB) was chosen as the model protein. CAB has a small compact structure (30,000 MW) which does not contain cysteine residues or carbohydrate groups in its nascent form as illustrated in Figure 5 (Sciaky et.al., 1976). CAB was chosen to investigate the protein aggregation phenomena which occurs during refolding since it has intermediates on its refolding pathway and these intermediates have been well characterized (Stein & Henkens, 1978; Doligkh et.al., 1984; Semisotnov et.al., 1987; Semisotnov et.al., 1990). As shown in Figure 6, unfolded CAB in 5 M GuHCl will rapidly form the first intermediate which is a compact molten globule structure with exposed hydrophobic clusters. These clusters will then collapse to form a core in the second intermediate. Finally, the native protein

NH_2-S H H W G Y G K H D G P Q H W H K D F P I A N G E R Q S P V N I D T

K A V V Q D P A L K P L A L V Y G E A T S R R M V N N G H S F N V E Y D

D S Q D K A V L K D G P L T G T Y R L V Q F H F H W G S S D D Q G S E H

T V D R K K Y A A E L H L V H W N T K Y G D F G T A A Q Q P D G L A V V

G V F L K V G D A N P A L Q K V L D A L D S I K T K G K S T D F P N F D P

G S L L P N V L D Y W T Y P G S L T T P P L L E S V T W I V L K E P I S V S

S Q Q M L K F R T L N F N A E G E P E L L M L A N W R P A Q P L K N R Q

V R G F P K-CO_2H

FIGURE 5: Sequence of bovine carbonic anhydrase B (Sciaky et.al., 1976). Active site zinc ion is chelated with ^{93}His and ^{95}His. The nascent protein does not contain carbohydrate groups.

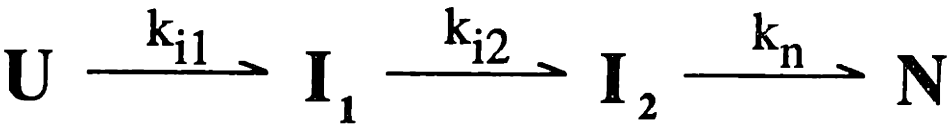


FIGURE 6: Reported pathway for refolding of CAB. The unfolded protein (U) rapidly forms the first compact intermediate (I_1) within 200 milliseconds (k_{i1} = 23.1 sec^{-1}). The first intermediate can then proceed to form the second intermediate (I_2, k_{i2} = 0.297 min^{-1}). The second intermediate continues to fold to form the native protein (N, k_n = 6.93 x 10^{-2} min^{-1}). [Reported for refolding from 3 M GuHCl to 1 M GuHCl by rapid dilution with 0.05 M Tris buffer and 2 M NaCl, pH 7.5 at 20°C (Stein & Henkens, 1978); refolding from 8.5 M urea to 4.1 M urea by rapid dilution with 0.1 M Tris buffer, pH 8 at 20°C (Semisotnov et.al., 1987)].

structure of CAB will completely form approximately one hour after dilution to refolding conditions (Stein & Henkens, 1978; Semisotnov et.al., 1987). The slow steps in the refolding process have been shown to occur as the result of *cis-trans* isomerization of the twenty proline residues in the protein sequence. In particular, it was postulated that the isomerization of two proline residues is the rate limiting step for the formation of native structure (Semisotnov et.al., 1990). The slow isomerization steps resulted in stable intermediates which could be readily studied. CAB also aggregates during refolding at high protein concentrations (mg/ml) and low denaturant concentrations (0.1 - 0.7 M GuHCl) (Ikai et.al. 1978). Since CAB formed stable intermediates and aggregated during refolding, it was ideally suited for the study of aggregation during refolding.

Recombinant Human Deoxyribonuclease (rhDNAse)

Two different forms of recombinant human deoxyribonuclease were studied in this thesis. Genentech, Inc. donated both the glycosylated form produced by Chinese Hamster Ovary (CHO) cells and the nonglycosylated form produced by *Escherichia coli.* rhDNAse is a small enzyme (29,339 MW) with two disulfide bonds (^{101}Cys-^{104}Cys and ^{173}Cys-^{209}Cys) and two glycosylation sites (^{18}Asn and ^{106}Asn) as shown in Figure 7. Structural analysis of the bovine analog indicates that Ca^{2+} and Mg^{2+} are required as cofactors for activity and calcium may be required for structure (Price, 1975; Okabe et.al., 1982; Suck et.al., 1984). Crystallography indicates that the bovine protein has a hydrophobic core consisting of two 6-stranded beta sheets (Suck et.al., 1984). Refolding studies have not been published for either the human or bovine form of the protein. However, equilibrium denaturation studies in GuHCl indicate that the bovine protein

FIGURE 7: Sequence of recombinant human deoxyribonuclease (rhDNAse) as provided by Genentech, Inc. The two disulfide bonds (^{101}Cys-^{104}Cys and ^{173}Cys-^{209}Cys) are illustrated (⌐¬) as well as the two glycosylation sites (↓) for the CHO derived form of the protein.

undergoes a transition in its near CD spectra (220 ηm) at approximately 2 M GuHCl (Okabe et.al., 1982). In addition, multiple conformational states were observed for bovine DNAse as a function of the calcium ion concentration at alkaline pH (Lizarraga et.al., 1979). The renaturation of sodium dodecyl sulfate treated bovine DNAse was successfully accomplished only by serial dilutions into 6 M guanidine hydrochloride (GuHCl) and buffer with a resulting overall dilution of greater than 100 fold (Liao, 1975). These results suggest that bovine DNAse has stable intermediates in its refolding pathway and that these species can fold to form the native structure.

Recombinant Tissue Plasminogen Activator (rtPA)

Another protein studied in this thesis was recombinant tissue plasminogen activator (rtPA). The glycosylated form produced by CHO cells was supplied by Genentech and used in refolding and aggregation studies. rtPA can be produced in two different forms: single chain (sctPA) or two chain (tctPA) tPA. The difference between the two forms is a single cleavage at ^{278}Arg as shown in Figure 8. The difference in structure and function between sctPA and tctPA was also elucidated in previous studies. sctPA changed conformation when bound to fibrin. The structure and activity of both forms were the same with fibrin. sctPA without fibrin was also shown to have only 26% of the protease activity of tctPA (Loscalzo, 1988). In addition, the CHO produced form contains glycosylation at three different sites (Asn residues 117, 184, and 448) and each site has a high degree of microheterogenity (Spellman et.al., 1989). rtPA glycosylated at all three sites is referred to as type I tPA and those forms without glycosylation at ^{184}Asn are referred to as type II tPA. The single chain protein has a molecular weight of 72,000 Da

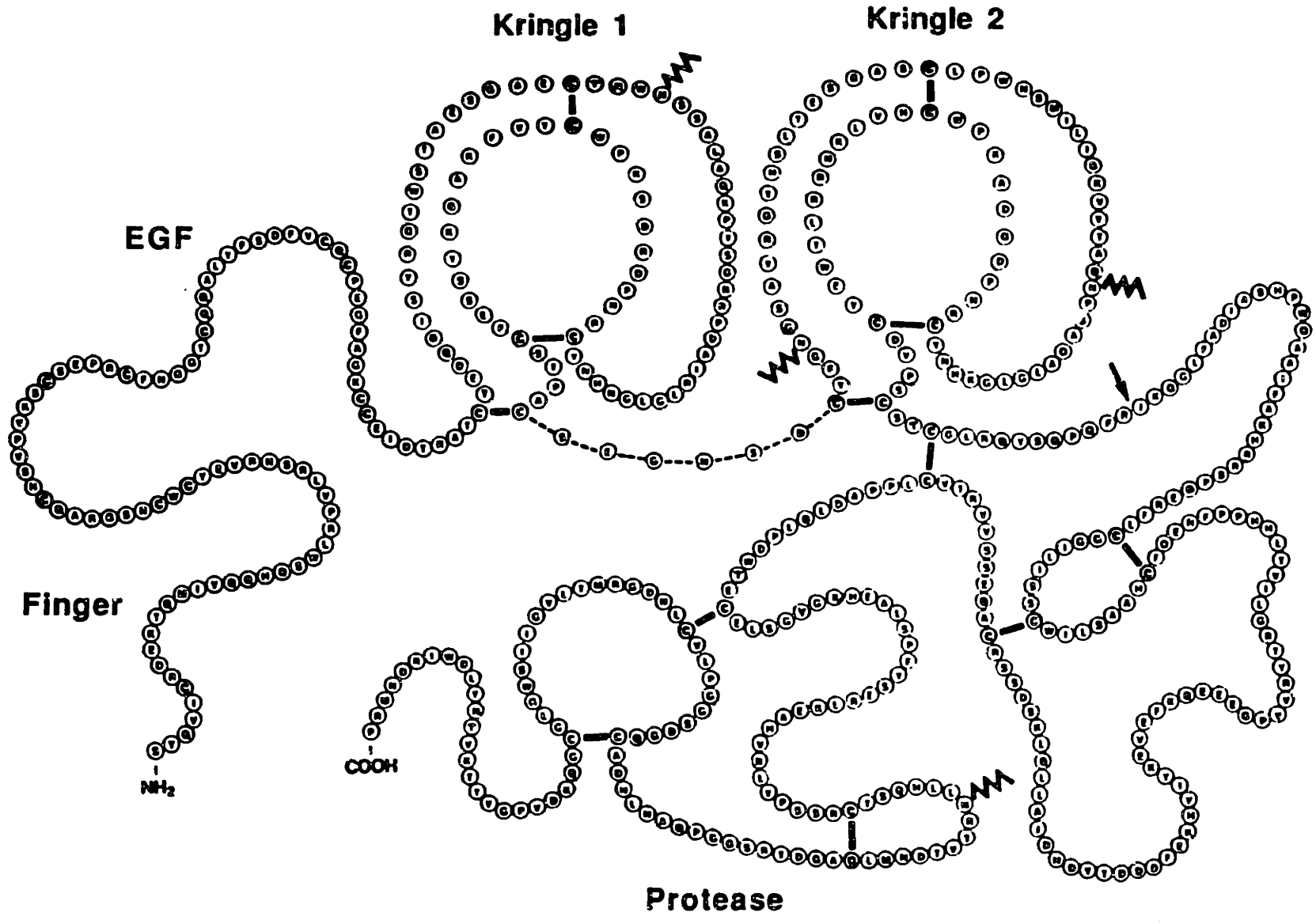


FIGURE 8: Sequence and domain structure of human tissue plasminogen activator (tPA) (Pennica et.al., 1983). The five domains of tPA are marked from the amino terminus as the finger, epidermal growth factor (EGF), kringle 1, kringle 2, and protease regions. tPA has 17 disulfide bonds (not shown) and one free thiol. The cleavage site for conversion from single chain (sctPA) to two chain (tctPA) is marked (↓) along with the glycosylation sites (/\/\).

with seventeen disulfide bonds as shown in Figure 8 (Rijken et.al., 1982). The thermal denaturation of the protein has been observed by differential scanning calorimetry (DSC). The DSC study revealed that their were two melting transitions and that the second stage represented a irreversible denaturation of the protein (Radek & Castellino, 1988). These stages may represent potential intermediate forms during the thermal unfolding. When produced in bacteria, tPA will form inclusion bodies. Solubilization and refolding of the inclusion body protein has resulted in low recovery yields (Wang et.al., 1989; Sarmientos et.al., 1989). Therefore, the protein is currently expressed in CHO cells for commercial production. Other equilibrium refolding or denaturation studies, if performed, have not been published.

Recombinant Gamma Interferon (rIFN-γ)

The other recombinant protein used in this thesis is recombinant human gamma interferon (immune interferon, rIFN-γ) produced by *E. coli* provided by Biogen, Inc. Denaturation using urea, GuHCl, or acid produces rIFN-γ monomers which aggregate during renaturation (Arakawa & Hsu, 1987; Arakawa et.al., 1985; Hsu & Arakawa, 1985; Patel, 1986; Yphantis & Arakawa, 1987). rIFN-γ does not contain disulfide bonds, but does have two amino terminal cysteine residues. The recombinant protein is not glycosylated, but the natural human interferon-γ (IFN-γ) has two glycosylation sites as shown in Figure 9 (Rinderknecht et.al., 1984). Therefore, only noncovalent interactions are involved in the aggregation process under reducing conditions. The exposure of hydrophobic aromatic groups in the monomer and not in the aggregate indicates the possible predominance of hydrophobic interactions (Arawaka et.al., 1985). rIFN-γ

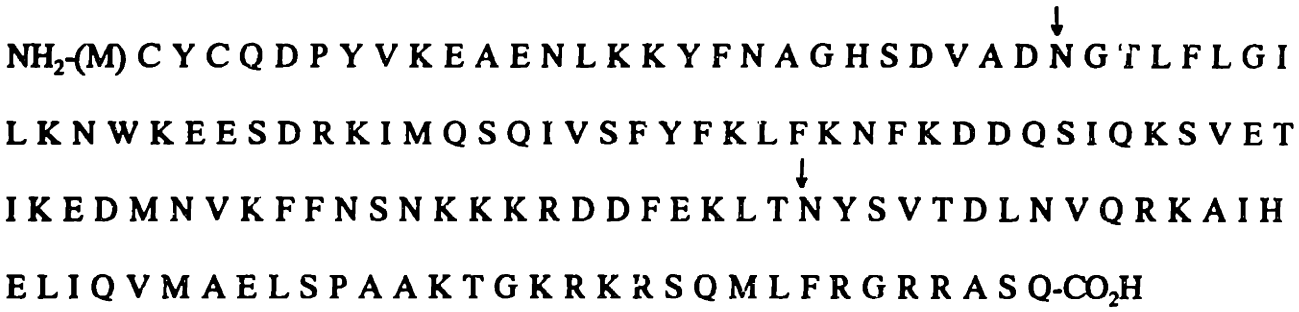


FIGURE 9: Sequence of Natural and Recombinant Gamma Interferon (rIFN-γ) (Rinderknecht et.al., 1987). The natural sequence does not include the terminal methionine residue (M). The glycosylation sites in the natural IFN-γ are marked (↓). Forms lacking the N-terminal C-Y-C peptide are also active (Hsu & Arakawa, 1985).

aggregate formation is also dependent upon the protein denaturant concentrations (Arakawa & Hsu, 1985; Patel, 1986). A heterogeneous population of aggregates is formed from the acid unfolding-refolding process (Yphantis & Arakawa, 1987; Arakawa & Hsu, 1987). During reactivation of acid denatured rIFN-γ, a thirty-three monomer aggregate (MW 0.5 x 10^6) is formed to represent 1% to 6.3% of the total protein at high concentrations and a sixteen monomer aggregate formed to represent greater than 2% of the total protein at low concentrations (Yphantis & Arakawa, 1987). The thermal denaturation of human rIFN-γ has also revealed that the protein will aggregate during unfolding. Furthermore, these thermal denaturation studies indicated that protonation of the histidine residues could be required to prevent aggregation (Mulkerrin & Wetzel, 1989). These results demonstrated that an aggregating intermediate occurs during refolding.

Cosolvent-Protein Interactions

Effect of Cosolvents on Native Proteins

An understanding of the interactions between native protein and solvent can be used to assist in interpreting the effects of solvent environment on protein refolding. Solvents have been shown to interact with proteins by two different mechanisms. First of all, solvents can bind to the protein and cause stabilization of different protein structures such as the unfolded state. Alternatively, the addition of cosolvent can cause the protein to become preferentially hydrated through an increase in the chemical potential of the system which is caused by an increase in the surface tension of the water environment. This preferential hydration mechanism favors protein structures with less accessible surface area reducing the free energy of the system. Therefore, compact molecules and aggregated species are favored by addition of hydrating solvents. These two effects can be measured and physically explained by the thermodynamics of the system. The change in chemical potential of the protein ($\delta\mu_p/\delta m_s$) upon solvent addition is directly proportional to the change in chemical potential of the cosolvent ($\delta\mu_s/\delta m_s$) and the preferential interaction of the cosolvent with the protein ($\delta m_s/\delta m_p$):

$$\left[\frac{\delta \mu_p}{\delta m_s}\right]_{T,P,m_p} = \left[\frac{\delta m_s}{\delta m_p}\right]_{T,\mu_w,\mu_s} \left[\frac{\delta \mu_s}{\delta m_s}\right]_{T,P,m_p} \tag{8}$$

where μ is the chemical potential, m is the molality, T is the system temperature, P is the system pressure, and the subscripts denote protein (p), cosolvent (s), and water (w). Figure 10 schematically describes the extent of preferential interactions ($\delta m_s/\delta m_p$) for

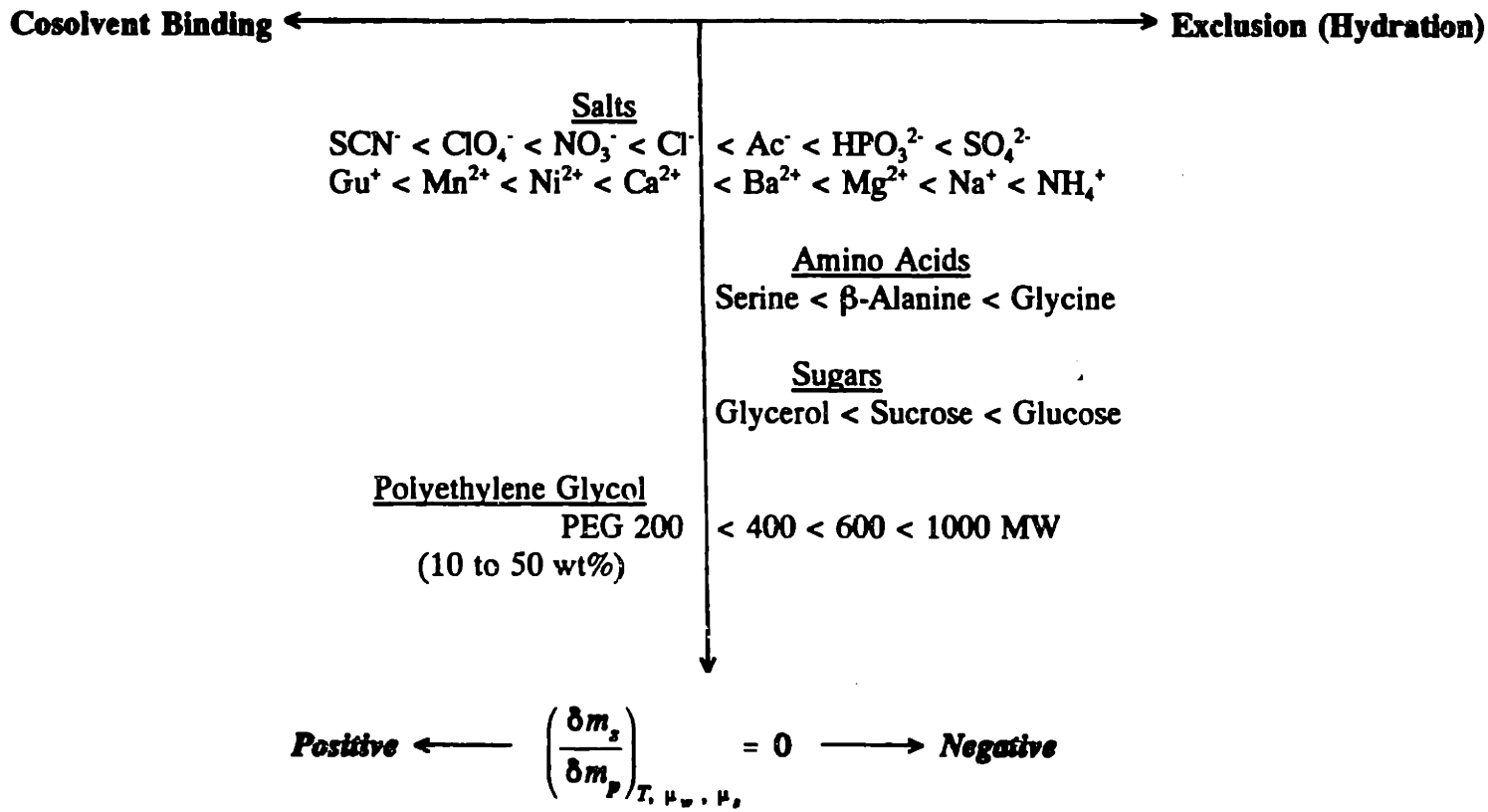


FIGURE 10: Cosolvent interactions with native proteins. The cosolvents left of the midpoint have been shown to bind nonspecifically to proteins (positive preferential cosolvent interaction). In constrast, the cosolvents to the right of the midpoint have been shown to act by preferential hydration of the protein which involves an increase in the concentration of water at the protein surface (negative preferential cosolvent interaction). The mechanism of interaction for polyethylene glycol (PEG) has been observed to be dependent upon the PEG concentration where high concentrations (10 to 50 wt%) result in exclusion and low concentrations (< 10 wt%) can result in PEG binding to the protein (Arakawa & Timasheff, 1985).

different solvent systems. In particular, sucrose and other sugars have been observed to act only by preferential hydration of the protein independent of the chemical nature of the protein surface (Arakawa et.al., 1990). On the other hand, polyethylene glycol (PEG) interactions with proteins are highly dependent on the solvent environment. PEG is excluded by stearic effects, but may bind to nonpolar regions and the binding is increased upon unfolding of the protein due to exposure of nonpolar groups (Arakawa et.al., 1990; Arakawa & Timasheff, 1985). In terms of protein association, the effect of cosolvents may be expressed as an equilibrium between a refolding intermediate, I, and its dimer, D, as:

$$\left[\frac{d \ln K_D}{d \ln a_s}\right]_{T, P, m_p} = 2\left[1/2\left[\frac{\delta m_s}{\delta m_p}\right]^{D}_{T, \mu_w, \mu_s} - \left[\frac{\delta m_s}{\delta m_p}\right]^{I}_{T, \mu_w, \mu_s}\right] \qquad (9)$$

where K_D is the equilibrium constant for association and a_s is the activity coefficient of the cosolvent. Therefore, the equilibrium is dependent upon the preferential interaction or binding of the cosolvent to each species. If the cosolvent binds to the intermediate, the intermediate is stabilized and does not form associated species since fewer cosolvent molecules can bind to the protein surface. Alternatively, the exclusion of the solvent from the protein surface would drive the equilibrium to a more compact species such as the dimer or the native protein. The balance between the competing process of refolding and aggregation would then require a cosolvent which interacted with the intermediate species to prevent association. To allow complete refolding, the cosolvent must be

excluded from the surface as the native structure is formed. Since PEG has the ability to interact with the protein by both mechanisms, it would be an efficient cosolvent for use in enhancing protein refolding.

Cosolvents and Protein Refolding

Several studies have been performed to assess the effects of various cosolvents on protein refolding. The use of sugars has been studied for the refolding of several proteins. When sucrose was used in the refolding of porphyrin c, the refolding of the protein was not altered (Brems et. al., 1982). In contrast, the first step in ribonuclease (RNAse) refolding occurs more slowly in the presence of sugars (Tsong, 1982). Glycerol and glucose have been observed to decrease the rate of refolding of octopine dehydrogenase (ODH) (Teschner et.al., 1987). Additional studies have also shown that sugars reduce the rate of refolding (Vaucheret et.al., 1987). On the other hand, PEG and poly-(vinylpyrrolidone) did not decrease the rate of refolding for ODH, but did alter the refolding pathway (Teschner et.al., 1987). Nondenaturing detergents have been successfully used to prevent aggregation during refolding of the membrane protein, rhodanese (Horwitz & Criscimagna, 1986). In particular, it has been proven that the detergent, lauryl maltoside, will interact with nonpolar regions of a rhodanese refolding intermediate and, thereby, prevent aggregation of the intermediate (Tandon & Horowitz, 1987). The application of less hydrophobic detergents or surfactants was also suggested to prevent aggregation of rhodanese (Tandon & Horowitz, 1986). These refolding studies along with the observed native protein interactions suggest that cosolvents such as PEG which weakly bind to partially folded protein will enhance refolding by blocking the

formation of aggregates.

MATERIALS AND METHODS

Chemical Materials

Bovine serum albumin (BSA), guanidine hydrochloride (GuHCl), Tris-sulfate, ethylenediaminetetraacetic acid (EDTA), p-nitrophenol acetate (pNPA), and ammonium acetate were molecular biology grade and purchased from Sigma Chemical Co. (St.Louis, MO). Polyethylene glycol (PEG), monomethoxy polyoxyethylene amine, bis(p-nitrophenol carbonate) polyoxyethylene (PEGpNP), salmon sperm DNA, methyl green, agarose, reduced and oxidized glutathione, β-mercaptoethanol, 1,1-carbonyldiimidazole, tetranitromethane (TNM), 3-carboxyl-PROXYL, 3-maleimido-TEMPO, N-acetyltyrosine, N-acetyltryptophan, N-α-acetyllysine, and all polyamino acids were also purchased from Sigma Chemical Co. (St. Louis, MO) and used as supplied. Dimethyl suberimidate dihydrochloride (DMS) and 2,4,6-trinitrobenzene sulfonic acid (TNBS) were obtained from Pierce (Rockford, IL). Triethanolamine, calcium chloride, magnesium chloride, and ethyl ether were purchased from Mallinckrodt (Paris, KY). HPLC grade acetonitrile, HPLC grade acetone, isopropanol and potassium iodide were from J.T. Baker (Phillipsburg, NJ). Decahydranapthalene (Decalin) was a racemic mixture obtained from Aldrich (Milwaukee, WI). The chromogenic peptide, S-2288 (H-D-Ile-Pro-Arg- p-nitroaniline), was purchased from Kabi Vitrum (Greenwich, CT). Imidazole was obtained from Fluka Chimika (Switzerland). The SDS-PAGE materials were obtained from Pharmacia LKB Biotechnology (Uppsala, Sweden). All buffers and samples were prepared with distilled water passed through a MilliQ water purification system (Millipore Corp., Bedford, MA).

Protein Materials and Preparation

Bovine Carbonic Anhydrase B (CAB): Lyophilized bovine carbonic anhydrase B (p*I* = 5.9) was purchased from Sigma Chemical Co. (St. Louis, MO). The purity of each lot was checked by size exclusion high performance liquid chromatography (HPLC) and gel electrophoresis with silver staining. In addition, the native protein in 50 mM Tris-sulfate and 5 mM EDTA at pH 7.5 was measured for activity to ascertain the formation of native protein structure (see Esterase Activity section). For all refolding experiments, the lyophilized protein was solubilized in 5 M GuHCl for at least 6 hours. Prior to experimental use, the denatured protein was filtered with 0.22 μm syringe filters (Gelman Sciences, Ann Arbor, MI) to remove any remaining insoluble protein.

Recombinant Human Deoxyribonuclease (rhDNAse): Two different forms of recombinant human deoxyribonuclease (rhDNAse) were provided by Genentech, Inc (S. San Francisco, CA). First of all, a 4 ml solution of native rhDNAse (1 mM $CaCl_2$, 150 mM NaCl, pH 7) produced from CHO cells was provided at a protein concentration of 4 mg/ml. The solution was divided into three separate aliquots. For the initial experiments, one aliquot was diafiltered with 8 M GuHCl buffer with 50 mM Tris, 10 mM $CaCl_2$, and 10 mM $MgCl_2$ at pH 8 by using ultrafiltration centrifuge tubes (10,000 MWCO; Millipore, Bedford, MA). One milliliter of the native protein was mixed with 1 ml of the 8 M GuHCl solution and placed in the ultrafiltration centrifuge tube. The tube was then centrifuged for 15 minutes in an IEC Centra 4 table top centrifuge (International Equipment Corporation, Needham Heights, MA) at 10,000 rpm. This

procedure resulted in a retentate of one half the original volume. One milliliter of the 8 M GuHCl solution was then added to the retentate and the solution was recentrifuged. This procedure was repeated until a final GuHCl concentration of 7 M was achieved in the retentate at a final volume of 1 ml. Half of the denatured rhDNAse in 7 M GuHCl was reduced by the addition of β-mercaptoethanol to a final concentration of 50 mM. The second aliquot (1 ml) of the native protein solution was diafiltered with 8 M urea by the same method resulting in the final conditions of 7.8 M urea, 50 mM Tris, 10 mM $CaCl_2$, and 10 mM $MgCl_2$ at pH 8. A portion of this protein solution was reduced with the redox couple of 10 mM GSH and 1 mM GSSG. The remaining native protein was utilized as a standard for the activity assay.

The second form of the protein was rhDNAse produced as inclusion bodies in *Escherichia coli.* The insoluble protein pellets of rhDNAse were solubilized in 8 M urea with 10 mM GSH, 1 mM GSSG, 50 mM Tris, 10 mM $CaCl_2$, and 10 mM $MgCl_2$ at pH 8. Gel electrophoresis of this final solution indicates that the inclusion body pellets are approximately 80% rhDNAse and that the inclusion body protein may not be completely reduced (see Appendix).

Recombinant Tissue Plasminogen Activator (rtPA): Recombinant glycosylated tissue plasminogen activator (rtPA) was expressed in CHO cells and provided as 100 mg of lyophilized protein by Genentech, Inc. (S.San Francisco, CA). Fifty milligrams of the lyophilized protein was denatured in 5 ml of 5 M GuHCl, 50 mM Tris-sulfate, and 5 mM EDTA at pH 7.5. Any unsolubilized protein was removed by filtration with 0.22 μm syringe filters (Gelman Sciences, Ann Arbor, MI). For use as an activity assay standard,

50 mg of lyophilized rtPA was resuspended in 50 ml of 50 mM Tris-sulfate and 5 mM EDTA at pH 7.5.

Recombinant Gamma Interferon (rIFN-γ): Two different forms of recombinant gamma interferon (rIFN-γ) were provided by Biogen, Inc. (Cambridge, MA). Both forms were produced from recombinant strains of *Escherichia coli.* A lyophilized powder of pure rIFN-γ was supplied and this material was resuspended with 4 M GuHCl, 100 mM phosphate buffered saline (PBS) at pH 7. The other form of rIFN-γ was the insoluble inclusion body pellets. These pellets were washed with PBS at pH 7 and solubilized by addition of the 4 M GuHCl. The remaining insoluble protein was removed by centrifugation at 10,000 rpm with a Sorvall centrifuge (DuPont, Wilmington, DE). The supernatant was filtered with 0.22 μm syringe filters (Gelman Sciences, Ann Arbor, MI) to remove any suspended aggregates. The resulting protein solution was analyzed by gel electrophoresis and size exclusion chromatography both of which indicate that the solution was approximately 60% rIFN-γ (see Appendix).

Protein Concentration

The protein concentration of CAB was measured by two different methods. Protein concentration for native CAB in 50 mM Tris-sulfate, 5 mM EDTA, at pH 7.5 was determined by absorbance at 280 ηm using an extinction coefficient of 1.83 $(mg/ml\ protein)^{-1}\ cm^{-1}$ and a molecular weight of 30,000 (Wong & Tanford, 1973). For CAB denatured in 5 M GuHCl, the concentration was determined by using a colorimetric dye binding assay, Bio-Rad reagent (Bio-Rad Laboratories, Richmond, CA) or BCA reagent

(Pierce Chemical Company, Rockford, IL), with bovine serum albumin (BSA) denatured in 5 M GuHCl as the standard. CAB in GuHCl concentrations of 2 M or greater was diluted with 8 M GuHCl to final concentration of 5 M GuHCl and analyzed with the BSA standards in 5 M GuHCl. The concentration of the other proteins was also determined by the colorimeteric dye binding assay with BSA dissolved in the same buffer as the standard.

Activity Assays

Esterase Activity (CAB): Enzymatic activity of CAB was determined by using the esterase reaction as described previously (Pocker & Stone, 1967). The unfolded CAB in 5 M GuHCl was rapidly diluted to the desired final protein and GuHCl concentrations and each aliquot of the refolded sample was analyzed for its enzymatic activity at various times after dilution. Each assay sample was diluted ten fold by 50 mM tris sulfate, 5 mM EDTA, pH 7.5, prior to addition of substrate, pNPA. The formation of pNP and decrease in pNPA were measured by absorbance at 348 ηm and 400 ηm, respectively, on a Model 8452 diode array spectrophotometer (Hewlett Packard, Mountain View, CA) using one second time intervals for two minutes after addition of pNPA. Recovery of activity was determined using the ester hydrolysis rate constant of the native protein at the same concentration in the dilution buffer (50 mM tris sulfate, 5 mM EDTA, pH 7.5).

The substrate-polymer analog, bis(p-nitrophenol carbonate) polyoxyethylene (PEGpNP, PEG 3350 MW), was also used to measure the recovery of active protein. CAB in 5 M GuHCl was diluted to the desired final protein and GuHCl concentrations

with a dilution buffer containing PEGpNP. The native protein was used as a control at the same final protein and PEGpNP concentrations. After dilution to the final conditions, the hydrolysis of the PEGpNP was determined from the formation of the product, pNP. The formation of pNP in each solution was measured by absorbance at 348 ηm and the absorbance was taken every 5 seconds for 30 minutes using a Model 8452 diode array spectrophotometer (Hewlett Packard, Mountain View, CA). The hydrolysis rate, $\Delta pNP/\Delta t$, was then calculated over each time interval for both the sample and the native protein. The fraction of active protein in the refolding solution was defined as the ratio of the hydrolysis rate of sample to the hydrolysis rate of the native protein over the same time interval. Therefore, the concentration of active protein, $[CAB]_{active}$, can be calculated from the hydrolysis rates and the total protein concentration, $[CAB]_f$, as follows:

$$[CAB]_{active} = \{\Delta pNP|_{Refold} / \Delta pNP|_{Native}\}_{\Delta t} [CAB]_f$$

The concentration of active protein was overestimated when the native protein became substrate limited since the hydrolysis rate of the native protein ($\Delta pNP/\Delta t$) decreased under substrate limiting conditions.

Methyl Green Plate Assay (rhDNAse): The methyl green DNA plate assay was used to determine the concentration of active rhDNAse. Methyl green DNA plates were prepared by first dissolving 2 g of salmon sperm DNA (sodium salt) in 1 L distilled water by stirring at 50 - 70°C. After dissolving the DNA, 100 mg of methyl green was added to the solution and stirred at the same temperature until completely dissolved. Seventeen grams of agarose was then mixed with the solution at 100°C until completely dissolved.

The solution was not boiled and heating to 100°C was assumed to sterilize the media. Finally, the solution was cooled to 60°C and 5 ml of 1 M $CaCl_2$ was added. The resulting solution was quickly added to agar plates at a volume of 25 ml per plate. For accurate repeatable assays, it was important to evenly distribute the media in the plate without surface deformations (eg. bubbles). All plates were stored at 4°C and warmed to 20°C prior to use.

The activity assay was performed by the dropwise placement of 5 µl of each sample or standard on the plate surface at equally distributed locations (20 samples per plate). On each plate, native rhDNAse was used as a standard at a concentration range of 5 to 100 µg/ml. The sample plates were then incubated at 37°C for a minimum of 15 hours. After incubation, the activity was determined by measuring the diameter of methyl green clearance for each sample and standard. A linear correlation between the logarithm of the standard rhDNAse concentration and its clearance diameter was used to calculate the concentration of active rhDNAse in each sample. Each sample and standard was assayed in duplicate and the duplicates did not deviate by more than ± 5 %.

Protease Activity (rtPA): The protease activity of recombinant tissue plasminogen activator (rtPA) was assayed as described previously (Friberger, 1982). Briefly, 100 µl of each sample or standard was diluted with 1.37 ml of 50 mM Tris and 5 mM EDTA, pH 7.5. A stock solution of the chromogenic peptide, S-2288, was prepared by dissolving the solid in the same buffer to a final concentration of 10 mM. After sample dilution, 30 µl of the S-2288 solution was mixed with the diluted sample. The formation of p-nitroaniline and decrease in S-2288 were measured by absorbance at 406 ηm and 316 ηm,

respectively, on a Model 8452 diode array spectrophotometer (Hewlett Packard, Mountain View, CA) using one second time intervals for two minutes after addition of S-2288. The concentration of protease active rtPA in each sample was determined by using the native protein as a standard over a protein concentration range of 0.01 to 1.0 mg/ml.

Cytotoxic Assay (rIFN-γ): To determine the activity of rIFN-γ after refolding, the cytopathic effect assay was used based on the procedure of Custer and Howe (Robbins, 1989). This procedure was performed by Biogen, Inc. (Cambridge, MA). Samples were prepared by dilution of the protein from 4 M GuHCl to a given final GuHCl and protein concentration. To allow complete refolding, the protein was incubated at 20°C for six hours and, then diluted 100 fold with phosphate buffered saline (PBS) at pH 7. After dilution, each sample was sterile filtered with 0.22 μm syringe filters (Gelman Sciences, Ann Arbor, MI). Sterile filtration was performed in a laminar flow hood and the samples were filtered directly into sterile screw top vials (Corning, St. Louis, MO). The rIFN-γ concentration of the filtrate was measured by the protein concentration assay described above. All samples were maintained frozen at -40°C prior to analysis.

Quasi-elastic Light Scattering (QLS)

System Description and Measurements: The quasi-elastic light scattering analysis was performed using two different systems. Initial QLS measurements were performed by using a Model N4 submicron particle analyzer (Coulter Electronics, Hialeah, FL) instrumented as described previously (Yarmush et.al., 1988). The photomultipler assembly was positioned at 90° to the incident laser beam (5 W argon ion, Model 2020-

05, Spectra Physics, Mountain View, CA). Samples were analyzed at a constant temperature of 20°C with a total volume of 200 µl. All buffers and protein solutions were prefiltered with 0.22 µm syringe filters (Gelman Sciences, Ann Arbor, MI) to remove any dust particles which would alter the QLS measurements. Glassware was rinsed in acetone and filtered distilled water to remove excess dust. QLS sample tubes (6 x 50mm, VWR Scientific, San Francisco, CA) were checked for imperfections prior to use.

Additional QLS measurements were performed using a Brookhaven light scattering system (Brookhaven Instruments, Holtsville, NY). The system consisted of a BI200SM goniometer with a photomultiplier positioned at 90° to the incident laser, 2 W argon ion at 488 ηm (Lexel, Fremont, CA). The goniometer assembly was temperature controlled at 20°C with an external water bath. In addition, the sample was placed in the goniometer with index matching fluid, *cis,trans* - decahydronapthalene, surrounding the sample. The photon data was collected using a BI2030 autocorrelator with 136 channels. A personal computer was used for system control and data storage. Prefiltered samples (≥ 1 ml) were placed in glass disposable culture tubes (12 x 75mm, VWR Scientific, San Francisco, CA) which were checked for imperfections and precleaned as described above prior to use.

The methods of determining particle size distributions from QLS autocorrelation function data have been well documented (Stock & Ray, 1985). The method of constrained regularization or CONTIN as described by Stock and Ray was used in the analysis of all autocorrelation data. The Model N4 system deploys a size distribution processor (SDP) system which utilizes CONTIN to calculate particle size distributions

(Coulter Electronics, 1984). The Brookhaven Instruments systems utilizes a personal computer version of the original CONTIN analysis protein to determine the particle size distribution as described in the Appendix (Provencher, 1983). The SDP and CONTIN analysis methods were utilized for CAB refolding and aggregation measurements.

After denaturation in 5 M GuHCl, CAB was measured by QLS to ascertain that only the unfolded random coil existed in solution. A mean hydrodynamic diameter of 10 ηm was obtained for each denaturation. The theoretical hydrodynamic diameter of the unfolded protein was calculated as 10.4 ηm using the reported intrinsic viscosity (Wong & Tanford, 1973; Corbett & Roche, 1984). Refolding was carried out by rapid dilution with a dilution buffer containing 50 mM Tris-sulfate and 5 mM EDTA, pH 7.5, to the desired final protein and GuHCl concentrations. For kinetic studies, the sample was analyzed by QLS immediately after dilution. The earliest possible time point obtained by QLS measurements was thirty seconds which is the minimum time needed for the autocorrelation channels to stabilize. The sample time used to collect autocorrelation data for each experiment was thirty seconds and the midpoint of this time period was used to determine the rate of association. For more rapid association kinetics, a sample time of twenty seconds was used to obtain the autocorrelation data. Each experiment was repeated a minimum of three times with the same time points to assure repeatability and to reduce the difficulties of signal to noise. Continuous measurements were made during the refolding until an equilibrium state was achieved or until micron size aggregates dominated in the sample. Each experiment was repeated several times to provide a significant number of time points for kinetic analysis. For the equilibrium studies, each

sample was analyzed at least two to three times to ascertain the repeatability of the particle size distribution at equilibrium. The particle size distributions for each solution showed a high degree of reproducibility.

Multimer Model of Submicron Aggregates: To determine the unit composition of the particles initially formed during aggregation, a model of the multimeric state of the protein was developed based on the worm-like chain model (Kratky & Porod, 1949). This model has been shown to successfully characterize the composition of antigen-antibody complexes which were analyzed by QLS (Murphy, 1989; Murphy, et.al., 1990). For purposes of this study, the multimers were assumed to consist of individual unit monomers. The unit monomer hydrodynamic diameter, D_h, was determined from the hydrodynamic diameter of the native structure at the same final protein and denaturant concentrations. The native protein structure was observed to change under different final concentrations and over long time scales to eventually reach the same equilibrium state as the refolding experiments (see Results and Discussion). Therefore, if the measured native protein hydrodynamic diameter was greater than that of the first observed intermediate, the diameter of the first intermediate was employed in the calculation of multimer size. For kinetic experiments, the mean hydrodynamic radius of the monomer, native or first intermediate, was used as the persistence length in the worm-like chain model. The hydrodynamic radius of CAB at 0.10 mg/ml protein (3.3 μM) in 2 M GuHCl was used as the persistence length for the equilibrium studies.

Two multimeric states were calculated utilizing the chain model as detailed in the Appendix. First, the dimer was calculated assuming a contour length, L_i, ranging from

two contacting spheres with 25% of their volume overlapped (1.5 D_h) and a linear chain of solid spheres with no excluded volume (2 D_h). The trimer of CAB was the other multimeric specie which was assumed to range in contour length from a three sphere cluster (2 D_h) to a linear chain of solid spheres (3 D_h). In addition, the linear form of the trimer was considered energetically more favorable than the triangular aggregates as determined from previous studies on the potential energy relations of native protein aggregates (Reithel, 1963). Higher order multimers were not evaluated and modelled since the maximum observed diameter did not exceed the linear trimer specie. Since the true multimer diameters are not known, a normal distribution about the mean diameter was used to calculate the actual concentration of the multimeric species (Murphy, 1989; Murphy, et.al., 1990; also see Appendix). The validity of this model for CAB association was confirmed by studies using size exclusion chromatography (see Results and Discussion, Chapter 2). These chromatography studies yielded the same multimer concentrations as those calculated using the worm-like chain model. The particle structure factor calculated using other models also does not vary significantly over the range of hydrodynamic radii observed in these experiments (Murphy, 1989). Therefore, alternative model systems may yield similar multimer concentrations.

Cross-Linking Analysis

To confirm the validity of the multimer model, cross-linking experiments were performed during the refolding process. The cross-linking agent, DMS, was found to have no effect on the native protein under the same final protein and denaturant conditions. Denatured CAB in 5 M GuHCl was rapidly diluted with buffer (50 mM Tris

sulfate, 5 mM EDTA, pH 7.5) to a final protein and denaturant concentration as desired for the experiments. DMS at a final concentration of 20 mM was added after a given time period to cross-link the multimers (Swaney & O'Brien, 1978; Hadju et.al., 1976). After one minute, the cross-linking was quenched by the addition of 1 M ammonium acetate to yield a final concentration of 100 mM. Further quenching was also performed by the addition of concentrated GuHCl to yield a final concentration of 5 M GuHCl which prevented further aggregation. Samples were analyzed by SDS-PAGE using 8-25% gradient polyacrylamide gels with SDS buffer strips on a Phast Gel Electrophoresis System (Pharmacia LKB Biotechnology, Uppsala, Sweden). The gels were stained with Coomassie Blue using the development section of the Phast System.

High Performance Liquid Chromatography

All HPLC analyses were performed with a model HP1090 analytical HPLC equipped with a diode array detector (Hewlett Packard, Mountain, View, CA). Size exclusion and hydrophobic interaction chromatography were used to study protein association and PEG binding, respectively. A Protein PAK 3000 SW column (Waters, Bedford, MA) was used for each size exclusion experiment. Prior to operation, the column was equilibrated with ten column volumes (100 ml) of elution buffer at the same denaturant concentration as the sample. A sample volume of 25 µl was applied to the column and eluted at a flow rate of 1.0 ml/min to facilitate rapid separation. For equilibrium experiments, each sample was equilibrated for three to eight hours before column separation. In PEG studies, the elution buffer contained the same concentration of PEG (8000 MW) and GuHCl as the sample.

For the size exclusion experiments, elution times for native CAB and BSA in the standard dilution buffer (50 mM Tris sulfate, 5 mM EDTA, pH 7.5) were used to calculate the size of the resultant associated species. In addition, CAB and BSA at low protein concentrations (< 0.10 mg/ml) in 2 M GuHCl were also utilized to determine the effect of denaturant on elution time. A similar calibration was performed for different PEG concentrations in 2 M GuHCl to observe the effect of PEG on the elution time. The results obtained were comparable to those determined previously for proteins in different denaturants (Corbett & Roche, 1984). Since the extinction coefficient at 280 ηm for CAB in 2 M GuHCl was not observed to change with protein concentration, the extinction coefficients of the monomer, dimer, and trimer species were assumed to be equal for all equilibrium experiments. With this assumption, the peak areas from absorbance measurements at 280 ηm were used to calculate the concentration of each species.

For PEG equilibrium binding studies, a hydrophobic interaction chromatography column with PEG as the weak hydrophobic ligand was utilized (Hydropore 5-HIC, Rainin Instrument Company, Woburn, MA). The column was equilibrated in the same concentration of GuHCl as the sample prior to operation. CAB in 5 M GuHCl was diluted to different concentrations of GuHCl and 0.20 mg/ml (6.7 μM) protein. After sample equilibration, a sample volume of 25 μl was loaded onto the column at a flowrate of 0.50 ml/min with buffer at the same GuHCl concentration. After one column volume of running buffer, bound protein was eluted by a step change to 50 mM Tris-sulfate buffer (pH 7.5). The concentration of bound protein was calculated based on the peak area for absorbance measurements at 280 ηm.

Fluorescence Measurements

Intrinsic Fluorescence: To measure the conformation changes in each protein, the intrinsic tyrosine or tryptophan fluorescence of each protein was measured. Each sample was prepared by rapid dilution to the desired final protein and denaturant concentration. After equilibration, each sample was placed in a quartz cuvette and the fluorescence was measured with a Model LS50 fluorimeter (Perkin-Elmore, Newton Centre, MA). The fluorimeter was controlled with a personal computer to facilitate data collection. At high final protein concentrations, an attenuating filter which reduces the emission intensity by 10 fold was used. The slit width for each reading was 10 ηm for the excitation beam and 5 ηm for the excitation beam. To measure the combined effect of tyrosine and tryptophan residues, the excitation wavelength was set to 280 ηm and the emission wavelength was 340 ηm (Creighton, 1984). An excitation wavelength of 296 ηm was utilized with the same emission wavelength to measure the change in the tryptophan environment (Stein & Henkens, 1978; Rodionova et.al., 1989). Each measurement was performed for the equilibrated refolded protein solutions and the native protein in the appropriate buffer. The relative fluorescence or change in fluorescence upon denaturation was defined as the ratio of the sample fluorescence to the native protein fluorescence (F/F_o) which was measured at the same wavelengths. In addition, control experiments were performed with N-acetyltryptophan in different concentrations of GuHCl and PEG. The fluorescence of these samples was not significantly different than the fluorescence of N-acetyltryptophan in Tris-sulfate buffer at pH 7.5. Fluorescence emission spectra were also measured to confirm the shift in the fluorescence maxima resulting from

refolding of the protein (see Appendix).

Fluorescence Quenching: Fluorescence quenching experiments were performed to measure the solvent accessibility of tryptophan residues in CAB. CAB has seven tryptophan residues and the native state has one tryptophan, ^{207}Trp in the active site, exposed to solvent (Kannan et.al., 1972). The six tryptophan residues which are incorporated in the native structure are exposed in 2 M GuHCl (Yazgan & Henkens, 1972). Iodide was chosen as the quenching agent since it is a large quenching molecule which does not readily diffuse into the interior of the protein. CAB was refolded by rapid dilution to a series of different final GuHCl concentrations and 0.10 mg/ml (3.33 μM) protein with and without 30 g/l PEG (8000). Separate stock solutions of potassium iodide were prepared at a concentration of 5.0 M in the same concentrations of GuHCl and PEG. After equilibrating each refolding sample for six to eight hours, the appropriate stock solution of potassium iodide was added to a final concentration of 100 mM. Each sample was then equilibrated an additional two hours. To obtain the relative quenching, duplicate samples were prepared without iodide at the same final protein and GuHCl concentrations as those used prior to iodide addition. After equilibration, the final protein concentration in each solution was corrected for the volume change from the iodide addition which occurred in the previous samples and, with this correction, both sets of solutions had the same overall final protein concentration. The fluorescence of each solution was analyzed by using a LS50 fluorimeter (Perkin-Elmer, Newton Centre, MA) set at an emission wavelength (5 ηm slit width) of 340 ηm and an excitation wavelength (10 ηm slit width) of 296 ηm. The relative fluorescence quenching was then defined as the fluorescence

without iodide divided by the fluorescence with iodide (F_o/F). Additional studies on the fluorescence quenching effect of acrylamide and the Stern-Volmer plots of these results are described in the Appendix.

Electron Spin Resonance

Spin Labelling Methodology for CAB: To effectively measure conformational changes or specific protein surface properties, a spin probe was chosen which labelled a limited number of residues and became incorporated into the interior of the protein. Since CAB has eighteen lysine residues, a typical amine reactive spin label such as maleimido-TEMPO was observed to randomly label the protein and labelling could not be performed in GuHCl. Therefore, to specifically label residues in the interior of the native protein, a tyrosine spin label was chosen. N-(2,2,5,5, tetramethyl-3-carbonyl-pyrroline-1-oxyl) imidazole (N-imidazole-3-carbonyl-PROXYL) labels only tyrosine residues on the protein. The label was synthesized from 3-carboxy-PROXYL and 1,1'-carbonyl diimidazole as described previously (Barratt et.al., 1969). It should be emphasized that the synthesis of this label was only successful when extreme precautions were taken to remove residual water from the acetone and all chemical apparatus. The final synthesis product was stored desiccated at 0°C until use. To confirm the successful synthesis of this label, poly-L-tyrosine (27,000 MW) and poly(tyrosine,glutamate) (1:1, sodium salt, 30,000 MW) were labelled by the method of Barratt et.al.(1969).

Prior to labelling the protein, it was necessary to determine the number of tyrosines exposed in each structural state: native, first intermediate (2 M GuHCl), and

unfolded (5 M GuHCl). The number of tyrosines exposed in each state was titrated with the tyrosine specific reagent, tetranitromethane (TNM) (Sokolovsky et.al., 1966). N-acetyltyrosine was used as a titration standard. TNM labelling of all samples was performed for 2 hours with 100 fold excess TNM at 20°C. The extent of nitration for each sample was determined by measurement of the byproduct, nitroformate ion, at 350 ηm (Sokolovsky et.al, 1966). To remove excess TNM, each solution was then dialyzed with ten fold excess buffer overnight at 4°C. The resulting nitrotyrosine residues gave the solutions a distinctive yellow color and the difference in color between each labelled state could be visually observed. Titration of the native protein in phosphate buffer (pH 7) indicated the prescence of a single solvent accessible tyrosine residue which is likely to be ^{7}Tyr in the active site (Eriksson et.al., 1988; for a discussion of the homology between human CA II and CAB see Structural Analysis Programs below). The first intermediate state (2 M GuHCl) has six solvent exposed tyrosines and nitration of the unfolded protein (5 M GuHCl) resulted in the nitration of all eight tyrosines. These results indicated that it would be necessary to block the solvent accessible tyrosine in the active site to prevent labelling at this tyrosine. When labelling of CAB was performed without blocking ^{7}Tyr, the ESR spectra did not change significantly during refolding or the rate of change was more rapid than the time scale for analysis which was approximately 1 minute. In addition, previous studies using CAB spin labelled at amino groups were successful in measuring the change from the unfolded to first intermediate state, but these studies were not successful in distinguishing first intermediate from the native protein (Semisotnov et.al., 1987; Ebert et.al., 1990).

Since the studies of amine spin labelled CAB indicated that residues which are exposed in the native state will dominate the spectra causing only slight observable changes for refolding from the first intermediate to the native protein, it was necessary to block the single exposed tyrosine in native CAB with TNM prior to spin labelling. The native protein (2 mg/ml) in phosphate buffer at pH 7 was mixed with 100 fold molar excess of TNM for 2 hours. The solution was then dialyzed overnight at 4°C with 10 fold excess phosphate buffer. After removal of excess TNM, 5 M GuHCl in phosphate buffer (pH 7) was added to a final concentration of 2 M GuHCl. The protein concentration was then adjusted to 0.20 mg/ml with 2 M GuHCl to avoid protein association during labelling. The tyrosine spin label, N-imidazole-3-carbonyl-PROXYL, was then added in 10 to 20 fold molar excess using a stock solution of the label in methanol at 8 mg/ml. The protein-spin label mixture was incubated with stirring at 20°C for a minimum of five hours. After incubation, unreacted label was removed by successive overnight dialyses against 2 M GuHCl with 5% isopropanol and 5 M GuHCl. The final protein solution was concentrated to at least 5 mg/ml CAB in 5 M GuHCl prior to use. The number of spin labels per protein as determined by TNM titration was three. The spin-labelled protein (spCAB) was refolded by rapid dilution to 1 M GuHCl and 0.50 mg/ml protein to assure the recovery of active protein. The protein recovered activity with the similiar kinetics as those observed with unlabelled CAB. At equilibrium, the protein only recovered 80% its fully active form. However, control experiments with CAB nitrated in the native state without spin labelling resulted in the same recovery of activity. The labelling of ^{7}Tyr in the active site reduces the esterase activity of the native protein. Therefore, it was

assumed that spCAB had the same folding pathway and structure as the unlabelled protein.

Spin Labelling for PEG: For determination of the specificity and stoichiometry of interaction between PEG and CAB, spin labelled PEG was prepared. Monomethoxy polyoxyethylene amine (5000 MW) was first dissolved at 2 g/l in 100 mM phosphate buffer, pH 7.5. The spin label reagent, 3-maleimido-TEMPO, was then added to this solution in a 100 fold molar excess. The spin label and polymer were mixed for 24 hours at 4°C. After mixing, undissolved spin label reagent was removed by filtration with 0.22 μm syringe filters (Gelman Sciences, Ann Arbor, MI). The filtrate was dialyzed against phosphate buffer with 10% isopropanol for 24 hours at 4°C. A portion of the dialyzed solution was removed for amine analysis and the remaining solution was dialyzed against Tris-sulfate buffer (pH 7.5) for 24 hours at 4°C. The aliquot removed after the first dialysis was analyzed for free amino groups by titration with 2,4,6,-trinitrobenzene sulfonic acid (TNBSA) with N-α-acetyllysine as the standard (Fields, 1971). This analysis indicated that the polymer had been completely labelled. The ESR spectra of the resulting polymer was measured to confirm the presence of spin label.

Experimental System and Data Analysis: ESR experiments were performed by diluting the spin labelled CAB (spCAB) in 5 M GuHCl to 2 M GuHCl and 0.10 mg/ml (3.33 μM) protein with different concentrations of PEG (8000 MW). For the spin labelled PEG (spPEG) experiments, unlabelled CAB in 5 M GuHCl was diluted to 2 M GuHCl and 0.10 mg/ml (3.33 μM) protein with different final concentrations of spPEG. After

sample equilibration for six hours, the X-band ESR spectra of each sample was measured by using either a Varian E-line or Bruker ESP 300 spectrometer both of which were set at a field modulation of 100 kHz. The spectrometer settings were maintained constant at a modulation amplitude of 1.78 Gauss, a scan time of 671 seconds, and a recording time constant of 0.65 seconds. Microwave power was set at 5 mW. Each spectrum was obtained by signal averaging over at least five scans. Digitized spectra were transferred to a VAX 6800 computer for further analysis. Spectral deconvolution and analysis was then performed by Professor Ted Randolph at Yale University (New Haven, CT).

Analysis of the spin labelled CAB samples indicated that the ESR spectra were composed of overlapping spectra, which is indicative of multiple spin label environments. Therefore, spectral deconvolution by factor analysis was performed on 10 spectra at a time with computer analysis programs based on the methods developed by Press et.al. (Malinowski & McCue, 1977a; Press et.al., 1989). Spectra were filtered with a low-pass Fourier filter before factor analysis. However, the results were not altered significantly when unfiltered data were used directly. Two of the ten eigenvalues were clearly non-zero based on the criteria of Malinowski and exceeded by almost two orders of magnitude the next largest eigenvalues, indicating that two major populations could be observed (Malinowski, 1977b). These two populations can be considered analogous to a combination of nitroxide radicals in a freely rotating, aqueous environment and in a nonpolar, highly spin-exchanged environment. Spectra reconstructed from the two associated eigenvectors provided good approximations to the experimental spectra. Double integration of the resulting basis spectra enabled estimates to be made of the

ratios of the two populations which produced the overlapping spectra.

In contrast, the ESR spectra obtained from analysis of the spPEG solutions yield a single population. Rotational diffusivities were computed for each sample by a least squares fitting of simulated spectra to the experimental data. Spectral simulations were calculated by using the modified programs from Schneider and Freed for slow motional ESR studies (Schneider & Freed, 1989). Rotational diffusivity was assumed to be described by Brownian motion. The literature values for the g tensor of the 2,2,6,6-tetramethyl 4-piperidinyl (TEMPO) radical were used in the fitting program (Capiomont et.al., 1974). The ESR spectra from these samples could be accurately simulated by varying only the rotational correlation time. Least-squares fitted spectra were in good agreement with the experimental data, and well within the noise level. Final values of rotational correlation times were insensitive to initialization conditions.

Structural Analysis Programs

Secondary Structure and Hydrophilicity Predictions: The secondary structure and hydrophilicity predictions of each protein studied was performed by applying a computer analysis program developed by the Genetics Computer Group at the University of Wisconsin, Madison (see Appendix for program listing). The program segment referred to as PeptideStructure was used to predict the secondary structure of each protein from its amino acid sequence. This program utilizes the prediction algorithms developed by Chou and Fasman as well as the overlapping probability procedure designed by Nishikawa (Chou & Fasman, 1978; Nishikawa, 1983). The Chou-Fasman algorithm was modified

to neglect the probability conditions: P_{bound} (P=probability) must be greater than 1.0 and P_α must be greater than P_β for helix predictions. In addition, a minimum of five residues was required for beta sheet formation (Devereux, 1989). The program also calculated the hydrophilicity profile of each protein according to the method of Kyte and Doolittle with a window of seven residues (Kyte & Doolittle, 1982). The application of this program has been described and demonstrated for several proteins (Jameson & Wolf, 1988).

A plot of the results from the PeptideStructure program was made by using the PlotStructure program (Devereux, 1989). This program generated two-dimensional structure plots of the predicted secondary structure and the hydrophilicity of the protein (Jameson & Wolf, 1988). The final plot presented the secondary structure as a series of different wave forms and the hydrophilicity as circles (O) for hydrophilic residues and diamonds (◇) for hydrophobic residues with the size of these symbols based upon a relative scale of hydrophilicity. Helices, beta sheets, turns, and coils were represented by sine waves, sharp saw-tooth waves, 180 degree turns, and dull saw-tooth waves, respectively. The potential glycosylation sites were also marked on the plots.

Protein Homology Analysis: To determine the homology between different protein species, the BestFit program from the Genetic Computer Group Sequence Analysis Software Package was used (Devereux, 1989; also see Appendix). This program was based upon the local homology algorithm of Smith and Waterman and the evolutionary similarity methods of Gribskov and Burgess (Smith & Waterman, 1981; Gribskov & Burgess, 1986). The algorithm fitted the two different protein sequences by inserting sequence gaps to obtain an optimal alignment of the best region of similarity (Devereux,

1989). For all analyses, a gap weight of 3.0 and a length weight of 0.10 were used. The final output indicated the percent similarity and percent identity as well as a sequence plot of the alignment. The sequence plot showed the relative match between the sequences symbolically. Identical residues, residues of high similarity, and residues of low similarity were depicted as a solid line, two dots, and a single dot, respectively, between the sequences. As an example, this program was applied to human CA II (human carbonic anhydrase C) and its species homolog, bovine CAB. The results of this comparison revealed an 89.6% similarity and 79.2% identity with no gaps required for the fit.

RESULTS AND DISCUSSION

To develop an understanding of protein aggregation and design improved refolding methodologies, the study of protein aggregation during refolding was performed on a model protein, bovine carbonic anhydrase B (CAB). In Chapters 1 and 2, the kinetics and thermodynamics of protein aggregation during refolding were studied in detail to provide the basis for the development of a model for aggregation. The model of refolding and aggregation for CAB was then postulated and its validity was confirmed through a comparison of the model prediction with the experimental results (Chapter 3).

With the knowledge of the aggregation pathway for CAB, modifications in the solvent environment were devised to prevent association (Chapter 4). These studies indicated that polyethylene glycol (PEG) had unique properties which reduced the aggregation process. Therefore, detailed studies were performed to elucidate the mechanism of PEG interaction with CAB during the refolding process. Both the specific mechanism of PEG interaction and its effect on equilibrium protein association were studied and described in Chapters 5 and 6. The results from these studies led to additional investigations on the interaction between PEG and CAB during the refolding process (Chapter 7). Using the knowledge obtained from these studies as well as the model pathway for CAB aggregation and refolding, a pathway for PEG enhanced refolding of CAB was developed. The model for this pathway was compared to the observed recovery of the biological activity and discussed in Chapter 8.

To determine the generality of the model developed in Chapter 8, PEG was applied in the refolding of other proteins. In Chapter 9, four different proteins were

refolded in the presence and absence of PEG at conditions where aggregation normally occurs. The analysis of these protein systems led to the general applicability of PEG and additional insight into the mechanism of PEG in the enhancement of refolding. The observed PEG enhanced protein refolding was correlated to the properties of each protein to obtain a knowledge of the possible general rules which govern refolding with PEG (Chapter 10). Hopefully, these rules as well as the model for PEG enhanced refolding will provide insight into the development of improved refolding methodologies.

Chapter 1. Kinetic Analysis of Carbonic Anhydrase B (CAB) Aggregation and Refolding

To understand the mechanisms of aggregation during refolding, a kinetic study of aggregation was performed at different final solution conditions. For each set of final conditions, the model protein, bovine carbonic anhydrase B (CAB), was first denatured in 5 M GuHCl. CAB in 5 M GuHCl was then diluted to different final protein and GuHCl concentrations and the aggregate distribution at the final conditions was determined. When the onset of precipitation was observed to occur on the order of minutes, the rate of association was then measured by quasi-elastic light scattering (QLS). These association rate studies provided a relationship between the final conditions and the association kinetics. Finally, the association kinetics at different final conditions were utilized to determine the pathway of aggregation.

1.1 Aggregate Distribution

In order to perform the kinetic analyses on the aggregation process, it was necessary to determine the distribution of aggregates and the final conditions where these aggregates could be observed by quasi-elastic light scattering (QLS). Therefore, refolding of CAB in 5 M GuHCl was performed by rapid dilution to different final protein and denaturant concentrations. For each refolding experiment, an initial compact structure was immediately formed upon dilution of the unfolded protein in 5 M GuHCl to low GuHCl concentrations ($\leq$ 1.0 M). As shown in the first particle size distribution (1.8 minutes) in Figure 1.1.a, the mean hydrodynamic diameter of this compact structure was approximately the same as the diameter of the native protein molecule (D_h = 4.8 $\pm$ 0.5

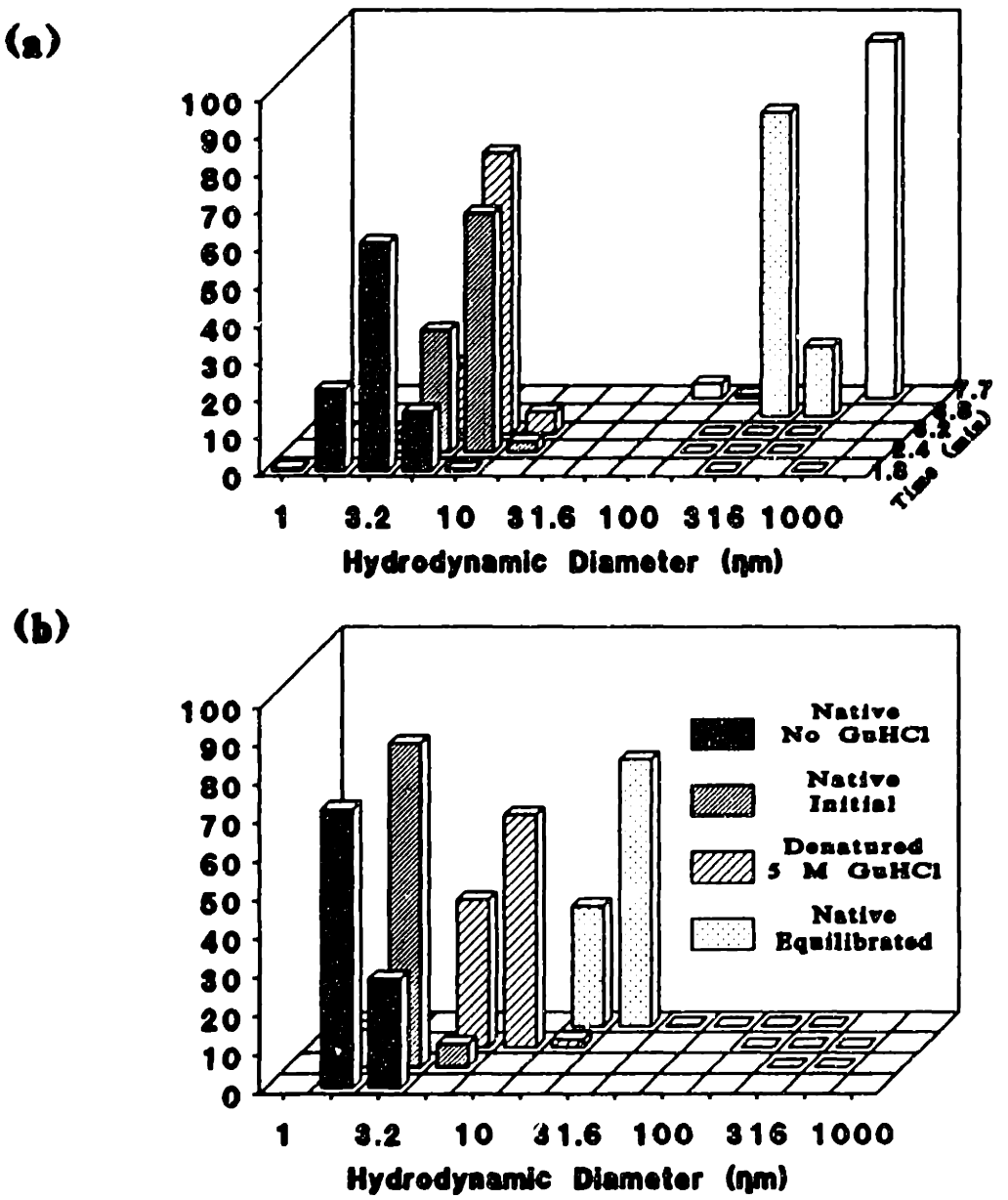


FIGURE 1.1: Size distributions for refolding, native, and denatured states of CAB. The z-average weight percent of each particle is shown over a range of hydrodynamic diameters as determined by CONTIN analysis of the QLS results. (a) Size distribution of CAB during refolding from 5 M GuHCl is shown as a function of time for refolding by rapid dilution to 0.50 mg/ml CAB and 0.70 M GuHCl. (b) Native CAB size distribution dependence on GuHCl concentration is shown. Native CAB at 0.50 mg/ml CAB and 0.60 M GuHCl was analyzed both immediately after addition of GuHCl (native initial) and after equilibration (native equilibrated). The size distribution of the native protein without GuHCl and the denatured protein in 5 M GuHCl are also included.

nm; Figure 1.1.b). For each final solution condition which resulted in aggregation, it was observed that this initial compact structure was the aggregating unit monomer which initiated the association process. As shown in Figure 1.1.a, when CAB in 5 M GuHCl was diluted to 0.70 M GuHCl and 0.50 mg/ml protein, the initial compact protein species grew in hydrodynamic diameter over time and eventually formed micron-sized aggregates. A similar effect was observed in a study of the native protein at various protein and GuHCl concentrations and a significant change in the size distribution was observed for conditions where multimers were observed to form. The addition of concentrated GuHCl to the native protein resulted in the initial formation of a structure which was larger than the native protein in the absence of GuHCl (Second distribution shown in Figure 1.1.b). The formation of this structure was caused by the destabilization of the tertiary structure of the protein as reported previously (Rodionova et.al., 1989). After equilibration in low GuHCl concentrations ($\leq$ 1.0 M), the native protein was found to have a particle size larger than the unfolded protein in 5 M GuHCl (D_h = 9.8 $\pm$ 0.8 nm). The difference in the particle size distribution between CAB in 5 M GuHCl and native CAB in low GuHCl concentrations is illustrated in Figure 1.1.b for native CAB equilibrated in 0.70 M GuHCl. In many cases, the native protein formed micron-sized aggregates after extended periods of exposure to moderate GuHCl concentrations (0.30 - 0.80 M). Since both the native protein and unfolded protein will form aggregates when placed in low concentrations of GuHCl ($\leq$ 1.0 M), both the unfolding and refolding pathways must proceed through an intermediate which associates. Therefore, these studies indicated that the initial intermediate structure formed in the refolding process (t= 1.8 minutes, Figure 1.1.a) could

be considered the unit monomer for association to multimers and larger aggregates.

1.2 Final Protein and Denaturant Concentration Effects on Aggregation

After measuring the unit monomer size distribution and the formation of submicron aggregates, multimer concentrations for each CAB refolding study were calculated based on the QLS results. These results indicated that both a dimer and trimer species exist prior to the formation of micron-sized aggregates (see Materials and Methods). The concentration of monomers, dimers, and trimers could only be determined for final conditions where the formation of micron size particles occurred on a time scale of greater than 1 minute. In many cases, refolding from 5 M GuHCl to low final GuHCl concentrations ($\leq$ 1.0 M GuHCl) and high protein concentrations (mg/ml) resulted in the formation of precipitates during the dilution step. For each of these final conditions, the rate of association could not be accurately determined and this rapid formation of aggregates was distinct to a given set of final conditions. The next distinct type of aggregation was the rapid formation of multimers prior to micron-sized aggregate formation. When unfolded CAB in 5 M GuHCl was rapidly diluted to the final conditions of 0.50 mg/ml (16.7 μM) protein and 0.60 M GuHCl, the protein monomer concentration decreased from 0.50 to 0.11 mg/ml (3.8 μM) in 1.5 minutes with the subsequent formation of the dimer and trimer species as shown in Figure 1.2.a. For this case, micron-sized aggregates dominated the scattering after 1.5 minutes.

The onset of micron-sized aggregates was delayed by increasing the final GuHCl concentration to 0.70 M at the same final protein concentration (0.50 mg/ml, 16.7 μM) (Figure 1.2.b). Rapid dilution of denatured CAB in 5 M GuHCl to 0.50 mg/ml protein

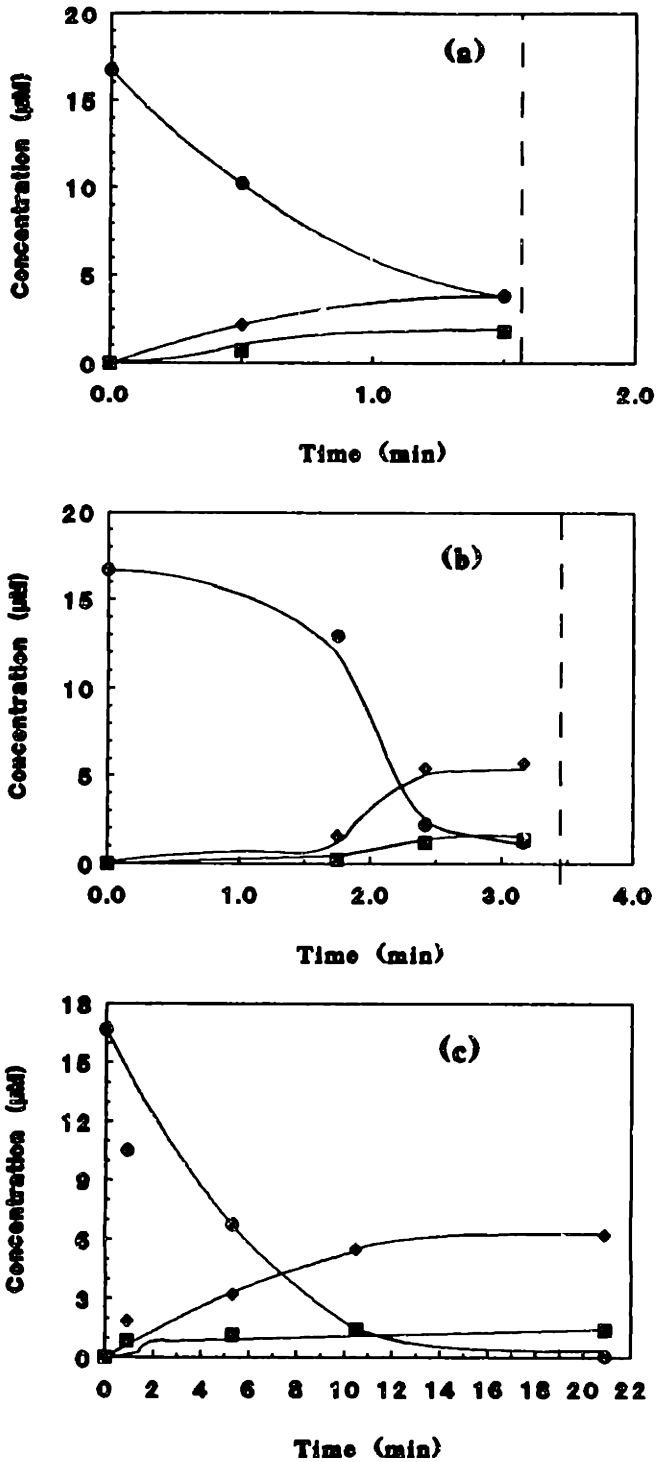


FIGURE 1.2: Refolding and aggregation of CAB was observed by QLS which provided monomer (●), dimer (♦), and trimer (■) concentrations as a function of time after dilution of CAB in 5 M GuHCl to 0.50 mg/ml (16.7 μM) CAB and different GuHCl concentrations. (a) Rapid formation of multimers (0.60 M GuHCl) (b) Moderate rate of multimer formation (0.70 M GuHCl) (c) Slow formation of multimers (0.80 M GuHCl). Dashed lines represent the time at which micron-sized aggregates were observed in the solution.

and 0.70 M GuHCl showed a decrease in the rate of multimer formation (Figure 1.2.b) when compared to refolding at 0.60 M GuHCl (Figure 1.2.a). For refolding at 0.70 M GuHCl, the final observed monomer concentration before large aggregate formation was significantly reduced (0.04 mg/ml, 1.2 μM).

The monomer concentration was observed to proceed to zero in the absence of large aggregates when the final GuHCl concentration was further increased to 0.80 M at the same final protein concentration (0.50 mg/ml, 16.7 μM). After dilution of the denatured CAB in 5 M GuHCl to 0.50 mg/ml (16.7 μM) and 0.80 M GuHCl, multimers formed relatively slowly and achieved a stable state or equilibrium after approximately 15 minutes (Figure 1.2.c). This stable state persisted for several hours and did not result in micron-sized aggregates. Several other final protein and denaturant conditions yielded the same type of distinct stages of multimer formation and micron-sized aggregation.

Combining the results of many similar experiments, an operating diagram was constructed (Figure 1.3). This diagram indicates the distinct regimes of aggregation and refolding which occurred at different final CAB and GuHCl concentrations. In each case, the denatured protein in 5 M GuHCl was rapidly diluted to the final conditions as shown in Figure 1.3. The aggregation regime was considered to be the region with final conditions which resulted in precipitation which was concomitant with the dilution process and, therefore, the rate of association could not be analyzed by QLS. For a constant final protein concentration and high final GuHCl concentrations, multimers were observed to form prior to the formation of large aggregates as shown previously in Figure 1.2.a/b. The final conditions where dimer and trimer achieved an equilibrium in the

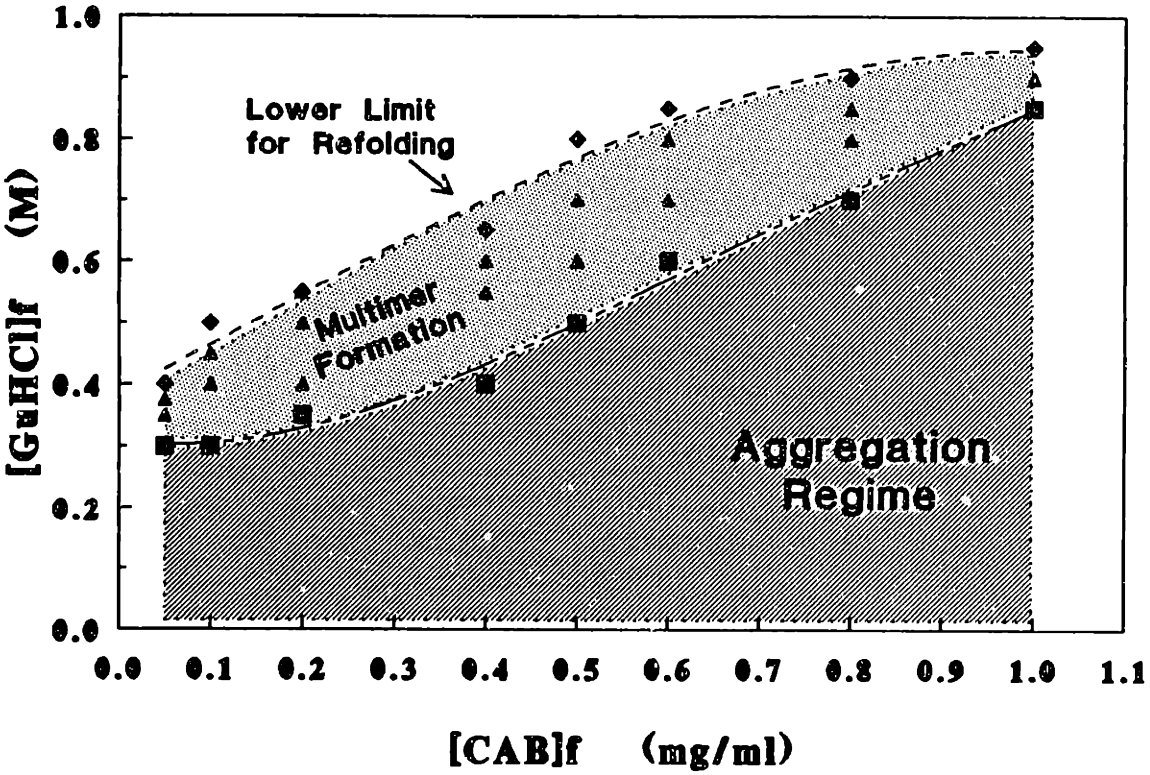


FIGURE 1.3: Regimes of refolding and aggregation of CAB. Each data point represents rapid dilution of CAB in 5 M GuHCl to a given final protein and GuHCl concentration. The aggregation regime is defined as the final solution conditions which result in the immediate formation of micron-sized aggregates. The upper boundary of the aggregation regime is defined by the lower data points (■). As depicted in Figure 1.2 (a/b), the cases where dimer and trimer species are observed prior to micron size aggregation constitute the multimer formation regime (▲). The lower limit of refolding is the distinct regime where multimers form, but do not proceed to form micron-sized aggregates (♦) (Figure 1.2.c). From the lower limit of refolding to 1 M GuHCl, the protein refolds to form either a stable intermediate or the native structure in the absence of aggregation.

absence of larger aggregates (Figure 1.2.c) was considered the lower limit of refolding where higher GuHCl concentrations at the same protein concentration resulted in refolding. The maximum final GuHCl concentration which allowed complete recovery of active protein was 1.0 M. Refolding at 1.0 M GuHCl has been previously observed to result in complete recovery of active protein (Dolgikh et.al., 1984; Stein & Henkens, 1978).

These studies indicated that refolding at high GuHCl concentrations (1.0 M) is required to avoid the formation of aggregates. The results shown in Figure 1.3 also revealed a correlation between the formation of associated species and the final conditions for refolding. The rate of association decreased as the final GuHCl concentration increased at a constant final protein concentration. If associated species form at high GuHCl concentrations, the association rate at these conditions would be much lower than that observed at the lower limit of refolding (Figure 1.2.c) and the overall association rate would likely be slower than the rate of refolding. It is therefore conceivable that a transient associated species is present during refolding at high GuHCl concentrations. Since this proposed associated species is only transient, the protein would eventually recover full biological activity.

1.3 Final Protein Concentration Effect on Refolding Conditions (1.0 M GuHCl)

If a transient association process occurs at 1.0 M GuHCl, the rate of refolding at this final condition should decrease with increasing protein concentration. When refolded from 5 M GuHCl to 1 M GuHCl, CAB was observed to refold to achieve complete recovery of activity (Stein & Henkens, 1978). However, the rate of refolding at 1.0 M

GuHCl was not previously determined for different final protein concentrations. Several refolding experiments were performed by rapid dilution of the unfolded protein in 5 M GuHCl to 1 M GuHCl at different final protein concentrations. As shown in Figure 1.4, the concentration of active protein was then measured as a function of time after dilution. The recovery of active protein occurred more slowly with increasing protein concentration. Therefore, the rate of refolding was dependent on the initial protein concentration and the refolding pathway at high protein concentrations ($\geq$ 0.20 mg/ml, 6.7 μM) may involve the formation of an associated species.

Although the rate of refolding was reduced with increasing protein concentration, refolding at 1.0 M GuHCl and each of the final protein concentrations shown in Figure 1.4 did result in the complete recovery of active protein after 1 hour. Therefore, if an associated specie did form at higher protein concentrations, it was only transiently populated. To determine if association was occurring at high protein concentrations, the protein was refolded to 1.0 mg/ml (33.3 μM) and 1.0 M GuHCl. At different times after dilution, samples of the refolding solution were analyzed by HPLC size exclusion to measure the formation of associated protein (Figure 1.5). At these final conditions, HPLC size exclusion results revealed the presence of a dimer species which decreased in concentration as a function of time after dilution. The association process was more rapid than the HPLC size exclusion analysis and, therefore, only the dissociation rate could be measured. Since the concentration of the dimer species decreased with time, the formation of this associated state was a transient process which only reduces the rate of refolding and not the final concentration of active protein. As a control, refolding was

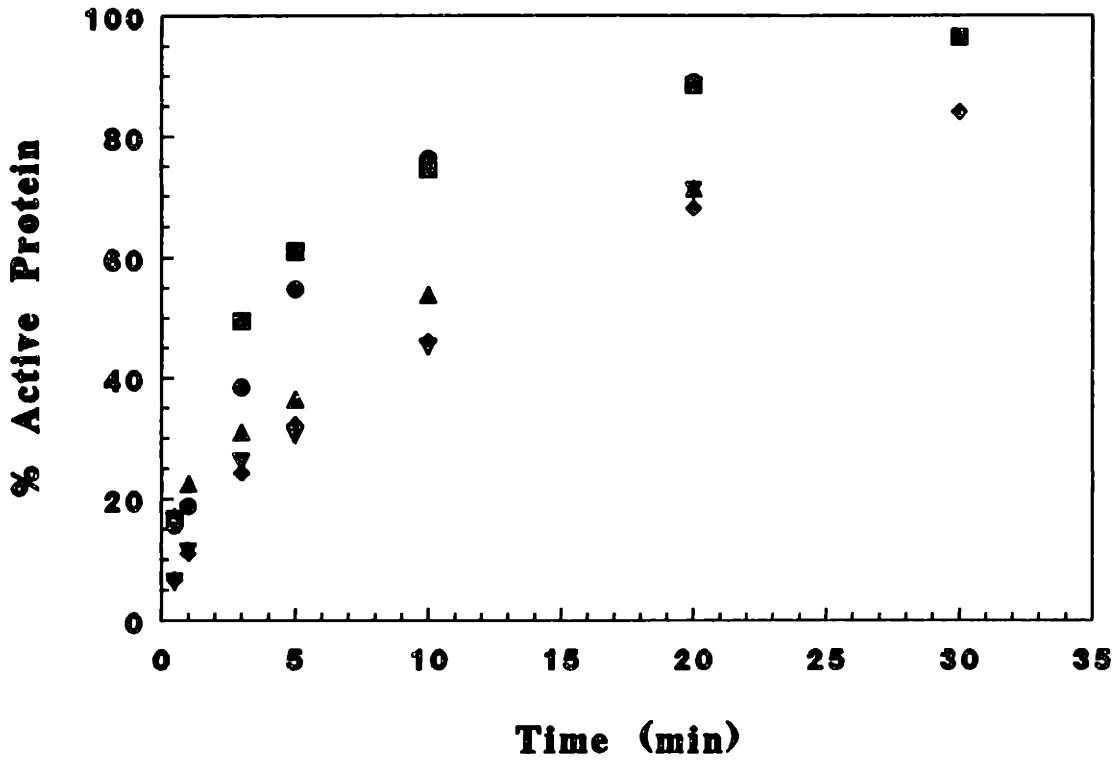


FIGURE 1.4: Refolding rate dependence on final protein concentration. CAB was refolded by rapid dilution from 5 M GuHCl to 1 M GuHCl and several different final protein concentrations. The recovery of activity was then measured as a function of time after dilution for each protein concentration: 0.10 (●), 0.20 (■), 0.35 (▲), 0.50 (♦), and 1.0 (▼) mg/ml.

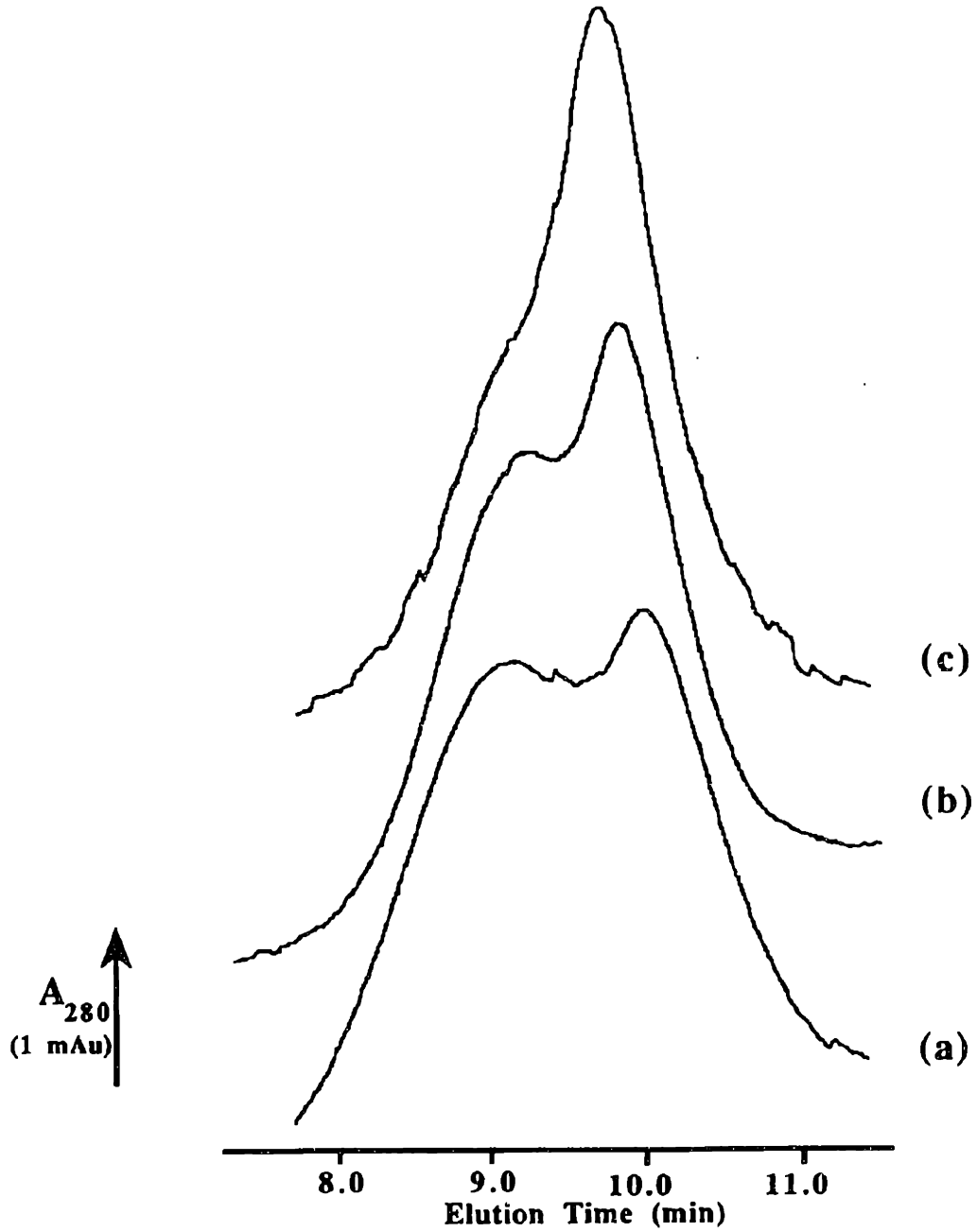


FIGURE 1.5: Measurement of reversible association during refolding. Refolding was performed by rapid dilution of CAB in 5 M GuHCl to 1.0 M GuHCl and 1.0 mg/ml (33.3 μM) protein. Aliquots of the refolding solution were injected onto a size exclusion HPLC column at a given time after dilution (see Materials and Methods). The monomer eluted at 10.1 minutes and the dimer eluted at 9.3 minutes. The chromatographs for samples which were injected at 4.25 (a), 10 (b), and 12 (c) minutes after dilution are shown.

performed at 0.10 mg/ml (3.33 μM) and 1.0 M GuHCl. The dimer species was not observed by HPLC size exclusion at these conditions. The formation of a transient dimer species is therefore dependent upon the final protein concentration.

The formation of a transient dimer for refolding at 1.0 M GuHCl validated the association dependence on the final GuHCl concentration which was observed in the previous experiments (Figure 1.3). Since the association process was reversible at 1.0 M GuHCl and association occurred rapidly, the dissociation rate may also be dependent on the final GuHCl concentration (Figure 1.5). As the final GuHCl concentration was increased, the rate of association was still relatively rapid, but the rate of dissociation was increased until the dissociation and refolding rates were comparable (1.0 M GuHCl). For conditions where the dissociation and refolding rates were similar, the observed rate of refolding was dependent upon the protein concentration since the association process was dependent upon the final protein concentration. In addition, since the recovery of active protein occurred at a slower rate at high protein concentrations, the transient dimer species must be formed from inactive protein.

1.4 Verification of Inactive Multimer Formation

To determine if the multimers observed in the previous studies were inactive, the recovery of the enzymatic activity was measured for each of the refolding experiments shown in Figure 1.3. In addition, the activity results from these experiments were used to confirm the QLS studies and to determine the extent of refolding. Reactivation of denatured CAB in 5 M GuHCl was performed by rapid dilution to a given final protein and GuHCl concentration. At a final protein concentration of 0.50 mg/ml, the recovery

of active CAB was complete after 5 minutes for final GuHCl concentrations ranging from 0.60 to 0.80 M (Figure 1.6). For refolding at 0.50 mg/ml protein and 0.80 M GuHCl, maximum renaturation was 30% active protein indicating the formation of inactive protein species. As a comparison of the relative rate and extent of recovery of active protein, refolding at 1.0 M GuHCl and 0.50 mg/ml protein is also shown in Figure 1.6. At these conditions, recovery of activity was greater than 80% after 30 minutes. These results correlate well with reported CAB refolding kinetics which were also performed at high protein concentrations and 1.0 M GuHCl (Stein & Henkens, 1978; Semisotnov et.al., 1990). The final activity for GuHCl concentrations of 0.60 M and 0.70 M was greater than expected due to presence of large aggregates which affect the absorbance measurements of the samples taken at times where micron-size aggregates were observed by QLS. For a final GuHCl concentration of 0.80 M, the dimers and trimers partially dissociated when diluted prior to the addition of substrate as observed by QLS and transmission electron microscopy (see Appendix for TEM results). The dilution effect would explain the lack of correlation between the final activity and the final monomer concentration for the low GuHCl experiments (0.60 to 0.80 M GuHCl). The dissociation upon dilution revealed that the association process is reversible as observed previously for refolding at 1.0 M GuHCl. In addition, these results further verify the formation of inactive associated species during the refolding process.

To confirm the formation of this inactive associated species as observed by QLS and size exclusion HPLC, cross-linking studies were performed to preserve the protein in the multimeric state. Refolding of CAB from 5 M GuHCl to 0.70 M GuHCl and 0.50

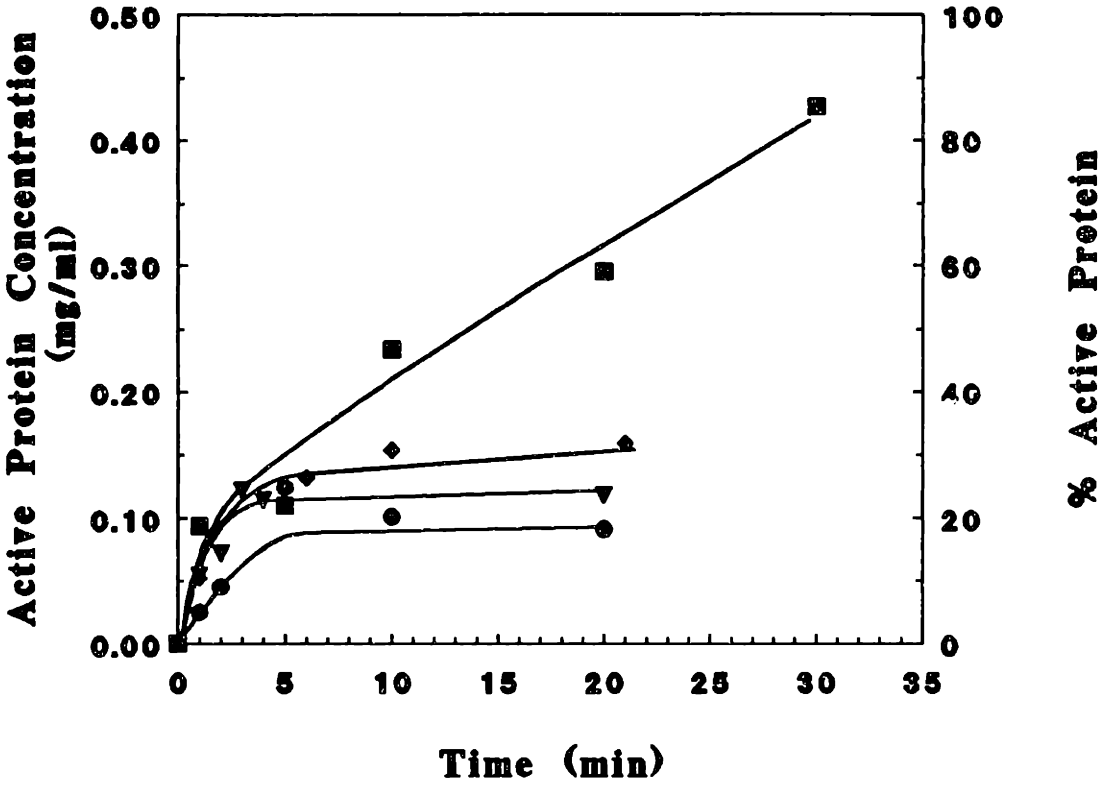


FIGURE 1.6: Concentration of active CAB as a function of time after rapid dilution from 5 M GuHCl. In each case, the protein is diluted to a final concentration of 0.50 mg/ml. The final GuHCl concentration for each case is 0.60 M (●), 0.70 M (▼), 0.80 M (◆) and 1.0 M (■). For the conditions where micron-sized aggregates are formed (0.60 M and 0.70 M GuHCl), the accuracy of the assay is greatly reduced due to light scattering.

mg/ml protein was performed by rapid dilution and aliquots of this refolding solution were cross-linked at 30 seconds intervals for 1.5 minutes after dilution. The resulting cross-linked proteins were then analyzed in 5 M GuHCl by gel electrophoresis (Figure 1.7). The dimer band intensity increased with the refolding time. The increase in dimer concentration with refolding time was similarly predicted by the QLS analysis as shown previously in Figure 1.2.b. The trimer species was probably not observed in this experiment due to low efficiency of the cross-linking agent and the low sensitivity of the staining technique which can detect typically 0.1 mg/ml protein (Pharmacia LKB Biotechnology, Uppsala, Sweden). When the cross-linked protein solution was not increased to 5 M GuHCl immediately after quenching, large aggregates were rapidly formed indicating the stabilization of an aggregating species.

1.5 Aggregating Species Determination

The intermediate species which was shown to associate in the previous studies was next determined by using the refolding pathway for CAB which has been described in the literature (see Literature Review) (Semisotnov et.al., 1987; Stein & Henkens, 1978). From the refolding kinetics of this pathway, it is possible to determine which intermediate structure forms the dimer and larger aggregates. Unfolded CAB in 5 M GuHCl rapidly folds into a compact molten globule structure when diluted to 1 M GuHCl (Dolgikh et.al., 1984). The unfolded protein does not exist in solution after 200 msec and the first intermediate with exposed hydrophobic cluste- is completely converted to the second intermediate after 15 minutes. Finally, the second intermediate will fold to form the native protein with a half time of 10 minutes (Semisotnov et.al., 1987; Semisotnov et.al.,

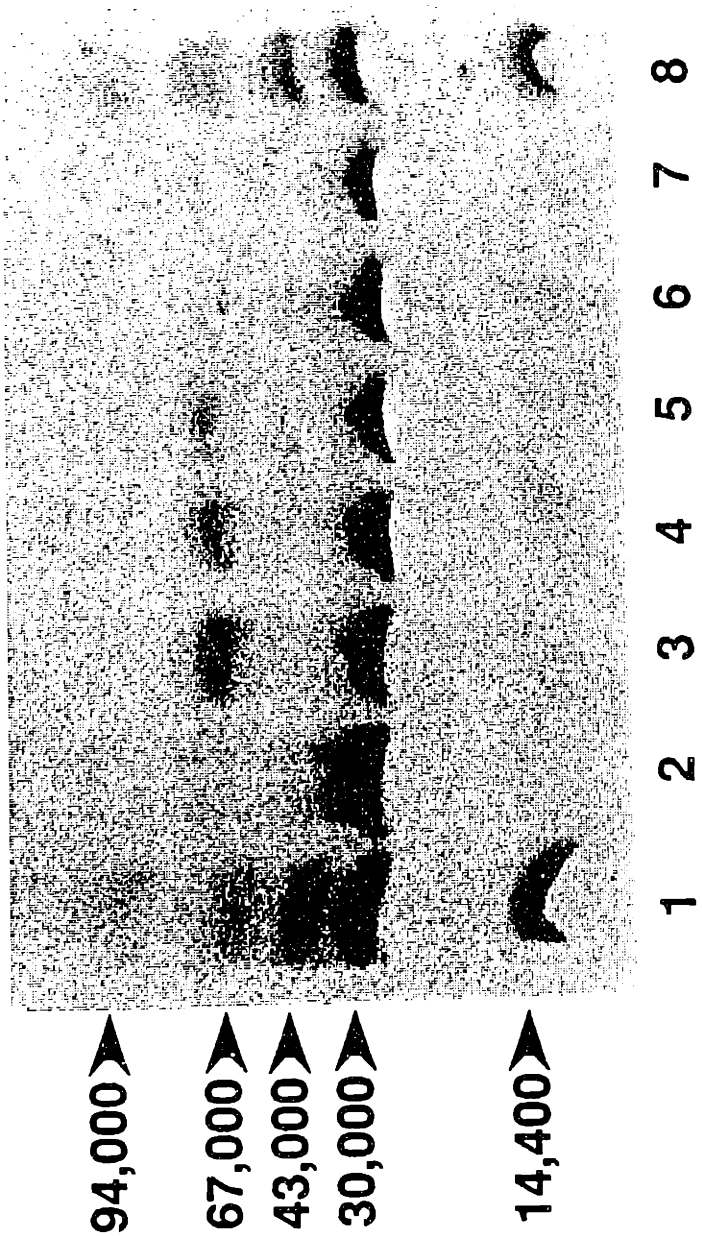


FIGURE 1.7: SDS-PAGE results of cross-linking during CAB refolding and aggregation. DMS is added to a solution of CAB which has been diluted from 5 M GuHCl to 0.70 M GuHCl with a final protein concentration of 0.50 mg/ml (see Materials and Methods for details). Each sample is brought to a GuHCl concentration of 5 M by addition of 8 M GuHCl. Lanes 1 and 8 are low molecular weight markers in 5 M GuHCl (14,400, 30,000, 43,000, 67,000, and 94,000 MW). Lanes 2 and 7 contain the starting denatured protein solution in 5 M GuHCl. Lane 3 represents the results of cross-linking a refolding solution (0.50 mg/ml CAB, 0.70 M GuHCl) at 1.50 minutes. The results of cross-linking the refolding solution at 1.0 and 0.5 minutes are shown in lanes 4 and 5, respectively. Lane 6 is the control with native protein at 0.50 mg/ml in 0.70 M GuHCl and the cross-linking agent.

1990). The propensity of exposed hydrophobic groups on proteins to cause aggregation suggests that the first intermediate is the structure which forms the dimer.

To confirm this hypothesis, double jump dilution experiments were performed at several different incubation times. First of all, refolding to each of the two final conditions was performed to determine the extent of aggregation and refolding. The aggregation step was performed by rapidly diluting unfolded CAB in 5 M GuHCl to 0.50 mg/ml protein and 0.60 M GuHCl. As shown previously in Figure 1.6, the protein only recovered 20% activity under these conditions (lower dashed line in Figure 1.8). In contrast, refolding at 1.0 M GuHCl with the same final protein concentration resulted in complete recovery of the active protein after 1 hour (upper dashed line in Figure 1.8). Double jump experiments were performed by initially diluting CAB in 5 M GuHCl to 0.83 mg/ml protein and 1.0 M GuHCl. The solution was then incubated for a given delay time of 2, 6, 10, or 15 minutes. After incubation, the second dilution was performed to yield a final protein concentration of 0.50 mg/ml and 0.60 M GuHCl (solid lines in Figure 1.8). The recovery of active protein increased at longer delay times between dilutions. For a delay time of 15 minutes, the protein completely recovered activity after 1 hour. Since the first intermediate is completely converted to the second intermediate in approximately 15 minutes, the first intermediate is the associating specie (Semisotnov et.al., 1987). Therefore, all species which are in the first intermediate state prior to the second dilution will form inactive protein aggregates. The other species such as the second intermediate and the native protein will refold to become active enzyme. Based on these assumptions and the previously reported refolding kinetics, the final recovery of

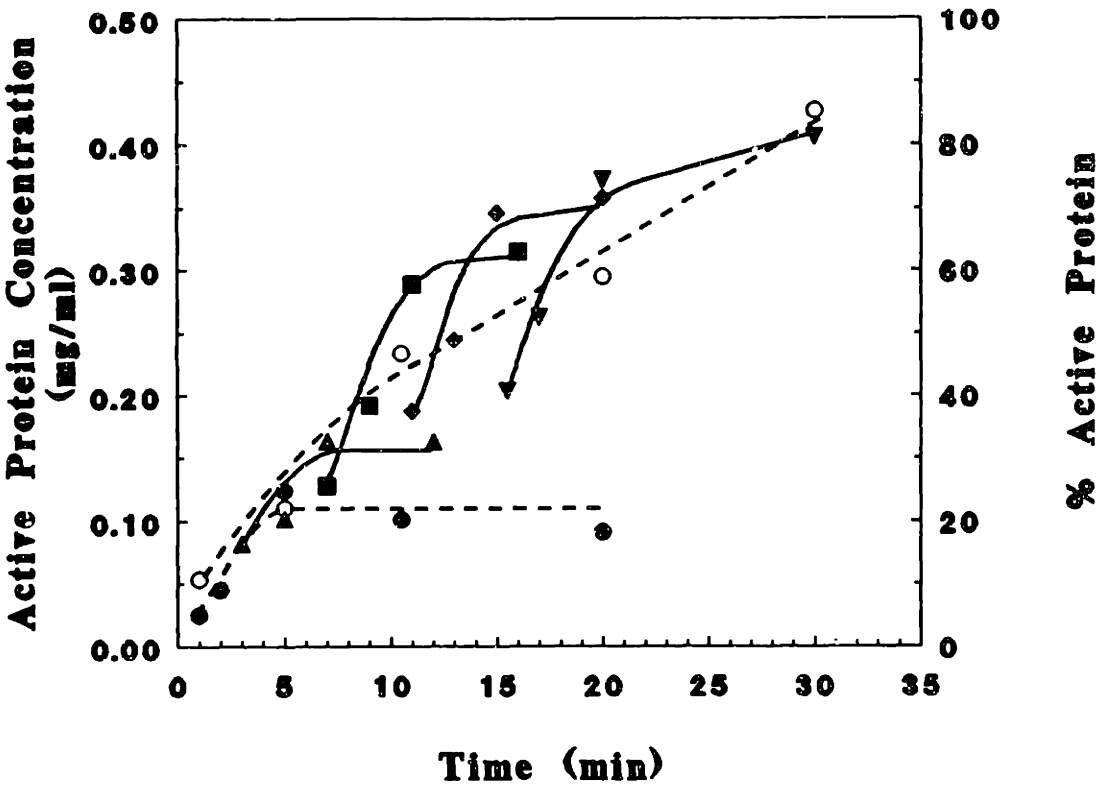


FIGURE 1.8: Double jump dilution experiments were performed to determine which intermediate in the refolding pathway will cause aggregation. To assess the final conditions after each dilution, the recovery of activity was measured for 30 minutes for single step dilutions to 0.50 mg/ml and 0.60 M GuHCl (●) and 1.0 M GuHCl (○). Double jump dilutions were conducted by dilution from 5 M GuHCl to 0.83 mg/ml CAB and 1.0 M GuHCl. Each solution was then incubated for a given delay time of 2 (▲), 6 (■), 10 (♦), or 15 (▼) minutes. The incubation was followed by dilution to 0.50 mg/ml CAB and 0.60 M GuHCl and the active protein concentration was measured as a function of time after the second dilution.

active protein was predicted for each delay time as shown in Figure 1.9 (Semisotnov et.al., 1987; Stein & Henkens, 1978). The final active protein concentration was determined by calculating the concentration of the second intermediate and native protein which are present in solution immediately before the second dilution step. These calculations were performed with the previously determined refolding rate constants (Semisotnov et.al., 1987; Stein & Henkens, 1978). A detailed discussion of the equations used for the unimolecular refolding pathway assumed in this case is presented in Section 3.1. The model prediction closely matches the experimental active protein concentration. These results suggest that the first intermediate is the monomer responsible for the formation of dimers and larger aggregates during the refolding of CAB.

1.6 Rate Analysis of Aggregation

The kinetic studies of refolding at different final conditions have shown that the first intermediate will associate to form the dimer and trimer species (Figures 1.2, 1.5, and 1.7). The rate of association of the first intermediate is dependent upon the final protein and GuHCl concentrations (Figures 1.3 and 1.4). The dimer and trimer species will aggregate to form micron-sized aggregates at low GuHCl concentrations as shown in Figure 1.3. To determine the rate of formation of the multimers and the larger aggregates, a model of the refolding and aggregation process must be developed.

Since the first intermediate associates to form multimers and large aggregates at low GuHCl concentrations and high protein concentrations (multimer formation regime, Figure 1.3), a refolding and aggregation model should involve the association pathway for this intermediate species as represented in Figure 1.10. The model is based on

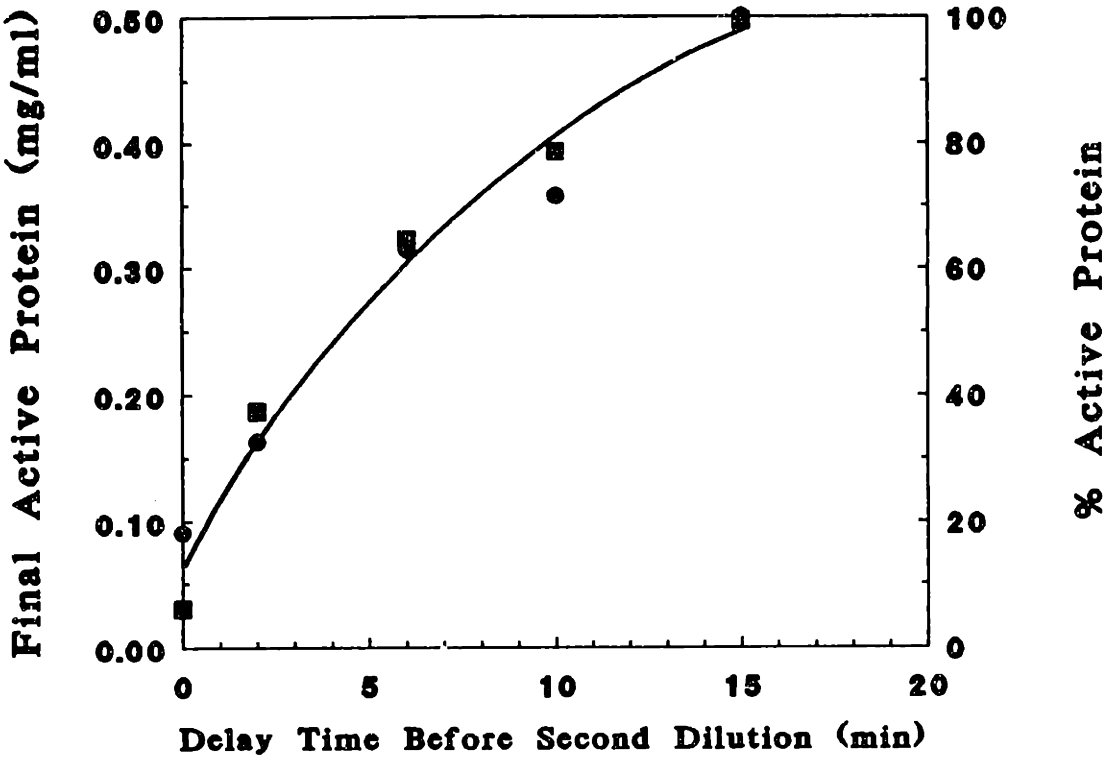


FIGURE 1.9: Kinetic model developed from double jump dilution experiments. For each double jump dilution experiment, the active protein concentration was measured after one hour at the final conditions (0.50 mg/ml CAB, 0.60 M GuHCl). The final active protein concentration was then plotted as a function of the delay time between the successive dilutions (●). The kinetic model was developed based on the assumptions that all protein in the first intermediate state prior to the second dilution will form aggregates and that the second intermediate and native protein will recover activity. For each delay time, the final activity was calculated based on these assumptions and the kinetic constants previously reported (■) (Semisotnov et.al., 1987).

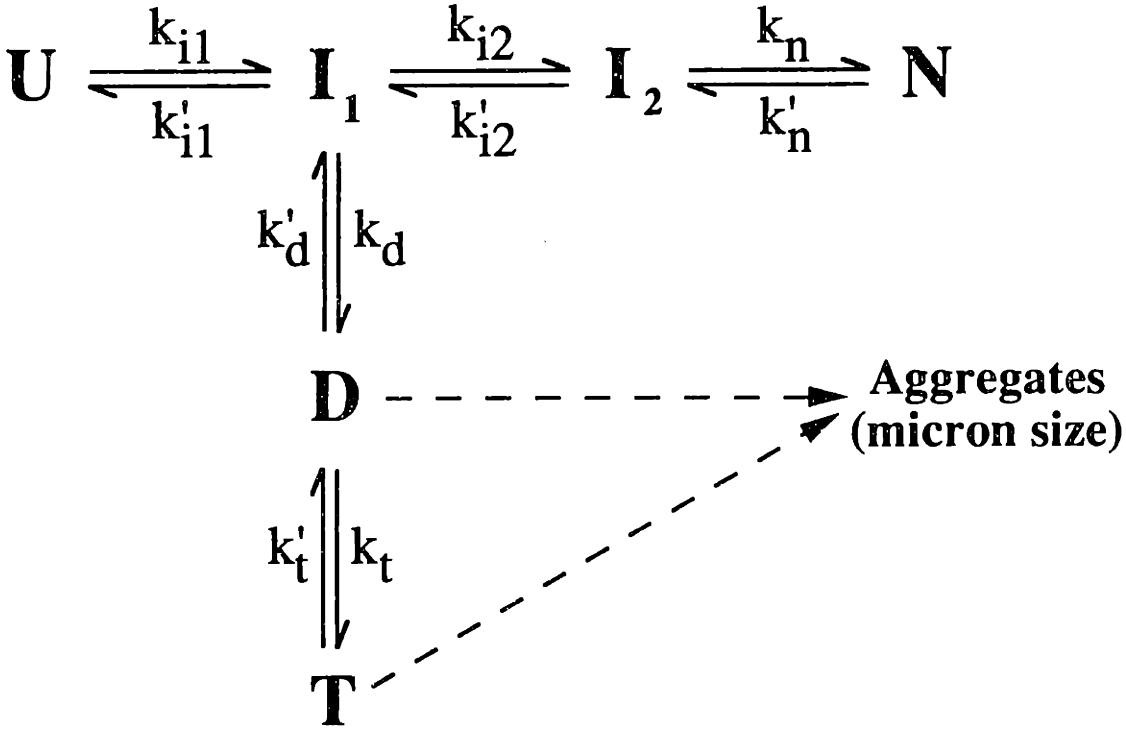


FIGURE 1.10: Proposed pathway for refolding and aggregation of CAB. The unfolded protein (U) collapses to form the molten globule first intermediate (I_1). The first intermediate can then proceed to form the second intermediate or the dimer species (D). The second intermediate continues to refold to form the native protein structure (N) as described previously (Stein & Henkens, 1978; Semisotnov et.al., 1987). The dimer may form the trimer species (T) with the addition of another intermediate protein (I_1) to the aggregate. The dimer, trimer, and first intermediate will then form the micron-sized aggregates under aggregating conditions. Each rate constant is dependent on the final protein and GuHCl concentration as discussed in the text.

refolding by rapidly diluting CAB in 5 M GuHCl to low GuHCl concentrations (< 1.0 M) and high protein concentrations (mg/ml). The formation of dimers, trimers, and large aggregates will occur when the protein is diluted to the final conditions of multimer formation or aggregation as depicted in Figure 1.3. The unfolded protein (U) rapidly forms the first intermediate (I_1) with a rate constant, k_{i1}, of 23.1 sec^{-1} (Semisotnov et.al., 1987). The first intermediate will then either proceed on the refolding pathway to form the second intermediate (I_2) with a rate constant, k_{i2}, of 0.297 min^{-1} (Semisotnov et.al., 1990) or form the dimer species with a rate constant, k_d, both of which are dependent upon the final protein and GuHCl concentrations. Finally the second intermediate will slowly fold into the native conformation (N) with a rate constant, k_n, of 6.93 x 10^{-2} min^{-1} (Semisotnov et.al., 1987). The slow formation of the native structure has shown to be the result of the cis-trans isomerization of the 19 proline residues in CAB (Semisotnov et.al., 1990). Proline isomerization has been discussed as the possible cause of slow refolding in other protein systems (Kim & Baldwin, 1982). Therefore, the rate of refolding from the second intermediate to the native protein may be difficult to increase as the result of the rate limiting proline isomerization.

The kinetics of dimer formation can now be studied based on the proposed model of aggregation and refolding for CAB. The reverse reaction rates shown in Figure 1.10 were assumed to be small relative to the forward reaction rates for the final conditions of low GuHCl concentration (< 1.0 M GuHCl) and high protein concentration (mg/ml). In addition, the formation of active protein species was very low relative to the extent of aggregation under these conditions and, therefore, the rate of dimer formation was

predominate for aggregating conditions as defined by the multimer formation and aggregation regimes in Figure 1.3. Based on these observations, the initial rate of dimer formation (R_D) for several different final conditions was then calculated from the initial slope of the dimer concentration as a function of time (Table 1.1). From Table 1.1, the dimer formation rate data at a constant final GuHCl concentration (0.80 M) and different final protein concentrations were used to determine the relationship between the protein concentration and the rate of dimer formation:

$$R_D = k_d [CAB]_f^{2.6} \qquad (1.1)$$

which approximates two monomer molecules in the first intermediate state associating to form the dimer. This relationship was valid for refolding at 0.80 M GuHCl and final protein concentrations ranging from 0.50 to 0.80 mg/ml CAB. Since there was an obvious dependence on the final GuHCl concentration at a constant final protein concentration, the rate constant, k_d, must be a function of the GuHCl concentration. Therefore, the rate of dimer formation as function of the final GuHCl concentration has been determined at a constant protein concentration (0.50 mg/ml) and at three different GuHCl concentrations from the data in Table 1.1. The rate of dimer formation (R_D) can be expressed as:

$$R_D = k^G_d [CAB]_f^{2.6} [GuHCl]_f^{-6.7} \qquad (1.2)$$

This strong inverse dependence on GuHCl may be explained by assessing the binding of GuHCl to the protein. For 0.50 mg/ml CAB in 0.60 to 0.80 M GuHCl, the number of

TABLE 1.1: Initial rate analysis of dimer formation for CAB aggregation[a]

$[CAB]_f$ (mg/ml)	$[CAB]_f$ (μM)	$[GuHCl]_f$ (M)	R_D (μM/min)	$[D]_f$ (μM)	t_f (min)
0.50	16.7	0.60	5.22	3.76	1.5
0.50	16.7	0.70	3.68	5.72	3.2
0.50	16.7	0.80	0.74	6.22	20.9
0.60	20.0	0.80	0.84	4.93	9.5
0.80	26.7	0.80	2.63	4.40	5.2

[a]Determination of dimer concentration described in Materials and Methods.

GuHCl molecules bound to the protein should range from 7 to 9 using relationships for GuHCl binding to native proteins (Lee & Timasheff, 1974; Arakawa & Timasheff, 1984). Finally, the initiation of precipitating aggregates may be a nucleation and growth phenomena which has been observed previously for protein crystallization (Reithel, 1963; Feher & Kam, 1985). The final observed dimer concentration, $[D]_f$, and the time of the final QLS measurement, t_f, are illustrated in Table 1.1. For each case, the final observed dimer concentration is consistently between 4 to 6 μM which is probably the critical nuclei concentration required for further aggregation. Additional studies of the formation of precipitating aggregates must be performed to determine if this aggregation can be modelled as a nucleation and growth process.

Overall, the kinetic analyses of refolding and aggregation have proven that multimeric protein species existed prior to the formation of large aggregates in solution. In addition, the kinetic studies have shown that the first intermediate in the refolding pathway aggregated to form these multimers which were a dimer and trimer species. The rate of formation of the dimer is dependent upon both the final protein and denaturant concentration. At low GuHCl concentrations (< 1.0 M) and high protein concentrations (mg/ml), the first intermediate rapidly associated to form the multimeric species and the recovery of active protein was drastically reduced. In contrast, refolding at 1.0 M GuHCl resulted in the transient formation of a dimer species which did not reduce the final recovery of active protein. These studies indicated that the refolding and aggregation can be generally described by the pathway shown in Figure 1.10. The rate constants in this pathway were all dependent upon the final GuHCl concentration. The multimolecular

rates were also dependent upon the final protein concentration.

Chapter 2: Thermodynamics of CAB Aggregation during Refolding

From the kinetic analysis of CAB aggregation and refolding, it was observed that the first intermediate on the refolding pathway aggregated irreversibly to form a dimer and trimer species at low GuHCl concentrations and high protein concentrations (Figure 1.3). However, the potential for reversible association was also observed for refolding at high protein concentrations and 1.0 M GuHCl (Section 1.3). Since the association of the first intermediate was observed to be reversible, studies on the equilibrium association of this intermediate were performed to obtain additional information on the mechanism of the association process. In addition, the relationship between denaturant concentration and associated species was determined to confirm the association of only the first intermediate in the refolding pathway. These equilibrium studies were also used to verify the formation of the dimer and trimer species. Finally, the rates of association and dissociation at equilibrium were determined to yield further insight into the relationship between the final solution conditions and aggregation.

2.1 Denaturant Concentration and Aggregating Protein Structure

First of all, to confirm the association of first intermediate, equilibrium refolding experiments were performed at different final GuHCl concentrations. The formation of the first intermediate in CAB equilibrium unfolding experiments was observed to have a dependence on the final GuHCl concentration (Rodionova et.al., 1989). At GuHCl concentrations ranging from 1.8 to 2.2 M, the first intermediate was highly populated during equilibrium unfolding as measured by binding of a hydrophobic probe, 8-anilino-1-naphthalene sulfonate (ANS) (Rodionova et.al., 1989). Since the first intermediate has

been observed to associate during refolding and this intermediate will bind ANS, association of CAB during equilibrium refolding should have the same dependence on the final GuHCl concentration as the binding of ANS. To test this hypothesis, a control experiment was performed by refolding from 5 M GuHCl to 0.10 mg/ml (3.3 μM) CAB and 2.0 M GuHCl. The species size distribution for the control was representative of a monomeric compact molten globule and was comparable to the native protein. This distribution was used as the monomer standard for calculating the concentration of monomer, dimer and trimer. To correlate the aggregation with the ANS binding results, denatured CAB in 5 M GuHCl was rapidly diluted to 33 μM (1.0 mg/ml) protein and different GuHCl concentrations ranging from 1.0 to 2.5 M. Each sample was allowed to equilibrate for three to eight hours and then analyzed by quasi-elastic light scattering (QLS). The resultant monomer, dimer, and trimer concentrations are shown in Figure 2.1. The greatest extent of association occurred between 1.8 and 2.2 M GuHCl. These results correlated well with the observed population of the first intermediate determined from ANS binding studies (Rodionova et.al., 1989). Therefore, the hydrophobic molten globule first intermediate which is present at 1.8 to 2.2 M GuHCl will associate at a concentration of 33 μM during equilibrium refolding.

2.2 Protein Concentration Effect on Equilibrium Association

To determine the effect of protein concentration on this association phenomenon, denatured CAB in 5 M GuHCl was diluted to 2.0 M GuHCl at several different protein concentrations ranging from 6.7 to 100 μM (0.20 - 3.0 mg/ml). A final GuHCl concentration of 2.0 M was chosen to maximize the association of the first intermediate

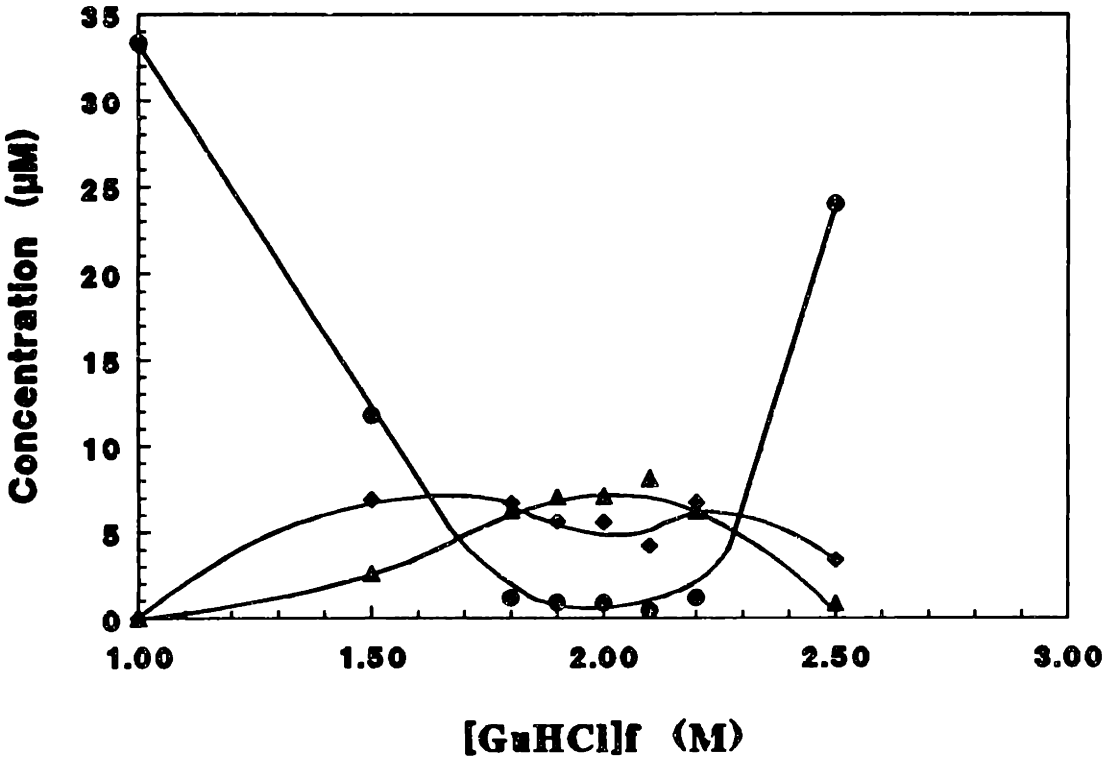


FIGURE 2.1: Equilibrium analysis of associating species as a function of GuHCl concentration. CAB in 5 M GuHCl was diluted to 1.0 mg/ml protein (33.3 μM) and a given GuHCl concentration. The solutions were then allowed to equilibrate for three hours. After equilibration, QLS analysis was performed on each sample. The resultant concentrations depicted are monomer (●), dimer (♦), and trimer (▲).

as already shown in Figure 2.1. After equilibration, each sample was analyzed by QLS and the distribution of monomer, dimer, and trimer was determined by using the control size distribution (0.10 mg/ml CAB, 2.0 M GuHCl). The intermediate did not associate at low final protein concentrations ($\leq$ 10 μM) (Figure 2.2). At higher final protein concentrations, CAB associated to form both dimer and trimer species. The multimer distribution shifts to the trimer as the protein concentration is increased to 100 μM (Figure 2.2). These results were confirmed by HPLC size exclusion chromatography. Each equilibrated sample was analyzed by HPLC as described in the Materials and Methods section and the distribution of monomer, dimer, and trimer are shown in Figure 2.3. The dependence of the association on final protein concentration as measured by HPLC was similar to the results obtained by QLS. Therefore, the HPLC results did not contain artifacts due to sample dilution or protein-column interactions. The HPLC results were very repeatable with consistent elution times and peak areas. Since the HPLC technique was simpler and yielded results similar to QLS, additional equilibrium association experiments were performed using size exclusion HPLC.

2.3 Equilibrium Model for Association in 2.0 M GuHCl

From these equilibrium results, a model for the equilibrium association of the first intermediate was postulated for refolding at 2.0 M GuHCl as follows:

$$2\ I_1 \rightleftharpoons D \qquad (2.1)$$

$$K_D = [D]/[I_1]^2 \qquad (2.2)$$

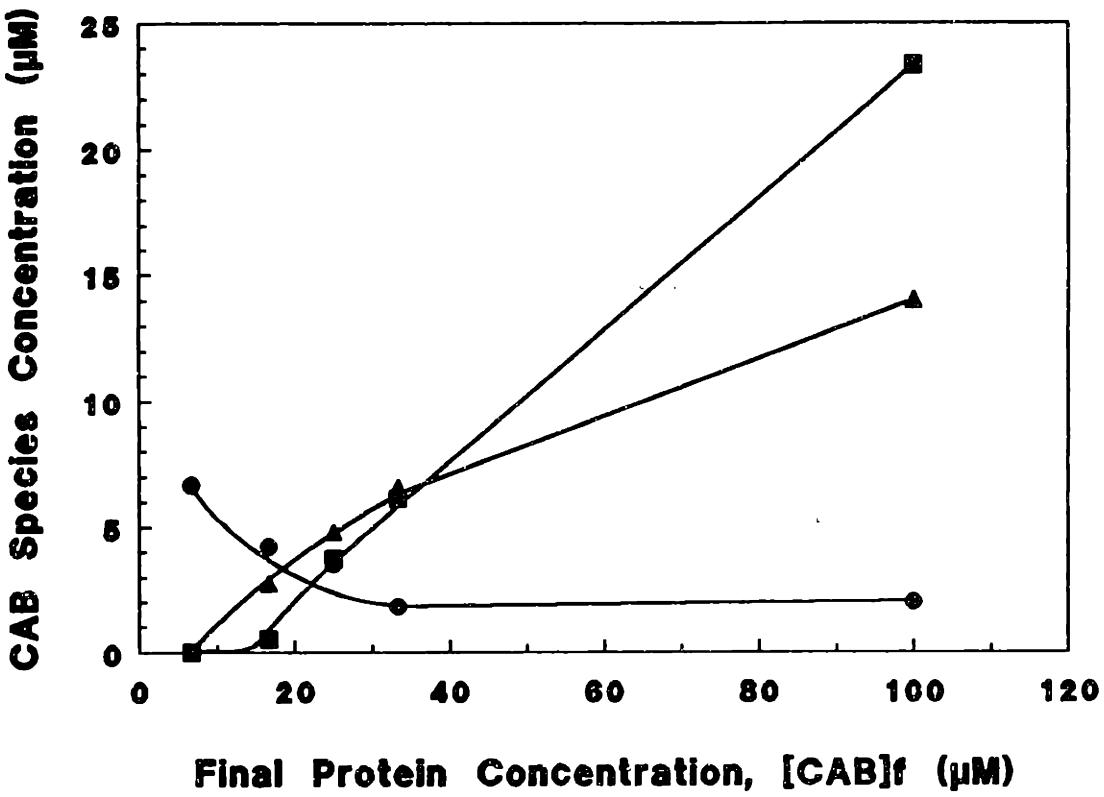


FIGURE 2.2: Equilibrium studies of the concentration dependence of association for the first intermediate. CAB in 5 M GuHCl was diluted to a final GuHCl of 2.0 M and a given protein concentration. After three hours of equilibration, the samples were analyzed by QLS. The change in distribution of monomer (●), dimer (▲), and trimer (■) are shown as a function of the final protein concentration.

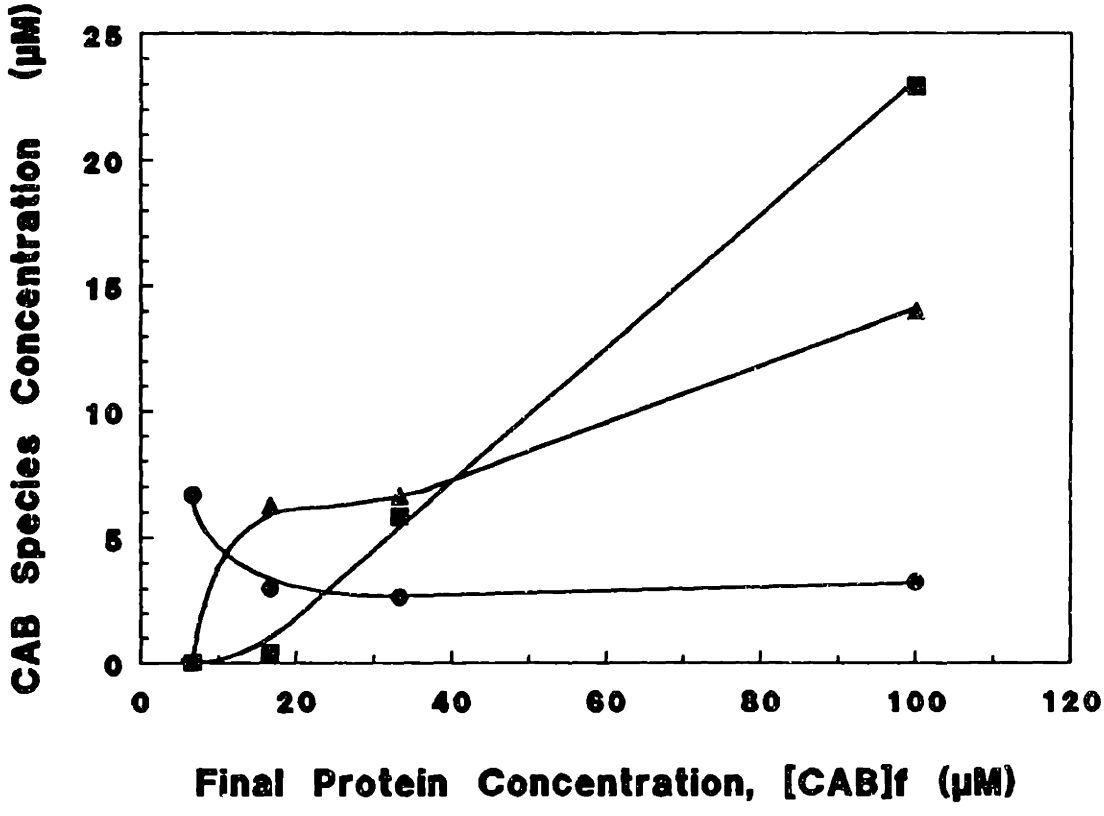


FIGURE 2.3: Equilibrium association of CAB as a function of final protein concentration ($[CAB]_f$) as measured by size exclusion HPLC. Equilibrium refolding was performed by rapid dilution of unfolded CAB in 5 M GuHCl to 2.0 M GuHCl and a range of final protein concentrations. After equilibration for three to eight hours, monomer (●), dimer (▲), and trimer (■) concentrations were determined for each final protein concentration as described in Materials and Methods.

$$D + I_1 \rightleftharpoons T \quad (2.3)$$

$$K_T = [T]/[D][I_1] \quad (2.4)$$

The first intermediate can associate to form the dimer with an equilibrium constant, K_D , for final protein concentrations greater than 10 μM . The dimer equilibrium constant, K_D, was 1.8 ± 0.2 μM^{-1} as calculated from the QLS data and 1.3 ± 0.1 μM^{-1} using the HPLC data. The largest multimer, trimer, was formed from the association of a dimer and a monomer with an equilibrium constant, K_T . For this association reaction, the equilibrium constant, K_T , was 0.53 ± 0.12 μM^{-1} and 0.42 ± 0.11 μM^{-1} for QLS and HPLC analyses, respectively. The equilibrium constants for dimer and trimer formation are comparable, but not equivalent, using these two analytical techniques.

With the knowledge of these equilibrium constants, the association rates were determined through additional HPLC experiments. To measure the association rate for dimer formation, denatured CAB in 5 M GuHCl was diluted to association conditions (33 μM protein and 2.0 M GuHCl) and immediately analyzed by size exclusion HPLC. Samples from the refolding solution were then analyzed at different times over a period of fifty minutes and converted to concentrations of monomer and dimer as shown in Figure 2.4. The trimeric specie was not observed over this time period and the relative change in peak areas was significantly different for each measurement. Therefore, the rate constant for dimer formation, k_d , was calculated from the initial slope of the dimer concentration as a function of time as shown in Figure 2.4. The association rate constant for dimerization, k_d, was 5.16 x 10^{-3} μM^{-1} min^{-1} with a half time for association of 134.3

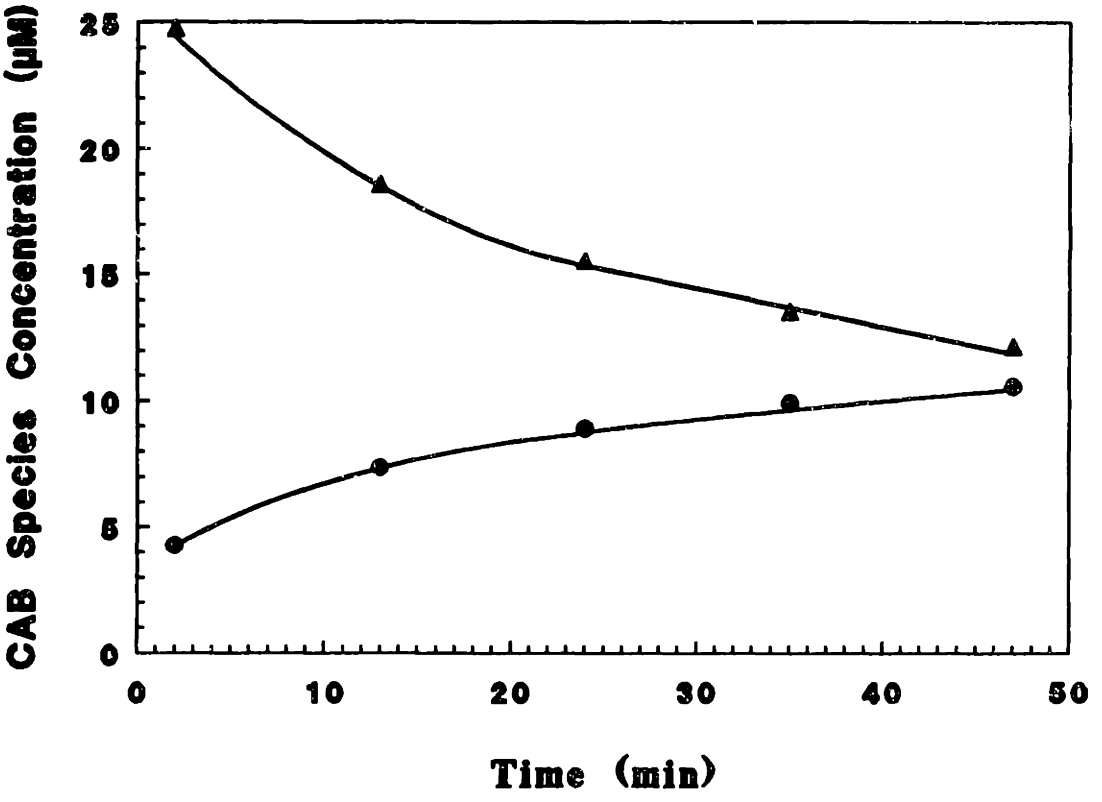


FIGURE 2.4: Association of the first intermediate during refolding as measured by size exclusion HPLC. Denatured CAB in 5 M GuHCl was rapidly diluted to association conditions (33 μM CAB, 2.0 M GuHCl) and analyzed at different times for a period of fifty minutes. The monomer (●) and dimer (▲) concentrations are shown as a function of time after dilution to the final conditions.

minutes. The forward rate constant, k_d , is directly related to the equilibrium constant.

$$K_D = k_d / k'_d \qquad (2.5)$$

Using this relationship, the dissociation rate constant, k'_d, was calculated to be 3.91×10^{-3} min^{-1} with a resultant half time of 177.3 minutes. In order to measure the rate constants for trimer formation, dissociation experiments were performed by rapid dilution of an equilibrated protein solution at association conditions (33 μM protein and 2.0 M GuHCl) to monomer conditions (3.3 μM protein, 2.0 M GuHCl). The diluted protein solution was analyzed by HPLC size exclusion to follow the dissociation over time (Figure 2.5). Aliquots of the diluted protein solution were analyzed as function of time after dilution and the distribution of monomer, dimer, and trimer was calculated for each measurement. The size distribution shifts to the monomer and dimer after ten minutes and the dimer slowly dissociates to form the monomer. The rate constant for trimer dissociation, k'_t, was calculated to be 0.316 min^{-1} ($t_{1/2}$ = 2.19 minutes) from the rate of decrease in trimer concentration. To determine the trimer association rate, the relationship between the rate constants and the equilibrium constant was utilized.

$$K_T = k_t / k'_t \qquad (2.6)$$

Based on this equilibrium model, the association rate constant for trimer formation, k_t , was 0.133 μM^{-1} min^{-1} with a half time of 5.22 minutes. Since the trimer equilibrium was rapid relative to the dimer equilibrium, the association was difficult to detect at low protein concentrations using the HPLC technique which required a total elution time of

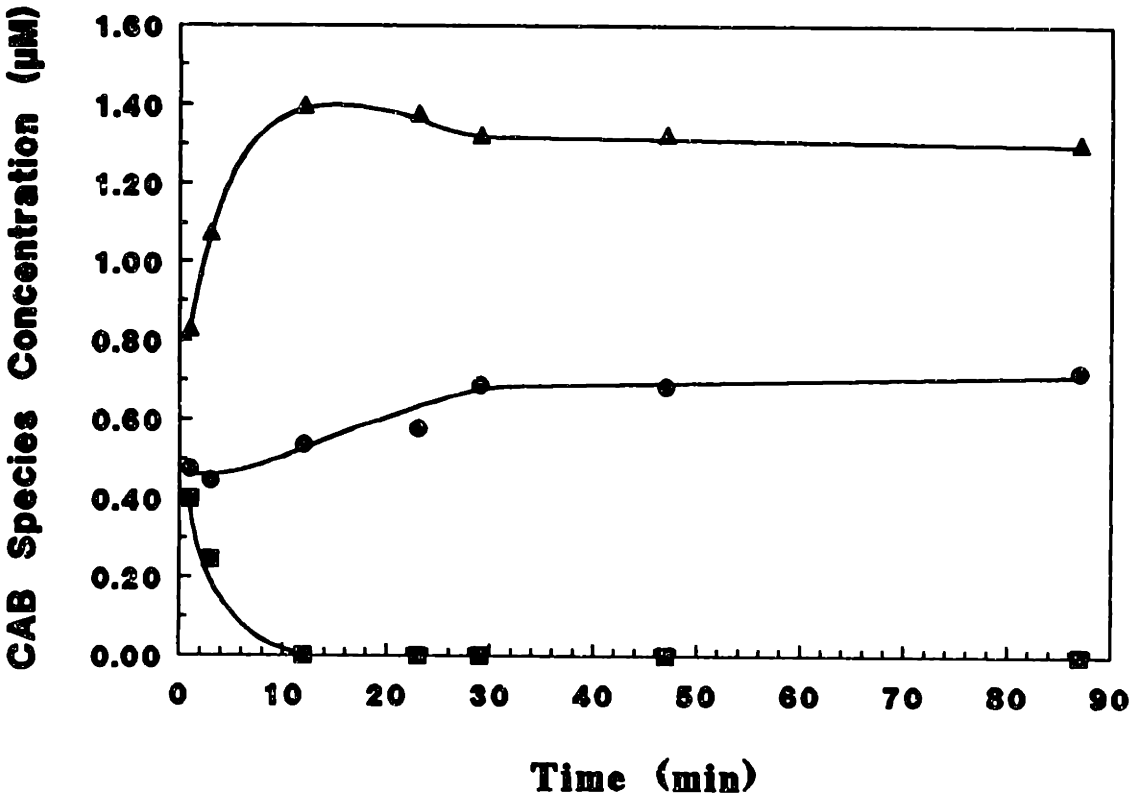


FIGURE 2.5: Dissociation of equilibrium associated protein solution. An equilibrium refolding solution at association conditions (33 μM CAB and 2.0 M GuHCl) was diluted to monomer conditions (3.3 μM and 2.0 M GuHCl) and analyzed by size exclusion HPLC. The monomer (●), dimer (▲), and trimer (■) concentrations are shown as a function of time after dilution to monomer conditions.

10 minutes.

To confirm the achievement of an equilibrium state, the equilibrated solutions in 2.0 M GuHCl were diluted to either the refolding or aggregation conditions which were observed in the kinetic studies. To verify the dissociation and refolding, each equilibrated solution was diluted to a final GuHCl concentration of 1.0 M to allow the protein to refold. After incubation for one hour, the protein completely recovered its activity and did not associate to form multimers. These results are comparable to the kinetic studies of refolding by rapid dilution from 5 M GuHCl to 1 M GuHCl. Therefore, the equilibrium associated species can dissociate to allow complete refolding of the protein. On the other hand, dilution of the associated CAB solutions (≥ 10 μM CAB, 2.0 M GuHCl) to the aggregation conditions (≥ 16.7 μM CAB, 0.50 M GuHCl) resulted in precipitation and recovery of approximately 20% activity. The precipitation and low recovery of activity were comparable to that observed by direct dilution from 5 M GuHCl to the same aggregation conditions. Therefore, the observed equilibrium association is a true reversible phenomenon.

These thermodynamic studies revealed the equilibrium association of the first intermediate in the refolding pathway of CAB. In addition, the formation of a dimer and trimer species was observed to occur at aggregating conditions. The extent of association was dependent upon the final protein and denaturant concentrations. For equilibrium refolding in 2.0 M GuHCl, the first intermediate was the predominate specie and this intermediate associated to form dimers and trimers at high protein concentrations (≥ 10 μM). The equilibrium constants were calculated for these aggregation conditions to

determine the extent of association (Table 2.1). In addition, the rate of association and dissociation was determined for both the dimer and trimer species in 2.0 M GuHCl as shown in Table 2.1. The dimer association rate was much greater than the dissociation rate and was also dependent upon the final protein concentration. The dependence on the final protein concentration was similar to that observed in the kinetic studies. Therefore, the association of the first intermediate decreased the concentration of monomeric intermediate species at equilibrium as well as during refolding and the rate and extent of the association was dependent upon the final solution conditions.

TABLE 2.1: Equilibrium constants for dimer and trimer formation with the dissociation and association rate constants.[a]

Analysis Technique:[b]	HPLC	QLS
Dimer Equilibrium Constant, K_D, (μM^{-1})	1.3 ± 0.1	1.8 ± 0.2
Dimer Association Constant, k_d, ($\mu M^{-1}min^{-1}$)	5.16×10^{-3}	
Dimer Dissociation Constant, k'_d, ($\mu M^{-1}min^{-1}$)	3.91×10^{-3}	
Trimer Equilibrium Constant, K_T, (μM^{-1})	0.42 ± 0.11	0.53 ± 0.12
Trimer Association Constant, k_t, ($\mu M^{-1}min^{-1}$)	0.133	
Trimer Dissociation Constant, k'_t, ($\mu M^{-1}min^{-1}$)	0.316	

[a] Constants were derived from Equations 2.1 through 2.6 as discussed in the text.
[b] Analysis techniques are described in the Materials and Methods Section.

Chapter 3: Model Pathways for the Aggregation and Refolding of CAB

The kinetic and thermodynamic analyses of aggregation and refolding established that the first intermediate in the refolding pathway associates to form dimers and trimers and the rate and extent of association were dependent upon the final solution conditions. To address the effect of the final conditions on refolding and aggregation, different models were developed to describe the folding process. Three models were developed based on the observed phenomena. The first model was developed for refolding at low GuHCl concentrations (< 1.0 M) where rapid aggregation of the protein occurred and the recovery of active protein was low (Section 1.2). This model was discussed in Section 1.6 (Figure 1.10). This pathway was relatively simple to model since the association process was very rapid and irreversible. On the other hand, more complex pathways were needed to describe the other two observed phenomena. Refolding at low protein concentration (< 10 μM) and 1.0 M GuHCl resulted in a more rapid recovery of activity than previously reported and, therefore, a detailed analysis was required to obtain the model parameters for these final conditions. If refolding was performed at higher protein concentrations (> 10 μM) and 1.0 M GuHCl, the rate of refolding was greatly reduced. To describe the decreased rate of refolding, it was necessary to develop a complex model involving a reversible bimolecular process. Refolding at both low and high protein concentrations in 1.0 M GuHCl resulted in pathways which, unlike the first model, required a detailed mathematical analysis to fully describe the folding process.

3.1 Model of Refolding at Low Prctein Concentration in 1.0 M GuHCl

Refolding of CAB at low protein concentrations (< 10 μM) in 1.0 M GuHCl

resulted in a more rapid recovery of active protein than refolding at higher protein concentrations (Figure 1.4). In addition, CAB in 2.0 M GuHCl did not associate at low protein concentrations (< 10 μM, see Chapter 2). Protein aggregation during refolding at 1.0 M GuHCl and these low protein concentrations was not observed by size exclusion HPLC studies (Section 1.3). Therefore, the pathway for refolding of CAB at low protein concentrations in 1.0 M GuHCl can be described by a basic series of unimolecular reactions.

$$U \xrightarrow{k_{i1}} I_1 \xrightarrow{k_{i2}} I_2 \xrightarrow{k_n} N \qquad (3.1)$$

where the reverse reaction rates were assumed to be small relative to the forward rates. When diluted to 1.0 M GuHCl from 5.0 M GuHCl, the unfolded protein, U, will fold to form the first intermediate, I_1, with a rate constant of k_{i1}. The first intermediate will continue to fold to form the second intermediate, I_2. Finally, the native protein, N, is formed from the second intermediate.

The pathway depicted in Equation 3.1 can be modelled by a series of rate expressions for each species. The rate of decrease in the unfolded protein (d[U]/dt) may be expressed as:

$$\frac{d[U]}{dt} = -k_{i1}[U] \qquad (3.2)$$

The initial concentration of unfolded protein is equivalent to the total protein concentration ($[CAB]_t$) and the concentration profile of the unfolded protein can be

determined from integration of Equation 3.2.

$$[U] = [CAB]_f \exp(-k_{i1}t) \tag{3.3}$$

The kinetics for first intermediate can be described by the rate expression:

$$\frac{d[I_1]}{dt} = k_{i1}[U] - k_{i2}[I_1] \tag{3.4}$$

By using Equations 3.3 and 3.4, the first intermediate concentration as a function of time may be written as:

$$[I_1] = \left(\frac{k_{i1}[CAB]_f}{k_{i2} - k_{i1}}\right)\{\exp(-k_{i1}t) - \exp(-k_{i2}t)\} \tag{3.5}$$

Previous refolding studies indicated that the first intermediate was formed very rapidly and that the rate was not dependent upon the initial protein concentration (k_{i1} = 23.1 sec^{-1} $t_{1/2}$ = 0.03 sec) (Semisotnov et.al., 1987). The rate of formation of the first intermediate was much greater than rate of formation of the second intermediate or the native protein ($k_{i1} >> k_{i2}$ or k_n) as reported previously (Semisotnov et.al. 1990; Semisotnov et.al. 1987; Stein and Henkens, 1978). Therefore, Equation 3.5 can be reduced to a simple exponential decay dependent upon the rate constant for conversion to the second intermediate.

$$[I_1]=[CAB]_f \exp(-k_{i2}t) \tag{3.6}$$

The formation of the second intermediate followed by its conversion to the native protein can be described by the rate expression:

$$\frac{d[I_2]}{dt} = k_{i2}\,[I_1] - k_n\,[I_2] \tag{3.7}$$

Using this rate expression and Equation 3.6, the concentration profile for the second intermediate can then be determined:

$$[I_2] = \left(\frac{k_{i2}[CAB]_f}{k_n - k_{i2}}\right)\{\exp(-k_{i2}t) - \exp(-k_n t)\} \tag{3.8}$$

Finally, the second intermediate folds to form the native protein as shown in Equation 3.1 and the rate of native protein formation may be represented by:

$$\frac{d[N]}{dt} = k_n\,[I_2] \tag{3.9}$$

Solving this expression using Equation 3.8, the concentration profile for native protein can be expressed as:

$$[N] = [CAB]_f\left\{\left(\frac{k_{i2}}{k_n - k_{i2}}\right)\exp(-k_n t) - \left(\frac{k_n}{k_n - k_{i2}}\right)\exp(-k_{i2}t) + 1\right\} \tag{3.10}$$

Equations 3.6, 3.8 and 3.10 were used to completely describe the pathway in Equation 3.1.

Since previous refolding kinetics were studied at high protein concentrations where the protein has been shown to form a transient dimer species, refolding at low protein concentrations will result in a greater refolding rate than that reported previously (Semisotnov et.al., 1990). One study conducted at low protein concentrations (5-10 μM) and 0.70 M GuHCl indicated a more rapid recovery of activity than at high protein concentrations (Henkens et.al., 1982). The increase in refolding rate should only be reflected by a change in the rate of conversion from the first intermediate to the second intermediate (k_{i2}) since the equilibrium association occurred through the first intermediate. As shown previously, the second intermediate has approximately 25% of the native protein activity (Wong & Tanford, 1973; Rodionova et.al., 1989). With this approximation, Equations 3.6, 3.8, and 3.10 were used with the previously reported rate constant for the conversion of the second intermediate to the native protein (k_n = 6.93 x 10^{-2} min^{-1}, $t_{1/2}$ = 10 min) to determine the rate constant for formation of the second intermediate, k_{i2} (Semisotnov et.al., 1987). The value of k_{i2} was adjusted to fit the model to the activity data obtained for refolding at 0.10 mg/ml (3.33 μM) and 1.0 M GuHCl as shown in Figure 1.4. The rate constant for conversion from the first intermediate to the second intermediate, k_{i2}, at low protein concentrations and 1.0 M GuHCl was thereby calculated to be 1.386 min^{-1} ($t_{1/2}$ = 30 sec) which was much greater than that reported previously (k_{i2} = 0.297 min^{-1}, $t_{1/2}$ = 120 sec; Semisotnov et.al., 1987). The difference in these rates was the result of the side reaction of association which was not considered in

the previous studies.

3.2 Model of Refolding at High Protein Concentration and 1.0 M GuHCl

In contrast to the pathway described in Equation 3.1, refolding at high protein concentrations (> 10 μM) and 1.0 M GuHCl resulted in slower refolding of the protein and the observed formation of a transient dimer species (Section 1.2). The dimer has been shown to occur from an association of the first intermediate in the refolding pathway (Section 1.5). Therefore, the first intermediate will reversibly associate to form a dimer under refolding conditions at high protein concentrations. The model for refolding at high protein concentrations in 1.0 M GuHCl would then become:

$$U \underset{}{\overset{k_{i1}}{\rightarrow}} I_1 \underset{}{\overset{k_{i2}}{\rightarrow}} I_2 \underset{}{\overset{k_n}{\rightarrow}} N \qquad (3.11)$$
$$k'_d \downarrow\uparrow k_d$$
$$D$$

In addition, the association of the first intermediate has been shown to occur with little reversibility at low GuHCl concentrations (< 1.0 M) and several protein concentrations (0.10 - 1.0 mg/ml) as shown in Figure 1.10 where the reverse rate constants were small relative to the forward rates ($k_{i1}>>k'_{i1}$, $k_d>>k'_d$, $k_t>>k'_t$, $k_{i2}>>k'_{i2}$, and $k_n>>k'_n$). The irreversibility of the pathway at these aggregation conditions and the rapid rate of dimer formation allowed for the calculation dimer formation rate at low GuHCl concentrations (< 1.0 M) as discussed in Section 1.6.

With the values for the rate constants at refolding in the absence of association

and the knowledge of transient dimer formation during refolding, the model parameters for refolding with the association pathway (Equation 3.11) were elucidated. For this model, the rate of decrease in the unfolded protein was represented by Equation 3.2. Since the rate of formation of the first intermediate was very rapid relative to the other folding steps and independent of the initial protein concentration (see derivation of Equation 3.6), the initial concentration of the first intermediate was equal to the total protein concentration ($[I_1]_{t=0} = [CAB]_f$) and the rate expression for the first intermediate reduces to:

$$\frac{d[I_1]}{dt} = k'_d [D] - k_{i2} [I_1] - k_d [I_1]^2 \qquad (3.12)$$

To compare the results obtained for refolding at high protein concentrations with those reported in the literature, Equation 3.12 can be simplified by grouping the parameters into an apparent rate constant, k_{app}:

$$k_{app} = -k'_d \frac{[D]}{[I_1]} + k_{i2} + k_d [I_1] \qquad (3.13)$$

As an approximation, the apparent rate constant can be used to determine the concentration as a function of time for each species if the dependence of k_{app} on the first intermediate and the dimer was neglected. This assumption resulted in the same equations for each species as before where k_{i2} would be replaced by k_{app} (Equations 3.6, 3.8, and 3.10). When this apparent rate approach was applied to fit the activity data from

refolding at 1.0 mg/ml (33.3 μM) CAB and 1.0 M GuHCl, the apparent rate constant was determined to be 0.337 min^{-1} ($t_{1/2}$ = 124 sec) which was very similar to the previously reported rate constant (k_{i2} = 0.297 min^{-1}, Semisotnov et.al., 1987). To confirm these rates, refolding was performed by dilution CAB in 5 M GuHCL to 1.0 M GuHCl and 0.50 mg/ml (16.7 μM) protein. After dilution to the final conditions, the refolding was measured by the change in absorbance at 280 ηm as shown in Figure 3.1. The curve shown in Figure 3.1 can be modelled with two single exponential functions. The exponentials had rate constants of 0.30 min^{-1} ($t_{1/2}$= 120 sec) and 7.24 x 10^{-2} min^{-1} ($t_{1/2}$= 9.5 min). These results were comparable to the previously reported kinetics (k_{i2} = 0.297 min^{-1} and k_n = 6.93 x 10^{-2} min^{-1}; Semisotnov et.al., 1990; Semisotnov et.al. 1987; Stein and Henkens, 1978).

The assumption of an apparent rate constant did not consider the effect of the protein concentration on the rate of refolding. It was therefore necessary to develop a more rigorous analysis of the pathway described in Equation 3.12. First of all, the rate of formation of dimer species has been shown to occur rapidly at low GuHCl concentrations and the association rate in 2 M GuHCl is also a rapid process (k_d = 5.16 x 10^6 min^{-1} M^{-1}). When refolding at high protein concentrations, the forward rate of dimer formation occurred very rapidly since the initial concentration of the first intermediate was equal to the total protein concentration. Therefore, the rate constants for the formation of dimer and the first intermediate were greater than those for the other steps on the pathway ($k_{i1} > k_d >> k'_d$, k_{i2}, and k_n). The rate expression for dimer formation can be expressed by:

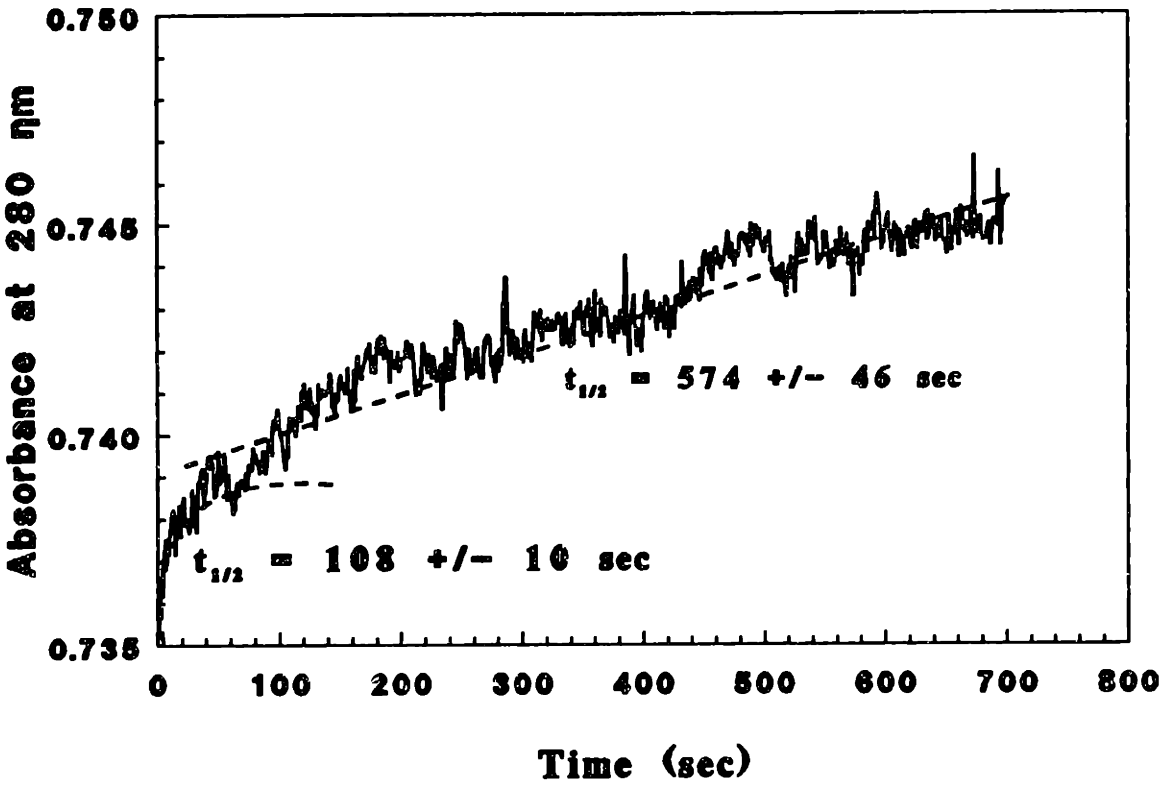


FIGURE 3.1: Refolding kinetics measured by absorbance at 280 ηm. Absorbance measurements were made within 10 seconds of dilution from 5 M GuHCl to 1 M GuHCl and 0.50 mg/ml CAB. The curve can be modelled by two exponentials with half times of 108 and 574 seconds as shown (Semisotnov et.al., 1990).

$$\frac{d[D]}{dt} = k_d [I_1]^2 - k'_d [D] \tag{3.14}$$

The equilibrium constant for the association of the first intermediate is defined as

$$K_D = \frac{[D]}{[I_1]^2} \tag{3.15}$$

and, therefore, Equation 3.14 becomes:

$$\frac{d[D]}{dt} = \left(\frac{k_d}{K_D}\right) [D] - k'_d [D] \tag{3.16}$$

Assuming a rapid equilibrium association ($K_D >> 1$), the rate expression for the dimer can be reduced to:

$$\frac{d[D]}{dt} = -k'_d [D] \tag{3.17}$$

and the concentration profile for the dimer can be calculated from the solution.

$$[D] = [D]|_{t=0} \exp(-k'_d t) \tag{3.18}$$

The initial concentration of dimer can be calculated from the rapid equilibrium condition at the initial condition as described by Equation 3.15 assuming the initial concentration

of the first intermediate ($[I_1]_{t=0}$) was equal to the total protein concentration ($[CAB]_f$) for the rapid folding of the unfolded protein ($k_{i1} >> k_d$ or k_{i2}, see Section 3.1).

After determining the concentration profile for the dimer, the concentration profile of the first intermediate was calculated assuming that the conversion from the first intermediate to the second intermediate was slow relative to the equilibrium association. With this assumption, the rate equation for the first intermediate as shown in Equation 3.12 can be reduced to:

$$\frac{d[I_1]}{dt} = k'_d [D] - k_{i2} [I_1] \tag{3.19}$$

The solution to this expression for the concentration of the first intermediate ($[I_1]$) yielded a relationship between the dissociation rate and the formation of the second intermediate as follows:

$$[I_1] = \left(\frac{k'_d [D]_{t=0}}{k_{i2} - k'_d} \right) \exp(-k'_d t) + \left([CAB]_f + \frac{k'_d [D]_{t=0}}{k'_d - k_{i2}} \right) \exp(-k_{i2} t) \tag{3.20}$$

Next, the second intermediate rate expression was a simple relationship between the rate of formation from the first intermediate and the rate of folding to the native protein.

$$\frac{d[I_2]}{dt} = k_{i2} [I_1] - k_n [I_2] \tag{3.21}$$

Equations 3.20 and 3.21 were then used to determine the concentration profile for the second intermediate.

$$[I_2] = a \exp(-k'_d t) + b \exp(-k_{t2} t) + c \exp(-k_n t) \qquad (3.22)$$

$$a = \frac{k'_d [D]_{t=0}}{(k_n - k'_d)(k_{t2} - k'_d)}$$

$$b = \frac{[I_1]_{t=0} - \left(\frac{k'_d [D]_{t=0}}{k_{t2} - k'_d}\right)}{k_n - k_{t2}}$$

$$c = \frac{k'_d [D]_{t=0}}{(k'_d - k_n)(k_{t2} - k'_d)} + \frac{[I_1]_{t=0} - \frac{k'_d [D]_{t=0}}{k_{t2} - k'_d}}{k_{t2} - k_n}$$

After calculating the concentration of the first and second intermediates as well as the dimer at each time, a simple mass balance on the total protein yielded the concentration profile for native protein.

$$[N]_t = [CAB]_f - [I_1]_t - [I_2]_t - 2[D]_t \qquad (3.23)$$

Equations 3.18, 3.20, 3.22, and 3.23 completely describe the pathway shown in Equation 3.11.

With the equations necessary to model the pathway with reversible association, the parameters for refolding were next determined. First of all, the rate constants for formation of the second intermediate (k_{i2}) and the native protein (k_n) were assumed to be the same as those calculated from the previous model (k_{i2} = 1.386 min^{-1} and k_n = 6.93 x 10^{-2} min^{-1}). Therefore, the only adjustable parameter in the series of equations is the rate constant for the dissociation of the dimer (k'_d). To determine this parameter, activity data from refolding at 1.0 mg/ml (33.3 μM) and 1.0 M GuHCl was used (Figure 1.4). As an initial estimate of the dissociation rate constant, the dissociation constant calculated from equilibrium studies was modified to reflect the different GuHCl concentration by using the relationship developed in Section 1.6.

$$k'_d = k'_d|_{eq,\ 2\ M\ GuHCl} \left(\frac{1\ M\ GuHCl}{2\ M\ GuHCl}\right)^{-6.7} \qquad (3.24)$$

This equation resulted in an initial estimate of 0.407 min^{-1} for k'_d. Next, the initial concentrations of the first intermediate and dimer were calculated as 3.3 μM and 15 μM, respectively, by using Equation 3.15 with an equilibrium constant of 1.3 μM^{-1} which was determined from the equilibrium studies (Chapter 2). These values were then used to calculate the concentration profile for the active protein as a function of time. The value

of k'_d was iteratively changed until a reasonable fit (± 5%) to the activity data was achieved as shown in Figure 3.2. The dissociation rate constant, k'_d, for this fit was 0.107 min^{-1}. The dimer concentration was then plotted and compared to the dimer concentration data as discussed in Section 1.3. The close match between model prediction and both the active protein and dimer concentration data was successfully achieved by modification of only one parameter (k'_d). In addition, the final association rate constant, k_d, was calculated to be 0.138 min^{-1} μM^{-1} based on the equilibrium relationship and the value of k'_d. These results also support the assumptions which were made to develop the model. The first assumption was the rapid equilibrium association ($K_D >> 1$) which was true for the conditions used to fit this model (33.3 μM). Since the equilibrium association was rapid at these conditions, the equilibrium prediction of the initial concentrations (Equation 3.15) was also a valid assumption. However, the complete validation of this assumption would require the measurement of dimer concentrations at very early refolding times (< 1 min). Overall, the model equations (Equations 3.18, 3.20, 3.22 and 3.23) accurately represented the reversible association pathway (Equation 3.11) and the observed decrease in the refolding rate at high protein concentrations (< 10 μM) and 1.0 M GuHCl.

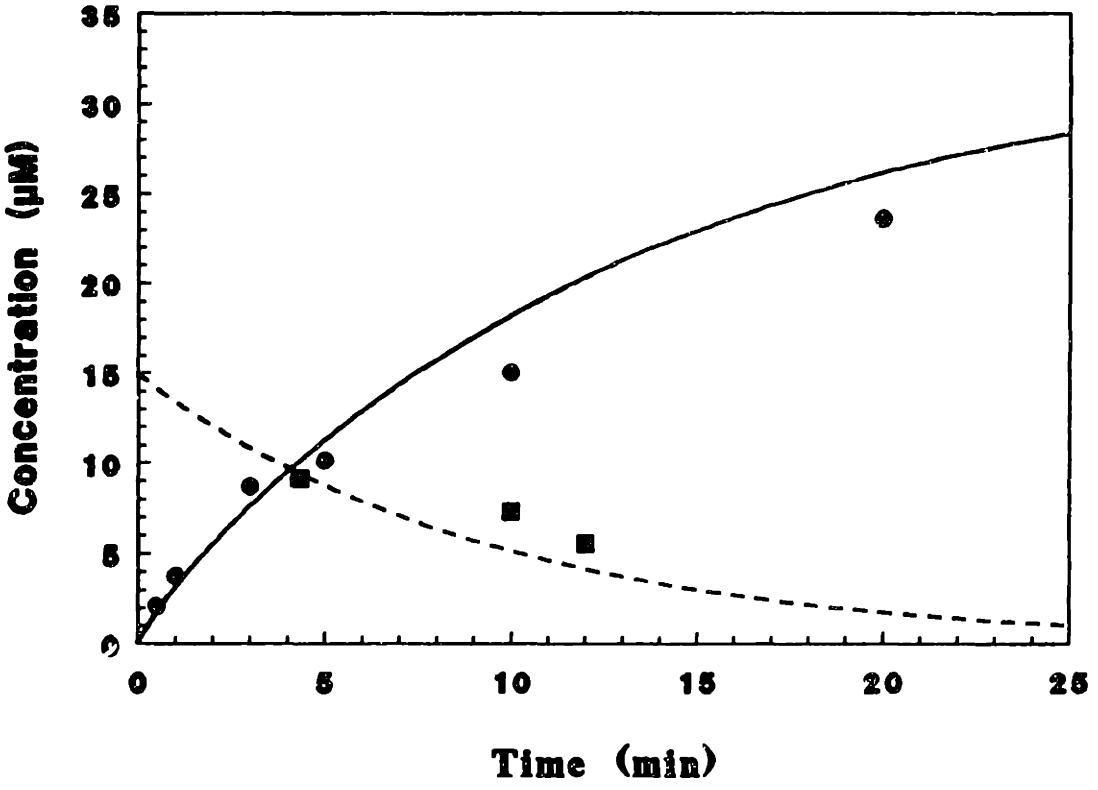


FIGURE 3.2: Reversible association pathway model comparison to refolding data. Refolding was performed at 1.0 mg/ml CAB and 1.0 M GuHCl. The recovery of active protein is plotted as a function of time (●) for comparison with the model prediction (–) which is calculated as described in the text. In addition, the concentration of dimer species was calculated from the data in Figure 1.5 (■). The model prediction of the dimer concentration (- -) was also calculated as described in the text.

Chapter 4: Cosolvent Effects on CAB Aggregation and Refolding

The model pathways for aggregation and refolding of CAB demonstrated that association occurred through the hydrophobic first intermediate on the refolding pathway. Therefore, a reduction in aggregation can be achieved by preventing the association of the first intermediate. One possible method for reducing the aggregation of the first intermediate would involve the use of a cosolvent which would prevent the association. However, this cosolvent should not interfere with the formation of the second intermediate or the native protein. Two different types of cosolvent systems have been used in an attempt to prevent aggregation of the first intermediate during refolding. First of all, sugars were used to determine if cosolvents which increase the hydration of the protein would cause a reduction in aggregation. Since sugars have been used in protein drug formulations to stabilize the native protein structure, these cosolvents may be useful in inhibiting aggregate formation. Polymers such as polyethylene glycol (PEG) were utilized as the other cosolvents. These cosolvents can interact with the protein through either hydration of the protein or binding to the protein surface (see Literature Review for detailed discussion).

4.1 Sugars: Sucrose and Glucose

The effects of increased hydration on protein aggregation during refolding were first determined by refolding in the presence of sucrose or glucose. Sucrose and glucose dissolved in the standard dilution buffer at 238.9 g/l (22% wt/vol) were used to measure the effect of sugars on aggregation and refolding. Dilution of CAB in 5 M GuHCl to 0.24 mg/ml CAB and 0.30 M GuHCl with the sugar solutions did not result in the

formation of large aggregates as observed by QLS. For each case, the multimer concentration was calculated from QLS analysis of each sample over time after dilution to the final conditions. The initial rate of dimer formation, R_D, was determined from the initial slope of the dimer concentration as a function of time. In addition, the recovery of active protein was measured as a function of time after dilution with each sugar solution. The initial rate of refolding, R_{Ref}, was calculated from the initial slope of the activity as a function of time. A comparison of refolding with the glucose (MW 180) and sucrose (MW 360) revealed that the initial rate of dimer formation, R_D, was similar in both cases, but the initial rate of refolding, R_{Ref}, was greater when sucrose was used as illustrated in Figure 4.1. Also, to determine the final recovery yield for each case, the recovery of active protein was measured after 1 hour. In both cases, the final recovery of active protein was not significantly greater than the final recovery of active protein in the absence sugar. Refolding under aggregation conditions of 1.0 - 1.2 mg/ml CAB and 0.5 M GuHCl in the presence of 22% wt/vol sucrose yielded similar refolding rates and little improvement in the final recovery of activity.

Although the final recovery of activity was not increased by the addition of sugars, the initial rate of refolding, R_{Ref}, was increased for each case. This increase in the initial refolding rate was likely the result of an increase in the hydration of the protein. Preferential hydration of the protein results in the formation of more compact states with hydrophilic surface characteristics such as the native state and also enhances the stability of compact structures such as the dimer and trimer. Therefore, sugars will prevent precipitation by stabilizing the native and multimeric states (Figure 4.2). As shown in

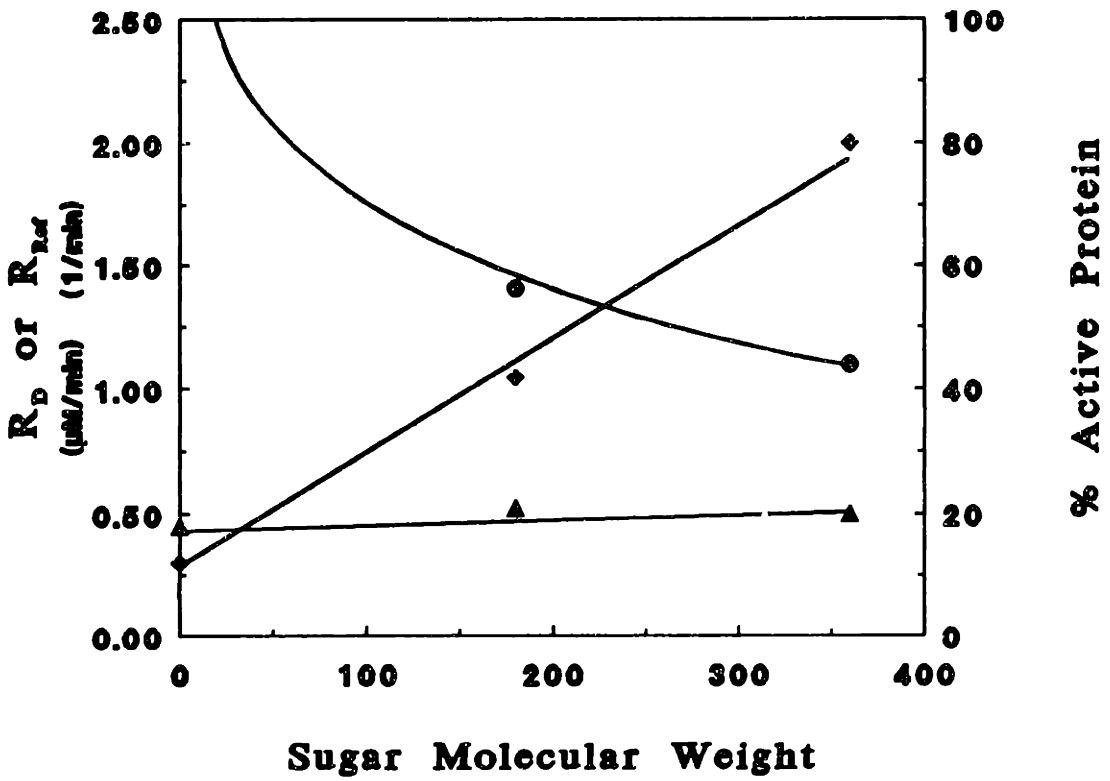


FIGURE 4.1: Sugar effects on CAB refolding and aggregation. The denatured protein in 5 M GuHCl was diluted with either glucose (MW 180) or sucrose (MW 360) at 238.9 g/l (22% wt/vol) dissolved in the standard buffer. The final aggregating conditions were 0.24 mg/ml CAB and 0.30 M GuHCl. The rate of dimer formation, R_D, was calculated from QLS analysis of each solution (●). The rate of refolding, R_{Ref}, was determined based on the initial rate of recovery of activity (♦). In addition, the percentage of total activity recovered after 1 hour was measured for each solution (▲).

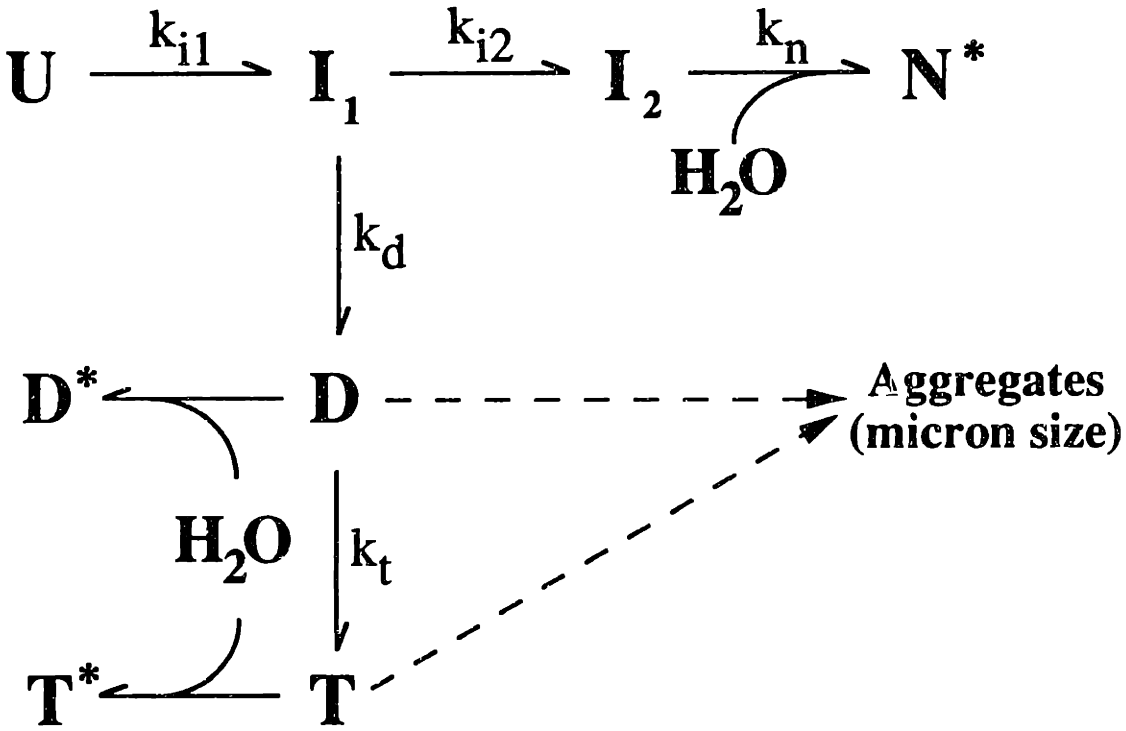


FIGURE 4.2: Proposed CAB refolding pathway in the presence of sugars. The normal refolding and aggregation pathways occur upon rapid dilution of CAB from 5 M GuHCl (Figure 1.10). Sugars are excluded from the protein surface and, thereby, cause the stabilization of the native and multimeric species through preferential hydration. A water layer forms around the protein species due to this hydration effect which is depicted here by water addition to the native, dimer and trimer to form the hydrated molecules, N^*, D^* and T^*. It is likely that several water molecules interact at the surface and only one molecule is shown here for affect.

Figure 4.2, the stabilization process will occur by the addition of a hydration layer around the surface of the compact species and this hydration layer would have a high surface tension. Since sugars did not increase the final recovery of active protein, the reduction in the rate of dimer and trimer formation may be the result of a slower rate of aggregation to large precipitates. In addition, the reduced rate of large aggregates formation may have been caused by the increase in viscosity. This viscosity increase may have effectively raised the critical nuclei concentration required for the nucleation and growth behavior observed for the precipitation of this protein (Section 1.6). These results indicate that cosolvents such as sugars which operate only by hydration of the protein will not be effective in enhancing the recovery of active protein.

4.2 Polymers: Polyamino acids and Polyethylene Glycol (PEG)

Cosolvents such as sugars which operate by a preferential hydration mechanism were unsuccessful in increasing the yield of active protein. Therefore, polymers which may act by both preferential hydration and binding were examined for their ability to reduce aggregation and improve recovery of active protein. The generality of the interaction of these polymers with the different protein species may be measured by applying several different polymers which exhibit similar properties. First of all, soluble polyamino acids with molecular weights ranging from 2500 to 3500 Daltons were used in an attempt to enhance refolding and prevent aggregation. Polyamino acid solutions at 3 g/l dissolved in the standard dilution buffer were used to dilute CAB in 5 M GuHCl to 1.0 mg/ml protein and 0.50 M GuHCl (aggregation conditions, see Figure 1.3). The recovery of active protein for each case was then measured as a function of time after

dilution as shown in Figure 4.3. Since polyglycine (3500 MW) was not soluble at 3 g/l in the dilution buffer, refolding with this polyamino acid was performed with a 1 g/l solution. The use of polyglycine (3500 MW) and polyalanine (3000 MW) resulted in a rate of refolding similar to that observed for refolding at the same final conditions in the presence of 3 g/l PEG (MW 3350) (Figure 4.3). The polylysine (2500 MW) solution yielded a greater rate of refolding, but the extent of refolding measured after 1 hour was the same for each polymer. Precipitation was not observed for any of the polyamino acid experiments. In each case, the protein rapidly reached a high level of activity and, then, slowly continued to refold to the fully active form. These results corresponded to the rapid formation of the second intermediate in the absence of first intermediate association and this rapid step was followed by the slow formation of the native state. Hydrophobic polyamino acids such as polyalanine and polylysine may have increased the rate of refolding through specific interactions with the hydrophobic first intermediate. Since these polymers can form some secondary structure, they may specifically interact with a similar structure on the protein. For example, polylysine will form a beta sheet structure in solution and this structure can interact with the hydrophobic beta sheet structure on the surface of an intermediate species. These polymers may therefore require a specific intermediate surface structure to be effective in preventing aggregation.

Unlike the polyamino acids, polyethylene glycol does not form a rigid structure in solution at low concentrations (< 10 % wt/vol). The refolding rate with polyethylene glycol, PEG (3350 MW), was similar to that observed with the polyamino acids (Figure 4.3). Refolding with PEG will therefore yield information on the two possible

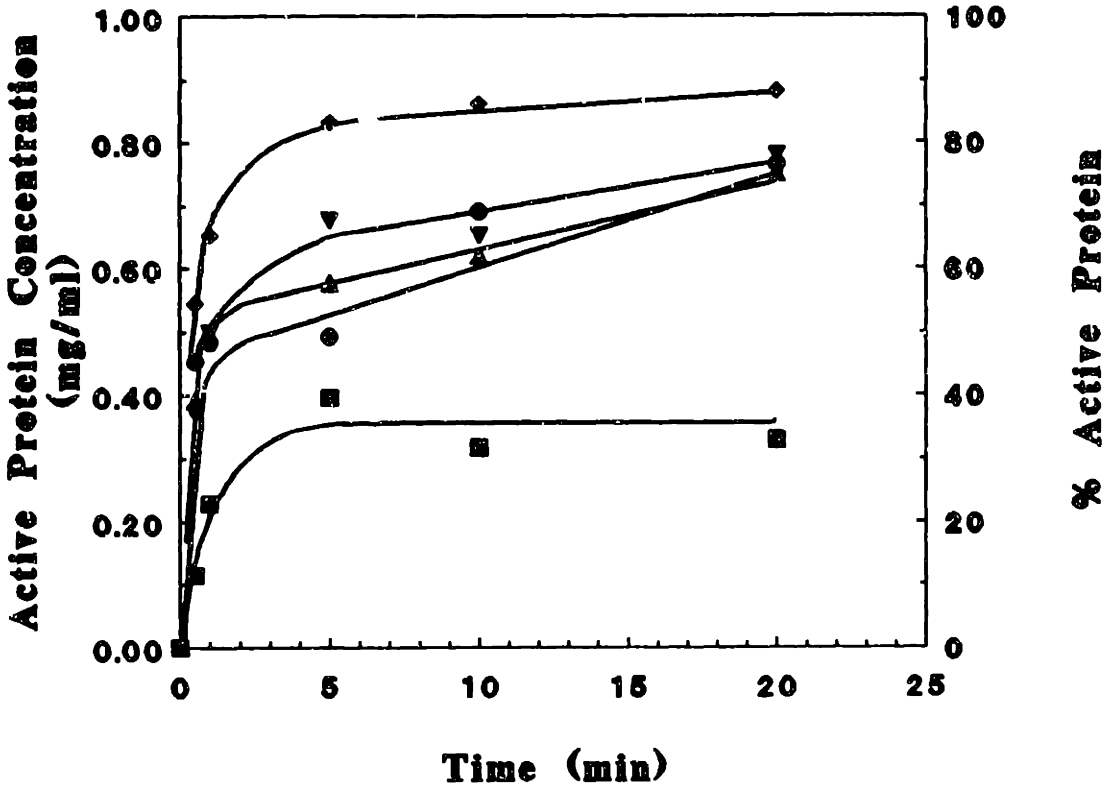


FIGURE 4.3: Effect of polyamino acids on the CAB aggregation during refolding. The protein in 5 M GuHCl was diluted to 1.0 mg/ml CAB and 0.50 M GuHCl (aggregation conditions) with different polyamino acid solutions. Polyalanine (3000 MW; ▼) and polylysine (2500 MW; ♦) were used at final concentration of 3 g/l. In addition, polyglycine (3500 MW; ●) was used at its solubility limit which was approximately 1 g/l. For comparison, the recovery of activity is displayed for refolding in 3 g/l PEG (3350 MW; ▲) and in the standard dilution buffer (■).

mechanisms of cosolvent interaction, hydration or binding, without a dependence on the structure of the polymer in solution. To measure the two potential mechanisms of interaction, refolding experiments were performed with different PEG concentrations at aggregation conditions. The effect of PEG at 3 g/l and different molecular weights was first examined by dilution of CAB from 5 M GuHCl to 1.0 mg/ml CAB and 0.50 M GuHCl with each different PEG solution. The recovery of active protein was then measured as a function of time after dilution as presented in Figure 4.4. In the absence of PEG, the protein recovered only 30% of its activity at these conditions and the maximum recovery was achieved within 5 minutes. On the other hand, refolding in the presence of 3 g/l PEG resulted in significant increase in the recovery of activity for each PEG molecular weight. The highest molecular weight PEG (8000 MW) at 3 g/l yielded the greatest rate of recovery resulting in the formation of fully active protein within 20 minutes. Although the lower molecular weight polymers (1000 and 3350 MW PEG) resulted in a slower rate of recovery of active protein, the protein completely recovered activity after 1 hour for refolding with each different PEG. A higher PEG concentration (30 g/l) was then used to observe the effect of PEG concentration on refolding at the aggregation conditions of 0.20 mg/ml protein and 0.30 M GuHCl. Using the standard dilution buffer, these final protein and GuHCl conditions result in the formation of aggregates as observed by quasi-elastic light scattering (QLS). For refolding at these conditions, the concentration of multimers was measured as a function of time by QLS. The recovery of active protein was also determined for several different times after dilution to the final conditions. The initial rate of dimer formation and the initial rate of

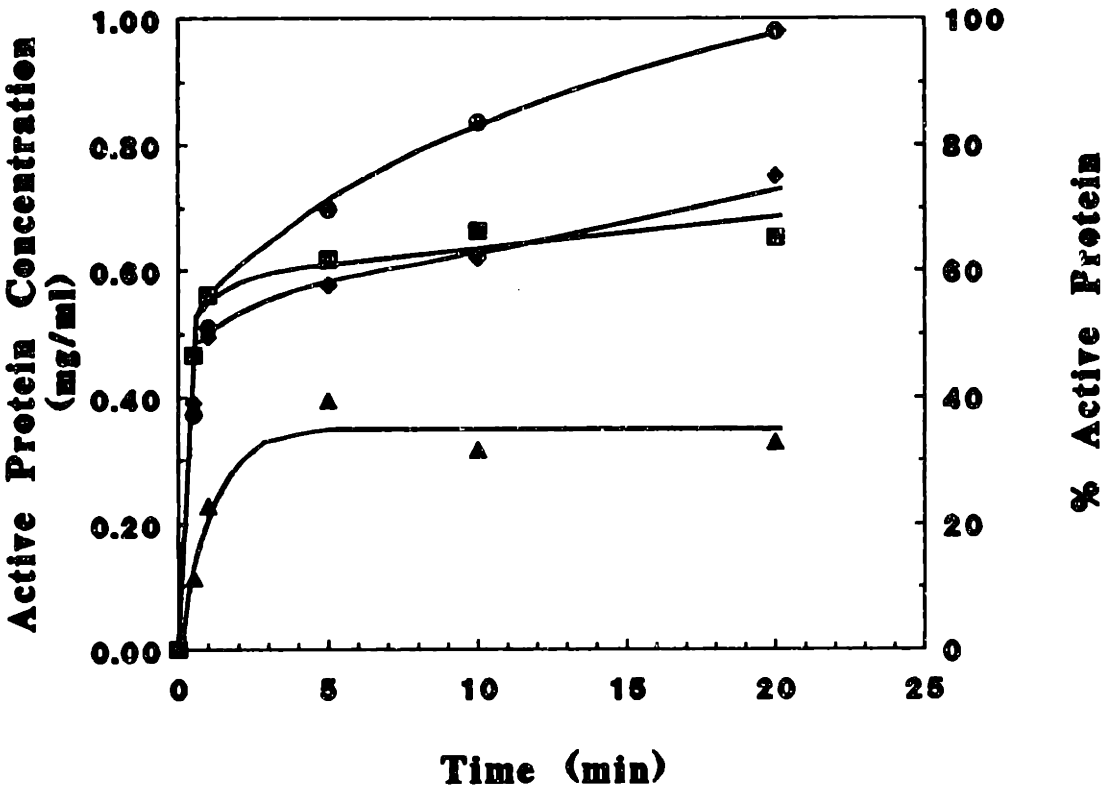


FIGURE 4.4: CAB aggregation and refolding in the presence of different molecular PEG solutions. CAB in 5 M GuHCl was diluted with 3 g/l PEG to 1.0 mg/ml protein and 0.50 M GuHCl (aggregation conditions). The recovery of activity as a function of time is plotted for the three different molecular weight PEG solutions, 8000 MW (●), 3350 MW (♦), and 1000 MW (■), along with the results for refolding in the absence of PEG in the dilution buffer (▲).

refolding were calculated from the initial slopes of their respective concentrations as a function of time. As displayed in Figure 4.5, the initial rate of refolding, R_{Ref} , was significantly increased in the presence of PEG at molecular weights between 1000 and 8000 daltons. The refolding rate dependence on molecular weight was similar to that observed for refolding at 3 g/l (Figure 4.4). In addition, the initial rate of dimer formation, R_D , as measured by QLS was greatly reduced when PEG ranging in molecular weight from 1000 to 8000 daltons was used in the dilution buffer. The use of PEG with molecular weights less than 1000 resulted in the formation of precipitates, but these precipitates formed at a slower rate than in the absence of PEG. Although the concentration of active protein recovered after 20 minutes was different for the 1000 to 8000 MW PEG as shown in Figure 4.5, the complete recovery of active protein was achieved in each case.

These results revealed that dilution buffers containing PEG with a molecular weight between 1000 to 8000 Daltons will decrease the rate and extent of aggregation. The use of 30 g/l PEG (8000 MW) has been observed to cause precipitation during refolding at aggregation conditions of 1.0 mg/ml CAB and 0.50 M GuHCl. Therefore, concentrations of PEG greater than 3 g/l should be avoided at higher final protein concentrations. The observed reduction in the dimer formation rate for both cases can not be accounted for by nonspecific interactions such as viscosity since the solution properties are not significantly altered by the addition of 3 or 30 g/l PEG (μ = 1.05 - 1.21 cp). These results were, however, indicative of the two mechanisms of PEG, protein binding which may prevail at low PEG concentrations (0 to 1.0 % wt/vol) and preferential

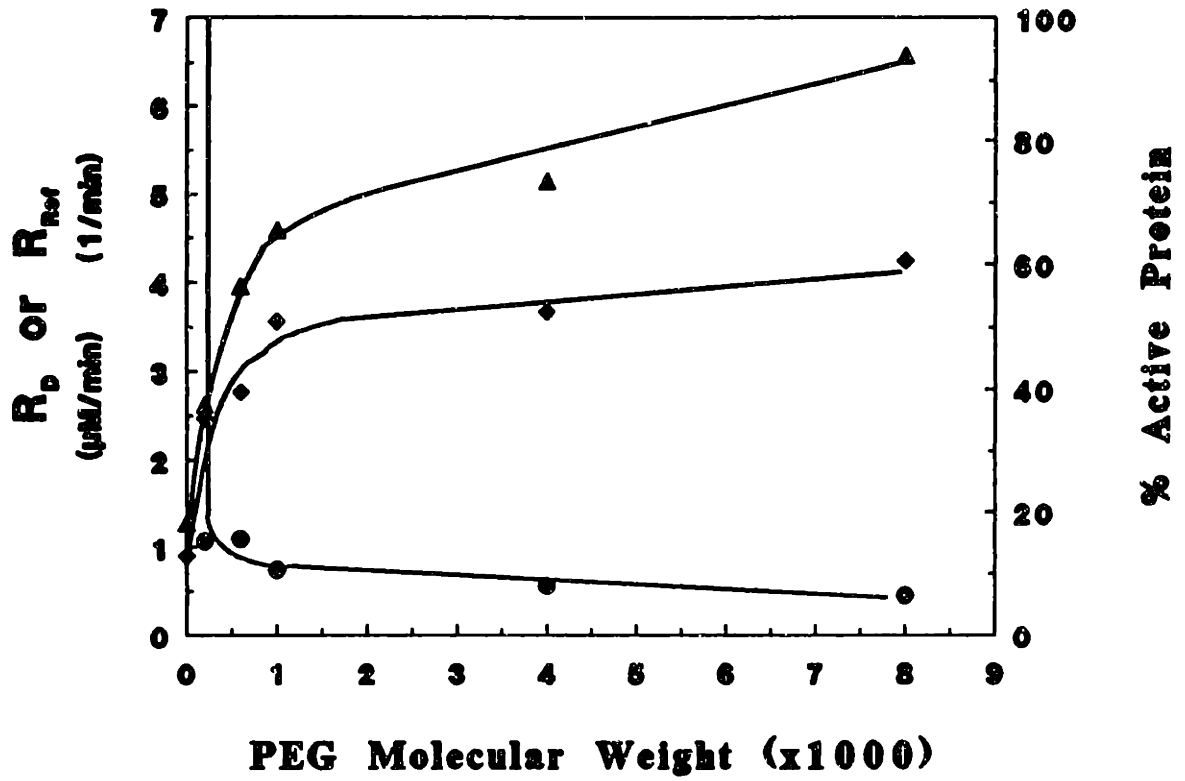


FIGURE 4.5: Rate of refolding and dimer formation as a function of PEG molecular weight. CAB in 5 M GuHCl was rapidly diluted with 30 g/l PEG to 0.20 mg/ml CAB and 0.30 M GuHCl (aggregation conditions). The rate of dimer formation, R_D, was calculated from QLS analysis of each solution (●). The rate of refolding, R_{Ref}, was determined based on the initial rate of recovery of activity (♦). In addition, the percentage of total activity recovered after twenty minutes was measured for each solution (▲).

hydration which is dominant at high PEG concentrations (1 to 50 % wt/vol) and dependent upon the polymer molecular weight (Arakawa et.al.,1985; Arakawa & Timasheff, 1990).

The study of different cosolvents to prevent association of the first intermediate showed that cosolvents such as PEG which can hydrate or bind to the protein were effective in reducing aggregation. On the other hand, cosolvents such as sugars which hydrate the protein surface decreased the rate of aggregation, but did not increase the recovery of active protein. Therefore, refolding at aggregation conditions with cosolvents such as PEG will increase the recovery of active protein. These results also revealed an increase in the rate of refolding for conditions where aggregation was previously observed. The cosolvent must then be increasing the rate of refolding by altering the unimolecular refolding pathway (see Equation 3.1) or preventing the formation of the associated species (see Figure 1.10). To discriminate between these two alternatives, a more detailed analysis of PEG enhanced refolding was required.

Chapter 5: Polyethylene Glycol (PEG) Enhancement of Refolding

Since PEG was shown to decrease aggregation and increase the recovery of active protein, additional studies were performed to determine whether PEG altered the refolding pathway or prevented the formation of associated species. First of all, the effect of PEG on the refolding pathway at 1.0 M GuHCl and high protein concentrations (≥ 10 μM) was examined to determine the ability of PEG to prevent the transient association of the first intermediate (Section 3.2). In addition, the relationship between the PEG concentration and the rate of refolding was determined to provide insight into the specific stoichiometry and mechanism of PEG interaction. To further characterize the interaction of PEG with the refolding protein, refolding was attempted with PEG at different molecular weights. These studies provided additional information on the mechanism of PEG enhanced refolding.

5.1 Intermediate Species Interaction with PEG during Refolding

In order to ascertain the effect of PEG on the refolding pathway, refolding was performed by dilution of denatured CAB from 5 M GuHCl to 1.0 M GuHCl and 0.50 mg/ml (16.7 μM) protein. These conditions resulted in the formation of a transient dimer species during the refolding process as discussed in Chapter 3. The recovery of active protein was measured as a function of time after dilution for refolding with 3 g/l PEG (3350 MW) or standard dilution buffer as shown in Figure 5.1. The rate of refolding was greatly enhanced by the addition of 3 g/l PEG to the dilution buffer and approximately 90% active protein was recovered in 10 minutes. In contrast, refolding without PEG in the dilution buffer resulted in the recovery of 80% active protein after 20 minutes (Figure

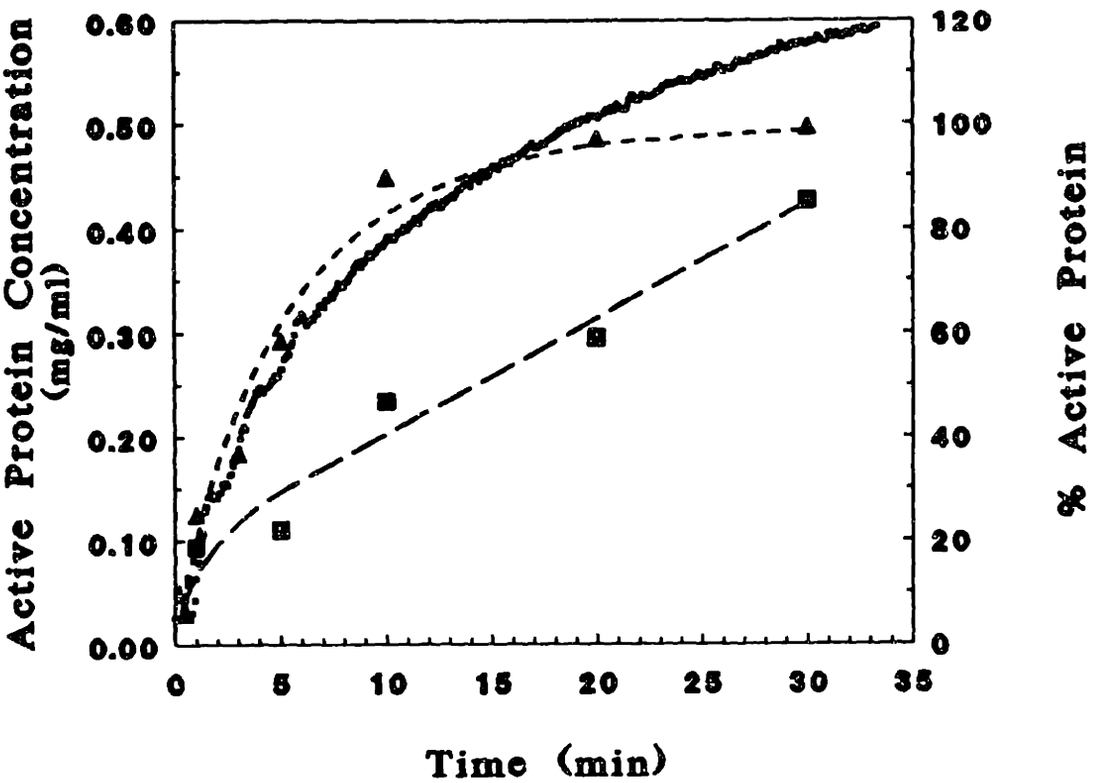


FIGURE 5.1: Recovery of activity for CAB refolding in PEG. CAB in 5 M GuHCl was diluted with either PEG (MW 3350) or PEGpNP at 3 g/l to a final protein concentration of 0.50 mg/ml (16.7 μM) and 1.0 M GuHCl (refolding conditions). The recovery of activity in PEG (- ▲ -) or PEGpNP (•) are shown as a function of time after dilution. For comparison, the refolding of CAB under the same conditions using standard buffer without PEG is also displayed (- ■ -).

5.1).

To measure the kinetics of the observed rate enhancement, the substrate-PEG analog, bis(p-nitrophenol carbonate) polyoxyethylene (PEGpNP, PEG 3350 MW), was used at 3 g/l in the dilution buffer for refolding to the same final conditions, 0.50 mg/ml (16.7 μM) protein and 1.0 M GuHCl. This substrate-PEG analog, PEGpNP, provided a means for continuous analysis of the refolding over time since the enzyme acted as an esterase to cleave the p-nitrophenol (pNP) from the PEGpNP. The hydrolysis product, pNP, was monitored spectrophotometrically at 5 second intervals. The hydrolysis rate was then used to calculate the concentration of active enzyme. The ratio of the hydrolysis rate for the refolding protein to hydrolysis rate of the native protein over each five second time interval (Δt = 5 sec) was considered equal to the fraction of active protein at each time after dilution with the buffer containing PEGpNP.

$$[CAB]_{active} = \{\Delta pNP|_{Refold}/\Delta pNP|_{Native}\}_{\Delta t}\,[CAB]_f \qquad (5.1)$$

For time periods greater than 20 minutes, the hydrolysis of PEGpNP became substrate limited for the native protein resulting in an underestimation of native protein hydrolysis rate ($\Delta pNP/\Delta t$). Therefore, the calculated percentage of active protein was greater than 100% at the end of the refolding process. These results indicated that PEG will enhance the rate of refolding at the final conditions, 0.50 mg/ml CAB and 1.0 M GuHCl, where the formation of a transient dimer species was previously observed (Section 3.2).

The enhanced refolding at conditions of transient dimer formation revealed that PEG may either be preventing the association or altering the refolding pathway. Either

of these mechanisms would require an interaction between PEG and an intermediate on the refolding pathway. To determine which intermediate specie was interacting with PEG, the substrate-PEG complex, PEGpNP, was used to probe potential surface interactions between PEG and the protein. If PEG binds to the protein during the refolding process, the substrate-PEG complex, PEGpNP, could be utilized to determine the effective concentration of PEGpNP at the surface of the protein during refolding. In addition, if PEGpNP binds to the protein, the effective concentration of the pNP substrate moiety would be greater at the active site than for the control case of PEG (3350 MW) with the normal ester substrate, p-nitrophenol acetate (pNPA), at the same concentrations. Since there was no significant difference in the rate of ester hydrolysis using the native protein in either PEGpNP or PEG with pNPA solutions, the effective concentration of substrate at the native protein surface was not significantly different for these two cases. In contrast, the rate of hydrolysis for the refolding protein at the early stages of refolding was greater for the PEGpNP solution than the PEG with pNPA solution. Since the second intermediate has approximately 25% of the native protein activity, the difference in hydrolysis rates could be accounted for by an increase in the concentration of substrate at the partially formed active site of the second intermediate (Wong & Tanford, 1973; Rodionova et.al., 1989). Therefore, PEG apparently does bind to the surface of the intermediate protein. The difference in the hydrolysis rates between refolding in PEGpNP and PEG with pNPA provided an estimate of the effective concentration of PEG at the surface of the protein.

$$\{ [pNP]_{PEGpNP} - [pNP]_{PEG+pNPA} \}_{\Delta t} = [pNP]_{eff} = [PEGpNP]_{eff} = [PEG]_{eff} \qquad (5.2)$$

To assess the effective concentration of PEGpNP at the protein surface during refolding, denatured CAB in 5 M GuHCl was diluted to the final conditions of 0.50 mg/ml CAB (16.7 μM) and 1.0 M GuHCl with either 3 g/l PEGpNP or 3 g/l PEG with pNPA ($[PEG]_f$ = 895 μM, $[\text{-pNP}]_f$ or $[pNPA]_f$ = 1.6 mM). The recovery of active protein was then calculated for the PEGpNP case by using Equation 5.1 and the effective PEG concentration was calculated from Equation 5.2 as shown in Figure 5.2. The ratio of effective PEG concentration ($[PEG]_{eff}$) to inactive protein ($[CAB]_{inact}$) was approximately 2 to 1 at the initial stages of refolding and this ratio decreased after the first 2 minutes. To determine which intermediate was interacting with the PEG, the previously reported rates of intermediate formation were used with Equations 3.6 and 3.8 to calculate the concentration of each intermediate over time as shown in Figure 5.2 (Semisotnov et.al., 1987; Stein & Henkens, 1978). The decrease in the ratio of effective PEG concentration to inactive protein ($[PEG]_{eff}/[CAB]_{inact}$) was concomitant with the decrease in the first intermediate indicating that the PEG interacted with the first intermediate in the refolding pathway and was excluded from the protein surface as the second intermediate was formed. These results showed that PEG may specifically interact with the first intermediate in the refolding pathway.

5.2 Stoichiometry of PEG Enhancement and Apparent Refolding Rates

As shown in Figure 5.2, the PEG interacts with the first intermediate and the ratio of effective PEG concentration to inactive protein concentration was 2 to 1. Therefore, the concentration of PEGpNP (3 g/l, 895 μM) used in the refolding experiment was greater than that required to interact with the inactive protein which was initially present

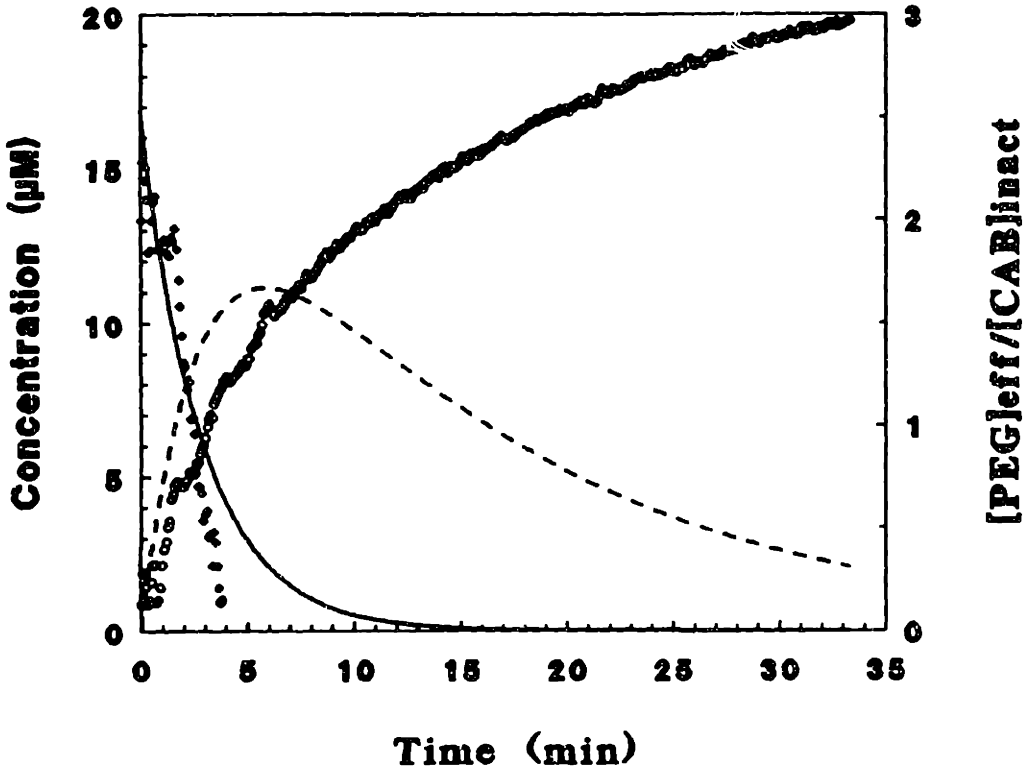


FIGURE 5.2: PEG interaction with intermediate species during refolding was assessed from analysis of the effective PEG concentration at the protein surface as a function of time ($[PEG]_{eff}/[CAB]_{inact}$, ♦). The active protein concentration as determined from PEGpNP measurements (○) was plotted with the calculated concentration profiles for the first intermediate (–) and the second intermediate (- -) based on the rate constants reported in the literature (Stein & Henkens, 1978; Semisotnov et.al., 1987).

at 0.50 mg/ml (16.7 μM) (Figure 5.2). To determine the relationship between PEG concentration and enhancement, refolding experiments were performed at several different molar ratios of PEG (3350 MW) to protein. Refolding of CAB in 5 M GuHCl was performed by rapid dilution to 0.50 mg/ml CAB and 1.0 M GuHCl with PEG in the dilution buffer. The initial rate of recovery of active protein, R_{ref}, was calculated from the slope of the activity as a function of time for each molar ratio (Figure 5.3). The optimum rate enhancement occurred at a molar ratio ($[PEG]_f/[CAB]_f$) between 2 and 3 and resulted in a three fold increase in the initial rate of refolding. This optimum corresponds to PEG concentrations between 0.15 and 0.20 g/l.

Using the assumption that PEG only interacts with the first intermediate in the refolding pathway, the increase in the refolding rate from the first to second intermediate was also calculated. The new apparent rate constant for formation of the second intermediate, k_{app}, was then determined for each case as described in Chapter 3 (Figure 5.3). The maximum apparent rate constant, k_{app}, was 1.08 min^{-1} which is comparable to the actual second intermediate rate constant (k_{i2} = 1.386 min^{-1}) determined for refolding at low protein concentrations and 1.0 M GuHCl where the first intermediate does not form a transient dimer species (Section 3.1). In addition, HPLC size exclusion analysis of the PEG enhanced refolding at 0.50 mg/ml (16.7 μM) CAB and 1.0 M GuHCl revealed that the transient dimer did not form during the refolding. Therefore, the PEG must be preventing the association of the first intermediate for refolding at the conditions where the formation of a transient dimer was previously observed (Section 3.2).

Since PEG was observed to prevent the formation of a transient dimer species at

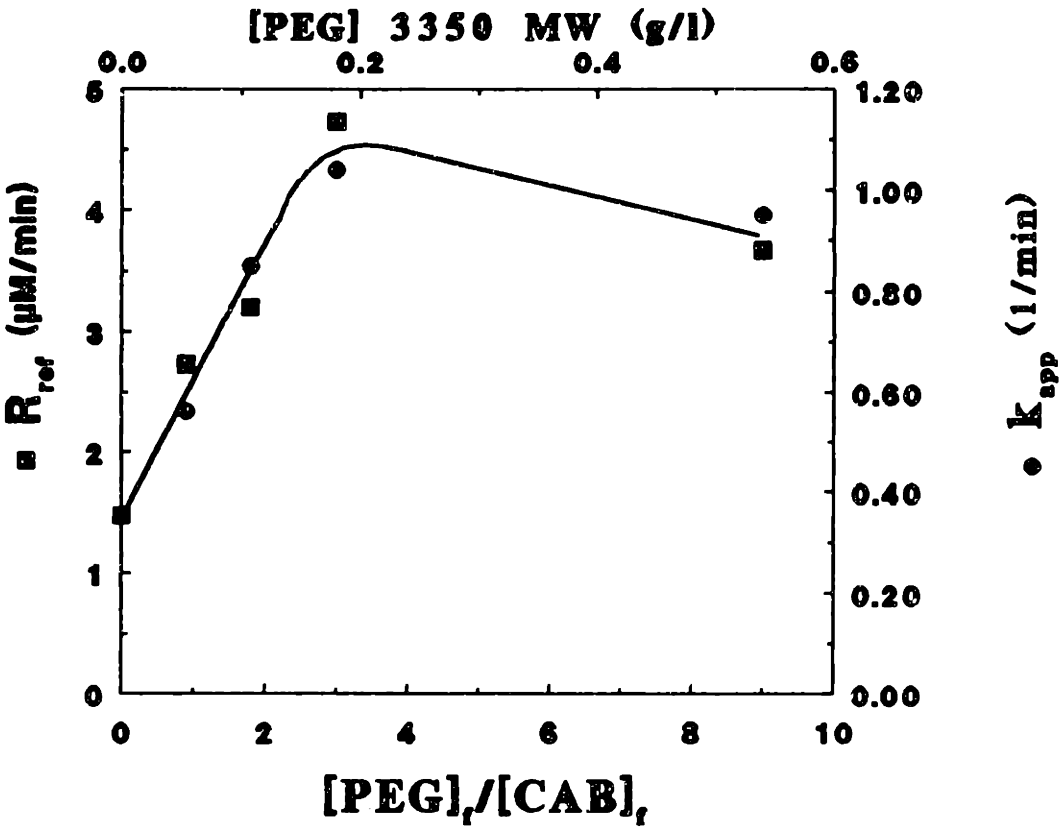


FIGURE 5.3: The PEG enhancement in refolding of CAB was determined by initial rate analysis of the activity as function of time. The initial rate of refolding, R_{ref}, was calculated from kinetic experiments of the protein activity at several different final molar ratios of PEG to CAB ([PEG]f/[CAB]f, ■). Assuming that PEG only interacts with the first intermediate to prevent association, the apparant rate constant for second intermediate formation, k_{app}, was calculated for each experiment by using the initial refolding rate data and the calculation method described in Chapter 3.

stoichiometric concentrations, it should also prevent the formation of dimers and trimers during refolding at aggregating conditions. To test this hypothesis, refolding at aggregating conditions of 1.0 mg/ml CAB and 0.50 M GuHCl was attempted at molar ratios of PEG to CAB ranging from 1 to 9. CAB in 5 M GuHCl was rapidly diluted to 1.0 mg/ml (33.3 μM) protein and 0.50 M GuHCl with PEG in the dilution buffer and the concentration of active protein was then measured as shown in Figure 5.4. The protein precipitated out of solution without PEG in the dilution buffer and resulted in the recovery of only 30% active protein. A one to one molar ratio of PEG to CAB provided only slightly higher recovery of active protein (35%) with precipitation of inactive protein. However, molar ratios ($[PEG]_f/[CAB]_f$) of 2, 3 and 9 resulted in complete recovery of active protein after 1 hour without the formation of precipitates. At these molar ratios, the protein also did not form multimers as observed by QLS. The minimum concentration of PEG (3350 MW) required to achieve complete recovery at these aggregating conditions was 0.35 g/l. These results along with the results of the refolding enhancement studies have indicated that a minimum molar ratio of PEG to CAB of 2 to 1 was required to prevent the formation of the multimers and, thereby, achieve increased recovery of active protein. Therefore, two PEG (3350 MW) molecules per CAB molecule were required to prevent aggregation or transient association of the first intermediate.

5.3 Effect of PEG Molecular Weight and Concentration

The previous studies showed a distinct stoichiometric relationship between PEG and the prevention of first intermediate association. To develop a further understanding of this relationship, refolding of CAB was performed by using PEG at different molecular

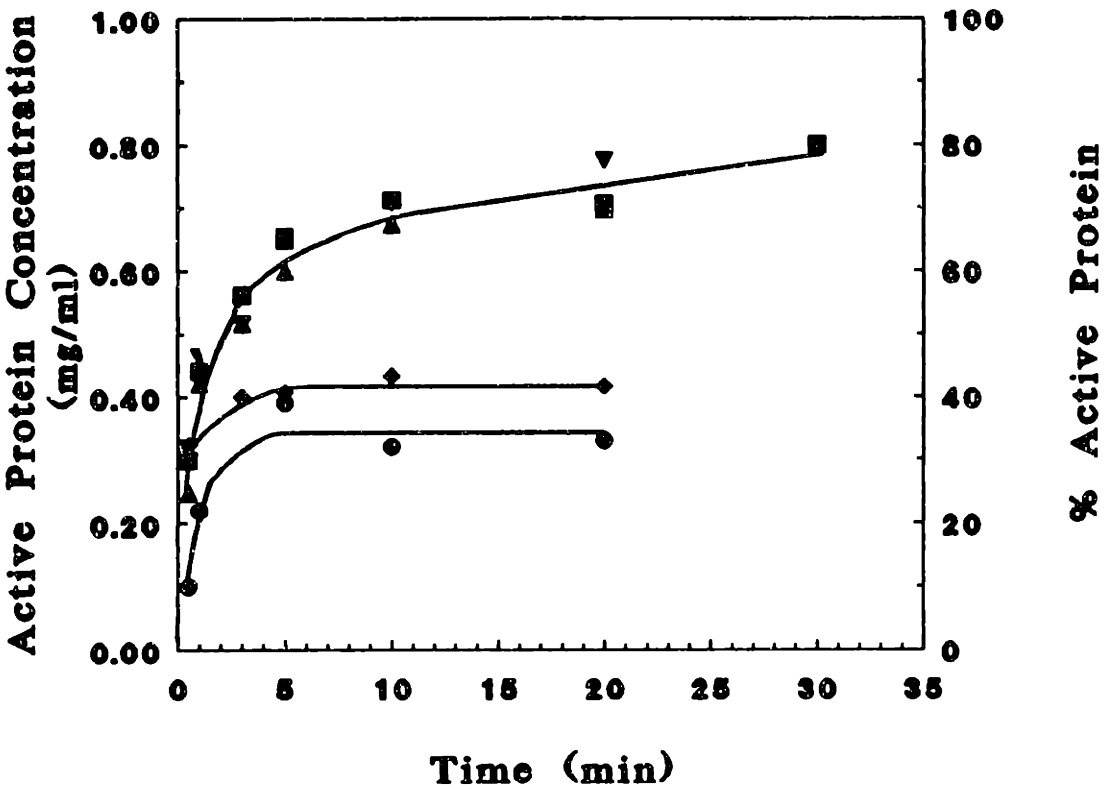


FIGURE 5.4: The ability of PEG to reduce aggregation during refolding was analyzed by rapid dilution of CAB in 5 M GuHCl to 1.0 mg/ml CAB and 0.50 M GuHCl with different molar ratios of PEG (3350 MW) to CAB. The esterase activity was measured as a function of time after dilution at a molar ratio of PEG to CAB ($[PEG]_f/[CAB]_f$) of 1 (♦), 2 (▼), 3 (▲), or 9 (■). The recovery of active protein is also plotted for refolding without PEG in the dilution buffer (●).

weights. Denatured CAB in 5 M GuHCl was refolded by rapid dilution to 1.0 mg/ml (33.3 μM) and 0.50 M GuHCl with PEG in the dilution buffer at a given molar ratio. Each solution was then allowed to equilibrate for 1 hour. After equilibration, the active protein concentration of each solution was measured as shown in Figure 5.5. With 8000 MW PEG in the dilution buffer, complete recovery of active protein was achieved for a narrow operating range of molar ratios (2 to 10). At a lower PEG molecular weight of 3350 Da, the range of molar ratios required for complete recovery was somewhat broader (2 to 20). This trend continued for the 1000 MW PEG which had a broader operating range (10 to 40) and required a higher minimum molar ratio (10) for complete recovery. When smaller molecular weight PEG (200 and 600 Da) was used in the dilution buffer, the protein did not completely recovery activity for any of the molar ratios attempted. In addition, when large molecular weight PEG (20,000 Da) was used in the dilution buffer, the recovery of active protein decreased and the extent of aggregation increased for all molar ratios (0.1 to 10).

These results corroborate the observed phenomenon of PEG adsorption to silica surfaces (Char et.al., 1989; Char et.al., 1988). For PEG adsorption on silica, the polymer will loop down and form a flat interface with the surface with a tail pointing into the solvent. The degree of surface packing has been shown to be dependent on the polymer molecular weight. In other words, as the chain length increases, the amount of polymer in contact with the surface increases resulting in less polymer required to cover the surface. After complete surface packing is achieved, addition of polymer results in a stacking effect where the segment length of the polymer in contact with the surface

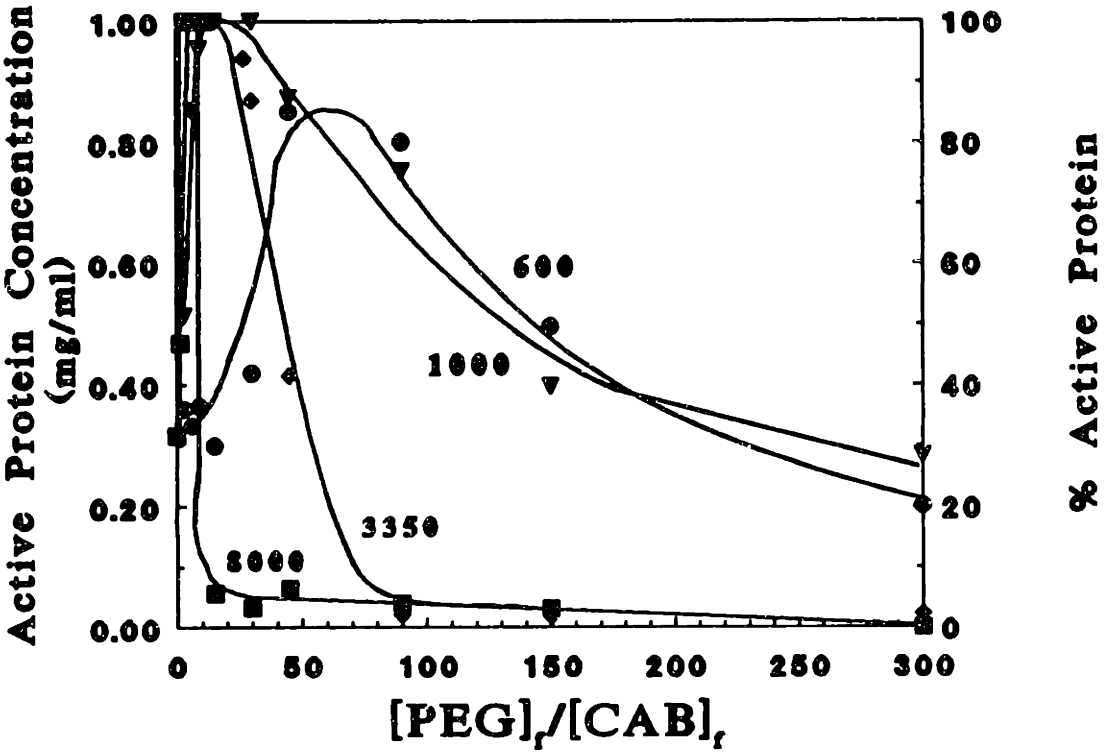


FIGURE 5.5: Effect of PEG molecular weight on refolding of CAB was assessed at final conditions which result in aggregation, 1.0 mg/ml CAB and 0.50 M GuHCl. The protein was refolded by rapid dilution to these final conditions with different final molar ratios ([PEG]f/[CAB]f) and different molecular weights of PEG (600 (●), 1000 (▼), 3350 (♦) and 8000 (■) MW). The activity of each solution was measured after incubation for one hour at the final conditions.

decreases and the number of polymers at the interface increases. For the low molecular weight polymers (200 and 600 Da), a stacking of the polymers at the surface may occur such that the polymer stacks on top of itself instead of spreading across the surface (Gast, 1990). This polymer adsorption theory would explain the molar ratios and stoichiometry observed in refolding of CAB with different PEG molecular weights. At high molecular weights (1000 to 8000 Da), PEG may follow a similar adsorption pattern of packing on the surface of CAB. Lower molecular weight PEG (200 and 600 Da) may not provide protection against aggregation since the polymer will not spread across the protein surface to cover the hydrophobic sections which are the likely cause of aggregation. A very large polymer such as the 20,000 MW PEG may adsorb to more than one protein molecule and cause an increase in the effective concentration of the protein and a subsequent increase in aggregation. Thus, refolding with PEG at different molecular weights would be predicted by polymer adsorption theory to occur with a similar stoichiometry as that shown in Figure 5.5.

The refolding experiments involving PEG at different molecular weights showed the specific stoichiometry of interaction between PEG and CAB. However, to uncouple the effect of the mass concentration and molecular weight, these results should be replotted as a function of the mass concentration of PEG. The data from Figure 5.5 was therefore replotted as the final recovery of active protein at each mass concentration of PEG (Figure 5.6). The combined results of experiments using PEG ranging from 200 to 20,000 Da revealed that mass concentrations between approximately 0.20 g/l and 3 g/l provided complete recovery of active protein under these final conditions. The data

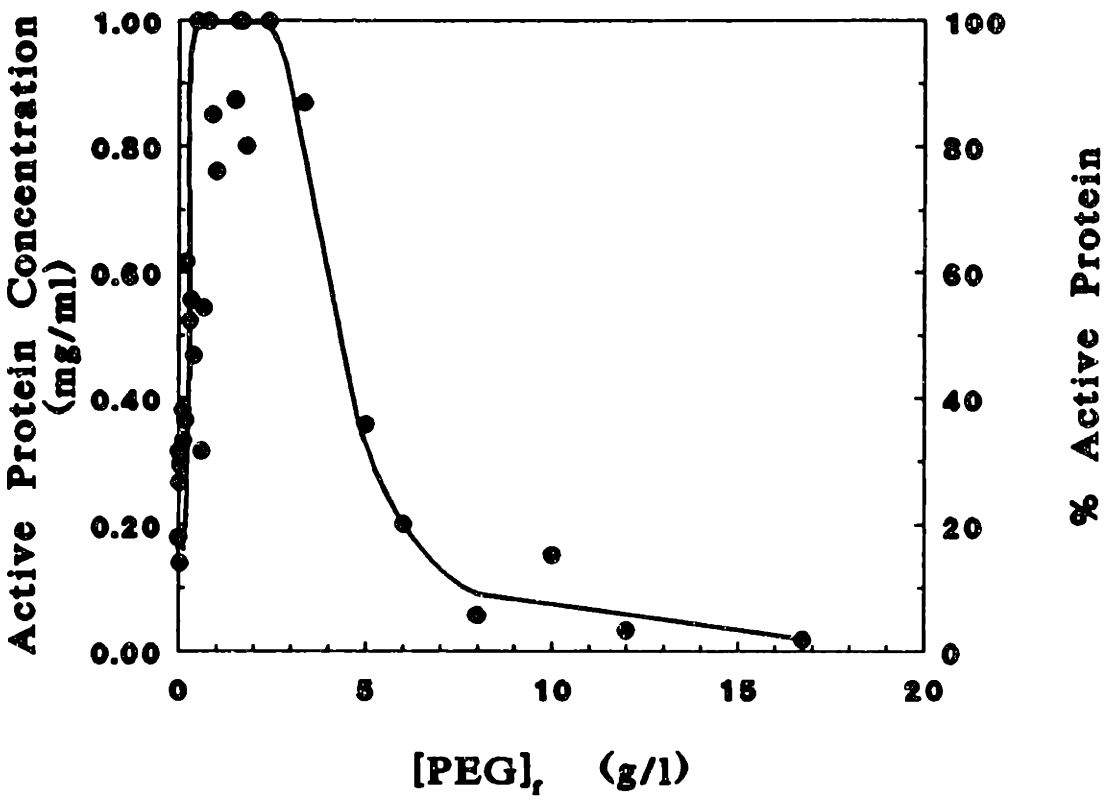


FIGURE 5.6: The final activity data from Figure 5.5 and additional experiments involving 200 and 20,000 MW PEG were plotted as a function of the final mass concentration of PEG in solution. All these experiments were plotted as a single set of data to show the relationship between mass concentration and recovery of active protein.

below the optimum mass concentration range for complete recovery of active protein as shown in Figure 5.6 represented the application of PEG molecular weights of 200, 600, and 20,000 Daltons where complete recovery was not achieved. The observed mass concentration effect may be explained by adsorption to the protein at low polymer concentrations and exclusion of the polymer from the protein surface at high polymer concentrations. A similar concentration effect was observed previously for PEG interaction with native proteins (Arakawa & Timasheff, 1985).

These refolding and aggregation studies have shown that the association of the first intermediate can be prevented through the application of specific molar ratios of PEG to CAB which are dependent upon the PEG molecular weight. PEG will interact with the first intermediate and prevent its association, but the distinct mechanism of this interaction was not elucidated from these experiments. PEG may alter the association process by dissociating the dimer through an increase the reverse rate constant, k'_d, binding to the first intermediate or causing the formation a new intermediate species which does not associate. To discriminate between each of these mechanisms, the effect of PEG on the equilibrium association of the first intermediate and the possible binding of PEG to the first intermediate must be determined.

Chapter 6: PEG Effect on Equilibrium Association of CAB

If PEG results in the formation of a new stable intermediate state or the dissociation of the dimer species, the equilibrium distribution of associated protein will be altered. First of all, if a new stable intermediate state is formed in the presence of PEG, the association dependence on the final GuHCl concentration should be altered since a higher GuHCl concentration would be required to destabilize the protein. On the other hand, if PEG only causes a dissociation of the multimer, a decrease in protein association should occur with increasing PEG concentration. Therefore, equilibrium protein association experiments were performed for different final conditions to elucidate the different phenomena.

6.1 Effects of Final Conditions on Equilibrium Association in PEG Solutions

To determine if PEG causes the formation of a new stable intermediate species, equilibrium refolding experiments were performed at 1.0 mg/ml CAB and final GuHCl concentrations ranging from 1.0 to 2.5 M in 30 g/l PEG (8000 MW) (see Chapter 2 for results without PEG). Each of these solutions was allowed to equilibrate for three to eight hours. After equilibration, the concentration of each species was determined by HPLC size exclusion analysis and plotted as a function of the final GuHCl concentration (Figure 6.1). The association of CAB had the same dependence on the final GuHCl as was observed without PEG (Figures 2.1 and 6.1). The association was the greatest for final GuHCl concentrations ranging from 1.8 to 2.2 M GuHCl. Since PEG did not shift the dependence on the final GuHCl concentration and the first intermediate at 2.0 M associated, the formation of a new stable intermediate state was not observed. However,

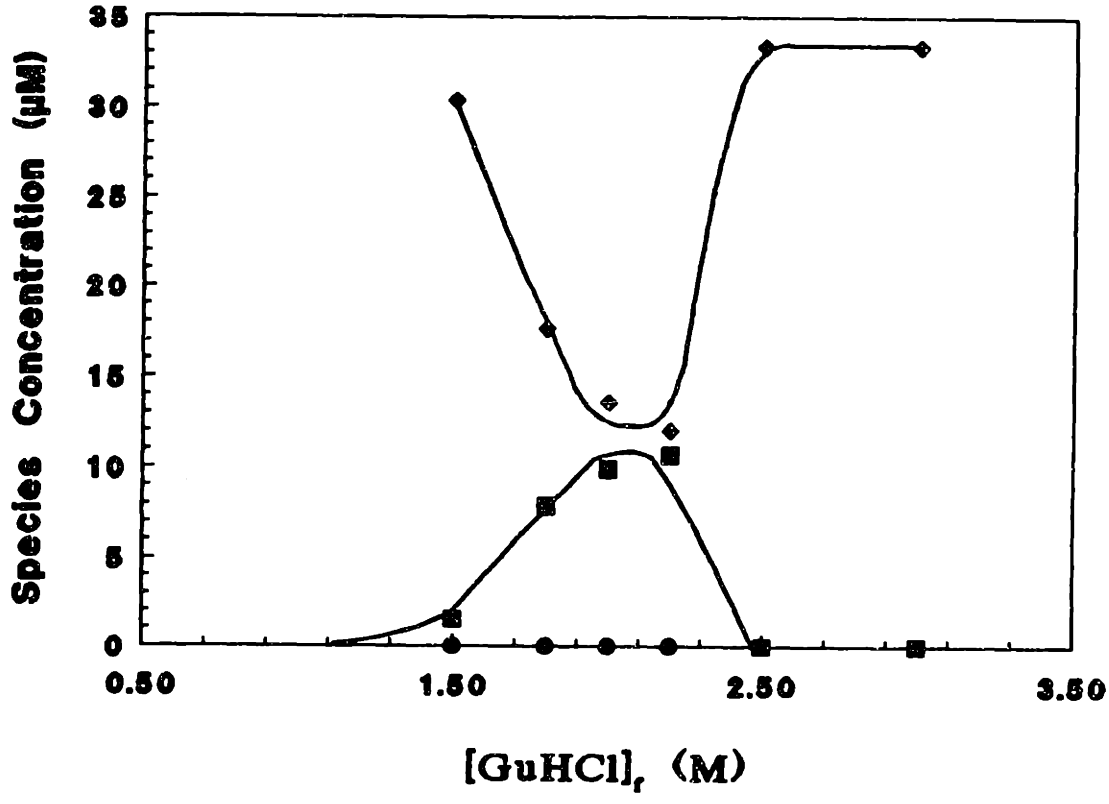


FIGURE 6.1: Equilibrium association as a function of GuHCl concentration in 30 g/l PEG (8000 MW). CAB in 5 M GuHCl was diluted to 2 M GuHCl and 1.0 mg/ml protein with PEG in the dilution buffer. The concentration of monomer (♦), dimer (■), and trimer (●) was determined by HPLC analysis. The trimer species (●) was not observed under these conditions.

the extent of association of the first intermediate at 2.0 M GuHCl was greatly reduced by the presence of PEG indicating a possible dissociation mechanism. The trimer species was not observed to form at any final GuHCl concentration at these conditions. PEG must then act by shifting the equilibrium association of the first intermediate at 2 M GuHCl to the monomer.

To verify the reduced association of the first intermediate caused by PEG, equilibrium experiments were performed at 2.0 M GuHCl and 1.0 mg/ml (33.3 μM) CAB where the first intermediate was observed to form an associated protein species (Figure 2.2). Denatured CAB in 5 M GuHCl was diluted to the association conditions with either 3 or 30 g/l of PEG (MW 8000) and allowed to equilibrate for three hours. After equilibration, the solutions were analyzed by QLS and the particle size distributions were converted into monomer, dimer, and trimer concentrations. The equilibrium distribution without PEG was 1.79 μM monomer (first intermediate), 6.52 μM dimer, and 6.17 μM trimer (Figure 2.2). With 3 g/l PEG in the dilution buffer, the distribution was only slightly altered. However, equilibration in the presence of 30 g/l PEG resulted in a shift in the size distribution to the monomeric state (13.56 μM monomer, 6.50 μM dimer, and 2.26 μM trimer) as observed by QLS analysis. The trimer species was not observed in the previous experiment which was performed by HPLC analysis since the time period for HPLC analysis was on the order of the equilibrium time between dimer and trimer (Figure 6.1 and Section 2.3). The equilibrium distribution of the associated species was therefore altered by PEG at high concentrations.

To confirm the formation of the associated state in the presence of PEG, CAB

denatured in 5 M GuHCl was refolded by dilution to 2.0 M GuHCl and two different final conditions, 5.0 mg/ml protein with 3 g/l PEG (8000 MW) (association conditions), and 1.0 mg/ml protein with 30 g/l PEG (8000 MW) (monomer conditions). These solutions were equilibrated and, then, diluted with the same concentration of PEG to aggregation conditions: 1.0 mg/ml CAB and 0.50 M GuHCl (3 g/l PEG MW 8000) and 0.20 mg/ml and 0.4 M GuHCl (30 g/l PEG MW 8000), respectively. Each solution was then assayed for activity after one hour at the final conditions. The protein in 3 g/l PEG throughout the experiment aggregated and recovered only 36% of its total activity compared to the direct dilution from 5 M GuHCl which results in 100% recovery of activity as shown in Figure 4.4. On the other hand, the solution in 30 g/l PEG did not aggregate and, after 30 minutes, recovered 88% of its total activity which is comparable to the direct dilution from 5 M GuHCl which results in 94% recovery of activity after 30 minutes (Figure 4.5). This solution also completely recovered activity after 1 hour. These results indicate that PEG prevents the formation of the associated state at equilibrium as well as during the refolding process.

6.2 Aggregate Distribution Dependence on PEG Concentration

To determine the relationship between the equilibrium distribution and the PEG concentration, equilibrium association studies were performed at 2.0 M GuHCl and 1.0 mg/ml CAB (33.3 μM) with several different PEG (8000 MW) concentrations. Denatured CAB in 5 M GuHCl was rapidly diluted to the final conditions with concentrated PEG solutions to yield final PEG concentrations from 1 to 60 g/l. Higher concentrations of PEG were not attempted since precipitation was observed at PEG concentrations greater

than 90 g/l. The solutions were allowed to equilibrate for three to six hours. After equilibration, the samples were analyzed by HPLC size exclusion with PEG and GuHCl at the same concentrations in the elution buffer (see Materials and Methods). The distribution of monomer, dimer, and trimer for each PEG concentration is shown in Figure 6.2.a. These results did not indicate the presence of a trimer species as observed by QLS analysis since the time period for HPLC analysis was longer than the equilibrium between the dimer and trimer species (Section 2.3). At PEG concentrations less than 30 g/l, the distribution shifts to the monomer. PEG concentrations greater than 30 g/l resulted in a shift in the distribution to the dimer species. These results are explained by the two mechanisms of PEG interaction. At low concentrations, PEG will interact with the first intermediate and prevent its association. PEG will be excluded from the protein surface (preferential hydration) at high concentrations and this exclusion will result in a return to association conditions.

To observe the apparent shift in the equilibrium distribution, the equilibrium constant for dimer formation was calculated and plotted as a function of the PEG concentration as shown in Figure 6.2.b. For this calculation, the overall monomer concentration was used instead of the first intermediate concentration since the first intermediate and PEG may bind to form a PEG-intermediate complex. The apparent equilibrium constant, K^P_D, was therefore defined to be:

$$K^P_D = [D]/[M]^2 \qquad (6.1)$$

where [M] and [D] are the concentrations of monomer and dimer as measured by HPLC

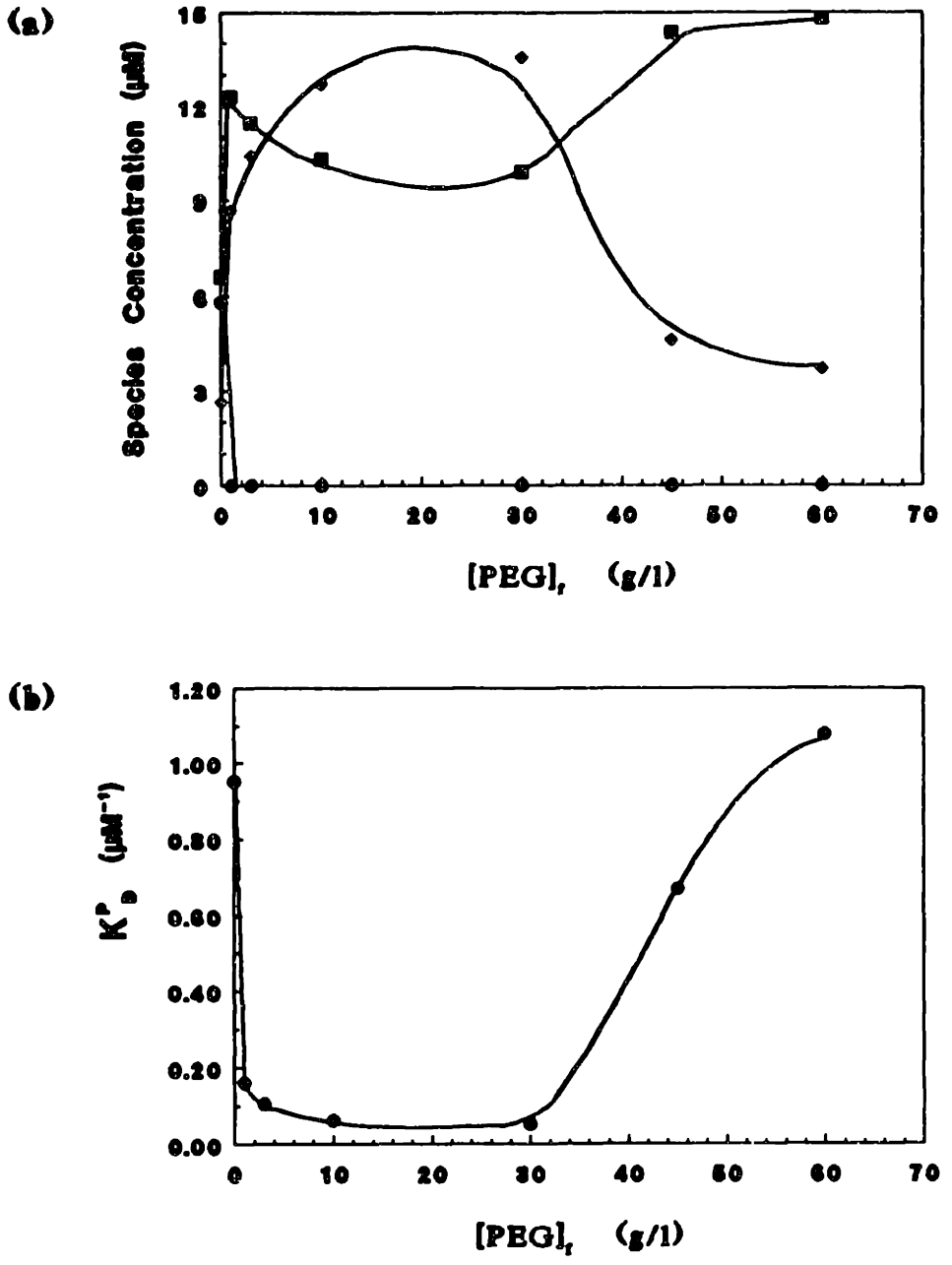


FIGURE 6.2: PEG concentration effect on equilibrium association of first intermediate. Equilibrium refolding experiments were performed by rapid dilution from 5 M GuHCl to 2 M GuHCl and 1.0 mg/ml (33.3 μM) CAB with different concentrations of PEG (8000 MW). (a) The concentration of monomer (♦), dimer (■) and trimer (●) was determined by HPLC analysis. (b) The apparent equilibrium constant for dimer formation, K^P_D, was calculated from the multimer concentration data in graph (a) by using Equation 6.1.

analysis (Figure 6.2.a). The equilibrium constant decreased dramatically with increasing PEG concentrations (Figure 6.2.b). This relationship confirmed the correlation between low PEG concentrations and reduced association as observed for the kinetic studies on refolding with PEG (Chapter 5).

Overall, equilibrium protein association studies in the presence of PEG have shown that the association of the first intermediate can be avoided at low PEG concentrations (< 30 g/l). PEG prevented the association of the first intermediate and the relationship between the final GuHCl concentration and the extent of association were not altered by PEG. Therefore, PEG did not form a new intermediate species which was more stable in GuHCl. These studies have shown that PEG will reduce association, but the mechanism of the reduction in association has not been demonstrated. It is conceivable that PEG will either cause the dissociation of the dimer species or prevent the association by binding to the protein surface.

Chapter 7: Mechanism of PEG Interaction with CAB during Refolding

Previous studies have shown that PEG reduced the association of the first intermediate during refolding and at equilibrium in 2.0 M GuHCl. Thus, PEG may act by either causing an increase in the dissociation of the dimer or a decrease in the association of the first intermediate by binding to the protein surface. Also, both the kinetic and equilibrium studies have shown that PEG will cause a reduction in association at stoichiometric molar ratios which are dependent on the PEG molecular weight. To distinguish between the two potential mechanisms of PEG interaction, dissociation or binding, several studies were performed to measure the interaction between PEG and the first intermediate species.

7.1 Equilibrium Binding Experiments

The potential mechanism of PEG binding to the first intermediate to prevent association was studied by equilibrium binding analyses. Previous studies using PEGpNP revealed that PEG may bind to the first intermediate during the refolding process (Chapter 5). However, it is necessary to prove conclusively that PEG will only bind to the first intermediate. Equilibrium binding experiments were therefore performed by using a hydrophobic interaction chromatography (HIC) column with PEG as the weak hydrophobic ligand attached to the column resin. The column was equilibrated at a given concentration of GuHCl and loaded with protein equilibrated at the same GuHCl concentration. The extent of retention was measured by elution with standard buffer without GuHCl or PEG as described in the Materials and Methods section (Figure 7.1). As shown in Figure 7.1, the maximum retention occurred at 2.0 M GuHCl. In

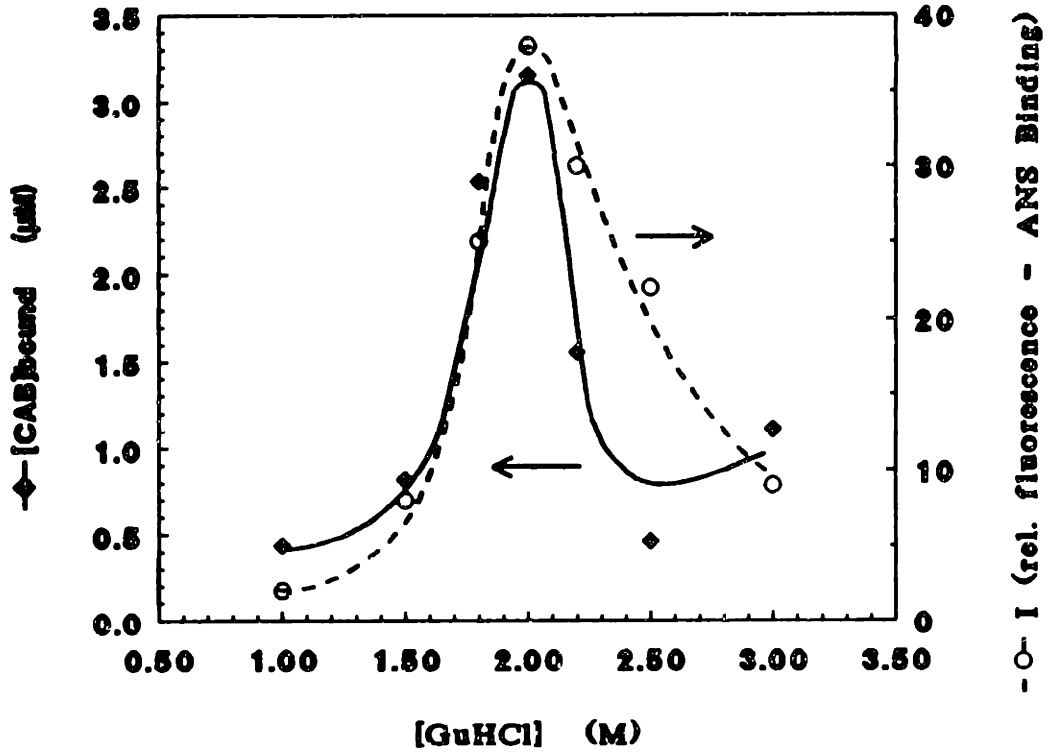


FIGURE 7.1: Equilibrium binding analysis was performed using a hydrophobic interaction chromatography column (HIC) with PEG as the hydrophobic ligand. The column was equilibrated at a given GuHCl concentration and the protein was loaded on the column (V_{load}= 25 µl) after equilibration at the same GuHCl concentration and 16.7 µM. The retention of CAB by the column was measured for each GuHCl concentration by elution with buffer (-♦-). Also, fluorescence of the hydrophobic probe, ANS, which binds to hydrophobic sections on CAB was plotted based on the previously reported equilibrium data (-O-) (Rodionova et.al., 1989).

comparison, the hydrophobic probe, 8-anilino-1-naphthalene sulfonate (ANS), bound to the greatest extent between 1.8 and 2.2 M GuHCl. As discussed previously, the binding of ANS has been shown to be concomitant with formation of the first intermediate in the refolding pathway (Rodionova et.al., 1989). Therefore, the first intermediate which is fully populated at 2 M GuHCl is retained by the PEG on the HIC column as shown in Figure 7.1. The second intermediate and native protein did not bind to the column and the extent of binding between the first intermediate and the column was relatively weak. Therefore, the first intermediate was the only species which bound to the PEG on the HIC column.

To confirm the binding of PEG to the first intermediate, equilibrium dialysis experiments were next performed. PEG at molecular weights ranging from 1000 to 8000 Daltons does not readily diffuse across dialysis membranes and the quantitative measurement of PEG concentration is also difficult. Therefore, a higher molecular weight PEG (5×10^6 MW) was used on one side a macroporous membrane dialysis cell which allowed diffusion of the protein without polymer diffusion (Abbott, 1990). PEG was dissolved at different concentrations in 2.0 M GuHCl and placed on one side of the membrane and CAB in 2 M GuHCl was added at 0.20 mg/ml (6.7 μM) to the other side of the cell. After equilibration for one week at 4 °C, the concentration of the protein on each side of the cell was assayed and the concentration of CAB bound to the polymer was calculated for several different molar concentrations of PEG (2 - 15 ηM) with a maximum mass concentration of 0.08 g/l (Figure 7.2). Control experiments showed that the native protein will not bind PEG. The equilibrium diffusion results along with the

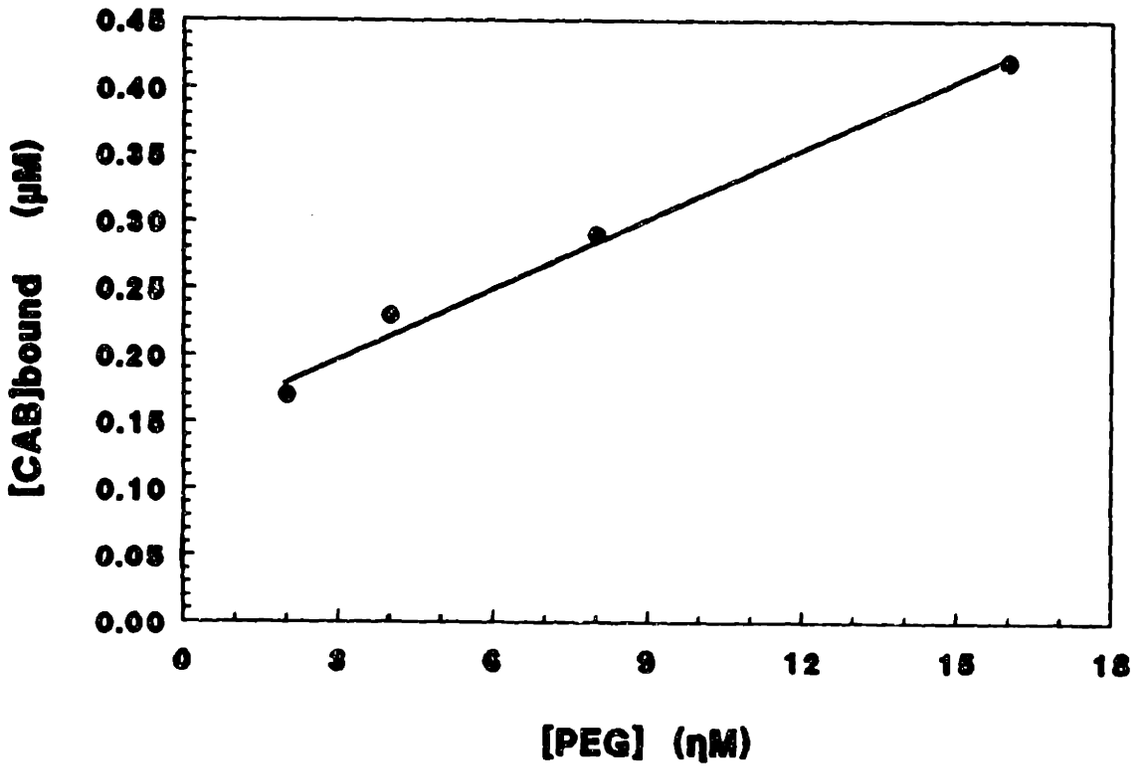


FIGURE 7.2: Equilibrium dialysis experiments were performed at 2.0 M GuHCl. A dialysis cell (5 ml) was used with a macroporous membrane (0.01 μm). The membrane allowed diffusion of CAB without diffusion of the 5 x 10^6 MW PEG (Abbott, 1990). PEG was placed in 2 M GuHCl on one side of the dialysis cell and CAB equilibrated at 2 M GuHCl and 6.7 μM protein was added to the o her side of the cell. Diffusion of CAB across the membrane was allowed to occur for on: week which was sufficient for equilibration. The final concentration of CAB on each side of the membrane was assayed and the concentration of CAB bound to the PEG was calculated for each different PEG concentration (2-15 ηM).

HIC results showed that the first intermediate is the only CAB species which will bind to PEG. Unfortunately, these binding studies did not provide insight into the specific binding of PEG to the first intermediate or the stoichiometry of the binding.

7.2 Fluorescence Studies of PEG Binding

The stoichiometry and strength of PEG binding to the first intermediate were determined by studying the changes in the intrinsic fluorescence of the protein. Since PEG binds to the first intermediate, the fluorescence properties of any exposed tryptophans on the surface should be altered as the result of this interaction. To determine if PEG will alter the tryptophan fluorescence of the protein, equilibrium refolding studies were performed by dilution to several different final GuHCl concentrations with standard buffer or with buffer containing PEG (8000 MW). After equilibration for three to six hours, the intrinsic tryptophan fluorescence of each solution was measured with an excitation wavelength of 296 ηm and an emission wavelength of 340 ηm (Rodionova et.al., 1989). Since CAB contains seven tryptophan residues with one exposed in the native state and six exposed in the first intermediate state, the change in fluorescence from the first intermediate to native state should be much greater than any variations caused by changes in the solvent (Yazgan & Henkens, 1972). Native CAB had the maximum fluorescence (F_o) and this value was used to normalize the other fluorescence readings, F/F_o, as shown in Figure 7.3. The equilibrium refolding fluorescence curve without PEG has a transition at 2 M GuHCl and samples analyzed at 0.1 M intervals from 1.5 to 2.5 M GuHCl indicate a slight plateau near 2 M GuHCl. When the final concentration of 30 g/l PEG (8000 MW) was used in the equilibrium

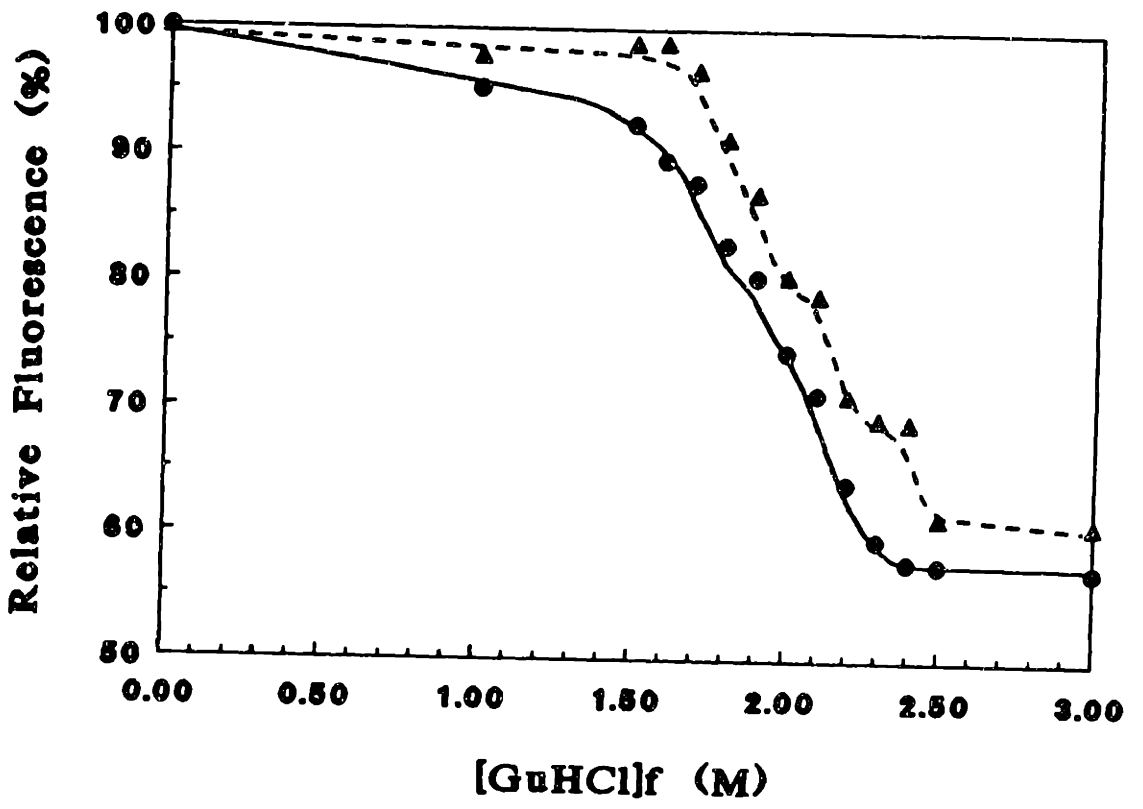


FIGURE 7.3: Fluorescence studies were performed for equilibrium refolding of CAB with and without PEG. The protein in 5 M GuHCl was diluted to a given final GuHCl concentration and 0.10 mg/ml (3.33 μM) CAB. After equilibration for four to six hours, the tryptophan fluorescence of each solution was measured at an excitation wavelength of 296 ηm and an emission wavelength of 340 ηm. The fluorescence of the native protein in buffer was used to normalize the fluorescence data (F_{GuHCl}/F_{native} *100%). This relative fluorescence was plotted as a function of the GuHCl concentration for conditions without PEG (●) and with 30 g/l PEG (8000 MW, ▲).

refolding solutions, the relative fluorescence (F/F_o) increased for samples with GuHCl concentrations greater than 1.0 M. The native protein did not increase in fluorescence in the presence of PEG and control experiments using tryptophan did not indicate a significant increase in fluorescence due to viscosity (< 5 % change for 90 g/l PEG). The equilibrium refolding curve in the presence of PEG has slight plateaus at 2.0 and 2.2 M GuHCl which may indicate the stabilization of the first intermediate species. In addition, the fluorescence curve with 30 g/l PEG showed the greatest relative deviation from the fluorescence curve without PEG at 2.0 M GuHCl. The shift to a higher fluorescence in 30 g/l PEG was also observed in the fluorescence emission spectra in the region between 310 and 370 ηm (see Appendix for fluorescence spectra).

These equilibrium refolding results demonstrated that PEG when bound to the protein will cause in an increase in the intrinsic fluorescence of the protein. Therefore, intrinsic fluorescence was used to measure the binding between PEG and the first intermediate. Equilibrium binding studies were performed by diluting CAB in 5 M GuHCl to 2.0 M GuHCl with concentrated PEG solutions. These equilibrium studies were performed at several different final PEG concentrations (0.1 to 90 g/l) and a low protein concentration (0.10 mg/ml, 3.33 μM) where protein association does not occur (Section 2.2). The fluorescence of each solution was measured after equilibration for six hours and the relative fluorescence increase was calculated from the difference in fluorescence between the sample with PEG, F, and the sample without PEG, F_o. A significant increase in the relative fluorescence of the first intermediate was observed as shown in Figure 7.4. The binding of PEG to the protein was calculated as described

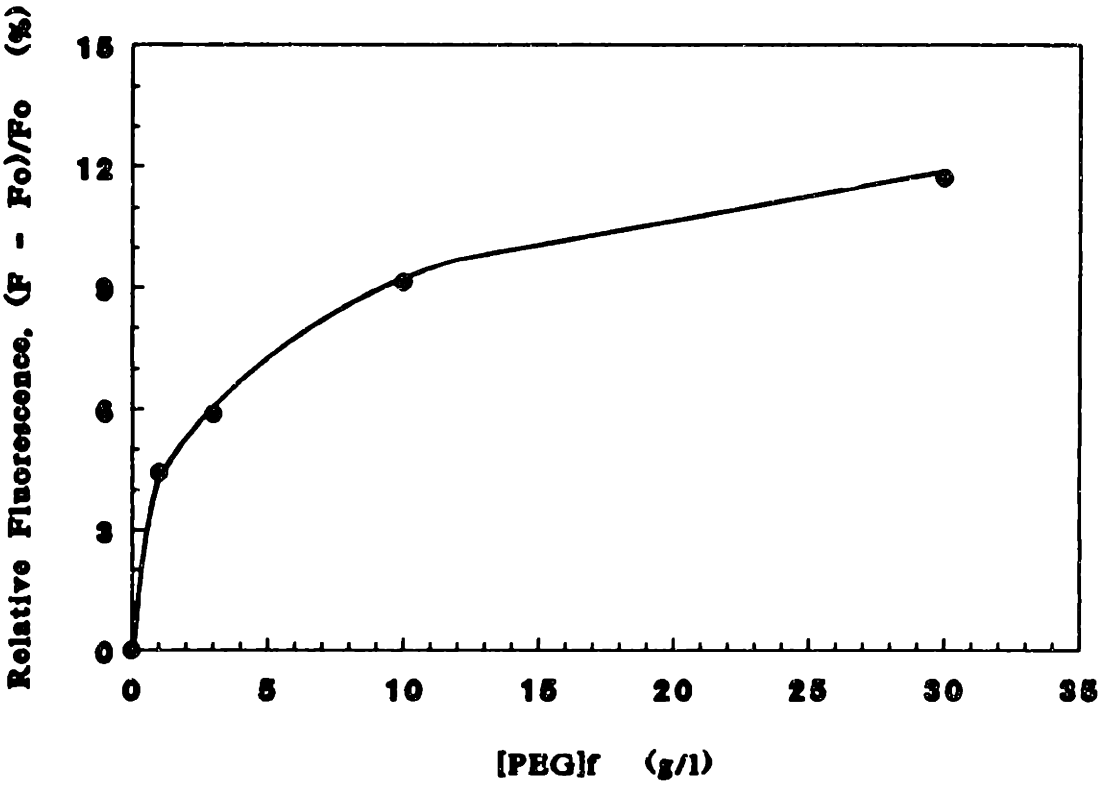


FIGURE 7.4: The effect of PEG on the tryptophan fluorescence of CAB in 2 M GuHCl was determined for several different PEG (8000 MW) concentrations. The fluorescence was measured at the same excitation and emission settings as in Figure 7.3 for CAB equilibrated at 2 M GuHCl and 0.10 mg/ml (3.33 μM) protein with PEG. Relative fluorescence in this case was defined as the increase in fluorescence caused by the presence of PEG where the sample with PEG has a fluorescence, F, and the sample without PEG has a fluorescence, F_o. The fluorescence of the native protein and tryptophan control samples increased by a maximum of 5% at PEG concentrations greater than 60 g/l.

previously with the assumption that PEG acts as the ligand (L) (Pesce et.al., 1971). Scatchard analysis revealed that the protein has one binding site for PEG as shown by the dashed line in Figure 7.5. In addition, the association constant for the binding site was 1.88×10^{-3} M^{-1}. Other additional binding sites may exist as indicated by the plateau shown in Figure 7.5. The observed plateau confirmed the previous results which indicated that several PEG molecules can bind to the surface and reduce aggregation (Chapter 5). The binding of PEG to the first intermediate was a weak association process involving one primary site on the protein surface.

Since PEG bound to the first intermediate and altered the intrinsic tryptophan fluorescence of the protein, a large quenching agent such as iodide (I^-) should be prevented from interacting with the tryptophans on the protein surface in the presence of PEG. To verify this hypothesis, equilibrium refolding was again performed at several different final GuHCl concentrations with and without 30 g/l PEG. The quenching of tryptophan fluorescence was then measured in the presence of 100 mM KI and the ratio of fluorescence without quenching agent to fluorescence with quenching agent (F_o/F) was calculated as described in Materials and Methods (Figure 7.6). As shown in Figure 7.6, the maximum fluorescence quenching occurred at 2 M GuHCl and decreased somewhat at higher GuHCl concentrations. The decrease in fluorescence beyond 2 M GuHCl may be the result of both the high Gu^+ concentration which could complex with free iodide ions and the formation of a random coil structure with fewer hydrophobic centers since this decrease was not observed in acrylamide quenching experiments (see Appendix). The presence of PEG at 30 g/l significantly decreases the quenching at 2 M GuHCl. Stern-

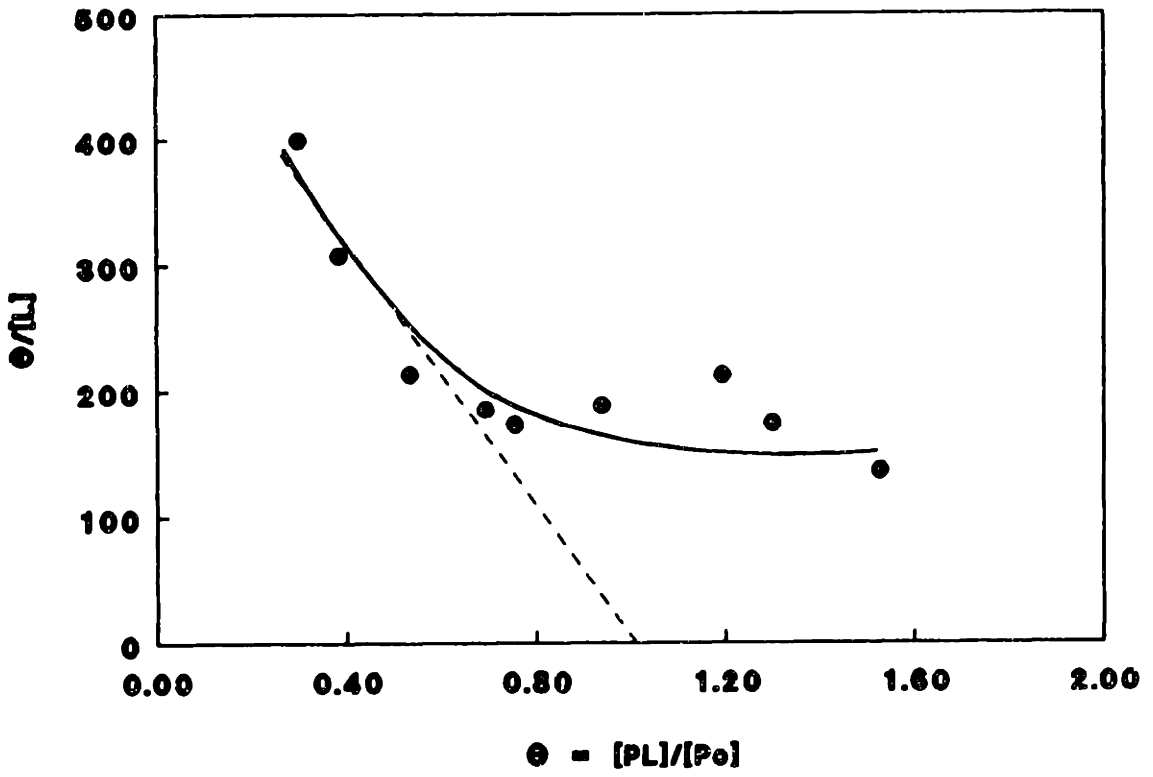


FIGURE 7.5: Binding analysis was performed by using the PEG fluorescence data in Figure 7.4 as well as additional measurements taken at several other PEG concentrations. The concentration of ligand (PEG) bound to protein divided by the total protein concentration (Θ = [PL]/[P_o]) and the free ligand concentration ([L]) were calculated as described previously (Pesce et.al., 1971). The Scatchard plot shown indicates that there is one primary binding site for PEG with an the equilibrium association constant (K_a) of 1.88 x 10^{-3} M^{-1}.

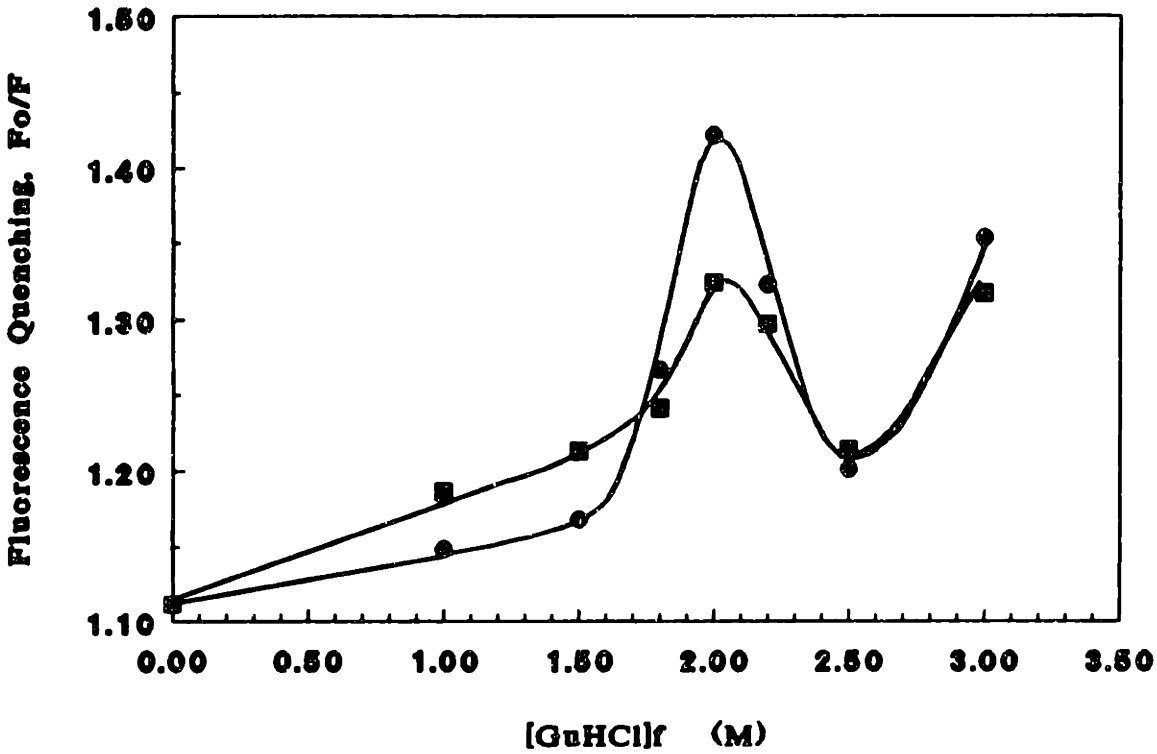


FIGURE 7.6: Quenching of CAB equilibrium refolding solutions was done through the addition of iodide. CAB in 5 M GuHCl was diluted to a given final GuHCl concentration and 0.10 mg/ml protein. Concentrated KI was added after equilibration to give a final concentration of 100 mM. The tryptophan fluorescence (λ_{ex} = 296 ηm, λ_{em} = 340 ηm) for each solution was then measured and the fluorescence quenching was calculated as the fluorescence without iodide, F_o, divided by the fluorescence with iodide, F. The fluorescence quenching for each sample was then plotted as a function of the final GuHCl concentration without PEG (●) or with 30 g/l PEG (8000 MW, ■).

Volmer analysis of the relative quenching in the presence of different PEG concentrations at 2 M GuHCl indicated a decrease in the static quenching constant with increasing PEG concentration. Therefore, PEG bound to the surface of the first intermediate and inhibited the quenching ability of iodide.

Fluorescence studies have shown that PEG will bind to one site on the first intermediate in 2.0 M GuHCl. When PEG was bound to this site, the intrinsic fluorescence of the first intermediate increased and quenching of the six surface tryptophans by iodide was significantly reduced. As PEG bound to the surface, it may have covered the exposed tryptophan residues preventing quenching and increasing the fluorescence. Alternatively, the binding of PEG to the first intermediate may result in a slight conformational change to a structure which has different fluorescence properties. It was not possible to distinguish between these two different phenomena with the fluorescence analysis. In addition, the fluorescence results did not indicate the specific site of the interaction between PEG and the first intermediate and the plateau in the Scatchard analysis as shown in Figure 7.5 revealed the potential for more than one PEG binding to the surface.

7.3 Electron Spin Resonance (ESR) Study of PEG Association

To distinguish between a change in conformational state and binding of PEG to the protein surface, electron spin resonance studies of PEG binding were performed using CAB which was spin labelled at tyrosine residues that are exposed in 2.0 M GuHCl (see Materials and Methods for labelling procedure). After labelling, CAB was found to contain an average of three spin labels per protein. When refolded by dilution from 5 M

GuHCl to 1.0 M GuHCl and 0.50 mg/ml protein, spin-labelled CAB (spCAB) recovered activity to the same extent and at the same rate as the control case without spin labelling as described in Materials and Methods. To determine if spCAB would provide additional information on the surface properties of the first intermediate, equilibrium refolding was performed by diluting spCAB in 5 M GuHCl to 2.0 M GuHCl and 0.10 mg/ml (3.33 μM) protein. After equilibration for six hours, the solution was analyzed by electron spin resonance (ESR) to measure the local environment of the spin labels on the protein surface (see Literature Review). The ESR analysis revealed two overlapping spectra which represented a combination of different label environments (see Materials and Methods). Two of the labels on the protein surface were interacting in close proximity to each other in a nonpolar environment. The third label was freely rotating on the surface of the first intermediate. The freely rotating label on the protein surface could act as a reporter of the local environment on the protein surface. Therefore, the single freely rotating label on the surface of spCAB in the first intermediate state was used to measure the effect of PEG binding.

Equilibrium refolding studies were conducted by diluting spCAB in 5 M GuHCl to 2.0 M GuHCl and 0.10 mg/ml (3.33 μM) protein with different concentrations of PEG (8000 MW). After equilibration for six hours, each solution was analyzed by ESR to determine the change in the freely rotating spin label environment. The freely rotating spin label moved into the environment of the other spin labels as the PEG concentration increased from 0.10 to 60 g/l. The decrease in the amount of freely rotating spin label on the protein surface was normalized to the conditions without PEG which was

considered to represent 100% freely rotating spin label. The normalized fraction of freely rotating spin label on the surface of the first intermediate was then plotted as a function of the PEG concentration as shown in Figure 7.7. The amount of freely rotating spin label on the protein surface decreases dramatically with increasing PEG concentration and apparently reaches a minimum at PEG concentrations greater than 30 g/l. The addition of PEG at low concentrations must therefore result in a change surface properties of the first intermediate such that the freely rotating spin label becomes constrained to the same environment as the other labels.

To understand this specific interaction and measure the stoichiometry for binding of PEG to the first intermediate, the spin labelled polymer, methoxy polyethylene glycol aminocarbonyl proxyl (spPEG), was synthesized as described in Materials and Methods. spPEG was then used to study the association between PEG and CAB. CAB in 5 M GuHCl was diluted to 2.0 M GuHCl and 3.33 μM protein with different concentrations of spPEG. The electron spin resonance of each sample was measured at equilibrium to study the change in rotational diffusivity of spPEG. The rotational diffusivity of spPEG without protein in solution was 2.08×10^9 sec^{-1} since the polymer was a random coil in this solution. As the ratio of protein to spPEG ($[CAB]_f/[spPEG]_f$) was increased, the rate of rotation of the polymer, rotational diffusivity, decreased through a binding of the polymer to the protein surface as shown in Figure 7.8. Any excess polymer which is not bound to the protein will have the diffusivity of the freely rotating spPEG (2.08×10^9 sec^{-1}). Therefore, the rotational diffusivity of spPEG measured by ESR was an average of the molecules bound to the protein and free in solution. In addition, when excess spPEG

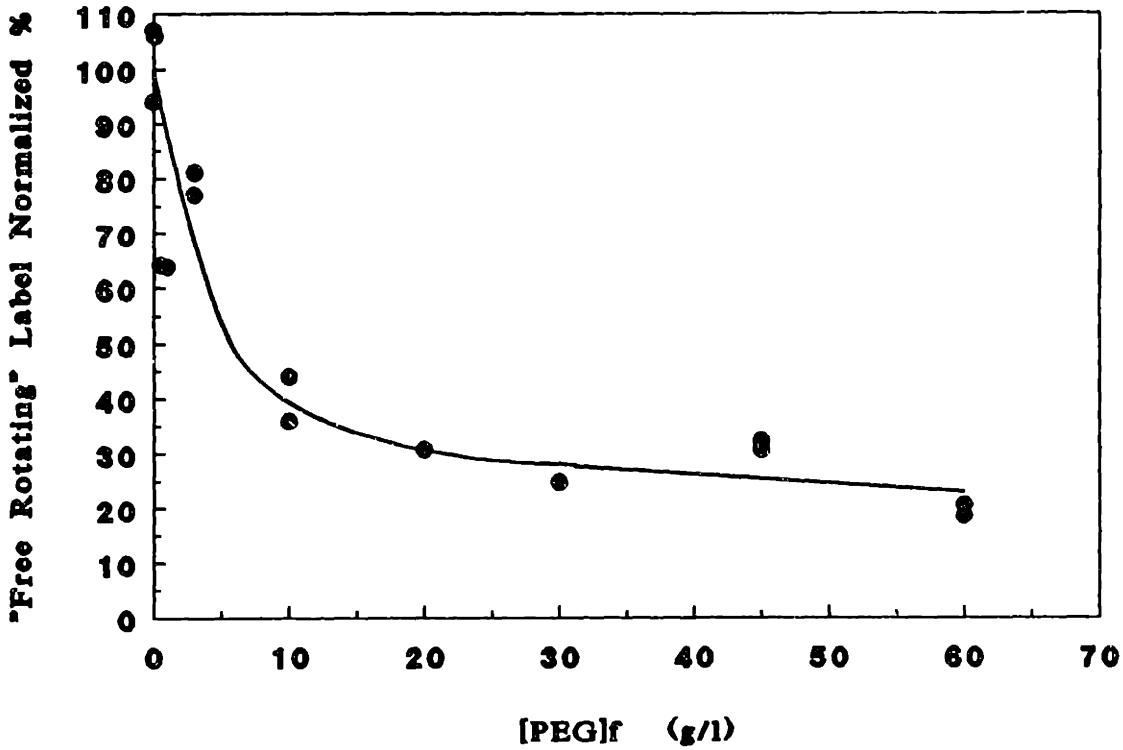


FIGURE 7.7: The relative fraction of freely rotating spin label on spCAB was calculated as described in the text. Equilibrium refolding was performed by rapid dilution from 5 M GuHCl to 2 M GuHCl and 0.10 mg/ml CAB with different concentrations of PEG (8000 MW). The spin resonance of each sample was measured after equilibration for six hours.

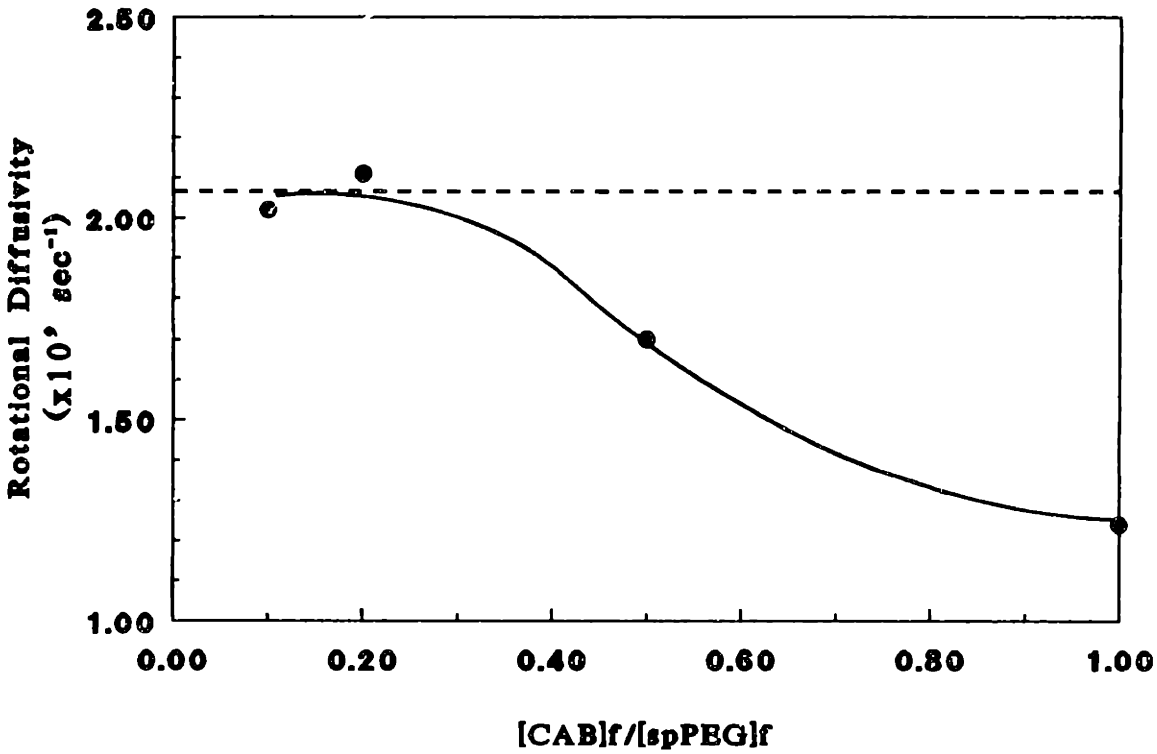


FIGURE 7.8: Electron spin resonance of spin labelled PEG (spPEG) was measured as a function of the molar ratio of spPEG to CAB. Equilibrium refolding was performed at 0.10 mg/ml (3.33 μM) protein and 2 M GuHCl at different molar ratios ($[CAB]_f/[spPEG]_f$). The rotational diffusivity of spPEG was calculated as described in Materials and Methods and plotted as a function of the molar ratio. The rotational diffusivity spPEG in 2 M GuHCl without CAB was 2.08 x 10^9 sec^{-1} as shown by the dashed line.

was added to the solution, the free spPEG dominated the ESR spectra resulting in a rotational diffusivity which was equivalent only to the free spPEG in solution as observed at low $[CAB]_f/[spPEG]_f$ (Figures 7.8). These results clearly indicated that spPEG will bind to the first intermediate in solution resulting in a decrease in its rotational diffusivity with an observed minimum of 1.25×10^9 sec^{-1} for a molar ratio of one spPEG molecule per CAB. This equilibrium study confirmed the fluorescence binding results which indicated that the first intermediate in the CAB refolding pathway has one primary binding site for PEG.

The ESR studies have revealed that the binding of PEG to the first intermediate occurred at a single binding site. In addition, studies with spCAB provided insight into the specific changes in the surface characteristics of the first intermediate upon binding of PEG. Therefore, PEG will bind to a single site on the first intermediate and alter its surface properties such that it will not associate to form multimers.

Chapter 8: Model Pathway for PEG Enhanced Refolding of CAB

8.1 Proposed Pathway for PEG Enhanced Refolding of CAB

To develop a pathway for PEG enhanced refolding of CAB, the results of several different analyses must be combined to provide a complete and accurate model. First of all, kinetic studies on the refolding and aggregation of CAB in the presence of PEG demonstrated that low molar ratios of PEG to CAB resulted in an increased recovery of active protein and a decrease in multimer formation (Chapter 5). These studies have also shown that low concentrations of PEG prevented the formation of a transient dimer species during refolding at high protein concentrations (> 10 μM) and 1.0 M GuHCl (Chapters 3 and 5). The ability of PEG to prevent association of the first intermediate was further confirmed by equilibrium studies at high protein concentrations (> 10 μM) and 2.0 M GuHCl. In these equilibrium studies, PEG (8000 MW) at low concentrations (≤ 30 g/l) greatly reduced the association of the first intermediate (Chapter 6). Therefore, the pathway for PEG enhanced refolding would include a reduction in association of the first intermediate at equilibrium and during refolding. The reduction could occur by a decrease in the dimer association rate constant, k_{r_i}, or an increase in the dimer dissociation rate constant, k'_d, as shown in Figure 1.10. Alternatively, the first intermediate could be shifted from the normal refolding pathway to a stable PEG-first intermediate complex which does not associate.

To determine the mechanism of this reduced association, the interaction of PEG with the first intermediate was determined. Equilibrium binding experiments revealed that PEG only bound to the first intermediate in the refolding pathway (Section 7.1). In

addition, it was shown that the first intermediate had one specific binding site for PEG (Sections 7.2 and 7.3). The binding of PEG to this site on the first intermediate resulted in changes in the properties of the intermediate (Sections 7.2 and 7.3). Furthermore, these studies indicated that the surface properties of the first intermediate were significantly altered from the binding of PEG. Therefore, the proposed model for refolding should include the binding of PEG to the surface of the first intermediate with concomitant changes in its surface properties such that it can not associate to form dimers or trimers.

To determine if the binding of PEG to the first intermediate resulted in an increased rate of refolding, refolding experiments were performed by rapid dilution of CAB in 5 M GuHCl to 1.0 M GuHCl and several different final protein concentrations with a final PEG (3350 MW) to protein molar ratio ($[PEG]_f/[CAB]_f$) of 3 to 1 for each case. The recovery of active protein was then measured after dilution as shown in Figure 8.1. The rate of refolding was the same for each case and was, therefore, independent of the protein concentration. Previous refolding studies at these conditions without PEG resulted in a decreased rate of refolding with increasing protein concentration as the result of transient dimer formation (Section 1.3, Figure 1.4). As a control, refolding was also performed by dilution of CAB in 5 M GuHCl to 1.0 M GuHCl and 0.10 mg/ml (3.33 μM) which are the conditions where association does not occur (Section 1.3). The rate of refolding for this control case was the same as the rate of refolding at the higher protein concentrations with PEG as shown in Figure 8.1. Therefore, PEG did not increase the rate of refolding, but it did prevent the transient association of first intermediate as observed previously (Chapter 5). The proposed pathway should then proceed from the

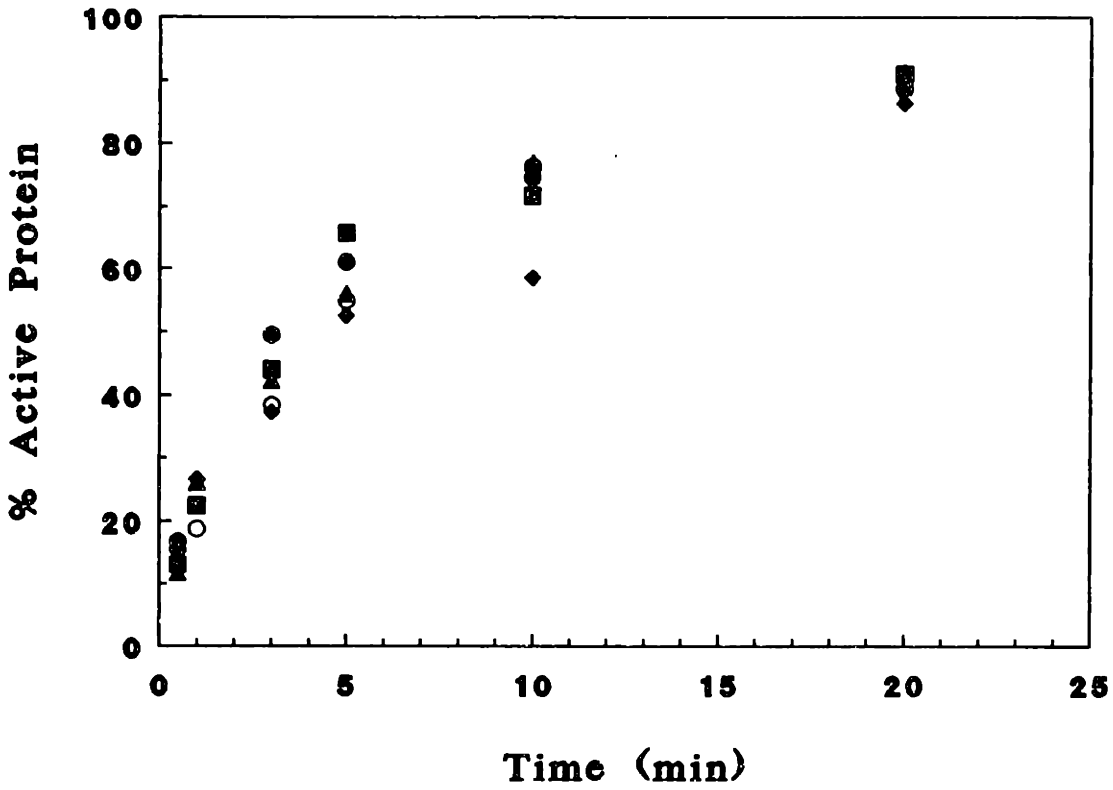


FIGURE 8.1: PEG enhanced refolding was performed at 1.0 M GuHCl and different protein concentrations at a PEG (3350 MW) molar ratio ($[PEG]_f/[CAB]_f$) of 3. The recovery of active protein is plotted as a function of time for each final protein concentration: 0.10 (●), 0.35 (■), 0.50 (▲), and 1.0 (♦) mg/ml. In addition, the recovery of active protein was plotted for refolding at 0.10 mg/ml and 1.0 M GuHCl (○) without PEG.

PEG-first intermediate complex to the second intermediate at the same rate as the refolding in the absence of aggregation.

To further confirm that the refolding rate from the first intermediate to the second intermediate did not increase in the presence of PEG, the rate constant for the formation of second intermediate was measured by the absorbance analysis technique described in Chapter 3 and this analysis was completely independent of the previous activity study. To determine the rate constant for second intermediate formation, refolding from 5 M GuHCl to 1.0 M GuHCl and 0.50 mg/ml (16.7 μM) protein was performed with PEG (3350 MW) in the dilution buffer at a final concentration of 0.17 g/l ($[CAB]_f/[PEG]_f$= 3). After dilution, the refolding was measured by the change in absorbance at 280 ηm as shown in Figure 8.2. As observed for the previous absorbance kinetic study (Figure 3.1), the change in absorbance could be modelled by two exponential phases. The rate constant for second exponential was 7.76 x 10^{-2} min^{-1} ($t_{1/2}$= 531 sec) which is within experimental error of the previously measured rate constant (k_n = 7.24 x 10^{-2} min^{-1}, Figure 3.1). The rate constant for the first exponential was the same as that determined for refolding at low protein concentrations (< 10 μM) and 1.0 M GuHCl (k_{i2} = 1.39 min^{-1}, $t_{1/2}$= 30 sec). Refolding in the presence of PEG must therefore proceed through a nonassociating first intermediate structure which folds to the second intermediate at the same rate.

To match each of the observed phenomena of PEG enhanced refolding of CAB, the pathway shown in Figure 8.3 was developed. When diluted to low GuHCl concentrations (≤ 2.0 M), the unfolded protein, U, in 5 M GuHCl will rapidly form the

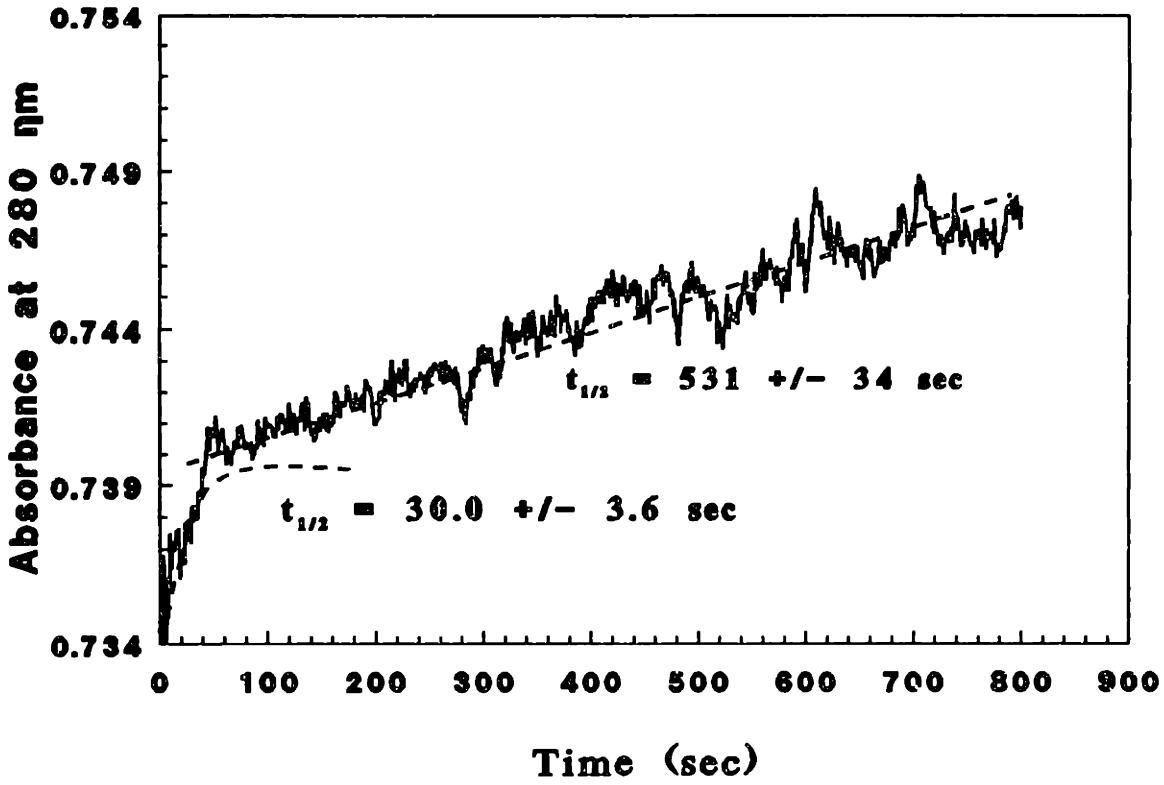


FIGURE 8.2: Refolding kinetics measured by absorbance at 280 ηm. Absorbance measurements were made within 10 seconds of dilution from 5 M GuHCl to 1 M GuHCl and 0.50 mg/ml (16.7 μM) CAB with a final PEG (3350 MW) concentration of 0.17 g/l (50 μM). The curve can be modelled by two exponentials with half times of 30 and 531 seconds as shown (Semisotnov et.al., 1990).

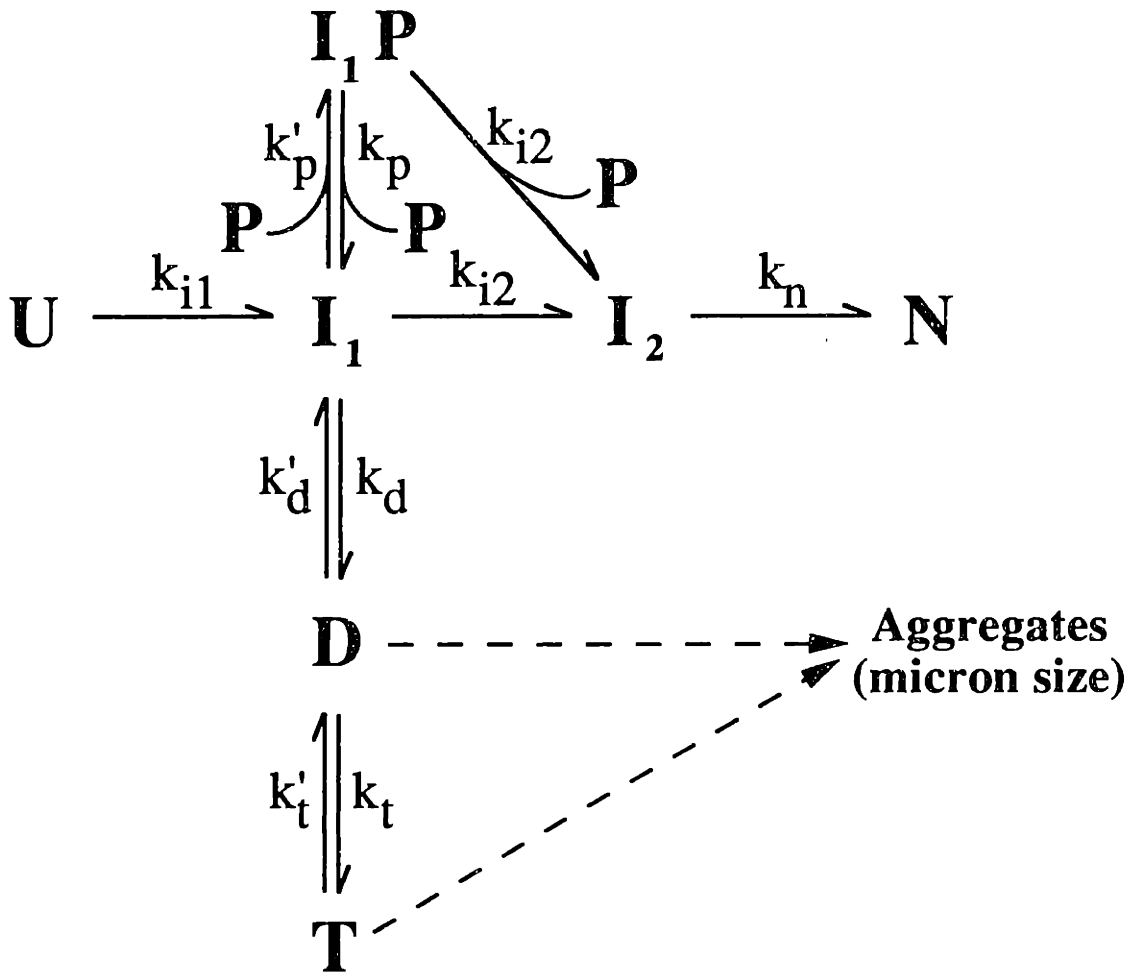


FIGURE 8.3: Model for PEG enhanced refolding of CAB. Upon rapid dilution from 5 M GuHCl, the unfolded protein (U) rapidly forms the first intermediate (I_1). This intermediate will then either proceed through the normal refolding and aggregation pathways (Figure 1.10) or it will bind to PEG, P, to form a nonassociating complex (I_1P). The refolding of the PEG-intermediate complex (I_1P) to the second intermediate occurs at the same rate as that observed for refolding without PEG at low protein concentrations (< 10 μM) and 1.0 M GuHCl.

first intermediate, I_1. The first intermediate may then proceed through three different pathways. Since PEG has been observed to bind to the first intermediate and prevent the association of this intermediate, a PEG-first intermediate complex, I_1P, which does not associate must be formed during the refolding process. Also, the surface properties of the first intermediate were altered in the presence of PEG as observed by fluorescence and ESR studies (Chapter 7). Therefore, the PEG-intermediate complex is distinct from the first intermediate which may also associate to form the dimer, D, or refold to the second intermediate, I_2. Since the rate of refolding to the second intermediate with PEG is the same as that observed for refolding at low protein concentrations (< 10 μM) in 1.0 M GuHCl, the folding of the PEG-first intermediate complex to the second intermediate has the same refolding rate constant (k_{i2}). The number of PEG molecules required to form the PEG-intermediate complex has been shown to be dependent upon the molecular weight of the polymer and, unlike the association reaction, the molar ratio of PEG to protein ($[PEG]_f/[CAB]_f$) required for enhancement was independent of the final GuHCl concentration (Chapter 5). Since PEG was required in at least a two fold molar excess for each PEG molecular weight (Chapter 5) and the equilibrium constant for dimer formation was significantly reduced in the presence of low PEG concentrations (Chapter 6), the equilibrium between the first intermediate and the PEG-intermediate complex must be shifted toward the nonassociating complex. The rate at which the nonassociating complex forms must be more rapid that the rate of association ($k'_p > k_d$) since the dimer was not observed for refolding in PEG at conditions where the transient dimer formed or conditions where irreversible association occurred (Chapter 5). Overall, this pathway

represents the rapid equilibration with PEG to form the nonassociating species, I_1P, followed by the folding of this species to the second intermediate at the same refolding rate as that observed in the absence of association.

8.2 Mathematical Model of Pathway and Refolding Rates

Kinetic and equilibrium refolding studies which were performed with PEG have shown that multimers will not form at aggregating conditions. Therefore, the rate of formation of the PEG-intermediate complex must be much greater than the rate of dimer formation for the final conditions studied in this research ($k'_p >> k_d$). The rate expression for the PEG-intermediate complex may then be written as:

$$\frac{d[I_1P]}{dt} = k'_p\,[I_1][P] - k_p\,[I_1P] - k_{i2}\,[I_1P] \qquad (8.1)$$

and the equilibrium between the complex and the first intermediate would be represented by:

$$K_{I_1P} = \frac{[I_1P]}{[I_1][P]} \qquad (8.2)$$

With this equilibrium expression, Equation 8.1 may be written as:

$$\frac{d[I_1P]}{dt} = \frac{k'_p\,[I_1P]}{K_{I_1P}} - k_p\,[I_1P] - k_{i2}\,[I_1P] \qquad (8.3)$$

The PEG-intermediate complex equilibrium must be more rapid than the equilibrium association to avoid the formation of dimers ($K_{I_1P} >> K_D$). The first intermediate will therefore quickly proceed to form the PEG-first intermediate complex as shown in Figure 8.3. The initial concentration of this complex would then be equivalent to the total protein concentration, $[CAB]_f$. Since the rate of formation of the complex must be rapid relative to the rate of second intermediate formation, the change in concentration of this complex with time can be expressed as:

$$\frac{d[I_1P]}{dt} = -k_{i2}\,[I_1P] \tag{8.4}$$

The concentration profile for the intermediate complex can then be written as:

$$[I_1P] = [CAB]_f \exp(-k_{i2}\,t) \tag{8.5}$$

Since the complex forms very rapidly, the concentration of the first intermediate which folds to form the second intermediate is negligible. The rate expression for the second intermediate could then be expressed as:

$$\frac{d[I_2]}{dt} = k_{i2}\,[I_1\,P] - k_n\,[I_2] \tag{8.6}$$

and the rate expression for the formation of the native protein is described by Equation 3.9. Using Equations 3.9, 8.5, and 8.6, the concentration profiles for the second

intermediate and the native protein were determined to be:

$$[I_2] = \left(\frac{k_{i2}[CAB]_f}{k_n - k_{i2}}\right)\{\exp(-k_{i2}t) - \exp(-k_n t)\} \qquad (8.7)$$

$$[N] = [CAB]_f\left\{\left(\frac{k_{i2}}{k_n - k_{i2}}\right)\exp(-k_n t) - \left(\frac{k_n}{k_n - k_{i2}}\right)\exp(-k_{i2}t) + 1\right\} \qquad (8.8)$$

Equations 8.5, 8.7, and 8.8 completely the describe the refolding pathway shown in Figure 8.3 and can be applied to all cases where the maximum PEG enhancement has been observed. Thus, this model can be applied to conditions where PEG prevents the formation of an associated species. If an associated species does form, the rates of both the association and the PEG-intermediate complex formation must be determined to derive an accurate model.

The concentration of each protein species can be calculated based on this mathematical model for refolding with PEG. When PEG enhanced refolding of CAB is performed at the optimum molar ratios ($[PEG]_f/[CAB]_f$) as described in Chapter 5, the formation of associated species did not occur. Therefore, refolding at these conditions can be modelled by the equations derived above. Equations 8.5, 8.7 and 8.8 were therefore used along with the rate constants derived from Figure 8.2 (k_{i2} = 1.39 min^{-1}, k_n = 7.76 x 10^{-2} min^{-1}) to model the refolding results shown in Figure 8.1. It should be emphasized that the rate constants obtained from Figure 8.2 are derived by independent

experiments which did not rely on the measurement of active protein. The model prediction as shown in Figure 8.4 matched the experimental active protein concentration to within the error of each measurement (± 5 %). Therefore, the assumption of the rapid formation of the PEG-intermediate complex, I_1P, was valid for these conditions. To further verify this assumption, the kinetics of complex formation would have to be determined. In addition, the assumption that the refolding of the complex to the second intermediate occurs at the same rate as the refolding without association was also shown to be valid for these conditions. In conclusion, the model depicted in Figure 8.3 can be described mathematically with Equations 8.5, 8.7, and 8.8 for conditions where PEG prevents the formation of multimers.

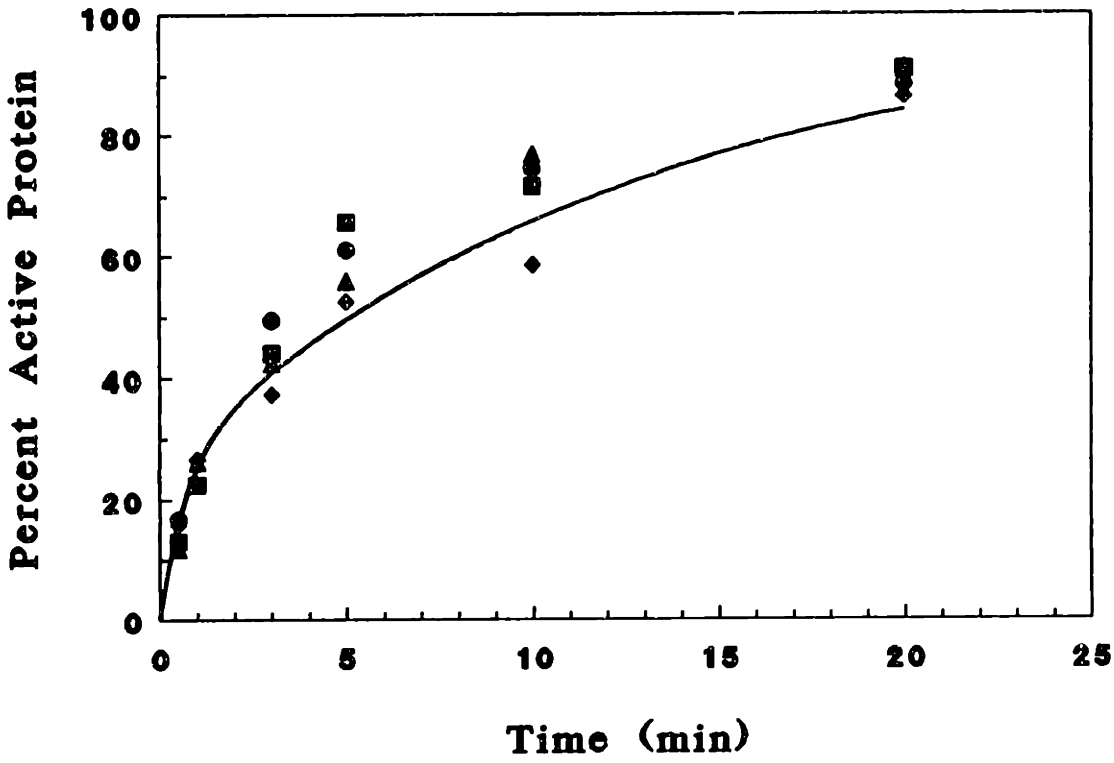


FIGURE 8.4: Model prediction of the recovery of active protein. The data from Figure 8.1 are plotted along with the predicted concentration of active protein which was calculated from Equation 8.8 (--). The data from Figure 8.1 represent refolding at 1.0 M GuHCl and a PEG molar ratio ($[PEG]_f/[CAB]_f$) of 3 for each protein concentration: 0.10 (●), 0.35 (■), 0.50 (▲), and 1.0 (♦) mg/ml.

Chapter 9: Extension of PEG Enhanced Refolding of Other Proteins

Since PEG prevents association of the hydrophobic first intermediate in the refolding pathway of CAB, it may also prevent the aggregation of intermediates during the refolding of other proteins. The aggregation of a hydrophobic intermediate during refolding has also been observed for bovine growth hormone as well as other proteins (Brems, 1988b; Mitraki & King, 1989). In addition, Ptitsyn and coworkers have recently postulated that a molten globule structure such as the first intermediate in CAB refolding may be common to many proteins independent of class or composition (Ptitsyn et.al., 1990). Thus, PEG could prevent aggregation of hydrophobic molten globule intermediates which are formed during the refolding of other proteins.

To assess the general applicability of PEG enhanced refolding, several other proteins were studied. These results should provide additional insights into the mechanism of PEG interaction and the general applicability of this technique. Unfortunately, the proteins discussed below are not well characterized in their refolding pathways. Therefore, a study of their refolding properties has also been included to improve the understanding of the PEG enhanced refolding results.

9.1 Glycosylated Recombinant Human Deoxyribonuclease (rhDNAse)

The general ability of PEG to enhance refolding was assessed by using recombinant human deoxyribonuclease (rhDNAse). Since the bovine analog of this protein has a hydrophobic beta sheet core similar to that of CAB, it may also have a hydrophobic intermediate on its refolding pathway (Suck et.al., 1984; Liljas et.al., 1972). To determine if a hydrophobic species exists on the refolding pathway for rhDNAse,

equilibrium refolding experiments were performed at different final denaturant and protein concentrations. The change in the protein properties and the aggregation behavior of the protein at each final condition was then measured to ascertain the presence of a hydrophobic intermediate.

Initial characterization of the glycosylated rhDNAse was performed by equilibrium refolding studies in GuHCl for comparison with the refolding of CAB. Since rhDNAse contains 16 tyrosines and 4 tryptophans in its polypeptide sequence, the refolding of this protein can be studied by analyzing the effect of environmental changes on these aromatic residues through fluorescence spectroscopy. In particular, the fluorescence measured at an excitation wavelength of 280 ηm and the emission wavelength of 340 ηm will provide insight into the changes occurring at both the tyrosine and tryptophan residues. On the other hand, the fluorescence measured at an excitation wavelength of 296 ηm with the same emission wavelength will primarily reveal the change in the environment around the tryptophan residues. To study the change in fluorescence during refolding, equilibrium refolding studies were performed by rapidly diluting denatured rhDNAse in 7 M GuHCl to different final GuHCl concentrations at a final protein concentration of 10 μg/ml. The refolding was performed by maintaining the protein at either oxidizing or reducing (50 mM β-mercaptoethanol) conditions. After equilibration for 3 to 6 hours, the fluorescence of each refolding solution was measured by excitation at either 280 ηm or 296 ηm and emission at 340 ηm. As shown in Figure 9.1, both the oxidized and reduced forms had a significant change in fluorescence at 2.0 M GuHCl and a slight plateau occurred between 2.0 and 2.5 M GuHCl. Previous studies on the denaturation of bovine DNAse

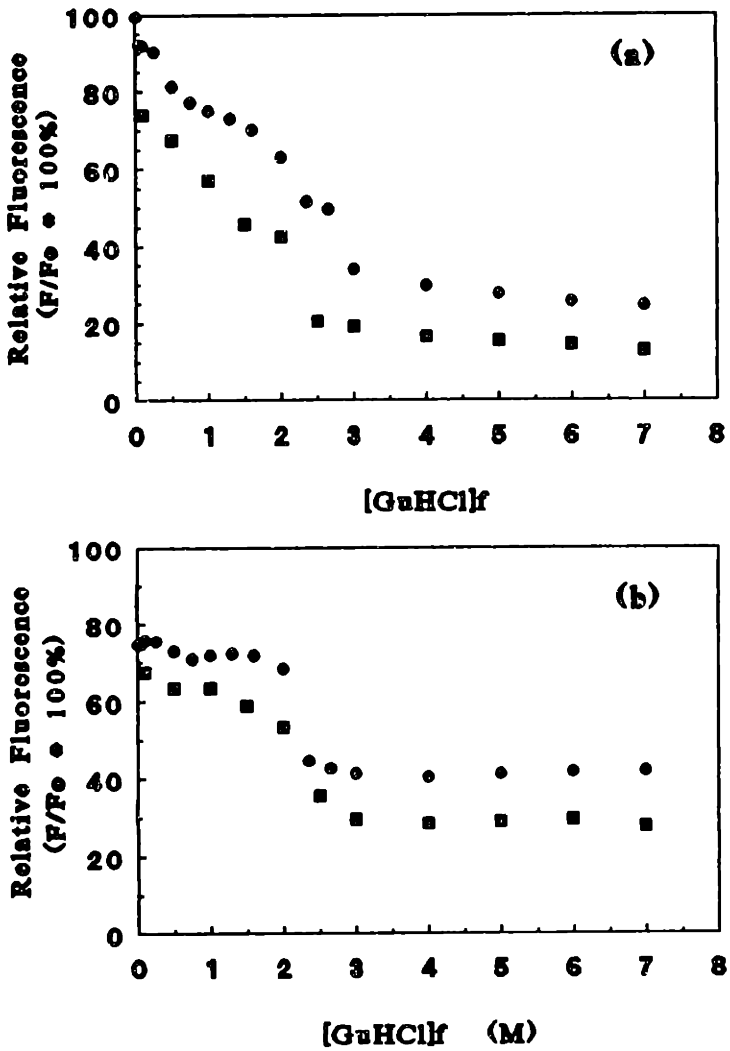


FIGURE 9.1: Equilibrium refolding of glycosylated rhDNAse was measured by fluorescence with an excitation wavelength of either 280 ηm (a) or 296 ηm (b) and an emission of 340 ηm. rhDNAse in 7 M GuHCl was refolded to a given final GuHCl concentration and 10 μg/ml protein. The protein was maintained in either oxidizing (●) or reducing (■) conditions during the refolding. The relative fluorescence (F/Fo) was calculated by normalizing all the fluorescence readings, F, by the fluorescence of the native protein in buffer, Fo.

have also revealed a large increase in the near circular dichroism (CD) spectra (220 ηm) at 2.0 M GuHCl. Therefore, rhDNAse may form a stable intermediate species at approximately 2.0 M GuHCl. In addition, the fluorescence for both the oxidized and reduced cases was relatively unchanged between 0.80 M and 1.2 M GuHCl as illustrated in Figure 9.1. At this range of GuHCl concentrations, another stable intermediate species may exist and lower GuHCl concentrations may be required to allow complete refolding of the protein. A GuHCl concentration of less than 0.30 M was required in order to regain greater than 90% of the native protein fluorescence for the oxidized protein (Figure 9.1). Even at low concentrations of GuHCl, the protein did not completely regain its native fluorescence for either the reduced or oxidized state. Refolding at higher protein concentrations (> 10 μg/ml) resulted in the formation of precipitates for final GuHCl concentrations ranging from 0.50 to 2.50 M GuHCl. These results indicated that an aggregating intermediate species existed at approximately 2.0 M GuHCl. As shown by fluorescence, several aromatic residues were exposed in this intermediate state. This intermediate which was observed to aggregate at 2.0 M GuHCl could be a hydrophobic intermediate on the refolding pathway for rhDNAse.

To study the formation of native structure in GuHCl, the activity of rhDNAse was measured at different GuHCl concentrations. Refolding was performed by rapid dilution from 7 M GuHCl to different GuHCl concentrations and 10 μg/ml protein at the oxidized conditions. After equilibration for 3 to 6 hours, the activity of each solution was determined as shown in Figure 9.2. The protein completely recovered its biological activity at 0.10 M GuHCl. The activity did not change significantly between 0.80 and

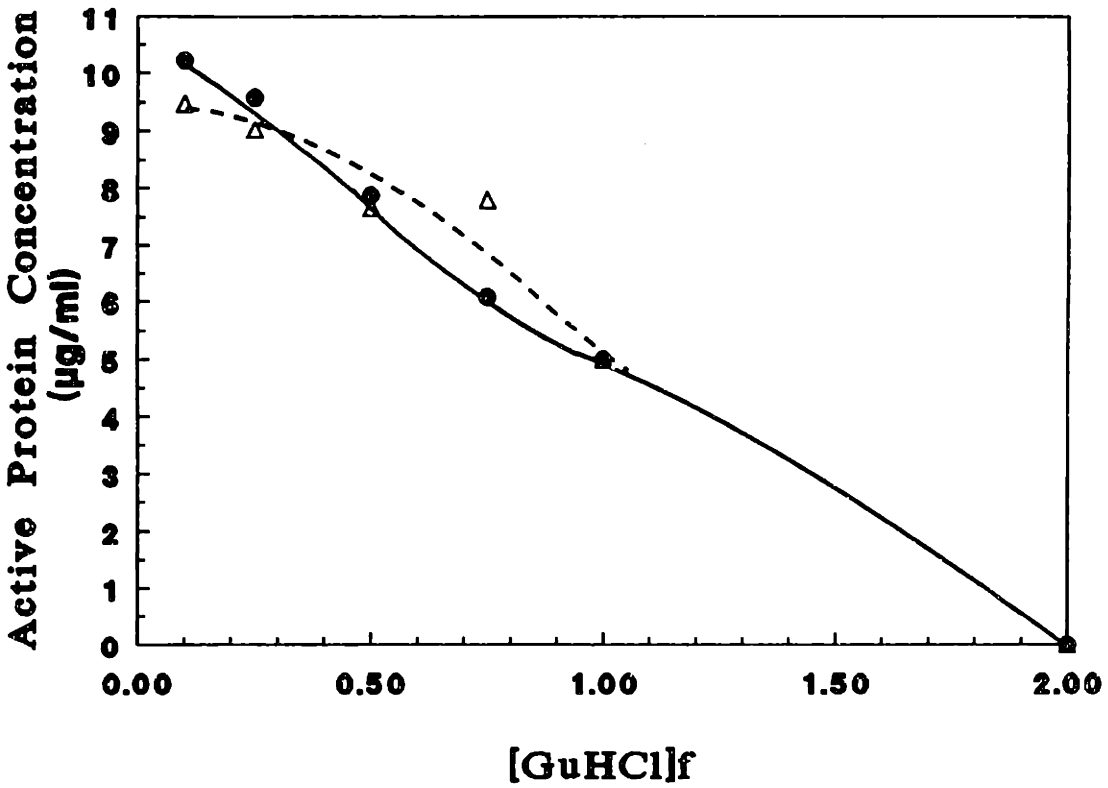


FIGURE 9.2: Effect of GuHCl concentration on recovery of activity for rhDNAse refolding. Recovery of active rhDNAse was determined for refolding from 7 M GuHCl to different GuHCl concentrations at final protein concentrations of 10 µg/ml (●) and 100 µg/ml (△).

1.0 M GuHCl. Since both the fluorescence and activity revealed a plateau in this GuHCl concentration range, a stable intermediate species with approximately 50% of the native protein activity could exist under these conditions. Refolding of rhDNAse at 100 μg/ml resulted in the formation of aggregates for all of the conditions tested and the protein recovered a maximum of 9.5% of its biological activity at 0.10 M GuHCl as depicted in Figure 9.2. The refolding of reduced rhDNAse was performed by rapid dilution from 7 M GuHCl and 50 mM β-mercaptoethanol to 0.10 M GuHCl and 2 mM β-mercaptoethanol at 10 μg/ml protein with a redox buffer containing 10 mM reduced glutathione (GSH) and 1 mM oxidized glutathione (GSSG). For refolding under these conditions, the protein recovered only 10% of its biological activity and the final concentration of reducing agent (2 mM β-mercaptoethanol) may need to be decreased to allow complete oxidation of the protein. Since one disulfide bond (^{173}Cys-^{209}Cys) is required for activity, it may be necessary to completely oxidize the protein in order to completely regain activity.

The activity of rhDNAse has been shown to be dependent upon both magnesium (Mg^{2+}) and calcium ions (Ca^{2+}) (Price, 1975). Therefore, it could be postulated that a binding competition existed between Gu^{+} and Ca^{2+} which must be bound to the protein for biological activity (Okabe et.al., 1982). Since the dilution buffer contains only 10 mM $CaCl_2$, the GuHCl concentration must be reduced significantly to allow calcium binding and subsequent formation of native structure. To completely refold rhDNAse, very low GuHCl concentrations ($\leq$ 0.10 M) were required for a significant recovery of biological activity.

To avoid the difficulties of refolding in the presence of denaturant ions, refolding

studied were next performed with a nonionic denaturant, urea. The native rhDNAse was unfolded in 7.8 M urea at either oxidized or reduced conditions. The reducing buffer consisted of 10 mM GSH and 1 mM GSSG as a redox couple. This redox buffer system was originally used to refold *Escherichia coli* derived rhDNAse (Winkler, 1990). To ascertain the formation of an aggregating intermediate, equilibrium refolding experiments were performed by dilution from 7.8 M urea to different final urea concentrations at 10 μg/ml protein. After equilibration for 3 to 6 hours, the relative fluorescence of each solution was measured by excitation at 280 ηm or 296 ηm and an emission of 340 ηm as shown in Figure 9.3. Two major changes in the relative fluorescence were observed for the oxidized form. The first large change in fluorescence occurred between 4.0 and 4.5 M urea. A plateau existed between 1.5 and 3.5 M urea which may reveal the presence of a stable intermediate state. The second major change in the fluorescence occurred for urea concentrations lower than 1.5 M urea. The reduced state had only one major change in its relative fluorescence and this change occurred between 4.0 and 4.5 M urea. The fluorescence of the reduced protein did not appear to change for urea concentrations between 3.5 and 1.0 M. Unlike the refolding results in GuHCl, the protein did not regain greater than 65% of its native fluorescence for any of the final conditions shown in Figure 9.3.

Since rhDNAse did not regain its native fluorescence for the conditions shown in Figure 9.3, it was necessary to determine if the protein would recovery activity in urea. Therefore, each equilibrium refolding solution which was used in the fluorescence studies was measured for activity as shown in Figure 9.4. The activity of both the reduced and

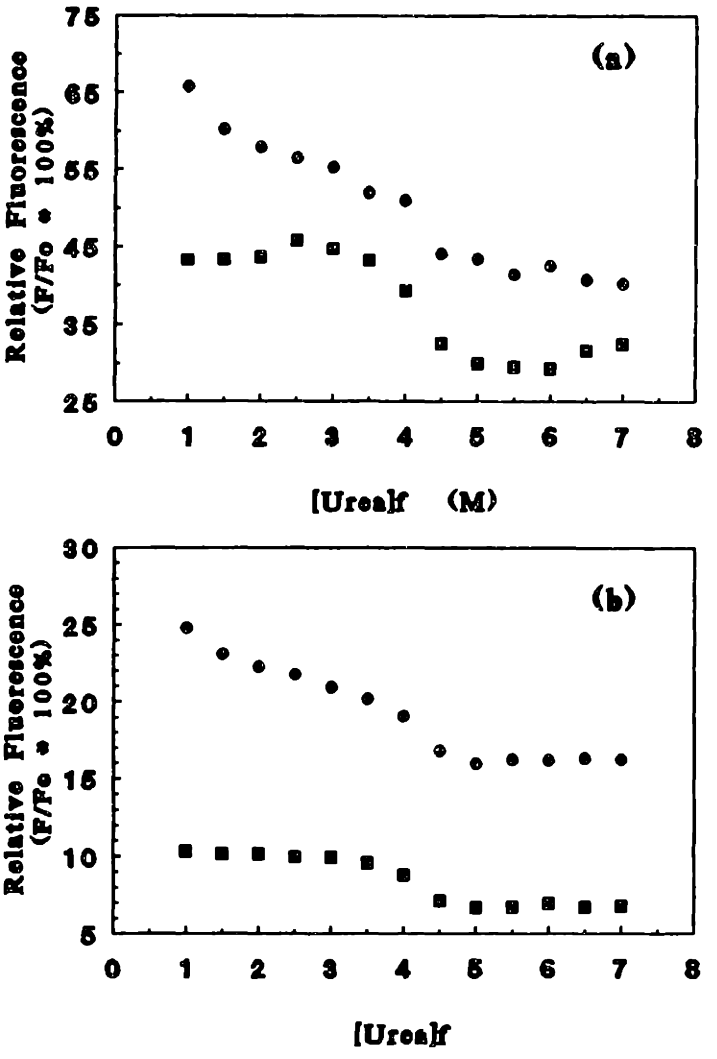


FIGURE 9.3: Equilibrium refolding of glycosylated rhDNAse was measured by fluorescence with an excitation wavelength of either 280 ηm (a) or 296 ηm (b) and an emission of 340 ηm. rhDNAse in 7.8 M urea was refolded to a given final urea concentration and 10 μg/ml protein. The protein was maintained in either oxidizing (●) or reducing (10 mM GSH/1 mM GSSG, ■) conditions during the refolding. The relative fluorescence (F/Fo) was calculated by normalizing all the fluorescence readings, F, by the fluorescence of the native protein in buffer, Fo.

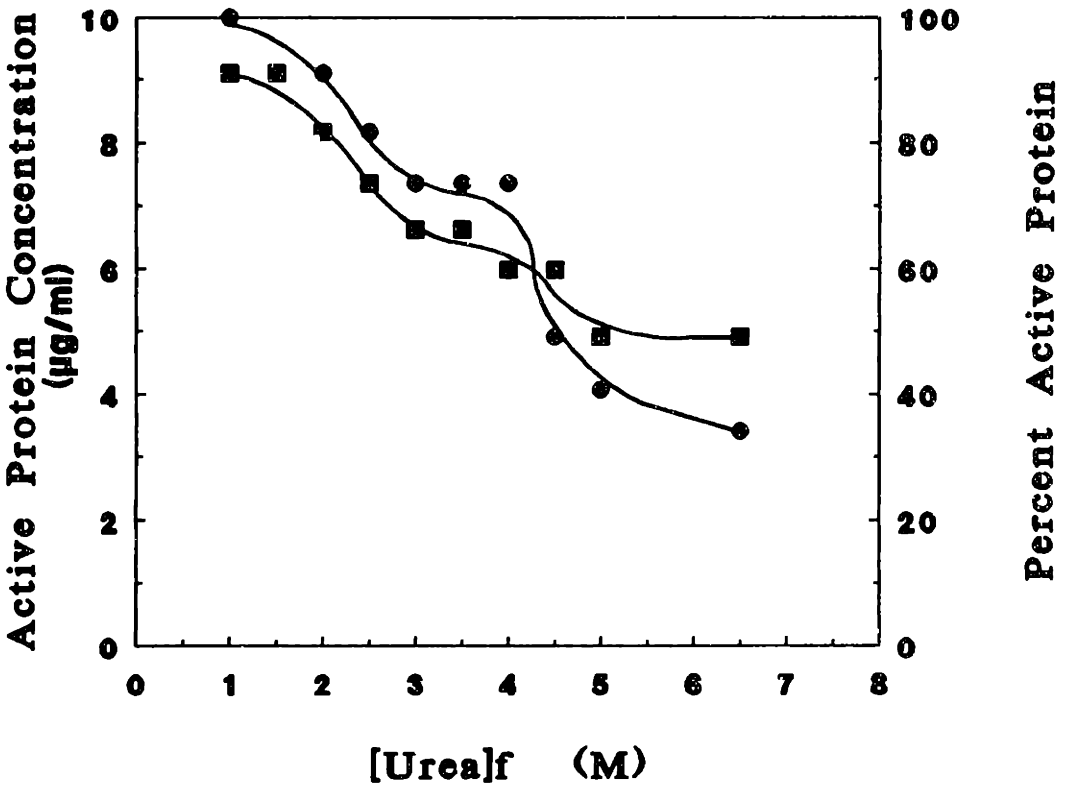


FIGURE 9.4: Effect of urea concentration on recovery of activity for rhDNAse refolding. Recovery of active rhDNAse was determined for refolding from 7.8 M urea to different urea concentrations at a final protein concentration of 10 µg/ml for both oxidizing (●) and reducing conditions (10 mM GSH/1 mM GSSG, ■).

oxidized forms increased dramatically between 4.0 and 5.0 M urea. Another large change in activity was observed between 1.5 and 2.5 M urea. A stable species with some activity (70% oxidized, 60% reduced) may exist between 3.0 and 4.0 M urea. The maximum recovery of active protein occurred at 1 M urea with a yield of 100% for the oxidized case and 90% for the reduced state. Refolding studies performed at higher protein concentrations (> 10 μg/ml) resulted in precipitation of the protein for final urea concentrations below 5.0 M. Therefore, an aggregating intermediate species may be formed during the refolding and may be stable at urea concentrations between 4.0 and 5.0 M as evidenced by the changes in both fluorescence and activity in this concentration range.

In order to reduce the aggregation observed at low urea concentrations and high protein concentrations, PEG enhanced refolding experiments were performed with rhDNAse in urea. Two different final conditions were chosen to study the effect of PEG. The final conditions were first characterized for refolding without PEG. First of all, refolding was performed by rapidly diluting rhDNAse in 7.2 M urea to 6.5 M urea at 200 μg/ml protein. Refolding at 6.5 M urea was previously shown to result in a reduction in precipitate formation for *E. coli* derived rhDNAse (Winkler, 1990). Therefore, a final condition of 6.5 M urea was considered to be nonaggregating conditions. In addition, refolding was performed at 4.0 M urea where precipitation was previously observed (Winkler, 1990). To achieve this aggregating condition, rhDNAse in 7.2 M urea was refolded by dilution to 4.0 M urea at 200 μg/ml protein. Both of the refolding experiments were performed with a redox couple of 10 mM GSH and 1 mM GSSG. As

mentioned previously, the high final urea concentration (6.5 M) was initially chosen in an attempt to avoid aggregation. However, refolding at these conditions (200 μg/ml rhDNAse, 6.5 M urea) resulted in 30% recovery of active protein and precipitation of inactive protein was observed. The refolding of rhDNAse at a low protein concentration (10 μg/ml) and 6.5 M urea had resulted in the recovery of 50% active protein as shown in Figure 9.4 and, therefore, multimer formation may have caused the reduced activity even at the low protein concentration. In addition, refolding at the same protein concentration (200 μg/ml) and 4.0 M urea resulted in 70% recovery of activity with aggregate formation.

In an attempt to avoid aggregation, refolding was performed at the same final conditions with redox buffer containing PEG (3350 MW). Refolding was performed by rapid dilution from 7.2 M urea to either 6.5 M or 4.0 M urea and 200 μg/ml protein with final PEG to protein molar ratios ($[PEG]_f/[rhDNAse]_f$) of 1 and 5. After equilibration for 6 hours, the recovery of activity was measured and plotted as a function of the final PEG to protein molar ratio as shown in Figure 9.5. Both molar ratios of PEG improved the recovery of active protein at both final urea conditions. Aggregation was not observed at a 5 to 1 molar ratio and complete recovery of activity was achieved in 4 M urea. Refolding at the same molar ratio in 6.5 M urea resulted in the same final recovery (50%) as observed in the equilibrium studies of the oxidized state (Figure 9.4). Therefore, PEG will increase the recovery of active rhDNAse during refolding at protein concentrations which normally result in aggregation.

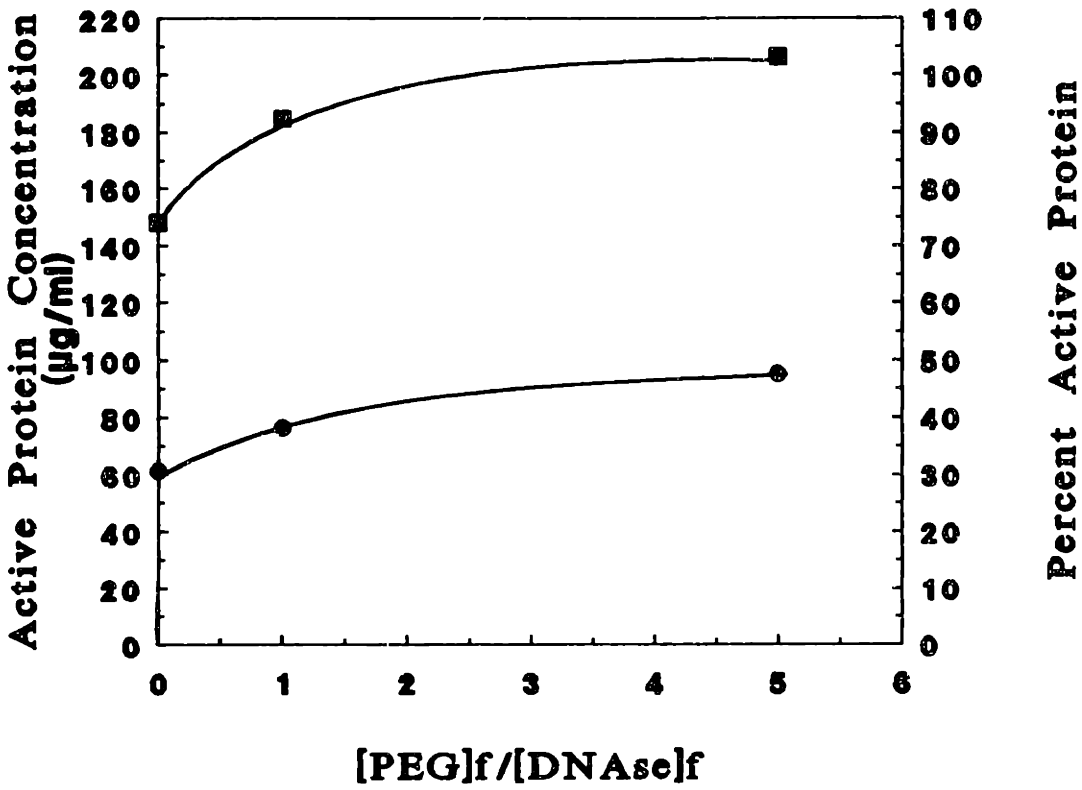


FIGURE 9.5: PEG enhanced refolding of glycosylated rhDNAse was performed at different molar ratios of PEG to rhDNAse ([PEG]f/[DNAse]f). rhDNAse in 7.2 M urea was refolded by rapid dilution to either 4.0 M (■) or 6.5 M (●) urea and 200 µg/ml protein with and without PEG (3350 MW) at oxidizing conditions (unfolded protein not reduced). After equilibration for three to six hours, the activity of each solution was measured.

9.2 rhDNAse from Escherichia Coli

Since PEG was successful in increasing the recovery of active glycosylated rhDNAse, it may be possible to enhance the recovery of active protein from impure inclusion bodies of rhDNAse. Previous refolding studies on the inclusion bodies of rhDNAse resulted in a recovery of only 19% active protein at 6.5 M urea with buffer containing 10 mM GSH, 1 mM GSSG, 50 mM Bicine, 2 mM $CoCl_2$, and 1 M sucrose at pH 8.8 (Winkler, 1990). For this study, the inclusion bodies were solubilized in 8 M urea with 10 mM GSH, 1 mM GSSG, 50 mM Tris, 10 mM $CaCl_2$, and 10 mM $MgCl_2$ at pH 8 for comparison with the glycosylated rhDNAse refolding studies.

Refolding of the solubilized inclusion bodies in 8 M urea was performed by dilution to 100 μg/ml total protein and either 4.0 or 6.5 M urea with the same redox buffer which was used to refold glycosylated rhDNAse. To prevent aggregation, PEG (3350 MW) was used in the redox buffer to provide final PEG to protein molar ratios ($[PEG]_f/[rhDNAse]_f$) of 2 and 10. The molar ratios were calculated by assuming that the inclusion bodies were 80% pure rhDNAse (see Appendix). After equilibration for 6 hours, the active protein concentration was measured for each sample with and without PEG as shown in Figure 9.6. Without PEG in the dilution buffer, the protein aggregated and recovered less than 20% activity at both final urea concentrations. In contrast, PEG increased the recovery of active protein for both cases. Assuming that the protein mixture was 80% rhDNAse, the recovery of active rhDNAse was 35% for refolding at 6.5 M urea and 50% for refolding at 4.0 M urea in the presence of PEG at a molar ratio of 10 to 1 (0.1 g/l PEG). Furthermore, visible aggregation was not observed for refolding in

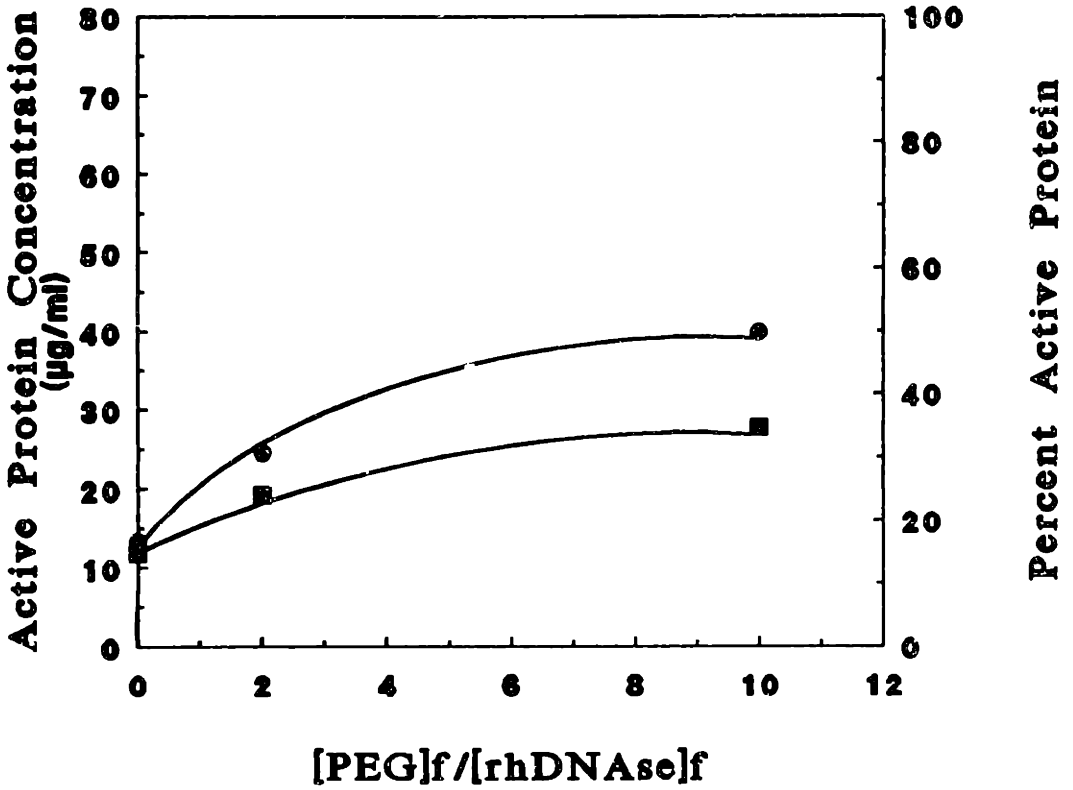


FIGURE 9.6: PEG enhanced refolding of *E. coli* derived rhDNAse from inclusion bodies was performed by rapid dilution from 8 M urea to either 4.0 M (●) and 6.5 M (■) urea and 100 µg/ml of total protein at different molar ratios of PEG to rhDNAse ($[PEG]_f/[rhDNAse]_f$). The activity of the final mixture was assayed and the relative percent active protein and molar ratios were calculated based on the assumption that the inclusion bodies were 80% rhDNAse.

4 M urea at the higher molar ratio ([PEG]f/[rhDNAse]f= 10). These results indicate that PEG will enhance recovery of rhDNAse from inclusion bodies and higher molar ratios of PEG may be required when additional aggregating protein is present.

9.3 Glycosylated Recombinant Tissue Plasminogen Activator (rtPA)

Since PEG will enhance the refolding of rhDNAse, it may be able to enhance the refolding of more complex proteins. Additional studies were therefore performed with glycosylated recombinant tissue plasminogen activator (rtPA). rtPA does not have a well characterized refolding or denaturation pathway. Thus, some preliminary equilibrium refolding experiments were also performed on this protein.

Initially, fluorescence measurements were used to study the structural changes which occurred from refolding rtPA from the denatured state in 5 M GuHCl. Equilibrium refolding studies were performed by diluting the denatured rtPA in 5 M GuHCl to different GuHCl concentrations and 10 μg/ml rtPA. For each experiment, the protein was not reduced in the unfolded state (see Appendix for redox studies). After equilibration for 6 hours, the fluorescence of each solution was measured with an excitation wavelength of either 280 ηm or 296 ηm and emission wavelength of 340 ηm. Each fluorescence measurement, F, was then normalized to the fluorescence of the native protein in buffer, F_o, to provide a relative fluorescence scale. As shown in Figure 9.7, the relative fluorescence had three separate phases. The first phase was represented by a large fluorescence difference between 3.0 and 3.5 M GuHCl with a 20% increase in the relative fluorescence. This large change was followed by a slight plateau between 2.5 and 3.0 M GuHCl. The next significant change in the relative fluorescence occurred between 2.0

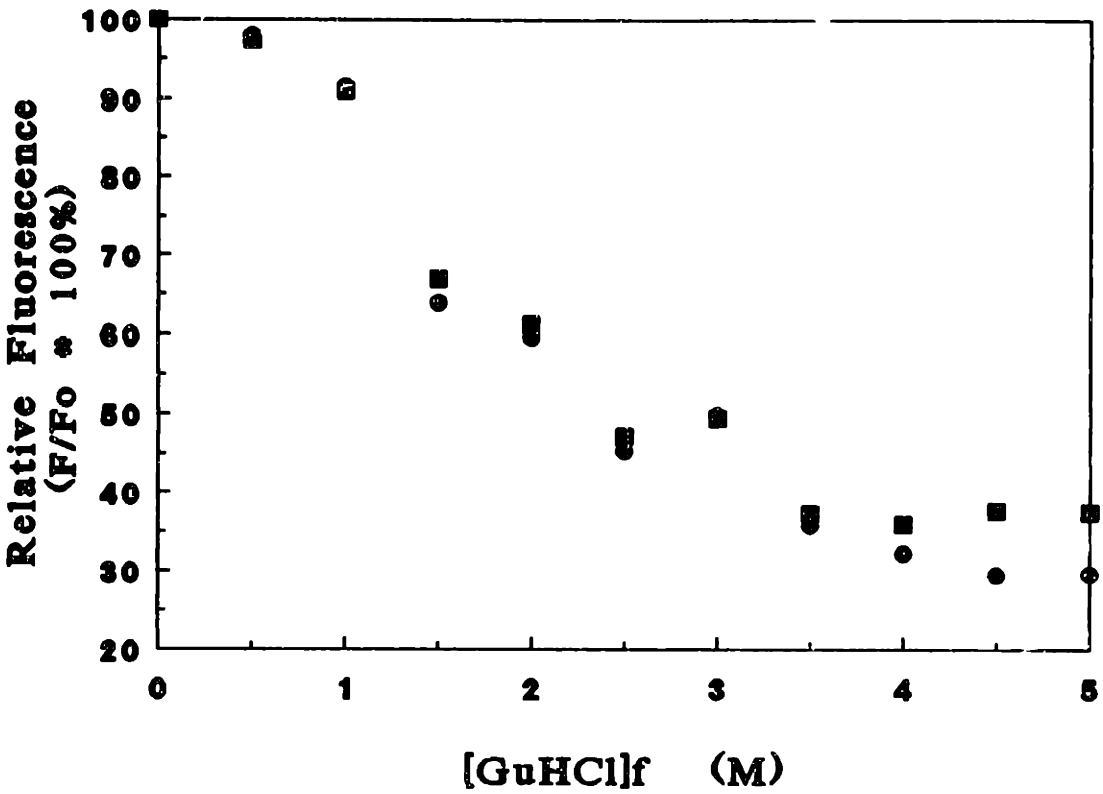


FIGURE 9.7: Equilibrium refolding of glycosylated rtPA was performed by dilution from 5 M GuHCl to a given final GuHCl concentration and 10 μg/ml protein without disulfide reduction. After equilibration for six hours, the fluorescence of each sample was measured with an excitation of either 280 ηm (●) or 296 ηm (■) and an emission of 340 ηm. The relative fluorescence was defined as the measured fluorescence of each sample, F, normalized by the fluorescence of the native protein in buffer, F_o.

and 2.5 M GuHCl. The large change in fluorescence was again followed by a small plateau which existed between 1.5 and 2.0 M GuHCl. The greatest change in the relative fluorescence occurred between 1.0 and 1.5 M GuHCl as shown in Figure 9.7. These results revealed that multiple intermediates may exist on the refolding pathway and these intermediates may be stable at different concentrations of GuHCl. In addition, to completely regain the native protein fluorescence, rtPA must be refolded at less than 0.50 M GuHCl. Refolding at higher protein concentrations resulted in the formation of precipitates for GuHCl concentrations less than 3.0 M. Therefore, the potential intermediates observed in the refolding study may be intermediate species which aggregate during the refolding process.

To determine if rtPA will recover its protease activity in GuHCl, the equilibrium refolding solutions which were used in the fluorescence studies were assayed for biological activity. Protease activity assays using the peptide substrate, S2288, were performed on each sample as described previously (Friberger, 1982). As illustrated in Figure 9.8, the recovery of protease activity was less than 90% for GuHCl concentrations of greater than 0.50 M. The most significant change in the recovery of active protein occurred at the same range of GuHCl concentrations, 1.0 to 1.5 M, as that observed in the fluorescence studies. To recover active protein and regain native fluorescence, it was shown that rtPA must be refolded at low GuHCl concentrations ($\leq$ 0.50 M).

To assess the ability of PEG to reduce aggregation at low GuHCl conditions, refolding was performed by dilution of rtPA in 5 M GuHCl to 0.50 M GuHCl and 100 µg/ml protein with and without PEG (3350 MW) in the dilution buffer. The protease

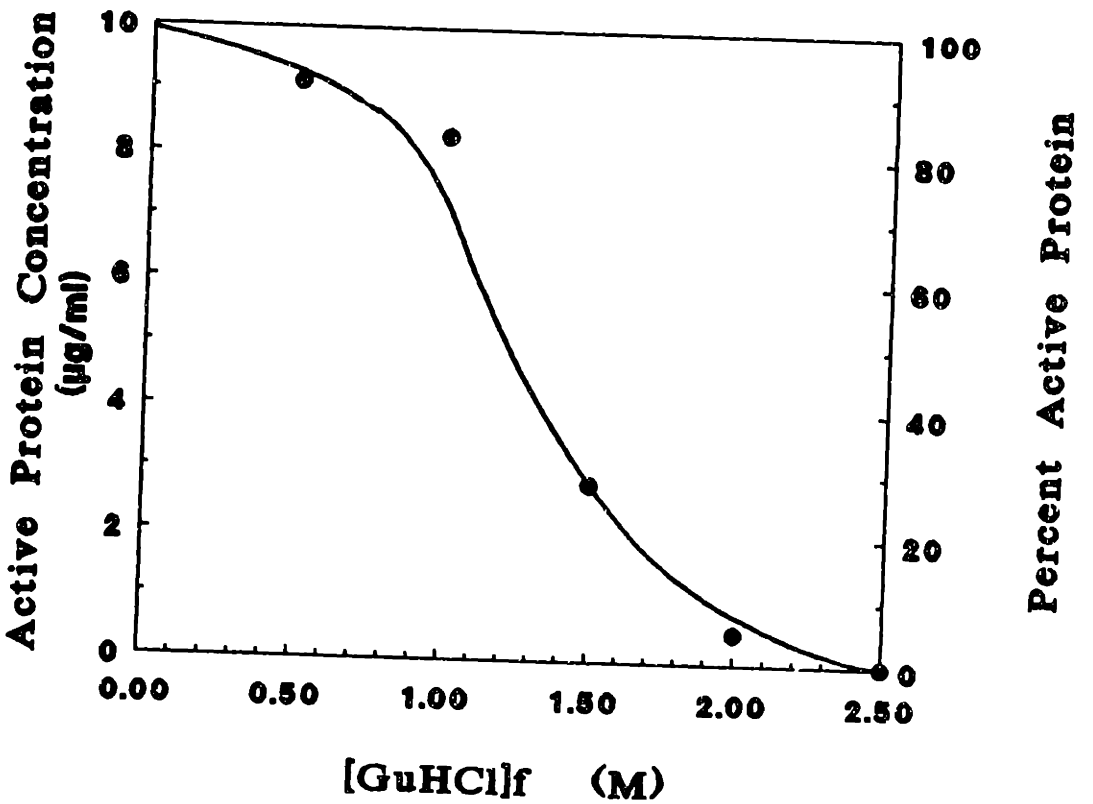


FIGURE 9.8: Equilibrium refolding of rtPA was performed as described in Figure 9.7. The recovery of active protein ([rtPA]f= 10 µg/ml) was determined from its protease activity as described previously (Friberger, 1982).

activity of each solution was measured after refolding for six to eight hours. In the absence of PEG, the protein recovered approximately 63% of the protease activity relative to the native protein in the same GuHCl concentration as displayed in Figure 9.9. The recovery of rtPA was improved by refolding at the same final conditions with final PEG molar ratios ($[PEG]_f/[rtPA]_f$) ranging from 2 to 20 (0.01 to 0.10 g/l PEG 3350 MW). A maximum recovery of only 70% active protein was obtained at a molar ratio ($[PEG]_f/[rtPA]_f$) of 20. Since aggregation was not observed in the presence of PEG at molar ratios greater than 10 for these experiments, aggregation may not be the dominate problem in refolding at 0.50 M GuHCl. To determine if a low GuHCl concentration would result in higher recovery, rtPA in 5 M GuHCl was refolded by rapid dilution to 0.20 mg/ml protein and 0.10 M GuHCl with PEG (3350 MW) in the renaturation buffer. After equilibration for 6 hours, the protease activity was measured for each solution as shown in Figure 9.10. PEG enhanced the recovery of active rtPA (protease activity) at an optimum molar ratio of 20 to 1. Greater molar ratios ($[PEG]_f/[rtPA]_f$) resulted in only slightly higher recovery of active protein. At these final conditions, PEG increased the recovery of protease activity by two fold resulting in recovery of 60% active rtPA. Therefore, PEG may also be useful in reducing aggregation during the refolding of rtPA.

9.4 Recombinant gamma interferon (rIFN-γ)

To further test the ability of PEG to enhance refolding, experimental studies were performed by using *E. coli* derived recombinant interferon-γ (rIFN-γ). Initial studies were performed with pure rIFN-γ in 4 M GuHCl at pH 7. For comparison with the previous refolding results, refolding was performed by rapid dilution from 4 M GuHCl

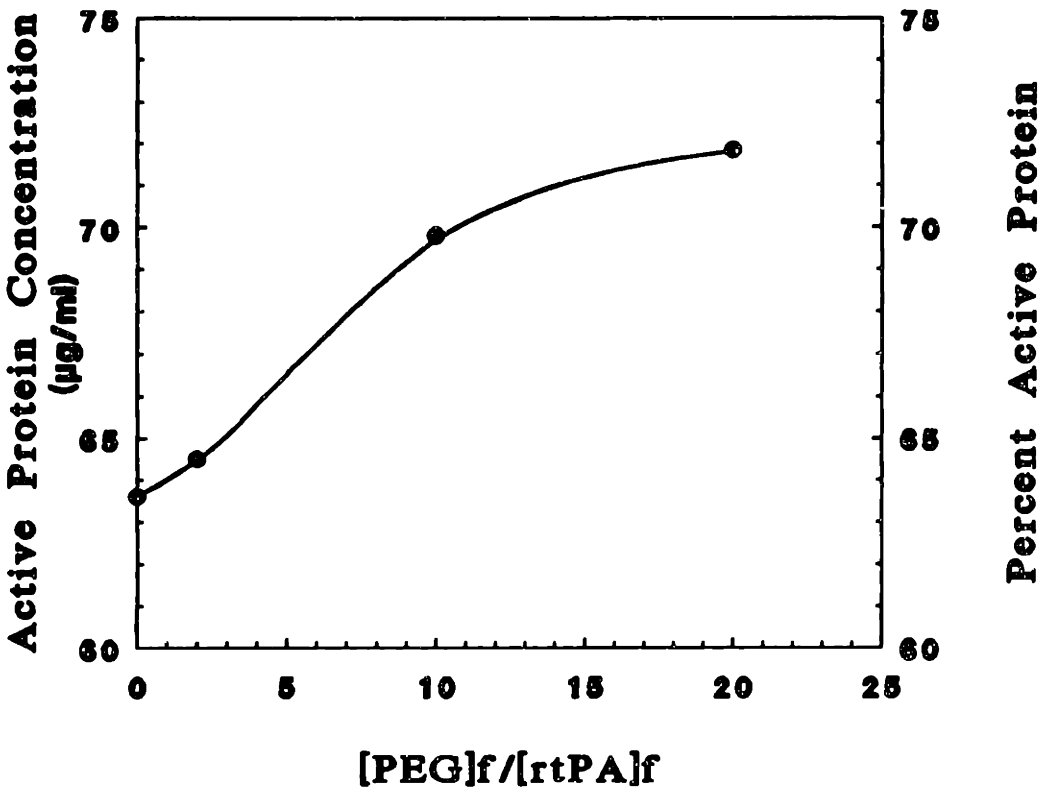


FIGURE 9.9: PEG enhanced refolding of glycosylated rtPA was performed by dilution from 5 M to 0.50 M GuHCl and 100 µg/ml protein at different molar ratios of PEG (3350 MW) to protein ([PEG]f/[rtPA]f) without disulfide reduction. The recovery of protease activity was measured after equilibration of each sample for at least six hours.

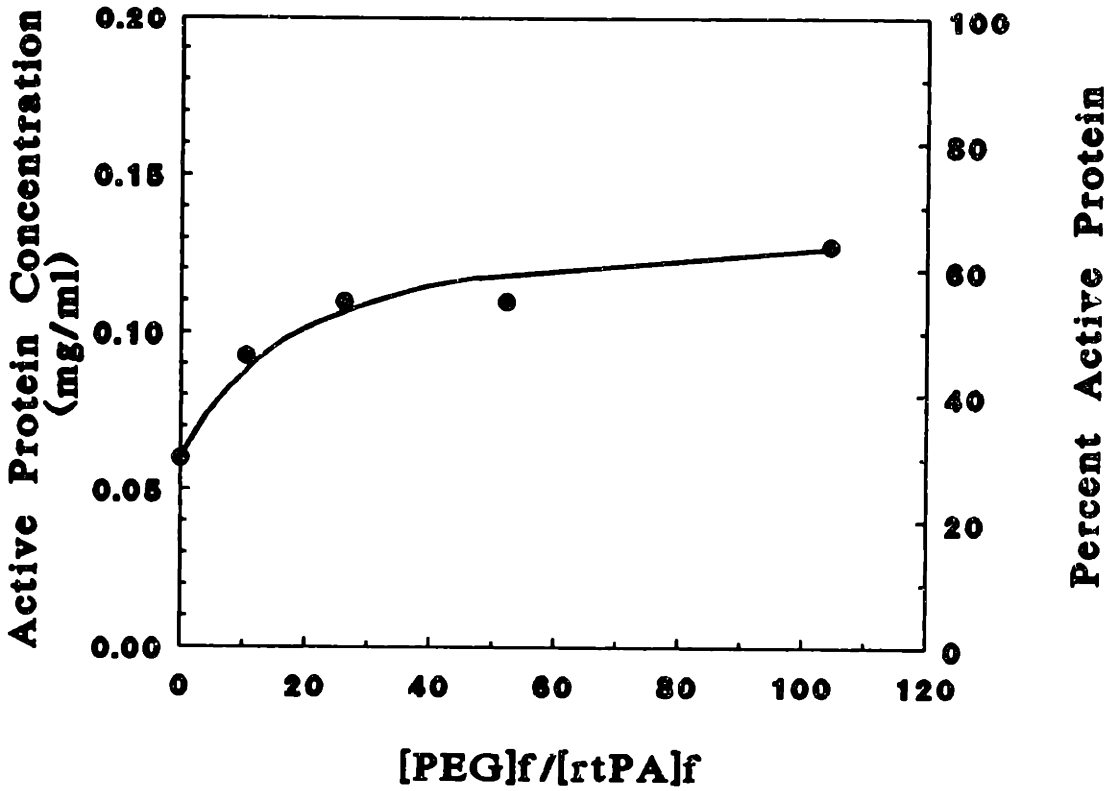


FIGURE 9.10: Refolding of rtPA at low GuHCl concentration with PEG. PEG enhanced refolding of glycosylated rtPA was performed by dilution from 5 M to 0.10 M GuHCl and 200 μg/ml protein at different molar ratios of PEG (3350 MW) to protein ([PEG]f/[rtPA]f) without disulfide reduction. The recovery of protease activity was measured after equilibration of each sample for at least six hours.

to 1 M GuHCl and 0.50 mg/ml rIFN-γ (Patel, 1986). Aggregation was observed at these conditions and the extent of aggregation was defined as:

$$\% \text{ Aggregated Protein} = \{[\text{rIFN-}\gamma]_{\text{soluble}}/[\text{rIFN-}\gamma]_f\}*100\% \qquad (9.1)$$

where $[\text{rIFN-}\gamma]_{\text{soluble}}$ is the protein concentration after sterile filtration (0.22 μm filter) and $[\text{rIFN-}\gamma]_f$ is the total final protein concentration. For these final conditions, 80% of the protein aggregated and only 20% recovery of activity was measured by cytotoxic assay (see Materials and Methods). These results indicated that all the inactive protein was in the form of large precipitates. In an attempt to avoid aggregation, refolding at these conditions was performed with 1 g/l PEG (8000 MW) which corresponded to a molar ratio ($[\text{PEG}]_f/[\text{rIFN-}\gamma]_f$) of 10. After equilibration for 6 hours, the active protein was measured by cytotoxic assay. The protein recovered 30.6 % of its biological activity and 61.1% of the total protein aggregated to form large precipitates. Therefore, PEG did reduce the formation of precipitates under these final conditions, but did not significantly enhance the recovery of active protein as displayed in Table 9.1. These results may indicate that the protein will not refold completely at 1.0 M GuHCl. To assess the effect of GuHCl on refolding, refolding was performed by diluting rIFN-γ in 4 M GuHCl to 0.50 M GuHCl and 1.0 mg/ml protein with buffer or 1 g/l PEG (8000 MW). Refolding at these conditions resulted in rapid precipitation of the protein and little recovery of activity for either case (Table 9.1). Therefore, additional PEG may be required to prevent aggregation under these conditions.

Additional studies were performed with rIFN-γ inclusion bodies which were

TABLE 9.1: Refolding of recombinant interferon-γ (rIFN-γ) from 4 M GuHCl.

[rIFN-γ]f (mg/ml)	[GuHCl]f (M)	% Active[a]	% Aggregated[b]	Dilution Buffer
1.0	0.50	1.6	91.0	Phosphate
1.0	0.50	2.8	82.5	1 g/l PEG
0.50	1.0	20.0	80.0	Phosphate
0.50	1.0	30.6	61.1	1 g/l PEG
1.0	0.50	52.8	71.6	22% Sucrose
1.0*	0.50	1.2	82.5	Phosphate
1.0*	0.50	1.0	59.0	1 g/l PEG
1.0*	0.50	12.4	32.1	22% Sucrose

a. % Active protein = (measured specific activity / 2 x 10^7 IU/mg)*100%
b. % Aggregated protein = ([rIFN-γ]soluble / [rIFN-γ]f)*100%

* These protein concentrations correspond to total protein in inclusion body preparation which is approximately 60% rIFN-γ. The data have not been altered to reflect the impurities. All other samples were pure rIFN-γ.

solubilized by the addition of 4 M GuHCl. The soluble protein consisted of approximately 60% rIFN-γ (see Appendix). The rIFN-γ inclusion body mixture in 4 M GuHCl was diluted to 0.50 M GuHCl and 1.0 mg/ml total protein with buffer or 1 g/l PEG (8000 MW). After equilibration, the concentration of active and precipitated protein was measured. As shown in Table 9.1, refolding at these conditions also resulted in a low recovery of activity and precipitation indicating that additional PEG may be required to enhance refolding. As a comparison, refolding was performed with a 22% wt/vol sucrose solution at the same final conditions. The final activity recovered was slightly greater (12.4%), but the extent of aggregation was significantly reduced (32.1%). Therefore, sucrose, as previously observed with CAB, will prevent the formation of precipitates, but will not greatly enhance the recovery of active protein. Since PEG did result in a slight improvement in recovery of active protein for refolding at 0.50 mg/ml rIFN-γ and 1.0 M GuHCl, it may be possible to improve recovery of active protein at lower GuHCl concentrations by using PEG in the dilution buffer at an optimal concentration.

Chapter 10: Relationship between PEG Enhancement and Protein Structure

Since PEG has been shown to enhance refolding of several proteins, there could be a general relationship between PEG enhancement and the protein properties. In addition, the optimum molar ratio ($[PEG]_f/[CAB]_f$) required for enhanced refolding was different for each protein. Therefore, it would be useful to develop a relationship between the optimum molar ratio required for enhancement and the protein properties. Since PEG has been shown to bind to a specific structural intermediate, for example, the molten globule first intermediate in the CAB refolding pathway, PEG may also bind to molten globule intermediates on the refolding pathway of other proteins. Recent refolding studies of several other proteins have shown that a molten globule intermediate forms during the refolding of these proteins (Ptitsyn et.al., 1990). This type of compact intermediate is characterized by a rigid structure with hydrophobic surfaces. The proteins which have this molten globule intermediate on their refolding pathway would also have hydrophobic cores in their native state. Thus, an analysis of the native protein structure for each protein should reveal insight into the relationship between the concentration of PEG required for enhancement and the protein structure.

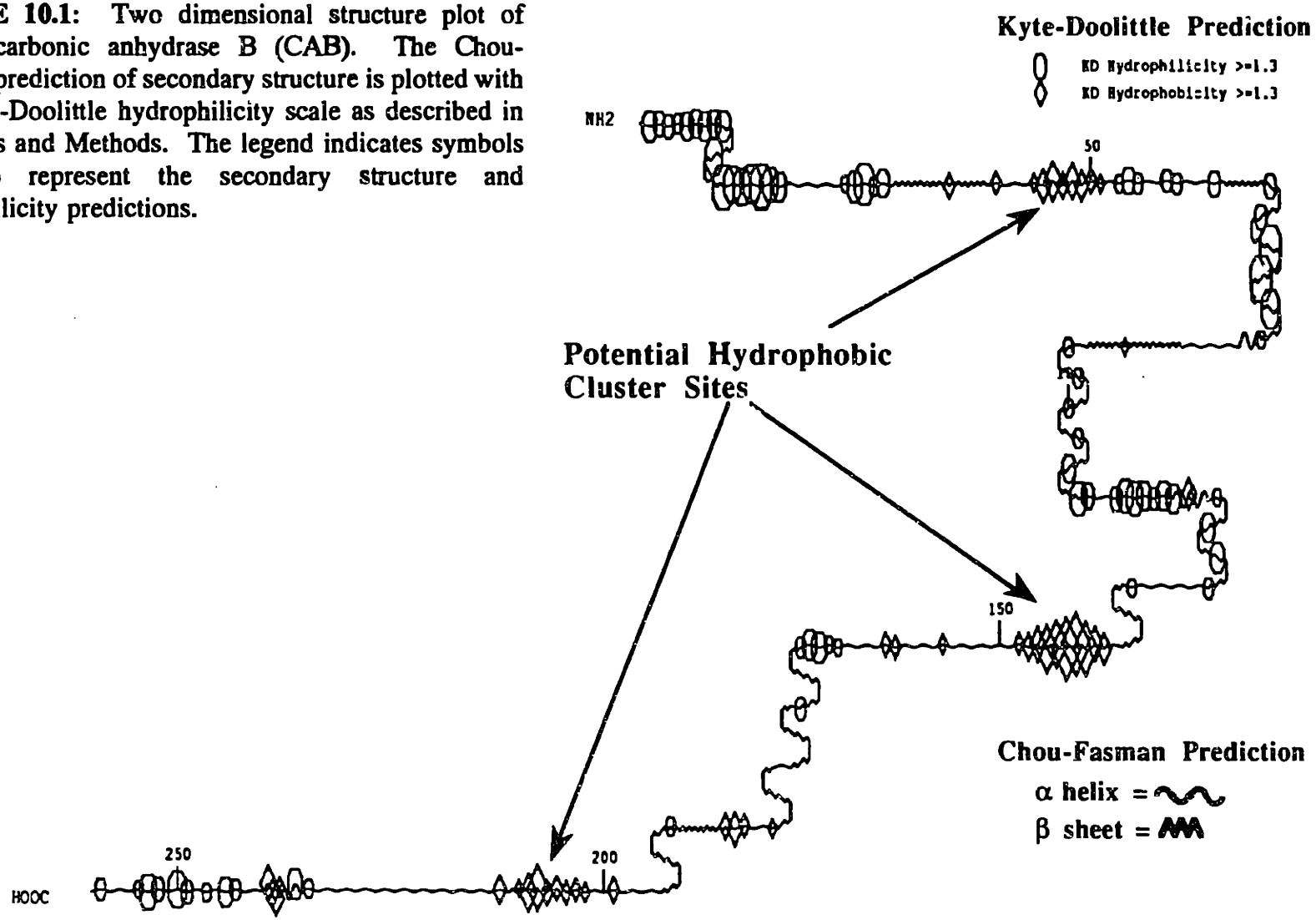
10.1 Hydrophobicity and Beta Sheet Structures

A common feature of the molten globule intermediates which have been observed during the refolding of several proteins is a high degree of hydrophobicity and beta sheet structure (Ptitsyn et.al., 1990). The propensity of PEG to bind to a molten globule intermediate, for example, the first intermediate on the CAB refolding pathway, may be an indication that PEG will also bind to the hydrophobic beta sheet surface on other

molten globule intermediates. Therefore, an analysis of native protein structures for hydrophobic beta sheets should provide a basis for the development of a model for the general ability of PEG to enhance refolding.

In most cases, the native structure of recombinant proteins is not well characterized and the three dimensional crystal structure often is not known. To develop the necessary relationship between hydrophobic beta sheets and PEG enhancement, a structure prediction method might be used. The most common method of predicting secondary structure was developed by Chou and Fasman (Chou & Fasman, 1978). This structure prediction algorithm was based on the structural analysis of several proteins. The analysis of different proteins led to a method which predicts the secondary structure based on the probability of an amino acid residue existing in a beta sheet, alpha helix, turn or random coil structure (Chou & Fasman, 1978). The primary sequence of the protein can then be related to the secondary structure by applying the prediction algorithm. In addition, studies by Kyte and Doolittle have resulted in a prediction method for determining the relative hydrophilicity of each amino acid in a protein (Kyte & Doolittle, 1982). Combining both prediction techniques, the structure of each protein studied was analyzed for its degree of beta sheet structure and the relative hydrophobicity of each amino acid in its sequence. The predicted secondary structure and relative hydrophilicity of CAB was plotted as a two dimensional structure diagram as described in Materials and Methods (Figure 10.1). As marked in Figure 10.1, the groups of beta sheet structure with hydrophobic residues were defined as potential hydrophobic cluster sites on the protein. Each of these clusters has been marked on the three dimensional

FIGURE 10.1: Two dimensional structure plot of bovine carbonic anhydrase B (CAB). The Chou-Fasman prediction of secondary structure is plotted with the Kyte-Doolittle hydrophilicity scale as described in Materials and Methods. The legend indicates symbols used to represent the secondary structure and hydrophilicity predictions.



structure plot of CAB (Figure 10.2; Creighton, 1984). It has been reported that three hydrophobic clusters were observed in the crystal structure of human carbonic anhydrase II (79.2% homologous to bovine CAB) and that these clusters are at an equal distance from the center zinc ion (Liljas et.al., 1972). As marked in Figure 10.2, two of the three predicted clusters are located at an equal distance from the zinc ion. This comparison indicates the potential for utilizing simple two dimensional structure plots for analysis of hydrophobic clusters. To further validate the hydrophobic cluster prediction method, a two dimensional structure plot of bovine DNAse (bDNAse) was compared to the published crystal structure (Figures 10.3 and 10.4; Suck et.al., 1984). The prediction method shows that their are six hydrophobic clusters distributed throughout the protein and the crystal structure indicates a hydrophobic core with two 6-stranded beta sheets as shown in Figure 10.3. Six of these strands are located in the center of the protein and correlate well with the hydrophobic clusters (Figure 10.4). Both CAB and bDNAse have a hydrophobic core structure consisting of beta sheets which may be exposed to the surface during refolding.

The recombinant proteins, rhDNAse and rtPA, have not been crystallized and, therefore, their three dimensional structures are not known. However, the crystal structure of the bovine analog of DNAse (bDNAse) could be similar to that of rhDNAse. To determine the homology between the two DNAse species, a comparison was made between each residue in the two sequences and the best match in the sequences is shown in Figure 10.5 (see Materials and Methods for homology plot method). From this homology analysis, rhDNAse was shown to have a 86.7% similarity to the sequence of

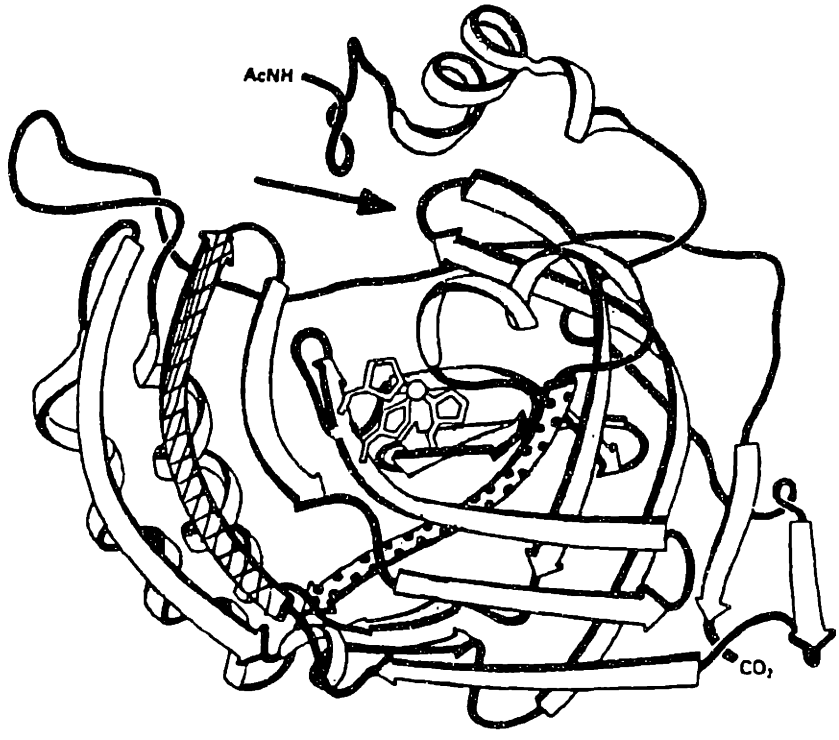


FIGURE 10.2: Ribbon diagram of three dimensional structure of human carbonic anhydrase II (Creighton, 1984). Human carbonic anhydrase II is 89.6% homologous to bovine carbonic anhydrase B (CAB). The hydrophobic cluster regions predicted from the two dimensional structure plot (Figure 10.1) are marked in the diagram. One beta sheet (dotted beta sheet ribbon) and a beta turn which is marked by the arrow are an equal distance from the zinc ion and are predicted as hydrophobic clusters. The third cluster is a beta sheet further from a the zinc ion (hatched beta sheet).

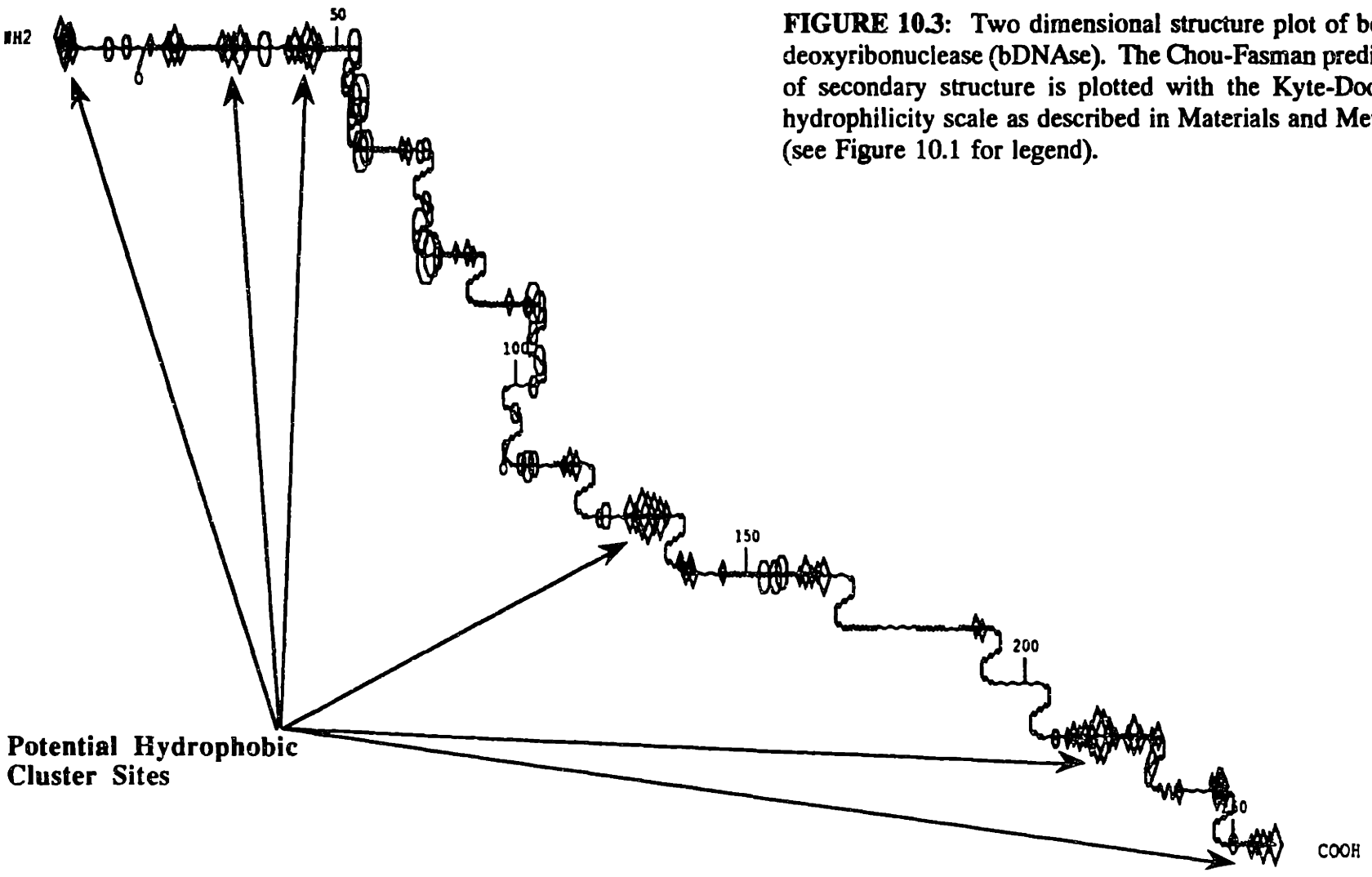


FIGURE 10.3: Two dimensional structure plot of bovine deoxyribonuclease (bDNAse). The Chou-Fasman prediction of secondary structure is plotted with the Kyte-Doolittle hydrophilicity scale as described in Materials and Methods (see Figure 10.1 for legend).

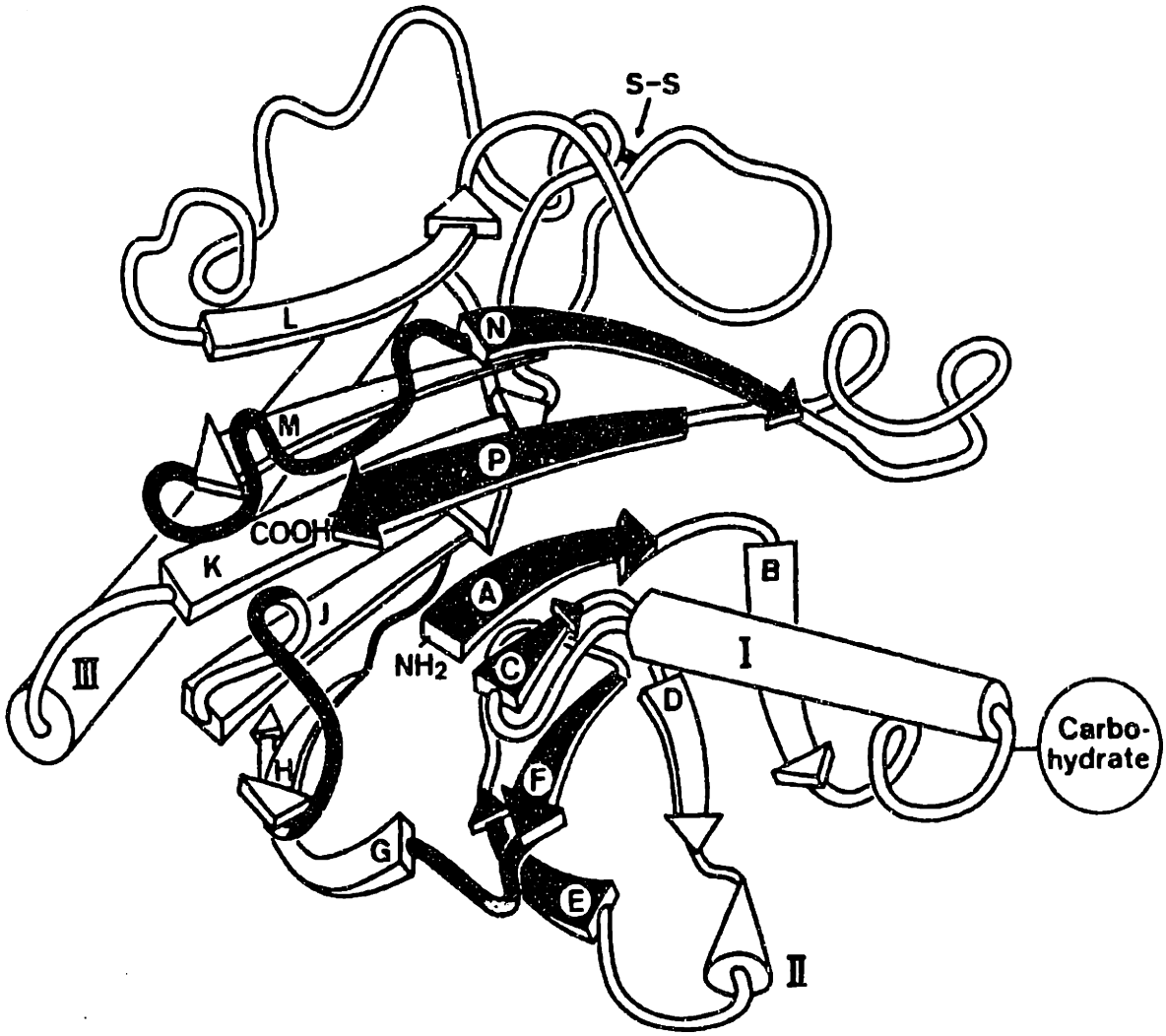


FIGURE 10.4: Ribbon diagram of three dimensional structure of bovine deoxyribonuclease (bDNAse) (Suck et.al., 1990). The hydrophobic core consists of two 6-stranded beta sheets. One half of this beta sheet region is marked by the darkened beta strand ribbons. The disulfide bond required for activity and the single carbohydrate at ^{18}Asn are also shown.

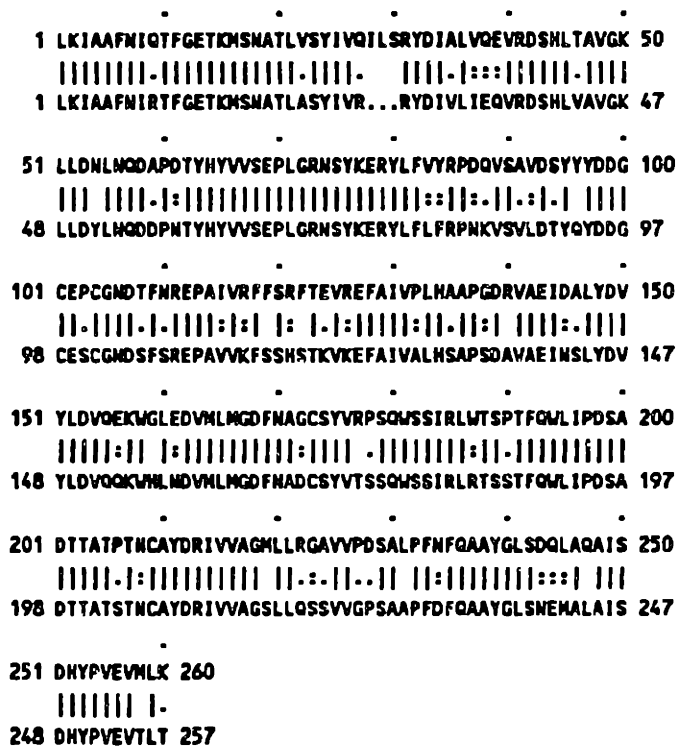


FIGURE 10.5: Homology plot for bovine DNAse (bottom sequence) and recombinant human DNAse (top sequence) determined as described in Materials and Methods. The solid lines indicate identical residues. Two dots indicate high similarity between residues and single dots indicate low similarity between residues. In addition, rhDNAse is glycosylated at both ^{18}Asn and ^{106}Asn.

bDNAse. Since rhDNAse has a sequence which is homologous to bDNAse which contains a hydrophobic core, rhDNAse could also have a hydrophobic core structure. Therefore, the stable species which was observed to aggregate in the equilibrium refolding studies of rhDNAse could be a hydrophobic intermediate with exposed beta sheets (Chapter 9). To confirm this assumption and determine if hydrophobic clusters exist in the structure of rhDNAse, the structure prediction methodology was applied to rhDNAse. The resulting two dimensional structure plot of rhDNAse revealed six potential hydrophobic cluster regions at approximately the same locations as observed for bDNAse. The presence of hydrophobic clusters in both CAB and rhDNAse analyses indicate that a relationship may exist between the PEG concentration required for enhancement and the number of hydrophobic clusters. Assuming a simple linear relationship between the hydrophobic clusters and the optimum ratio of PEG (3350 MW) to protein ($\{[PEG]_f/[Protein]_f\}_{opt}$), the optimum molar ratio of PEG to protein can be predicted for each protein:

$$\left(\frac{[PEG]_f}{[Protein]_f}\right)_{opt} = \left(\frac{[PEG]_f}{[CAB]_f}\right)_{opt}\left(\frac{\#\ of\ Clusters\ in\ Protein}{3\ Clusters\ in\ CAB}\right) \quad (10.1)$$

where $\{[PEG]_f/[CAB]_f\}_{opt}$ was determined as 3 from the previous studies (Section 5.2). This equation yields an optimum molar ratio ($[PEG]_f/[Protein]_f$) for rhDNAse of 6 to 1 which is close to the experimental optimum of 5 to 1 (Section 9.1). To further test this analysis method, the structure prediction algorithms were applied to rtPA and the two

dimensional structure plot of rtPA revealed seven potential hydrophobic clusters. Using Equation 10.1, the expected optimum molar ratio of PEG (3350 MW) to rtPA would then be 14 to 1. This result is much lower than the observed optimum ratio of 20 to 1. However, rtPA does have several regions of extended hydrophobic beta sheet structure which would likely require additional PEG to cover its surface. As a control, the structure prediction method was used for analysis of human pancreatic ribonuclease (RNAse). The analysis of RNAse indicated that it does not have any potential hydrophobic cluster sites and, therefore, may not interact with PEG. Since RNAse does not form a hydrophobic molten globule during refolding, the postulated relationship between the existence of hydrophobic clusters in the two dimensional plots and the formation of the molten globule intermediates could be a valid. In general, it may be possible to use the structural prediction techniques developed previously in conjunction with Equation 10.1 to determine the concentration of PEG required for enhancement (Chou & Fasman, 1978; Kyte & Doolittle, 1982).

10.2 Hydrophobic Index Correlation with PEG Enhancement

Since the hydrophobic cluster analysis method involves complex structure prediction algorithms and assumptions of specific protein structure interactions with PEG, a simpler model relating the hydrophobicity of each protein to its optimum PEG to protein molar ratio was also examined. The most common feature observed in each structural analysis was the grouping of hydrophobic residues as illustrated by the hydrophobic clusters marked in Figure 10.1. To generalize this observation, the overall hydrophobicity of each protein can be utilized for a relationship to the optimum molar

ratio. The overall hydrophobicity of the protein can be calculated from the hydrophobic index of each residue which has been determined by several different researchers (Eisenberg et.al, 1982). The hydrophobic index of each protein was calculated by using the primary sequence of each protein along with three different hydrophobicity scales for the amino acid residues (Janin, 1979; Eisenberg et.al, 1982; Nozaki & Tanford, 1971). The index of each protein was calculated by summing the hydrophobic contributions of each residue in the sequence. For each scale, the resulting hydrophobic indices indicated that rtPA was the most hydrophobic protein studied followed by rhDNAse which was more hydrophobic than CAB. The optimum molar ratio of PEG to protein determined from previous studies was then plotted as a function of the hydrophobicity index of each protein as shown in Figure 10.6. Each hydrophobicity index yielded a linear relationship between the protein hydrophobicity and the molar ratio required for enhanced refolding. Therefore, as a first approximation, the concentration of PEG required to enhance refolding can be estimated from the hydrophobic index of the protein. These results as well as the hydrophobic cluster analysis reveal a strong correlation between the hydrophobicity of the protein structure and the ability of PEG to prevent association of the protein.

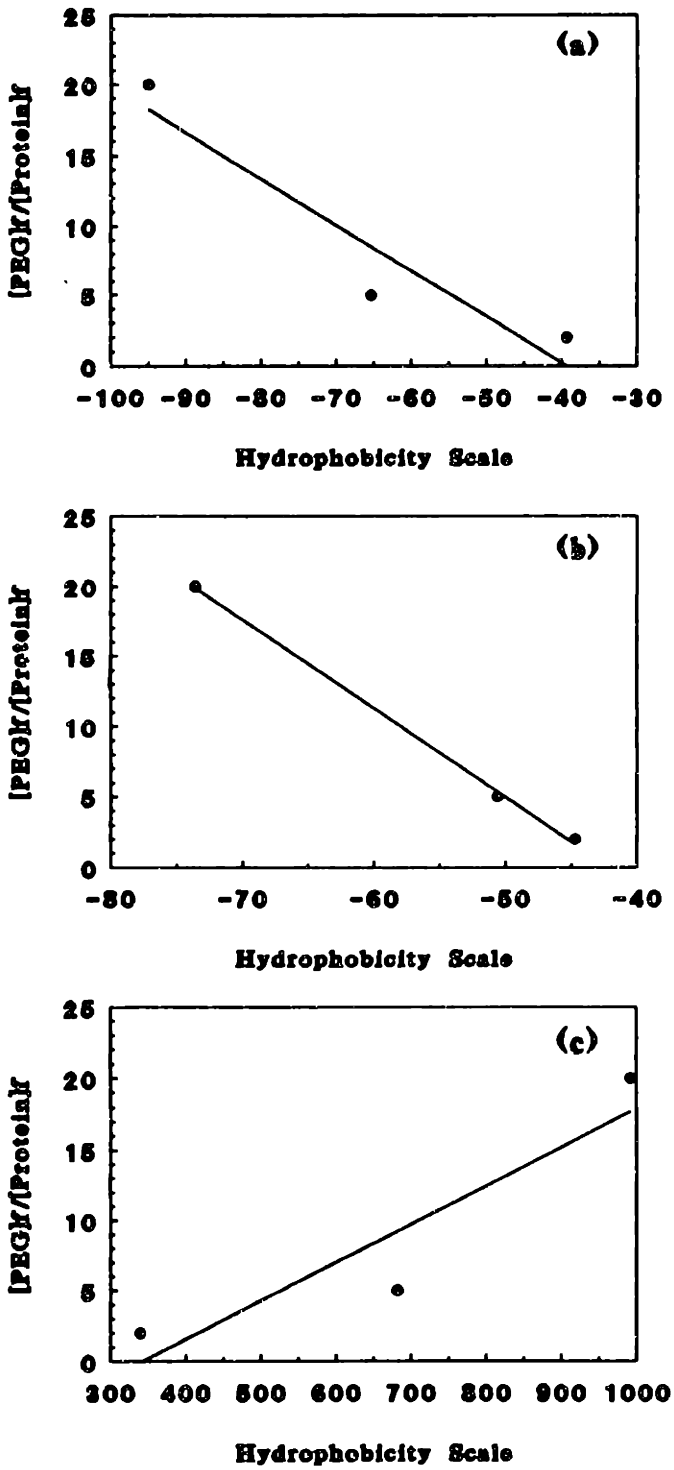


FIGURE 10.9: Optimum molar ratio of PEG (3350 MW) to protein ($[PEG]_f/[Protein]_f$) for CAB, rhDNAse and rtPA was 2, 5, and 20, respectively, and shown as a function of the hydrophobic index of each protein which was calculated by using three different scales: (a) Eisenberg et.al., 1982 (b) Janin, 1979 and (c) Nozaki & Tanford, 1971.

CONCLUSIONS

The pathway for aggregation and refolding was determined for the model protein, bovine carbonic anhydrase B (CAB). Kinetic and thermodynamic studies were used to elucidate the mechanisms of aggregation during refolding. From the kinetic analyses, the refolding and aggregation was shown to be dependent on the final conditions. Four different regimes of aggregation and refolding were observed and these regimes were constrained to be within a range of final protein and GuHCl concentrations. In the first regime, the formation of large aggregates occurred concomitant with dilution of denatured CAB in 5 M GuHCl to low GuHCl concentrations (< 0.50 M) and high protein concentrations (mg/ml) but the rapid aggregation kinetics could not be measured. Refolding at moderate GuHCl concentrations (0.50 - 1.0 M) and high protein concentrations (mg/ml) resulted in a slower rate of aggregation. In this regime, the formation of multimeric species was observed prior to precipitation. At these conditions, the first intermediate on the refolding pathway for CAB aggregates to form dimers and trimers both of which associated to produce micron-sized aggregates. When the final GuHCl concentration was increased to 1.0 M, two other regimes of aggregation and refolding were observed. For refolding at 1.0 M GuHCl and high protein concentrations (> 10 μM), the first intermediate formed a transient dimer and complete recover of activity was achieved. The final regime occurred at the refolding conditions of 1.0 M GuHCl and low protein concentrations (< 10 μM) where the protein recovers activity and does not associate. Association of the first intermediate was also observed in thermodynamic studies. Dimers and trimers of the first intermediate were formed for

equilibrium refolding at 2.0 M GuHCl and high protein concentrations (> 10 μM). The dimer equilibrium was slow ($t_{1/2}$ ~ 120 min) and the trimer equilibrium was more rapid ($t_{1/2}$ ~ 5 min). The kinetic and thermodynamic studies of aggregation and refolding provided insight into the pathways and kinetic models of aggregation during refolding.

Three different pathways for refolding and aggregation of CAB were developed to describe each of the regimes where the kinetics of the process were measured. For the regime of multimer formation prior to precipitation, the pathway was assumed to proceed by a rapid formation of the dimer from the first intermediate. Since the recovery of active protein was minimal at these conditions, folding of the first intermediate to the second intermediate was negligible compared to the association reaction. For the regime where a transient dimer was observed, the refolding pathway consisted of an equilibrium association of the first intermediate. This equilibrium was assumed to occur rapidly and the rate limiting step in refolding was the dissociation of the dimer. These assumptions were validated by using only one variable, the dissociation rate constant, to fit both the experimental activity and dimer concentration data. Finally, the refolding pathway in the absence of aggregation was modelled as a series of unimolecular reactions where the first intermediate folded to the second intermediate more rapidly than previously reported (Semisotnov et.al., 1990). In general, the models successfully predicted refolding and first intermediate aggregation which resulted in either a reduction of the rate or extent of folding to active protein and, in some cases, precipitation.

To prevent the aggregation of the first intermediate, polyethylene glycol (PEG) was utilized. PEG at stoichiometric concentrations (< 3 g/l) prevented association of the

first intermediate and, thereby, enhanced refolding. The concentration of PEG required to prevent aggregation was shown to be a function of the PEG molecular weight. In addition, only PEG ranging in molecular weight from 1000 to 8000 Daltons was effective in completely preventing aggregation. To understand the mechanism of this reduced aggregation, the interaction between PEG and the first intermediate was determined by equilibrium analyses. PEG bound only to the first intermediate in the refolding pathway and the binding occurred at a single site on the protein. In addition, when PEG bound to the first intermediate, an apparent change in the protein surface properties was observed. Equilibrium association of the first intermediate in 2.0 M GuHCl was also reduced in the presence of low PEG concentrations (< 10 g/l). Association of the first intermediate to transiently form a dimer during refolding was completely inhibited by PEG. The rate of refolding in the presence of PEG was shown to be the same as the refolding rate at nonassociating conditions. These results led to the development of a model pathway which included the formation of a PEG-first intermediate complex which does not aggregate. The equilibrium formation of the complex was assumed to occur more rapidly than dimerization and the refolding of the complex to the second intermediate occurred at the same rate as that determined for the unimolecular pathway. These assumptions led to an excellent fit between the experimental results and the model predictions which used the independently derived rate constants.

The ability of PEG to inhibit aggregation of other proteins was assessed by using three recombinant human proteins: deoxyribonuclease (rhDNAse), tissue plasminogen activator (rtPA), and gamma interferon (rIFN-γ). Refolding of rhDNAse and rtPA was

enhanced by PEG. Refolding of glycosylated rhDNAse by rapid dilution to aggregating conditions resulted in the complete recovery of active protein for a PEG (3350 MW) to protein molar ratio ($[PEG]_f/[CAB]_f$) of 5 to 1. In addition, a three fold increase in recovery of active rhDNAse was also achieved in the refolding of *E. coli* inclusion bodies of rhDNAse at a PEG (3350 MW) to protein molar ratio of 10. For refolding of the inclusion body protein, additional PEG was required to prevent aggregation of the contaminant proteins in the inclusion body. The recovery of active rtPA was increased two fold when refolded without disulfide reduction in the presence of a PEG (3350 MW) to protein molar ratio ($[PEG]_f/[rtPA]_f$) of 20. Unfortunately, refolding of rIFN-γ was not enhanced by PEG for the conditions attempted in this study. Relationships between the concentration of PEG required for enhancement in refolding and the protein properties were developed from these results. The presence of hydrophobic clusters which were defined as segments of hydrophobic beta sheet was correlated to the minimum concentration of PEG needed for refolding enhancement. By using the results from the CAB aggregation studies, the molar ratio of PEG to protein was predicted from the assumption that a linear relationship existed between PEG concentration and the number of hydrophobic clusters. This model predicted a PEG to protein molar ratio for enhancement of 6 for rhDNAse and 14 for rtPA. Since the hydrophobic cluster model required an assumption of specific interactions between PEG and protein structures, a more general relationship was derived based on the overall hydrophobicity of the protein. A correlation between the PEG to protein molar ratio and the protein hydrophobicity was linear for each of three different hydrophobicity scales. Therefore, the protein

hydrophobicity and, possibly, structure could dictate the concentration of PEG required for enhancement in refolding.

In general, aggregation during refolding of many proteins probably occurs through a hydrophobic intermediate. Several proteins form compact hydrophobic intermediates during refolding and the folding of these intermediates to other less hydrophobic intermediates or the native state is usually the rate limiting step in the process (Kuwajima, 1989; Ptitsyn et.al., 1990; Kim & Baldwin, 1990). The accumulation of these hydrophobic intermediates can result in aggregation and reduced recovery. PEG could bind to these intermediates and prevent aggregation while still allowing the protein to refold to its native state. Additional studies on the specific interaction of PEG with partially folded proteins should yield general strategies for preventing aggregation during refolding.

RECOMMENDATIONS FOR FUTURE WORK

The research on refolding and aggregation discussed in this thesis has led to the hypothesis of several potential mechanisms. However, the mechanisms of protein-protein and PEG-protein interactions are still not completely understood and deserve additional analysis. In addition, general protein structural arguments for these mechanisms should be developed through the study on other proteins. After developing an understanding of these mechanisms, rational design of polymers or synthetic folding aides could then be attempted. These molecules as well as PEG could also be used to prevent aggregation of proteins in formulations. Overall, there are several intriguing possibilities for future work on refolding and aggregation.

General Study of Intermediate Aggregation

The aggregation of the hydrophobic intermediates during refolding has been observed in this study as well as other refolding studies (Seckler et.al., 1989). It has been implied that this aggregation occurs by specific structural interactions which are similiar to those formed intramolecularly (Mitraki & King, 1989). To improve the understanding of aggregation, the study of specific protein structures which result in aggregation should be undertaken. This study would involve characterization of hydrophobic intermediates which occur during the refolding of several proteins. The initial choice of proteins could be derived from a list of proteins which form molten globule structures during refolding (Ptitsyn et.al., 1990). The similiarity in structure between these intermediates and their aggregation behavior should provide insight into the specific structural interactions which

cause aggregation. The postulated structural relationships could be tested by site-directed mutagenesis methods which result in either a stabilization of the intermediate or prevent the formation of the aggregating structural segment (see Literature Review).

Relationship between PEG and Specific Protein Structures

The studies of PEG enhancement in refolding lead to the hypothesis that the properties of the protein may dictate the relationship between PEG concentration and a reduction in aggregation (see Chapter 10). Since PEG was shown to prevent aggregation in the refolding of several proteins and bind to a specific intermediate, PEG could also prevent the aggregation of other proteins which form similiar folding intermediates. First of all, the specific site for binding between PEG and the first intermediate in the refolding of CAB must be determined. Thus, a study should be performed to investigate the specific location of PEG binding and determine if similiar sites exist on the other proteins studied in this work. This study should indicate the specific aggregation sites on each protein and the affinity of PEG for these sites. In addition, the rate of PEG binding to these intermediates should be measured and compared to the rate of aggregation in order to develop a mechanistic model. The results of this study should reveal the type of intermediate structures which aggregate and the specific interaction of PEG with these structures.

The association and dissociation of PEG with the protein should also be further characterized. Although preliminary HPLC size exclusion and reverse phase studies seemed to indicate that all the PEG molecules were released from the protein, it is uncertain at this time whether PEG is quantitatively dissociatiated from the protein during

refolding. Radiolabelled PEG could be used to probe the refolded proteins to assess the fate of the PEG molecule.

Rational Design of PEG Analogs and Folding Aides

After developing a better understanding of the specific PEG-protein and protein-protein interactions, specific molecules which block aggregation might be developed. In particular, the knowledge of PEG-protein interactions should provide insight into modifications of PEG. These PEG analogs could be designed to bind with different affinities to the specific protein structures which cause aggregation. In addition, these analogs could incorporate some of the catalytic properties which were observed for molecular chaperones (see Literature Review). For example, the analog could block aggregation by binding to the hydrophobic surface and, then, catalyze disulfide bond formation. This analog could be easily made through the addition of various thiol groups to PEG. The thiols could be placed at the termini of the polymer or distributed throughout the polymer backbone. However, precautions should be taken to avoid modifications which significantly change the structural properties of PEG. If the specific polymer properties which are required for preventing aggregation are known, other polymers or molecules could be synthesized and used as folding aides. These folding aides would act in a similiar manner to PEG or its analogs, but they could be designed with different affinities. The rational design of folding aides could be performed with additional knowledge of refolding and aggregation. These designs would assist in determining the mechanisms of both folding and aggregation.

Application of PEG and Polymers to Protein Stability in Formulations

Aggregation of proteins also occurs in the area of formulations. For example, when protein solutions are stored for extended periods, aggregation of the solution can occur resulting in product loss. The aggregation is likely caused by the association of an intermediate protein structure which is in equilibrium with native state. In addition, reconstitution of lyophilized proteins often results in the formation of aggregates since the lyophilization process can partially denature the protein by subjecting it to extreme conditions (low temperature, pH, etc.). Therefore, it could be assumed that the methods used to block aggregation during refolding may also be applied to protein formulations. Unlike refolding and aggregation, the stabilizer must be biodegradable and/or accepted for use by the FDA. Therefore, biodegradable polymers with similiar properties to PEG could be used to enhance the stability of proteins both in solution and during lyophilization. If protein aggregation in formulations occurs by a mechanism which is similiar to that observed during refolding, the same general rules can be applied to improve the stability of protein formulations.

Molecular Modelling of Protein Structure and Cosolvent Assisted Protein Refolding

The present analysis on the protein structure with emphasis on the hydrophobic clusters and their interactions with cosolvents such as PEG has promised some preliminary, but interesting, data interpretation. It is recommended that future research be performed in this direction in order to provide useful guides and rules for cosolvent assisted protein refolding. For example, from the primary protein sequence analysis, one could hypothesize and test the relative strengths of cosolvent association and dissociation

with these hydrophobic clusters. If one can extend this present analysis to a larger number of proteins, predictive refolding enhancements with different cosolvents could be developed.

In view of the fact that the different hydrophobic clusters within a given protein have different degrees of affinity or association with the cosolvent, one might further ask whether multiple cosolvents should be examined for enhanced refolding. More specifically, what are the fundamental rules to assess in concert the interactions between the protein and cosolvent to aid in protein refolding.

NOMENCLATURE

a	Persistence length of multimer (ηm)
a_s	Activity coefficient of cosolvent
$[AAA]_f$	Concentration of acetazolamide after dilution (μM)
$[CAB]_{active}$	Active protein concentration (mg/ml or μM)
$[CAB]_{bound}$	Concentration of protein bound to polymer (μM)
$[CAB]_f$	Final protein concentration after dilution (mg/ml or μM)
d_s	Diameter of unit monomer (ηm)
$[D]$	Dimer concentration (μM)
$[D]_f$	Final concentration of dimer measured by QLS prior to precipitation (μM)
D_h	Hydrodynamic diameter of particle in solution (ηm)
D_R	Rotational diffusivity (1/sec)
D_s	Particle diffusivity in solution (cm/sec^2)
D_{si}	Diffusivity of particle i in solution (cm/sec^2)
$[D]_{t=0}$	Dimer concentration immediately after dilution (μM)
$[D]_t$	Dimer concentration at given time, t
E	Energy level of resonating electron
F	Fluorescence of sample (intensity)
F_o	Fluorescence of control sample (intensity)
F_p	Fluorescence of solution with PEG (intensity)
$G(\tau)$	Autocorrelation function from QLS analysis
G_{io}	Autocorrelation value for species i at $\tau = 0$

$[GuHCl]_f$	Final GuHCl concentration after dilution (M)
H_o	Magnetic field strength (Gauss)
$[I_1]$	Concentration of first intermediate in CAB refolding (μM)
$[I_1]_{t=0}$	CAB first intermediate concentration immediately after dilution (μM)
$[I_1]_t$	CAB first intermediate concentration at given time, t
$[I_1P]$	CAB first intermediate - PEG complex concentration (μM)
$[I_2]$	Concentration of second intermediate in CAB refolding (μM)
$[I_2]_t$	CAB second intermediate concentration at given time, t
k_{AAA}	Rate constant for binding of AAA to native protein (1/min)
k_{AAAi2}	Rate constant for refolding of CAB second intermediate with AAA (1/min)
k_{app}	Apparent rate constant for formation of second intermediate in CAB refolding pathway (1/min)
k_B	Boltzman's constant
k_d	Association rate constant for dimer formation ($\mu M^{-1}min^{-1}$)
k'_d	Dissociation rate constant for dimer (1/min)
k^G_d	Association rate constant for dimer formation independent of GuHCl concentration
k_{i1}	Rate constant for formation of first intermediate in CAB refolding (1/sec)
k_{i2}	Rate constant for formation of second intermediate in CAB refolding (1/min)
k_n	Rate constant for folding to native state (1/min)
k_p	Association rate constant for first intermediate-PEG complex
k'_p	Dissociation rate constant for first intermediate-PEG complex
k_t	Association rate constant for trimer formation ($\mu M^{-1}min^{-1}$)

k'_t	Dissociation rate constant for trimer (1/min)
K	System and solution constant for autocorrelation function $(sec^2/cm)^2$
K_D	Equilibrium constant for dimer formation (μM^{-1})
K^P_D	Equilibrium constant for dimer in the presence of PEG (μM^{-1})
K_i	Equilibrium constant for inhibitor binding (μM)
K_{I1P}	Equilibrium constant for CAB first intermediate-PEG association (μM^{-1})
K_{SV}	Collisional quenching constant (M^{-1})
K_T	Equilibrium constant for trimer formation (μM)
[L]	Free ligand concentration (mM)
L_i	Contour length of multimeric specie, i (ηm)
m_i	Molality of water (w), protein (p) or cosolvent (s)
[N]	Native protein concentration (μM)
$[N]_t$	Native protein concentration at given time, t
[P]	PEG concentration (mM)
P	Pressure of system
$[P_o]$	Total protein concentration (μM)
$[PEG]_{eff}$	Effective concentration of PEG at protein surface (μM)
$[PEG]_f$	Final concentration of PEG after dilution (g/l)
$[PEGpNP]_{eff}$	Effective concentration of PEGpNP at protein surface (μM)
$[PEGSA]_f$	Concentration of PEGSA after dilution (g/l)
[PL]	Concentration of ligand-protein complex (μM)
$[pNP]_{eff}$	Effective concentration of p-nitrophenol (pNP) at protein surface (μM)

ΔpNP_{Native}	Change in p-nitrophenol (pNP) concentration over time interval, Δt, for native protein
$[pNP]_{PEGpNP}$	p-nitrophenol (pNP) concentration for refolding with PEGpNP (μM)
$[pNP]_{PEG+pNPA}$	p-nitrophenol (pNP) concentration for refolding with PEG with pNPA (μM)
ΔpNP_{Refold}	Change in p-nitrophenol (pNP) concentration over time interval, Δt, for refolding protein
$[Protein]_f$	Protein concentration after dilution
$[Q]$	Concentration of quenching agent (M)
R_D	Initial rate of dimer formation (μM/min)
R_g	Radius of gyration for particle (ηm)
R_h	Hydrodynamic radius of a particle in solution (ηm)
R_{Ref}	Initial rate of refolding (1/min)
$[rhDNAse]_f$	Recombinant human deoxyribonuclease concentration after dilution
$[rIFN\text{-}\gamma]_f$	Recombinant human gamma interferon concentration after dilution
$[rtPA]_f$	Recombinant human tissue plasminogen activator concentration after dilution
$[SA]_f$	Sulfonamide inhibitor concentration after dilution (μM)
$[spPEG]_f$	Concentration of spin-labelled PEG after dilution (μM)
t	Time after dilution to final conditions (min or sec)
$t_{1/2}$	Half time for refolding reaction step (min or sec)
t_f	Time of final QLS measurement prior to precipitation (min)
$[T]$	Concentration of trimer (μM)
T	Temperature of solution (K)
$[U]$	Unfolded protein concentration (μM)

$[Urea]_f$	Concentration of urea after dilution (M)
V	Static quenching constant (M^{-1})
Γ_i	Time constant for decay of autocorrelation function for species i (1/sec)
η	Kinematic viscosity of the solution (cp)
θ	Angle of light scattering measurement relative to incident light (radians)
Θ	Concentration of protein bound to ligand ([PL]) divided by total protein concentration ($[P_o]$)
λ	Wavelength of incident laser light in QLS analysis (488 ηm)
λ_{em}	Emission wavelength used in fluorescence measurements (ηm)
λ_{ex}	Excitation wavelength used in fluorescence measurements (ηm)
μ_i	Chemical potential of water (w), protein (p) or cosolvent (s)
τ	Sample time for autocorrelation data (sec)
$\tau_{baseline}$	Sample time at which autocorrelation data reaches the baseline value (sec)
τ_c	Rotational correlation time (sec)

REFERENCES

Abbott, N. (1990) Personal Communication, Chemical Engineering Department, Massachusetts Institute of Technology, Cambridge. MA.

Al-Obeidi, A. M. and Light, A. (1988) *J. Biol. Chem.*, **263**, 8642-8645.

Altenbach, C., Marti, T., Khorana, G. H., and Hubbell, W. L. (1990) *Science*, **248**, 1088-1092.

Anglister, J. (1989) *In:* Bioological Magnetic Resonance, v. 8, L. J. Berliner and J. Reuben (eds.), Plenum Press, New York, 597-614.

Arakawa, T. and Timasheff, S. N. (1984) *Biochemistry*, **23**, 5924-5929.

Arawaka, T. and Timasheff, S. N. (1985) *Biochemistry*, **24**, 6756-6762.

Arakawa, T., Alton, N. K., and Hsu, Y.-R. (1985) *J. Biol. Chem.*, **260**, 14435-14441.

Arakawa, T. and Hsu, Y.-R. (1987) *Biochemistry*, **26**, 5428-5434.

Arakawa, T., Bhat, R., and Timasheff, S.N. (1990) *Biochemistry*, **29**, 1924.

Barratt, M. D., Dodd, G. H., and Chapman, D. (1969) *Biochim. Biophys. Acta*, **194**, 600-602.

Bashford, D., Cohen, F. E., Karplus, M., Kuntz, I. D., and Weaver, D. L. (1988) *Prot. Struct. Funct. Gen.*, **4**, 211-227.

Beckman, R. P., Mizzen, L. A., and Welch, W. J. (1990) *Science*, **248**, 850-854.

Berliner, L. J. (1981) *In*: Spectroscopy in Biochemistry, v. 2, CRC Press, Boca Raton, Florida, 1-56.

Bloomfield, V. A. and Lim. T. K. (1978) *Methods Enzymol.*, **48**, 415-494.

Bowden, G. A. and Georgiou, G. (1988) *Biotech. Prog.*, **4**, 97-101.

Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) *Science*, **247**, 1306-1310.

Brems, D. N., Lin, Y. C., and Stellwagen, E. (1982) *J. Biol. Chem.*, **257**, 3864-3869.

Brems, D. N., and Stellwagen, E. (1983) *J. Biol. Chem.*, **258**, 3655-3661.

Brems, D. N., Plaisted, H. A., Kauffman, E. W., Stodola, J. D., Easton, L. C., and White,R. D. (1985) *Biochemistry*, **24**, 7662-7668.

Brems, D. N., Plaisted, S. M., Kauffman, E. W., and Havel, H. A. (1986) *Biochemistry*, **25**, 6539-6543.

Brems, D. N., Plaisted, S. M., Dougherty, J. J., and Holzman, T. F. (1987) *J. Biol. Chem.*, **262**, 2590-2596.

Brems, D. N., Plaisted, S. M., Havel, H. A., and Tomich, C.-S. C. (1988a) *P.N.A.S.*, **85**, 3367-3371.

Brems, D. N. (1988b) *Biochemistry*, **27**, 4541-4546.

Builder, S. E. and Ogez, J. R. (1985) *U.S. Patent 4,511,502*

Capiomont, A., Chion, B., Lajzerowicz, J., and Lemaire, H. (1974) *J. Chem. Phys.*, **60**, 2530-2535.

Carlson, J. D., Kim, P. S., and Yarmush, M. L. (1990) *Personal Communication*

Char, K., Gast, A. P., and Frank, C. W. (1988) *Langmuir*, **4**, 989-998.

Char, K., Frank, C. W., and Gast, A. P. (1989) *Langmuir*, **5**, 1096-1105.

Chou, P. Y. and Fasman, G. D. (1978) *Adv. Enzymol.*, **47**, 45-147.

Corbett, R. J. T. and Roche, R. S. (1984) *Biochemistry*, **23**, 1888-1894.

Coulter Electronics (1984) Coulter Model N4 Manual, Hialeah, Florida.

Cousens, L. S., Shuster, J. R., Gallegos, C., Ku, L., Stempien, M. M. Urdea, M. S., Sanchez-Pescador, R., Taylor, A., and Tekamp-Olson, P. (1987) *Gene*, **61**, 265-275.

Creighton, T. E. (1984) *Proteins: Structures and Molecular Properties*, W.H. Freeman and Company, New York.

Creighton, T. E. (1985) *In:* Protein Structure Folding and Design, Oxender, D. L. (ed.), Alan R. Liss, Inc., New York, pp. 249-251.

Damodaran, S. (1987) *Biochim. Biophys. Acta*, **914**, 114-121.

Darnell, J. E., Lodish, H., and Baltimore, D. (1986) *Molecular Cell Biology*, Chapter 4 Scientific American Books, W.H. Freeman & Co., New York, NY.

Devereux, J. (1989) *GCG Sequence Analysis Software Package, Version 6.0*, Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, Wisconsin.

Dill, K. (1990) *Biochemistry*, **29**, 7133-7155.

Doligkh, D. A., Kolomiets, A. P., Bolotina, I. A., and Ptitsyn, O. B. (1984) *FEBS Letters*, **165**, 88-92.

Donovan, J. M., Benedek, G. B., and Carey, M. C. (1987) *Biochemistry*, **26**, 8116-8125.

Ebert, B., Semisotnov, G. V., and Rodionova, N. A. (1990) *Studia Biophys.*, **137**, 125-132.

Eftink, M. R. and Ghiron, C. A. (1981) *Anal. Biochem.*, **114**, 199-227.

Eisenberg, D., Weiss, R. M., Terwilliger, T. C., and Wilcox, W. (1982) *Faraday Symp. Chem. Soc.*, **17**, 109-120.

Endo, S., Saito, Y., and Wada, A. (1983) *Anal. Biochem.*, **131**, 108-210.

Eriksson, A. E., Jones, T. A., and Liljas, A. (1988) *Prot. Struct. Funct. Gen.*, **4**, 274-282.

Fane, B., Villafane, R., Mitraki, A., and King, J. (1991) *J. Biol. Chem., in press.*

Feher, G. and Kam, Z. (1985) *Meth. Enzymology*, **114**, 49-76.

Fields, R. (1971) *Biochem. J.*, **124**, 581-590.

Fish, W. W., Danielsson, A., Nordling, K., Miller, S. H., Lam, C. F., and Bjork, I. (1985) *Biochemistry*, **24**, 1510-1515.

Freedman, R. B. (1989) *Cell*, **57**, 1069-1072.

Friberger, P. (1982) *Scand. J. Clin. Lab. Invest.*, **42 (S162)**, 15-98.

Fuke, Y., Sekiguchi, M., and Matsuoka, H. (1985) *J. Food Sci.*, **50**, 1283-1289.

Garvey, E. P. and Matthews, C. R. (1989) *Biochemistry*, **28**, 2083-2093.

Gast, A.P. (1990) Personal Communication, Department of Chemical Engineering, Stanford University, Palo Alto, California

Georgiou, G., Telford, J. N., Shuler, M. L., and Wilson, D. B. (1986) *Appl. Environ. Mircobiol.*, **52**, 1157-1161.

Ghelis, C., and Yon, J. (1982) *Protein Folding*, Academic Press, New York.

Goldenberg, D. P. (1988) *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 481-507.

Goldenberg, D. P., Smith, D. H., and King, J. (1983) *P.N.A.S.*, **80**, 7060-7064.

Goloubinoff, B., Gatenby, A. A., and Lorimer, G. (1989) *Nature*, **337**, 44-47.

Gribskov, M. and Burgess, R. R. (1986) *Nucl. Acids Res.*, **14**, 6745-6763.

Haase-Pettingell, C. and King, J. (1988) *J. Biol. Chem.*, **263**, 4977-4983.

Hadju, J., Bartha, F., and Friedrich, P. (1976) *Eur. J. Biochem.*, **68**, 373-383.

Haffner, P. H. and Coleman, J. E. (1975) *J. Biol. Chem.*, **250**, 996-1005.

Hagen, A., Hatton, T. A., and Wang, D. I. C. (1990) *Biotech. Bioeng.*, **35**, 966-975.

Hardy, S. J. S. and Randall, L. L. (1991) *Science*, **251**, 439-443.

Hartl, F.-U. and Neupert, W. (1990) *Science*, **247**, 930-938.

Havel, H. A., Kauffman, E. W., Plaisted, S. M., and Brems, D. N. (1986) *Biochemistry*, **25**, 6533-6538.

Havel, H. A., Kauffman, E. W., and Elzinga, P. A. (1988) *Biochim. Biophys. Acta*, **955**, 154-163.

Henkens, R. W., Kitchell, B. B., Lottich, S. C., Stein, P. J., and Williams, T. J. (1982) *Biochemistry*, **21**, 5918-5923.

Holzman, T. F., Dougherty, J. J., Brems, D. N., and McKenzie, N. E. (1990) *Biochemistry*, **29**, 1255-1261.

Horne, R. W., Hobart, J. M., and Pasquali-Ronchetti, I. (1975) *J. Ultrastruct. Res.*, **53**, 319-330.

Horne, R. W. (1978) *J. Microscopy*, **113**, 241-256.

Horwich, A. L., Neuport, W., and Hartl, F.-U. (1990) *TIBTECH*, **8**, 126-131.

Horwitz, P. and Criscimagna, N. L. (1986) *J. Biol. Chem.*, **261**, 15652-15658.

Hower, J. F., Henkens, R. W., and Chestnut, D. B. (1971) *J. A.C.S*, **93**, 6665-6671.

Hsu, Y.-R. and Arakawa, T. (1985) *Biochemistry*, **24**, 7959-7966.

Ikai, A., Tanaka, S., and Noda, H. (1978) *Arch. Biochem. Biophys.*, **190**, 39-45.

Jameson, T. and Wolf, J. (1988) *CABIOS*, **4**, 181-186.

Janin, J. (1979) *Nature*, **277**, 491-493.

Jedziniak, J. A., Nicoli, D. F., Baram, H., and Benedek, G. B. (1978) *Invest. Ophthalmol. Visual Sci.*, **17**, 51-57.

Kadima, W., McPherson, A., Dunn, M. F., and Jurnak, F. A. (1990) *Biophys. J.*, **57**, 125-132.

Kane, J. F. and Hartley, D. L. (1988) *TIBTECH*, **6**, 95-101.

Karplus, M. and Weaver, D. L. (1976) *Nature*, **260**, 104-111.

Karplus, M. and Weaver, D. L. (1979) *Biopolymers*, **18**, 1421.

Kim, P. S. and Baldwin, R. L. (1982) *Ann. Rev. Biochem.*, **51**, 459-472.

Kim, P. S. and Baldwin, R. L. (1990) *Ann. Rev. Biochem.*, **59**, 631-660.

King, J. (1986) *Biotechnology*, **4**, 297-303.

King, J., Haase, C., and Yu, M. (1987) *Prot. Eng.*, **1**, 109-121.

King, J. (1989) *Chem. Eng. News*, **67**, 32-54.

Kosman, D. and Bereman, R. (1981) *In*: Spectroscopy in Biochemistry, v. 2, CRC Press, Boca Raton, Florida, 57-108.

Kratky, O. and Porod, G. (1949) *Rec. Trav. Chim. 68*, 1106-1122.

Kuwajima, K. (1989) *Prot. Struct. Funct. Gen.*, **6**, 87-103.

Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol.*, **157**, 105-132.

Lang, K., Schmid, F. X., and Fischer, G. (1987) *Nature*, **329**, 268-270.

Lang, K. and Schmid, F. X. (1988) *Nature*, **331**, 453-455.

Laurie, O. and Oakes, J. (1976) *J. Chem. Soc., Faraday Trans. I*, **72**, 2681-2683.

Lee, J. C. and Timasheff, S. N. (1974) *Biochemistry*, **13**, 257-265.

Liao, T.-H. (1975) *J. Biol. Chem.*, **250**, 3831-3836.

Likhtenshtein, G. I. (1976) *Spin Labeling Methods in Molecular Biology*, John Wiley & Sons, New York.

Liljas, A., Kannan, K. K., Bergstén, P.-C., Waara, I., Fridborg, K., Strandberg, B., Carlbom, U., Järup, L., Lövgren, S., and Petef, M. (1972) *Nature New Biology*, **235**, 131-137.

Lim, W. A. and Sauer, R. T. (1989) *Nature*, **339**, 31-36.

Lizarraga, B., Bustamante, C., Gil, A., and Melgar, E. (1979) *Biochim. Biophys. Acta*, **579**, 298-302.

Lodish, H. (1988) Personal Communication, Whitehead Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA.

Loscalzo, J. and Handin, R. I. (1984) *Biochemistry*, **23**, 3880-3886.

Loscalzo, J. (1988) *J. Clin. Invest.*, **82**, 1391-1397.

Malinkowski, E. R. and McCue, M. (1977a) *Anal. Chem.*, **49**, 284-287.

Malinkowski, E. R. (1977b) *Anal. Chem.*, **49**, 606-617.

Marston, F. A. O. (1986) *Biochem. J.*, **240**, 1-12.

Martenson, R. E. (1978) *J. Biol. Chem.*, **253**, 8887-8893.

Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature*, **346**, 440-445.

McDonnell, M. E. and Jamieson, A. M. (1976) *Biopolymers*, **15**, 1283-1299.

Meyer, D. I. (1982) *TIBS*, **2**, 320-321.

Mikol, V., Hirsch, E., and Giegé, R. (1989) *FEBS Letters*, **258**, 63-66.

Mitraki, A., Betton, J.-M., Desmadril, M., and Yon, J. M. (1987) *Eur. J. Biochem.*, **163**, 29-34.

Mitraki, A. and King, J. (1989) *Biotechnology*, **7**, 690-697.

Mitraki, A., Haase-Pettingeil, C., and King, J. (1991) *In:* Protein Refolding, A.C.S. Symposium Series, G. Georgiou (ed.), *in press.*

Morris, G. E., Frost, L. C., Newport, P. A., and Hudson, N. (1987) *Biochem. J.*, **248**, 53-57.

Mulkerrin, M. and Wetzel, R. (1989) *Biochemistry*, **28**, 6556-6561.

Murphy, R. M., Slayter, H., Schurtenberger, P., Chamberlin, R. A., Colton, C. K., and Yarmush, M. L. (1988) *Biophys. J.*, **54**, 45-56.

Murphy, R.M. (1989) Department of Chemical Engineering, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Murphy, R. M., Chamberlin, R. A., Schurtenberger, P., Colton, C. K., and Yarmush, M. L. (1990) *Biochemistry*, **29**, 10889-10899.

Mushak, P. and Coleman, J. E. (1972) *J. Biol. Chem.*, **247**, 373-380.

Napier, R. M., East, J. M., and Lee, A. G. (1987) *Biochim. Biophys. Acta*, **903**, 365-373.

Nicoli, D. F. and Benedek, G. B. (1976) *Biopolymers*, **15**, 2421-2437.

Nishikawa, K. (1983) *Biochim. Biophys. Acta*, **748**, 285-299.

Nordio, P. L. (1976) *In:* Spin Labelling: Theory and Applications, Berliner, L. J. (ed.), Academic Press, San Francisco, California, 5-52.

Nozaki, Y. and Tanford, C. (1971) *J. Biol. Chem.*, **246**, 2211-2217.

Oakes, J. and Cafe, M. C. (1973) *Eur. J. Biochem.*, **36**, 559-563.

Okabe, N., Fujita, E., and Tomita, K.-I. (1982) *Biochim. Biophys. Acta*, **700**, 165-170.

Park, S., Liu, G., Topping, T. B., Cover, W. H., and Randall, L. L. (1988) *Science*, **239**, 1033-1035.

Patel, S. S. (1986) Department of Applied Biological Sciences, M.S. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., and Collen, D. (1983) *Nature*, **301**, 214-221.

Pesce, A. J., Rosen, C.-G. and Pasby, T. L. (1971) *Fluorescence Spectroscopy: An Introduction For Biology and Medicine*, Chapter 7, Marcel Dekker, Inc. New York.

Pocker, Y. and Stone, J. T. (1967) *Biochemistry*, **6**, 668-678.

Porter, D. H., and Cardenas, J. M. (1980) *Biochemistry*, **19**, 3447.

Prakash, V., Loucheux, C., Scheufele, S., Gorbunoff, M. J., and Timasheff, S. N. (1981) *Arch. Biochem. Biophys.*, **210**, 455-464.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989) *Numerical Recipes*, Cambridge University Press, London, 335-449.

Price, P. A. (1975) *J. Biol. Chem.*, **250**, 1981-1986.

Provencher, S. W. (1983) *In:* Photon Correlation Techniques, Schulz-Dubois, E. O. (ed.), Springer-Verlag: New York, NY, 322-328.

Ptitsyn, O. B. (1987) *J. Protein Chem.*, **6**, 273.

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., Razgulyaev, O. I. (1990) *FEBS Lett.*, **262**, 20-24.

Radek, J. T. and Castellino, F. J. (1988) *Arch. Biochem. Biophys.*, **267**, 776-786.

Randall, L. L. and Hardy S. J. S. (1984) *In*: Modern Cell Biology, v. 3, 1-20, A.R. Liss, Inc. New York, NY.

Randall, L. L., Hardy, S. J. S., and Thom, J. R. (1987) *Annu Rev. Microbiol.*, **41**, 507-541.

Reithel, F. J. (1963) *Adv. Prot. Chem.*, **18**, 124-226.

Rijken, D. C., Hoylaerts, M., and Collen, D. (1982) *J. Biol. Chem.*, **257**, 2920-2925.

Rinderknecht, E., O'Connor, B. H., and Rodriquez, H. (1984) *J. Biol. Chem.*, **259**, 6790-6798.

Robbins, P. D. (1989) Department of Chemical Engineering, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 97-99.

Robinson, A. and Austin, B. (1987) *Biochem. J.*, **246**, 249-261.

Rodionova, N. A., Semisotnov, G. V., Kutyshenko, V. P., Uverskii, V. N., Bolotina, I. A., Bychkova, V. E., and Ptitsyn, O. B. (1989) *Mol. Biol. (Moskow)*, **23**, 683-692.

Rothman, J. E. (1989) *Cell*, **59**, 591-601.

Rudolph, R., Zettlmeissl, G., and Jaenicke, R. (1979) *Biochemistry*, **18**, 5572-5575.

Sabatini, D. D., Kreibich, G., Morimoto, T., and Adesnik, M. (1982) *J. Cell. Biol.*, **92**, 1-22.

Sarmientos, P., Duchesne, M., Denefle, P., Boiziau, J., Fromage, N., Delporte, N., Parker, F., Lelievre, Y., Mayaux, J.-F., and Cartwright, T. (1989) *Biotechnology*, **7**, 495-501.

Schein, C. H. (1989) *Biotechnology*, **7**, 1141-1149.

Schneider, D. J. and Freed, J. H. (1989) *In:* Bioological Magnetic Resonance, v. 8, L. J. Berliner and J. Reuben (eds.), Plenum Press, New York, 1-76.

Sciaky, M., Limozin, N., Filippi-Foveau, D., Gulian, J. M., and Laurent-Tabusse, G. (1976) *Biochimie*, **59**, 1071-1082.

Seckler, R., Fuchs, A., King, J., and Jaenicke, R. (1989) *J. Biol. Chem.*, **264**, 11750-11753.

Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J., and Ptitsyn, O.B. (1987) *FEBS Letters*, **224**, 9-13.

Semisotnov, G. V., Uversky, V. N., Sokolovsky, I. V., Gutin, A. M., Razgulyaev, O. I., and Rodionova, N. A. (1990) *J. Mol. Biol.*, **213**, 561-568.

Shalongo, W., Ledger, R., Jagannadham, M. V., and Stellwagen, E. (1987) *Biochemistry*, **26**, 3135-3141.

Shalongo, W., Jagannadham, M. V., Flynn, C. and Stellwagen, E. (1989) *Biochemistry*, **28**, 4820-4825.

Shortle, D. (1989) *J. Biol Chem.*, **264**, 5315-5318.

Smith, T. F. and Waterman, M. S. (1981) *Adv. Appl. Math.*, **2**, 482-489.

Sokolovsky, M., Riordan, J. F., and Vallee, B. L. (1966) *Biochemistry*, **5**, 3582-3589.

Spellman, M. W., Basa, L. J., Leonard, C. K., Chakel, J. A., and O'Connor, J. V. (1989) *J. Biol. Chem.*, **264**, 14100-14111.

Stein, P. J. and Henkens, R. W. (1978) *J. Biol. Chem.*, **253**, 8016-8018.

Stellwagen, E. and Wilgus, H. (1978) *Nature*, **275**, 342-347.

Stellwagen, E. (1979) *J. Mol. Biol.*, **135**, 217-229.

Stock, R. S. and Ray, W. H. (1985) *J. Polym. Sci.: Polym. Phys.*, **23**, 1393-1447.

Sturtevant, J. M., Yu, M., Haase-Pettingell, C., and King, J. (1989) *J. Biol. Chem.*, **264**, 10693-10698.

Suck, D., Oefner, C., Kabsch, W. (1984) *EMBO Journal*, **3**, 2423-2430.

Swaney, J. B. and O'Brien, K. (1978) *J. Biol. Chem.*, **253**, 7069-7077.

Tandon, S. and Horwitz, P. M. (1987) *J. Biol. Chemistry*, **262**, 4486-4491.

Tandon, S. and Horwitz, P. M. (1986) *J. Biol. Chemistry*, **261**, 15615-15618.

Teschner, W., Rudolph, R., and Garel, J.-R. (1987) *Biochemistry*, **26**, 2791-2796.

Thiyagarajan, P. and Johnson, M. E. (1983) *Biophys. J.*, **42**, 269-274.

Tsong, T. Y. (1982) *Biochemistry*, **21**, 1493-1497.

Vaucheret, H., Signon, L., Le Bras, G., and Garel, J.-R., (1987) *Biochemistry*, **26**, 2785-2789.

Vidgren, J., Liljas, A., and Walker, N. P. C. (1990) *Int. J. Biol. Macromol.*, **12**, 342-344.

Walter, P., Ibranimi, I., and Blobel, G. (1981) *J. Cell. Biol.*, **91**, 545-550.

Wang, L.-F., Hum, W. T., Kalyan, N. K., Lee, S. G., Hung, P. P. and Doi, R. H. (1989) *Gene*, **84**, 127-133.

Wee, V. T., Feldmann, R. J., Tanis, R. J., and Chignell, C. F. (1976) *Mol. Pharm.*, **12**, 832-843.

Winkler, M. (1990) Personal Communication, Genentech, Inc., S. San Francisco, CA.

Wong, K.-P. and Tanford, C. (1973) *J. Biol. Chem.*, **248**, 8518-8523.

Wright, P. E., Dyson, H. J., Lerner, R. A. (1988) *Biochemistry*, **27**, 7167-7175.

Wright, G. and Freedman, R. B. (1989) *Protein Eng.*, **2**, 583-588.

Yarmush, D. L., Morel, G., and Yarmush, M. L. (1987) *J. Biochem. Biophys. Methods*, **14**, 279-289.

Yarmush, D. M., Murphy, R. M., Colton, C. K., Fisch, M., and Yarmush, M. L. (1988) *Mol. Immun.*, **25**, 17-32.

Yazgan, A. and Henkens, R. W. (1972) *Biochemistry*, **11**, 1314-1318.

Yphantis, D. A. and Arakawa, T. (1987) *Biochemistry*, **26**, 5422-5429.

Zettlmeissl, G., Rudolph, R., and Jaenicke, R. (1979) *Biochemistry*, **18**, 5567.

APPENDICES

APPENDIX 1: INCLUSION BODY ANALYSIS

Recombinant Human Deoxyribonuclease (rhDNAse)

As described in Materials and Methods, insoluble inclusion body pellets of rhDNAse were dissolved in 8 M urea with 10 mM GSH, 1 mM GSSG, 50 mM Tris, 10 mM $CaCl_2$, and 10 mM $MgCl_2$ at pH 8. The aggregated protein was mixed in this solution for 5 hours. Any remaining insoluble protein was then removed by filtration (0.22 μM). The soluble protein was analyzed by gel electrophoresis. As shown in Figure A1.1, the soluble protein and the pure glycosylated protein were analyzed by gel electrophoresis using an 8 to 25% gradient acrylamide gel with SDS buffer strips (Pharmacia LKB, Sweden). The electrophoresis and Coomassie Blue staining were performed with a Phast System (Pharmacia LKB, Sweden). The gel shown in Figure A1.1 revealed that rhDNAse was the dominant protein. However, several proteolysis products, small peptides, were also present at a high fraction of the total protein as evidenced by the dense bands observed near the end of the gel. It was assumed from this analysis as well as a previous analysis which was performed by Genentech that the inclusion body pellets consisted of approximately 80% rhDNAse.

Recombinant Human Gamma Interferon (rIFN-γ)

Soluble inclusion body protein in 4 M GuHCl was analyzed by gel electrophoresis and size exclusion chromatography (see Material and Methods for solubilization). The purity of the solubilized protein was initially determined by using size exclusion chromatography. Two hundred microliters of the protein mixture was injected onto size

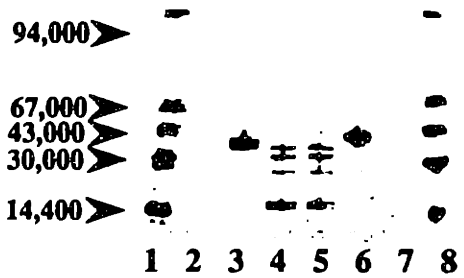


FIGURE A1.1: Gel electrophoresis results for rhDNAse inclusion body protein in 8 M urea. Low molecular weight protein markers in 8 M urea were run in lanes 1 an 7 (Pharmacia LKB, Sweden, see also Figure 1.7). Glycosylated rhDNase in either the oxidized or reduced (20 mM β-mercaptoethanol) state were placed in lanes 3 and 6, respectively. The solubilized rhDNAse inclusion body protein was analyzed in duplicate (lanes 4 and 5).

exclusion columns connected in series (Superose 6, Superose 6, and Superose 12; Pharmacia LKB, Sweden) all of which were preequilibrated with 4 M GuHCl. Separation was achieved by operating at a flowrate of 0.20 ml/min. Injection, sampling, and other operations were performed with LCC 500 control system interfaced to a complete Fast Protein Liquid Chromatography (FPLC) system (Pharmacia LKB, Sweden). The resulting chromatogram of the separation is shown in Figure A1.2. Fractions from the separation were analyzed by SDS-PAGE with silver staining and the digitized gels are shown in Figure A1.2.

Purification of the solubilized inclusion bodies was performed by several different methods. Samples in GuHCl were initially purified by size exclusion chromatography with either a Superose 6 or Superose 12 column (Pharmacia LKB, Sweden). Solubilization was also performed with urea. The inclusion body protein was dissolved in either 5 M or 8 M urea and any insoluble protein was removed by filtration (0.22 μM). In addition, solubilized protein in 4 M GuHCl was exhaustively dialyzed against 8 M urea. Samples in 8 M urea were purified by either size exclusion (Superose 6 or 12) or anion exchange (Mono Q, Pharmacia LKB, Sweden). The results of each solubilization and the puriest fractions obtained by each separation were analyzed by SDS-PAGE and the digitized gels are shown in Figure A1.3. The solubized protein in 4 M GuHCl was also analyzed for purity by gel electrophoresis (Figure A1.3). All electrophoresis analyses were performed by using an 8 - 25% acylamide gradient gel with SDS buffer strips (Pharmacia LKB, Sweden) and SDS as well as β-mercaptoethanol was added to urea samples prior to analysis. The electrophoresis and silver staining were performed with

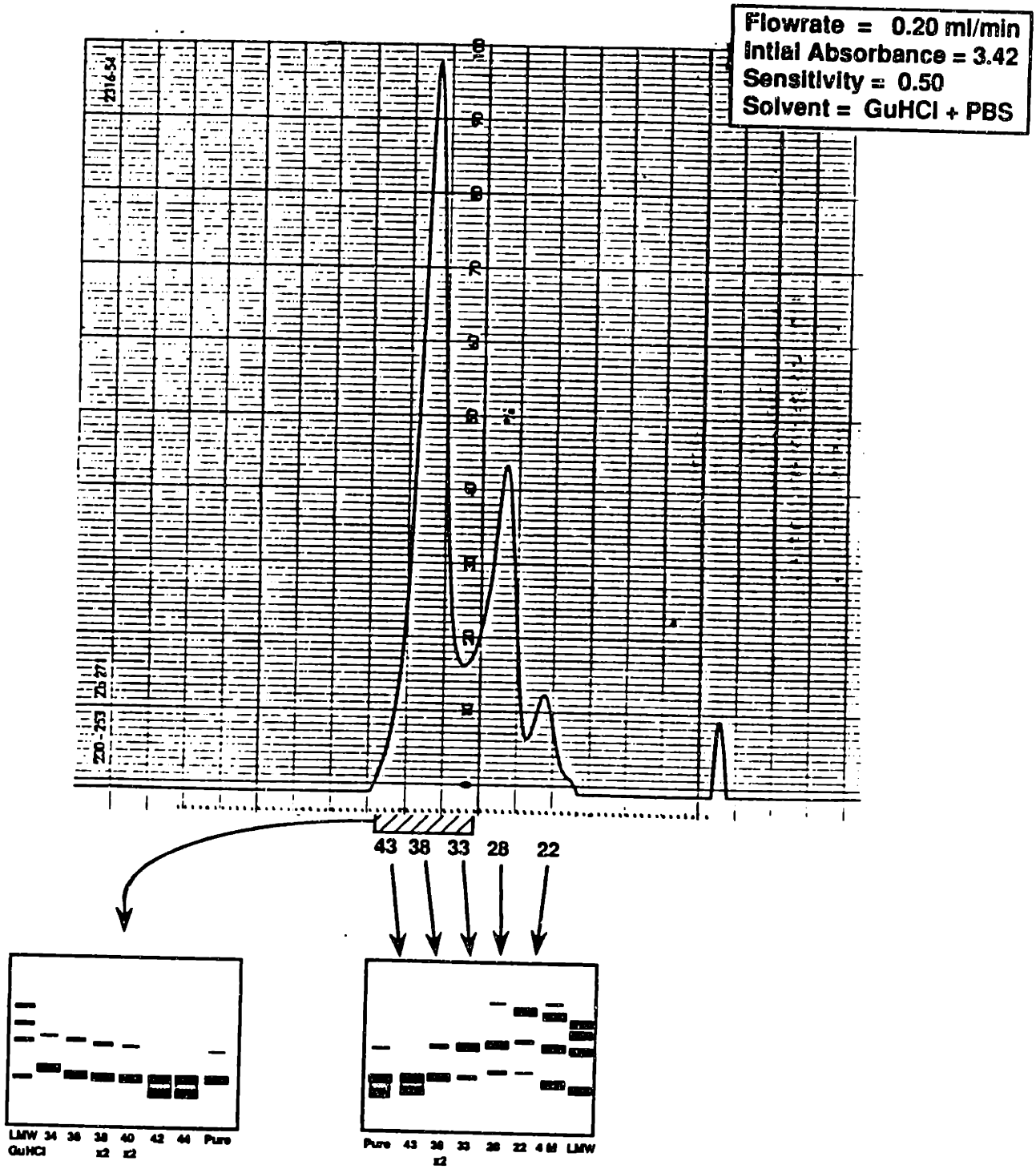


FIGURE A1.2: Separation of rIFN-γ from inclusion body mixture performed by using a series of size exclusion columns (Superose 6, Superose 12, and Superose 12) which were loaded with 200 µl of 10 mg/ml inclusion body protein and eluted with 4 M GuHCl at 0.20 ml/min. SDS PAGE analysis of fractions is also shown.

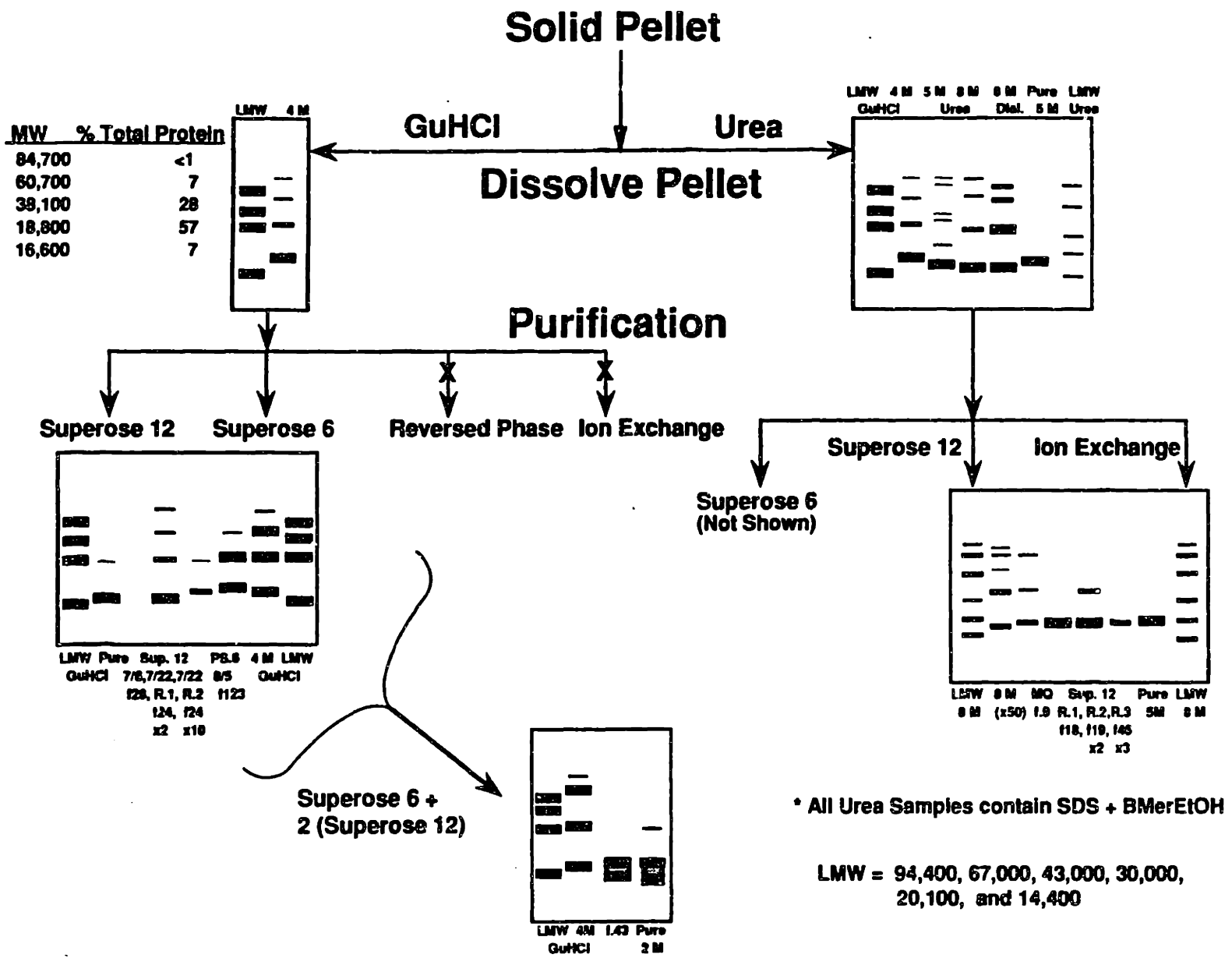


FIGURE A1.3: Digitized gel electrophoresis results for different purification and solubilization methods. Each sample was analyzed by gel electrophoresis with a 8-25% gradient gel and SDS buffer strips on a Phast System (Pharmacia LKB, Sweden). The gel was silver stained and scanned to produce the above results. The puriest fractions from each chromatographic separation are shown.

a Phast System (Pharmacia LKB, Sweden). As shown in Figure A1.3, the gel electrophoresis results indicated that rIFN-γ was the primary protein, but a large amount of an apparent dimer specie was also present. When the protein mixture was completely reduced in high concentrations of reducing agent (50 mM β-mercaptoethanol), the gel electrophoresis results were the same as those shown in Figure A1.3. Therefore, it is likely that the apparent dimer may either be another type of covalent dimer (C-C, C-N) or an unrelated protein. The size exclusion and gel electrophoresis results indicated that the inclusion bodies consisted of approximately 60% rIFN-γ (~ 18,800 MW).

APPENDIX 2: QUASI-ELASTIC LIGHT SCATTERING (QLS): MULTIMER CALCULATION METHOD

The particle size distributions which were obtained by CONTIN analysis of the autocorrelation data were used to calculate the multimer concentrations. The model used for multimer calculations was the worm-like chain model developed by Kratky and Porod (Kratky & Porod, 1949). For this model, the protein was assumed to form a continuous chain of spheres which could align as shown in Figure A2.1. The dimer and trimer of the first intermediate in the CAB refolding pathway were assumed to form with a persistence length, a, and a contour length, L_i. The hydrodynamic radius, R_h, was then calculated for each multimer based on a unit monomer with diameter, d_s, as described by Murphy (Murphy, 1989). For $L_i/a \leq 4.556$, the hydrodynamic diameter can be calculated from (Murphy, 1989):

$$R_h = \frac{0.5\ L_i}{A_1\ L_i^{0.5} + A_2 + A_3\ L_i^{-0.5} + A_4\ L_i^{-1.0} + A_5\ L_i^{-1.5}} \qquad \text{(A2.1)}$$

$$A_1 = 1.843$$

$$A_2 = -(1.0 - 0.01412\ d_s^2 + 0.00592\ d_s^4)\ \ln(d_s) - 1.0561 - 0.1667\ d_s - 0.1900\ d_s^2 - 0.0224\ d_s^3 + 0.0190\ d_s^4$$

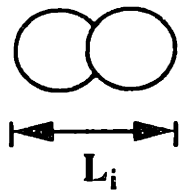
$$A_3 = 0.1382 + 0.6910\ d_s^2$$

$$A_4 = -(0.04167\ d_s^2 + 0.00567\ d_s^4)\ \ln(d_s) - 0.3301 + 0.5\ d_s - 0.5854\ d_s^2 - 0.0094\ d_s^3 - 0.0421\ d_s^4$$

$$A_5 = -0.0300 + 0.1209\ d_s^2 + 0.0259\ d_s^4$$

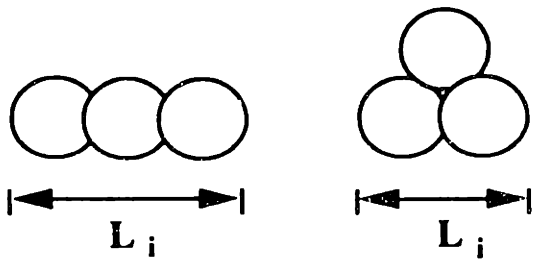
The diffusivity of each particle was then calculated by using the Stokes-Einstein equation

Dimer Model : L_i/a = 1.5 to 2.0



L_i/a = 1.5 25% Excluded Volume
2.0 Linear Solid Spheres

Trimer Model : L_i/a = 2.0 to 3.0



L_i/a = 2.0 Three Sphere Cluster
3.0 Linear Solid Spheres

FIGURE A2.1: Multimer model for analysis of particle size distribution data which is obtained by QLS measurements. Worm-like chain model of multimer formation as described by Murphy was used to determine the actual multimer size (Murphy, 1989).

(Literature Review, Equation 4). From the diffusivity of each specie, the autocorrelation function was next calculated from Equations 1, 2, and 3 in the Literature Review section with an assumption of the fraction of monomers, dimers and trimers in the solution. Then, the calculated autocorrelation function was compared to the actual autocorrelation function. The estimated multimer distribution was obtained by iteration. The assumed concentration of each specie was repeatediy changed until the calculated autocorrelation data matched the experimental autocorrelation function. This procedure was performed to achieve a close fit (± 10%) between the data and the prediction as described previously by Murphy (Murphy, 1989).

APPENDIX 3: STRUCTURAL ANALYSIS PROGRAMS

Program Settings and Protein Sequence Sources

The programs used in this thesis to study protein structure were the PeptideStructure and PlotStructure programs which were developed by the Genetics Computer Group at the University of Wisconsin, Madison (Devereux, 1989). The PeptideStructure program was run by inputing the protein sequences which were obtained from different data libraries as shown in Table A3.1. For each protein, the entire sequence provided by the database was used. The Kyte and Doolittle method for relative hydrophilicity of each protein was used by the PeptideStructure program to calculate the protein hydrophilicity profile (Kyte & Doolittle, 1982). Recombinant human deoxyribonuclease (rhDNAse) was input into the program by modifying the sequence of bovine DNAse (Table A3.1, see also Literature Review, Figure 7). After running the PeptideStructure program on each sequence, the PlotStructure routine was used to plot the results and calculate the secondary structure. As described in Material and Methods, the Chou and Fasman algorithm was used by the PlotStructure program to predict the secondary structure (Chou & Fasman, 1978). The PlotStructure routine was operated by chosing a "squiggle" plot which overlayed the secondary structure and hydrophilicity predictions. The hydrophilicity scale for the plot was chosen as the default setting which is a relative scale of 1.3. Additional discussion of the program parameters is listed in the next section.

The other program which was used in this thesis was the BestFit homology software package from the Genetics Computer Group at the University of Wisconsin,

TABLE A3.1: Sources of Protein Sequences for Structural Analysis.

Protein	Code Name	Database[a]
bovine carbonic anhydrase B	CRBO2	NBRF
human carbonic anhydrase II	CRHU2	NBRF
bovine pancreatic deoxyribonuclease I	DRN1$BOVIN	Swiss Protein
precursor, human pancreatic tissue plasminogen activator	UKHUT	NBRF
precursor, human gamma interferon	IVHUG	NBRF
human pancreatic ribonuclease	RNP$HUMAN	Swiss Protein

[a] NBRF refers to the National Biomedical Research Foundation's protein sequence database. Swiss Protein refers to the European Molecular Biology Laboratory's data library (Geneva, Switzerland).

Madison (Devereux, 1989). Bovine carbonic anhydrase was compared to human carbonic anhydrase II and bovine deoxyribonuclease was compared to recombinant human deoxyribonuclease by using the BestFit program (sequences from Table A3.1). The default settings were used for each analysis: Gap Width = 3 and Gap Length Weight = 0.10. The standard deviation was not determined for either run. A complete discussion of the parameters is listed in the next section.

Description of Programs

To better understand the programs used in this thesis, the relevant sections of the GCG Software Manual are included in the following pages (Devereux, 1989). Additional program information may be obtained by contacting the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin, 53705. For Massachusetts Institute of Technology users, the program manual can be obtained from the Whittaker Health Sciences Computer Facilities.

FUNCTION

PeptideStructure makes secondary structure predictions for a peptide sequence. The predictions include, in addition to alpha, beta, coil, and turn, measures for antigenicity, flexibility, hydrophobicity, and surface probability. **PlotStructure** displays the predictions graphically.

DESCRIPTION

PeptideStructure makes predictions of the following features of an amino acid sequence:

- Secondary structure according to the Chou-Fasman method,
- Secondary structure according to the Garnier-Osguthorpe-Robson method,
- Hydrophilicity according to either the Kyte-Doolittle or Hopp-Woods method,
- Surface probability according to the Emini method,
- Flexibility according to the Karplus-Schulz method,
- Glycosylation sites,
- Antigenic index according to the Jameson-Wolf method.

The results are written into a file for graphical presentation with **PlotStructure**.

PeptideStructure uses the original method of Chou and Fasman (Adv. in Enzymol., **47**; 45-148 (1978)) to predict helices, sheets, and turns. It resolves overlapping regions of helices and beta-sheets with the "overall probability" procedure introduced by K. Nishikawa (Biochim. Biophys. Acta, **748**; 285-299 (1983)). This same procedure also locates turns that are not in conflict with other secondary structures. The Chou-Fasman rules are slightly modified as follows: Helix: the condition that p(bound) be greater than 1.0 and that p(alpha) be greater than p(beta) are not used; Sheet: a minimum length of five residues is required.

PeptideStructure also predicts secondary structure according to a slightly modified method of Robson-Garnier (Garnier et al. J. Mol. Biol., **120**; 97 (1978)). The minimum length of a helix is six and of a beta-sheet is four. Regions without adequate predictions are replaced by the conformational state of the next best probability.

Hydrophilicity is calculated according to the algorithm of Hopp and Woods (Proc. Natl. Acad. Sci. USA, **78**; 3824-3828 (1981)). The window, normally set to seven residues, can be changed with the optional parameter `/HWINdow`.

Surface probability is calculated according to a formula of Emini et al. (J. Virol., **55(3)**; 836-839 (1985)), which is slightly modified for the end values of the protein chains. The single probabilities are taken from Janin et al. (J. Mol. Biol., **125**; 357-386 (1978)).

Glycosylation sites are predicted for sites where the residues have the composition NxT or NxS. When x is D, W, or P, the site is taken to be a weak glycosylation site, otherwise it is a strong glycosylation site.

The antigenic index (AI) is a measure of the probability that a region is antigenic. It is calculated by summing several weighted measures of secondary structure. The method is described by Jameson and Wolf (CABIOS, **4(1)**; 181-186 (1988)) and the formula is defined under the topic ALGORITHM below. Peaks of antigenic index can be smoothed with the `/BROAdening` option.

EXAMPLE

In this example, **PeptideStructure** calculates the secondary structure of the outer membrane protein F precursor from E. coli.

```
$ PeptideStructure

  PEPTIDESTRUCTURE for what peptide sequence ?  NBRF:Mmecf

                Begin (* 1 *) ?
              End (*   362 *) ?

  Calculate hydrophilicity according to

      H)opp-Woods or
      K)yte-Doolittle

  Please choose one (* K *) :

  What should I call the output file (* Mmecf.P2s *) ?

$
```

OUTPUT

Here is part of the output file:

```
PEPTIDESTRUCTURE of: Prot:Mmecf  check: 2147  from: 1  to: 362

Outer membrane protein F precursor - Escherichia coli

Hydrophilicity (Kyte-Doolittle) averaged over a window of: 7
Surface Probability according to Emini
Chain Flexibility according to Karplus-Schulz
Secondary Structure according to Chou-Fasman
Secondary Structure according to Garnier-Osguthorpe-Robson
Antigenicity Index according to Jameson-Wolf

                    Date: February 17, 1989  14:50

Pos  AA  GlycoS  HyPhil  SurfPr  FlexPr  CF-Pred GORPred AI-Ind ..

1    M       .   1.150   5.841   1.000        .       H   0.900
2    M       .   1.620   9.487   1.000        .       H   0.900
3    K       .   0.600   5.846   1.000        .       H   0.900

////////////////////////////////////////////////////////////////

360  Y       .  -1.050   0.754   1.000        B       B  -0.450
361  Q       .  -1.340   1.223   1.000        B       B  -0.300
362  F       .  -0.550   1.909   1.000        B       B  -0.300
```

RELATED PROGRAMS

PlotStructure plots the results from **PeptideStructure**. **PepPlot** plots parallel curves of the standard measures of protein secondary structure.

RESTRICTIONS

PeptideStructure requires a peptide sequence. The B (Asx), Z (Glx), and * characters (Appendix III) are not yet supported for this program.

ALGORITHM

See the papers cited under the topic DESCRIPTION. Do not attempt to interpret protein secondary structure predictions without reading the Robson-Garnier paper.

The antigenic index (AI) is calculated by summing several weighted measures of secondary structure. The method is described by Jameson and Wolf (CABIOS, 4(1); 181-186 (1988)). The formula for the antigenic index (AI) is:

```
AI = 0.3*[H] + 0.15*[S]+ 0.15*[F] + 0.2*[Cs] + 0.2*[Rs]

where:

[H] =                2          1          -1         -2          for
Hydrophilicity       >0.5       >0.        >-0.4      <-0.4

[S] =                1                     0                      for
Surface probability  >=1.0                 <1.0

[F] =                1                     0                      for
Flexibility          >=1.0                 <1.0

[Cs] =               2                     1              0       for
Chou-F.              strong turn           turn or coil   other

[Rs] =               2                     1              0       for
Robson-G.            strong turn           turn or coil   other
```

Peaks of antigenic index can be smoothed with the /BROAdening option. In each case where the antigenic index has a peak (except inside a strong helix), the index can be decremented (broadened) smoothly to at least 80%, 40%, and 20% of the peak in the three surrounding amino acids.

CONSIDERATIONS

You should realize that measures of protein secondary structure are only weakly correlated with actual structures. The Chou-Fasman method was designed to apply to soluble (globular) proteins.

COMMAND LINE SUMMARY

All parameters for this program may be put on the command line. Use the option /CHEck to see the summary below and to have a chance to modify the command line before the program executes. The capitalized letters in the qualifier names are the letters that you must type. Square brackets ([and]) enclose qualifiers or parameter values that are optional. **Appendix XI** describes command line control in detail.

```
Syntax: $ PEPTIDESTructure [/INfile=]SeqName.Seq /Default

Required Parameters:

/BEGin=1                    first base of sequence
/END=100                    last base of sequence
/MENu=k                     K)yte-Doolittle vs H)opp-Woods menu
[/OUTfile=]SeqName.P2s      output file name

Local Data Files: None
```

```
Optional Parameters:

/HWINdow=7    sets the window for the hydrophilicity calculation
/BROAdening   broadens peaks of antigenic index outside of strong helices
```

ACKNOWLEDGEMENT

PeptideStructure and **PlotStructure** were communicated to us by Dr. Hans Wolf for Drs. Susanne Modrow and Manfred Motz of the Max von Pettenkofer-Institut of the University of Munich and for Bradford Jameson of the California Institute of Technology. The programs were written for them by Dr. B. J. Foertsch and G. Herrmann and were modified for compatibility with GCG Version 5 by John Devereux. The graphic schematic and overlying symbols for hydropathy and glycosylation are based on the early work (1983) of Ellis Golub of the University of Pennsylvania. These programs have been described and used by Cohen et al. (J. Virol. **49**; 102-108 (1984)), Starcich et al. (Cell, **45**; 637-648 (1986)), Motz et al. (Gene, **42**; 303-312 (1986)) and by Modrow and Wolf (Proc. Natl. Acad. Sci. USA **83**; 5703-5707 (1986)). The method for calculating antigenic index is described by Jameson and Wolf (CABIOS, **4**(1), 181-186 (1988)). Any use of **PeptideStructure** or **PlotStructure** for published research should cite the Jameson and Wolf article.

LOCAL DATA FILES

None.

The residue-specific attributes will be moved into a data file in future versions.

OPTIONAL PARAMETERS

The parameters and switches below can only be set from the command line. Optional parameters that apply to all programs are described in the **Introduction** to the **Program Manual**.

/BROAdening

sets **PeptideStructure** to broaden the peaks of the antigenic index as described under the topic DESCRIPTION.

/HWINdow=7

lets you set the window of integration for the hydrophilicity measurement. The window is normally set to seven residues.

Printed: October 2, 1989 20:00

PLOTSTRUCTURE+

FUNCTION

PlotStructure plots the output file from **PeptideStructure**. The measures can be shown on parallel panels of a graph or with a 2-dimensional squiggly representation.

DESCRIPTION

PeptideStructure writes a file with several kinds of protein secondary structure predictions shown as a table. **PlotStructure** reads the **PeptideStructure** output file and displays it in any of several different ways.

There are two main options, each with several suboptions.

One main option generates a two-dimensional plot representing predicted secondary structures (helices, beta sheets, turns, and coils) with different wave forms. Helices are shown with a sine wave, Beta sheets with a sharp saw-tooth wave, turns with 180 degree turns, and coils with a dull saw-tooth wave. Any of four different quantitative attributes (hydrophilicity, surface probability, flexibility, or antigenic index) can be superimposed over the wave with special symbols wherever the attribute exceeds some set threshold. The size of the symbols is proportional to the value of the attribute. In addition, possible glycosylation sites can be marked on the two-dimensional plot.

The other main option for **PlotStructure** creates a one-dimensional multi-paneled plot. The residues are numbered on the x-axis and the attributes are represented as continuous curves in each of several different panels.

EXAMPLE

The output from an earlier session with **PeptideStructure** is plotted in this session with **PlotStructure**. The plots (in the two **Program Manual** figures below) show the predicted secondary structure of the E. coli outer membrane protein F precursor.

```
$ PlotStructure

 PLOTSTRUCTURE of what PEPTIDESTRUCTURE output file ?  Mmecf.P2s

                Plot Begin (* 1 *) ?
                Plot End (* 362 *) ?

 Do you want a

   1)-dimensional (panel graph) or a
   2)-dimensional (squiggly) plot

 Please choose one (* 2 *):

 Do you want to plot the

   C)hou-Fasman or the
   G)arnier-Osguthorpe-Robson predictions

 Please choose one (* C *):
```

```
Superimpose over the squiggly plot:

   H)ydrophilicity,
   S)urface probability,
   F)lexibility,
   A)ntigenic index or
   N)othing

Please choose one (* H *):

What hydrophilicity threshold (* 1.3 *) ?

When the HP7550 is ready press <rtn>

$
```

OUTPUT

If you are reading the **Program Manual**, the plot from this session and from a session selecting a one-dimensional plot for the same input data are shown in the two figures below.

RELATED PROGRAMS

PeptideStructure calculates the secondary structure of a peptide and writes a file suitable for input to **PlotStructure**. **PepPlot** plots measures of protein secondary structure as a set of parallel curves in different panels of the same plot.

RESTRICTIONS

Unknown. The input file for **PlotStructure** must be the output file from **PeptideStructure**. There are some unusual format requirements for this file discussed in the INPUT FILE topic below.

INPUT FILE

The input file for **PlotStructure** is the output file from **PeptideStructure**. **PlotStructure** reads the first two lines of the input file to identify and document the sequence that was analysed by **PeptideStructure**. It also searches for a line with the string ' Date: ' for the date of the session with **PeptideStructure**. This date is shown on all plots so that plots from the same input file may be associated. Here is some of the input file used for the example session.

```
PEPTIDESTRUCTURE of: Prot:Mmecf  check: 2147  from: 1  to: 362

Outer membrane protein F precursor - Escherichia coli

Hydrophilicity (Kyte-Doolittle) averaged over a window of: 7
Surface Probability according to Emini
Chain Flexibility according to Karplus-Schulz
Secondary Structure according to Chou-Fasman
Secondary Structure according to Garnier-Osguthorpe-Robson
Antigenicity Index according to Jameson-Wolf

                    Date: February 17, 1989  14:50
```

```
Pos   AA  GlycoS   HyPhil   SurfPr   FlexPr   CF-Pred GORPred AI-Ind ..

1     M       .    1.150    5.841    1.000          .      H   0.900
2     M       .    1.620    9.487    1.000          .      H   0.900
3     K       .    0.600    5.846    1.000          .      H   0.900

////////////////////////////////////////////////////////////////////

360   Y       .   -1.050    0.754    1.000          B      B  -0.450
361   Q       .   -1.340    1.223    1.000          B      B  -0.300
362   F       .   -0.550    1.909    1.000          B      B  -0.300
```

Ctrl-C

If you need to stop this program, type a `Ctrl-C`, and the program will reset your terminal and session as gracefully as possible. Searches and comparisons will write out the results from the part of the search that was complete when the `Ctrl-C` was typed. Plotters should stop and put the pen away or start plotting the next page; terminals should return to interactive mode.

DEVICES REQUIRED

The GCG Package must know the kind of graphics output device you have and the computer port to which that device is connected. *These configuration parameters must be set before you run a GCG program with graphics output.* See the **Graphics** section of the GCG **User's Guide**.

GCG graphics can be drawn on many different devices including Hewlett Packard HP7221, 7475, 7470, 7550, 7580, and ColorPro plotters, Digital Equipment VT240, 241, 330, 340 graphic terminals, LA50, 75, 100, 200, and LN03 PLUS printers, Tektronix and GraphOn 4014 and 4100 series graphic terminals, Apple, Dataproducts, and LN03 *ScriptWriter* laser printers, and VAX workstations. Our graphics will also plot through the VAX GKS device-independent graphics interface.

You can store GCG graphics output in a text file for modification and later input to the **Figure** program by adding `/FIGure=FileName` to the command line.

COMMAND LINE SUMMARY

All parameters for this program may be put on the command line. Use the option `/CHEck` to see the summary below and to have a chance to modify the command line before the program executes. The capitalized letters in the qualifier names are the letters that you must type. Square brackets ([and]) enclose qualifiers or parameter values that are optional. **Appendix XI** describes command line control in detail.

```
Syntax: $ PLOTSTructure [/INfile=]SeqName.P2s /Default

Required Parameters:

/BEGin=1    first residue to plot
/END=362    last  residue to plot
/MENu1=2    1)-dimensional (panel) or 2)-dimensional (squiggly) plot
/MENu2=C    plot either C)hou-Fasman or G)arnier-Robson prediction
/MENu3=H    superimposes: H)ydrophilicity, S)urface-probability,
                          F)lexibility, A)ntigenicity or N)othing
```

```
Optional Parameters:

/THReshold=1.3  sets threshold for superimposed measure (Menu3)
/NUMbering=50   numbers plot in intervals of 50 aa
/STRONgonly     shows only the strong Chou-Fasman structures
/GLYCosylation  show points where glycosylation could occur
/DENsity        sets the number of residues per 100 platen units
                     (per page)

All GCG graphics programs accept these and other switches.
See the Graphics section of the User's Guide for others.

/FIGure[=FileName] stores plot as a text file for later input to FIGURE
/NOTEXt            suppresses all text
/FONT=3            writes all text in font number 3 (see Appendix VIII)
/COLor=1           draws whole plot with color number one
/LINEWidth=0.5     draws lines 0.5 platen units wide (not all devices)
/GRId              draws graph (in platen units) behind the plot
/SCAle=1.2         enlarges the plot by 20 percent (zoom in)
/XPAN=10.0         moves plot to the right 10 platen units (pan right)
/YPAN=10.0         moves plot up 10 platen units (pan up)
/PORtrait          rotates the plot 90 degrees
```

ACKNOWLEDGEMENT

PeptideStructure and **PlotStructure** were communicated to us by Dr. Hans Wolf for Drs. Susanne Modrow and Manfred Motz of the Max von Pettenkofer-Institut of the University of Munich and for Bradford Jameson of the California Institute of Technology. The programs were written for them by Dr. B. J. Foertsch and G. Herrmann and were modified for compatibility with GCG Version 5 by John Devereux. The graphic schematic and overlying symbols for hydropathy and glycosylation are based on the early work (1983) of Ellis Golub of the University of Pennsylvania. These programs have been described and used by Cohen et al. (J. Virol. **49**; 102-108 (1984)), Starcich et al. (Cell, **45**; 637-648 (1986)), Motz et al. (Gene, **42**; 303-312 (1986)) and by Modrow and Wolf (Proc. Natl. Acad. Sci. USA **83**; 5703-5707 (1986)). The method for calculating antigenic index is described by Jameson and Wolf (CABIOS, **4(1)**, 181-186 (1988)). Any use of **PeptideStructure** or **PlotStructure** for published research should cite the Jameson and Wolf article.

LOCAL DATA FILES

None.

OPTIONAL PARAMETERS

The parameters and switches below can only be set from the command line. Optional parameters that apply to all programs are described in the **Introduction** to the **Program Manual**.

/NOGLY Cosylation

suppresses the display of possible glycosylation sites.

/NOSTRONgonly

Normally **PlotStructure** only shows the Chou-Fasman predictions that are strong. These predictions are shown as capital letters in the **PeptideStructure** output file. You can plot the weakly predicted features with **/NOSTRON**gonly.

/NUMbering=50

For the two-dimensional plot, numbering is normally enabled with a numbering interval of 50 residues. You can reset the interval with this optional parameter or you can suppress numbering altogether with **/NONUMbering**.

THReshold=1.3

Each superimposed statistic on the two-dimensional (squiggly) plot has a threshold that you can set.

/DENsity=1000

sets the number of bases or residues per 100 platen units (pu). This is usually equivalent to the number of bases or residues per page. Output from different GCG graphics programs that are run at the same density can be compared by lining up the plots on a light box. There is more information in the **Graphics** section of the **User's Guide** under the topic DENSITY CONTROL.

The options described below apply to all GCG graphics programs. These and others are described in more detail in the **Graphics** section of the **User's Guide**. Some graphics devices do not support all of these switches.

/FIGure=ProgramName.Figure

writes the plot as a text file of plotting instructions suitable for input to GCG's **Figure** program instead of drawing the plot on your plotter.

/FONT=3

draws all text characters on the plot using font 3 (see **Appendix VIII**).

/NOTEXt

suppresses all of the text on the plot.

/COLor=1

draws the entire plot with the pen in stall 1.

/LINEWidth=0.2

sets the line thickness to 0.5 platen units. Most graphic device drivers do not yet support this option.

/GRId

draws a grid, marked off in platen units, behind the plot.

/COPies=3

makes multiple copies of the plot on some plotters and printers.

/AUTOFeed

For multipage plots on plotters equipped with automatic paper feed, the **/AUTOFeed** switch advances the page automatically. Plotters that are "queued" will always behave as if this option were selected.

/SPEed=8.0

lets you set a pen speed on x-y plotters between 1.0 (slow) and 10.0 (fast) to achieve high quality plots for publication.

/NOCLIpping

Some programs clip lines and text at the edge of a masked plotting area. This switch disables clipping, showing line segments and characters that go outside of the masked area.

With the options described below you can expand (or reduce) the plot, move it in either direction or rotate it 90 degrees (zoom, pan, and rotate).

/SCAle=1.2

expands the plot by 20 percent by resetting the scaling factor (normally 1.0) to 1.2 (zoom in). You can expand the axes independently with **/XSCAle** and **/YSCAle**. Numbers less than 1.0 contract the plot (zoom out).

/PORtrait

Plots are normally displayed with the horizontal axis longer than the vertical (landscape). This option rotates the plot 90 degrees. Plots are reduced or enlarged, depending on the platen size, to fill the page.

/XPAN=30.0

moves the plot to the right by 30 platen units (pan right).

/YPAN=30.0

moves the plot up by 30 platen units (pan up).

Printed: October 2, 1989 20:00

FUNCTION

BestFit makes an optimal alignment of the best segment of similarity between two sequences. Optimal alignments are found by inserting gaps to maximize the number of matches using the *local homology* algorithm of Smith and Waterman.

DESCRIPTION

BestFit inserts gaps to obtain the optimal alignment of the best region of similarity between two sequences and then displays the alignment in a format similar to the output from **Gap**. The sequences can be of very different lengths and have only a small segment of similarity between them. You could take a short RNA sequence, for example, and run it against the whole mitochondrial genome.

SEARCHING FOR SIMILARITY

BestFit is the most powerful algorithm we know for identifying the best region of similarity between two sequences whose relationship is unknown.

EXAMPLE

Gamma.Seq contains an Alu family sequence somewhere in the first 500 bases. HsRep2 in the EMBL data library contains a generic human Alu family repeat. The two sequences are aligned and the best segment of similarity is found with **BestFit**.

```
$ Bestfit

 BESTFIT of what sequence 1 ?  Gamma.Seq

                  Begin (* 1 *) ?
                End (* 11375 *) ?  500
               Reverse (* No *) ?

to what sequence 2 (* Gamma.Seq *) ?  EMBL:HsRep2

                  Begin (* 1 *) ?
                End (*   207 *) ?
               Reverse (* No *) ?

What is the gap weight (* 5.00 *) ?

What is the gap length weight (* 0.30 *) ?

What should I call the paired output display file (* Gamma.Pair *)

Aligning ..........-..

         Gaps:     3
      Quality: 129.3
Quality Ratio: 0.625
   Similarity: 84.466
       Length:   209
```

OUTPUT

Here is the output file; notice how **BestFit** finds and displays only the best segments of similarity.

```
 BESTFIT of: Gamma.Seq  check: 6474  from: 1  to: 500

Human fetal beta globins G and A gamma
from Shen, Slightom and Smithies,  Cell 26; 191-203.
Analyzed by Smithies et al. Cell 26; 345-353.

 to: Hsrep2  check: 4238  from: 1  to: 207

ID   HSREP2     standard; DNA; 207 BP.
AC   V00580;
DT   27-AUG-1987  (an correction)
DT   11-APR-1983  (annotation)
DT   17-FEB-1981  (first entry) . . .

 Symbol comparison table: Gencoredisk:[Gcgcore.Rundata]Swgapdna.Cmp
 CompCheck: 5234

        Gap Weight:  5.000      Average Match:  1.000
     Length Weight:  0.300   Average Mismatch: -0.900
           Quality:  129.3             Length:    209
             Ratio:  0.625               Gaps:      3
Percent Similarity: 84.466   Percent Identity: 84.466

 Gamma.Seq x Hsrep2        December 29, 1988  13:29  ..

                .         .         .         .         .
    137 AGACCAACCTGGCCAACATGGTGAAATCCCATCTCTAC.AAAAATACAAA 185
        |||||| |||||||||||||||||  ||||||||||  ||||||||||
      1 \GACCAGCCTGGCCAACATGGTGAAACTCCATCTCTACTGAAAATACAAA 50
                .         .         .         .         .
    186 AATTAGACAGGCATGATGGCAAGTGCCTGTAATCCCAGCTACTTGGGAGG 235
        |||||| |||||||| ||    ||||||| |||||||||||||| |||||
     51 AATTAGCCAGGCATGGTGATGCGTGCCTGGAATCCCAGCTACTTAGGAGG 100
                .         .         .         .         .
    236 CTGAGGAAGGAGAATTGCTTGAACCTGGAAGGCAGGAGTTGCAGTGAGCC 285
        |||||  || |||||  ||| ||||  | |||  |  ||||||||||||
    101 CTGAGACAGAAGAATCCCTTAAACCAAG.AGGTGGAGGTTGCAGTGAGCC 149
                .         .         .         .         .
    286 GAGATCATACCACTGCACTCCAGCCTGGGTGACAGAACAAGACTCTGTCT 335
        ||||||  ||  |||||||||||||| |||||||| | ||||||  |||
    150 GAGATCGCACGGCTGCACTCCAGCCT.GGTGACAGAGCGAGACTCCATCT 198

    336 CAAAAAAAA 344
        |||||||||
    199 CAAAAAAAA 207
```

RELATED PROGRAMS

When you want an alignment that covers the whole length of both sequences, use **Gap**. When you are only trying to find the best segment of similarity between two sequences, use **BestFit**. **DotPlot** displays the entire surface of comparison revealing repeated structures. **GapShow** displays the pattern of differences between two aligned sequences. **Pretty** displays alignments of several sequences. **LineUp** is an editor for editing multiple sequence alignments. **CompTable** helps generate symbol comparison

tables for peptide comparison.

ALGORITHM

BestFit uses the "local homology" algorithm of Smith and Waterman (Advances in Applied Mathematics 2; 482-489 (1981)) to find the best segments of similarity between two sequences. **BestFit** uses symbol comparison values of 1.0 for each match and -0.90 for each mismatch to construct a path matrix that represents the entire surface of the comparison with a score at every position for the best possible alignment path to that point. Random alignments should have a path value that averages about zero. If the best path to any point has a negative value, a zero is put in that position.

After the path matrix is complete, the highest value on the surface represents the end of the best region of similarity between the sequences. The best path from this highest value backwards to the point where the values revert to zero is the alignment shown by **BestFit**. The alignment is the best segment of similarity between the two sequences.

The gap weight and gap length weight are set by the user and evaluated in exactly the same manner as for other gapping algorithms. For each alignment, the quality is equal to the number of matches, less 0.9 times the number of mismatches, less the gap weight times the number of gaps, less gap length weight times the total length of all the gaps.

```
Quality = 1.0 x Matches  -  0.90 x Mismatches
                  - (GapWeight x GapNumber)
          - (GapLengthWeight x TotalLengthofGaps)
```

The actual value of the coefficients for Matches and Mismatches are determined by the symbol comparison table you use (see Local Data Files below).

CONSIDERATIONS

BestFit Always Finds Something

BestFit will always find a solution for any two sequences you compare, even if there is no significant similarity between them!! You must evaluate the results critically to decide if the segment shown is not just a random region of relative similarity.

The Segments Shown Obscure Alternative Segments

BestFit only shows one segment of similarity, so if there are several, all but one will be obscured. You can approach this problem with graphic matrix analysis (see **Compare** and **DotPlot**). Alternatively, you can run **BestFit** on ranges outside the ranges of similarity found in earlier runs to bring other segments out of the "shadow" of the best segment.

The Best Fit is Only One Member of a Family

Like all fast gapping algorithms, the alignment displayed is a member of the family of best alignments. This family may have other members of equal quality, but will not have any member with a higher quality. The family is usually significantly different for different choices of gap weight and gap length weight. You should read the considerations in the **Program Manual** for the program **Gap** to learn more about how to assign gap weight and gap length weight.

The Surface of Comparison

The magnitude of the computer's job is proportional to the area of the surface of comparison. That area is determined by the product of the lengths of the two sequences compared. **BestFit** can evaluate a surface of up to one million elements. This surface would be large enough to compare

two sequences 1,000 symbols long, or one sequence 100 symbols long with another sequence 10,000 symbols long. When you have much longer sequences that are known to align well, you can use the command line switch `/LIMit` to use the surface more efficiently. (See the section on Command Line Switches below.)

The Public Symbol Comparison Table is Very Stringent

The table SWGapDNA.Cmp penalizes mismatches -0.9 so the segments found may be very brief. This penalty means that the alignment cannot be extended by three bases to pick one extra match. The table used by Smith and Waterman, when **BestFit** was first described, used -0.333 for the mismatch penalty. You can **Fetch** RandomDNA.Cmp and rename it SWGapDNA.Cmp to use these values, or use the NWSGapDNA.Cmp which has no mismatch penalty at all.

ALIGNING LONG SEQUENCES

This program can align very long sequences if you know roughly where the alignment of interest begins. Run the program with the command line option `/LIMit`. Then set the starting coordinates for each sequence near the point where the alignment of interest begins and set gap shift limits on each sequence. The program will then align the sequences from your starting point such that the sequences do not get out of phase by more than the gap shift limits you have set. If you started both sequences at base number one and set the gap shift limit for seqeunce one to 100 and for sequence two to 50, then base 350 in sequence one could not be gapped to any base outside of the range 300-450 on sequence two.

ALIGNMENT METRICS

BestFit and **Gap** display four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality, described above, is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment.

Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are "similar". Symbols that are across from gaps are ignored. A similarity is scored when the symbol comparison value for a pair of symbols is greater than or equal to 0.50, the "similarity threshold." This threshold is also used by the display procedure to decide when to put a ':' between two aligned symbols. You can reset it from the command line with the second optional parameter of `/PAIr`. For instance the expression `/PAIr=1.0,0.5` would set the similarity threshold to 0.5.

THE SIMILARITY AND IDENTITY METRICS ARE NOT OPTIMIZED BY ALIGNMENT PROGRAMS SO THEY SHOULD NOT BE USED TO COMPARE ALIGNMENTS!!

We do not know of any good metric that estimates the probability of the alignment occurring by chance.

PEPTIDE SEQUENCES

If this program thinks your sequences are peptide sequences, it will use a symbol comparison table with matches equal to 1.5 and mismatches based on the evolutionary distance between the amino acids as measured by Dayhoff and normalized by Gribskov (Gribskov and Burgess (1986) Nucl. Acids Res. **14(16)**; 6745-6763). You can insist that your sequences are really peptide sequences with the command line option `/PROtein`.

RESTRICTIONS

This program cannot evaluate a surface of comparison larger than one million bases squared. A 100 x 10,000 comparison is possible as well as a 1,000 x 1,000 comparison. Input sequences may not be more than 30,000 symbols long. You can have your system manager increase the maximum surface of comparison if your VAX has enough virtual memory.

COMMAND LINE SUMMARY

All parameters for this program may be put on the command line. Use the option /CHEck to see the summary below and to have a chance to modify the command line before the program executes. The capitalized letters in the qualifier names are the letters that you must type. Square brackets ([and]) enclose qualifiers or parameter values that are optional. **Appendix XI** describes command line control in detail.

```
Syntax: $ BESTfit [/INfile1=]Hpr.Seq [/INfile2=]Hpf.Seq /Default

Required Parameters:
/BEGin1=1   /BEGin2=1     beginning of each sequence
/END1=2966  /END2=2740    end of each sequence
/NOREV1     /NOREV2       strand of each sequence
/GAPweight=5.0            gap weight
/LENgthweight=0.3         gap length weight
[/OUTfile1=]Hpr.Txt       output file for alignment

Local Data Files: [/DATa=]SWGapDna.Cmp  comparison table for nucleic acids
                  [/DATa=]SWGapPep.Cmp  comparison table for peptides

Optional Parameters:
/OUTfile2=Hpr.Gap          new sequence file for sequence 1 with gaps added
/OUTfile3=Hpf.Gap           "     "       "   "      "    2   "   "     "
/LIMit1=1 /LIMit2=240      limit the surface of comparison
/HIGhroad                  makes the top alignment for your parameters
/LOWroad                   makes the bottom alignment for your parameters
/PAIr=1.0,0.5,0.1          thresholds for displaying '|', ':', and '.'
/WIDth=50                  the number of sequence symbols per line
/PAGe=60                   adds a line with a form feed every 60 lines
/NOBIGGaps                 suppresses abbreviation of large gaps with '.'s
```

ACKNOWLEDGEMENTS

Gap and **BestFit** were originally written for Version 1.0 by Paul Haeberli from a careful reading of the Needleman and Wunsch (J. Mol. Biol. **48**; 443-453 (1970)) and the Smith and Waterman (Adv. Appl. Math. **2**; 482-489 (1981)) papers.

Limited alignments were designed by Paul Haeberli and added to the Package for Version 3.0. They were united into a single program by Philip Marquess for Version 4.0. Default gap penalties for protein alignments were modified according to the suggestions of Argos for Version 6.0 (personal communication; CABIOS in press).

LOCAL DATA FILES

The files described below supply data to this program. The program will automatically read them from a public data directory unless 1) you have a data file with exactly the same name in your current default directory, or 2) you name a file on the command line with an expression like /DATa1=MyData.Dat. The concept of a local data file is described in more detail in **Appendix X**.

If the first sequence you name is a nucleic acid, **BestFit** will use the symbol comparison table in the public file SWGapDNA.Cmp. (SW stands for Smith and Waterman.) If the first sequence you name is a peptide sequence, **BestFit** will read SWGapPep.Cmp instead. The presence of these files in your current default directory will cause **BestFit** to read your version instead. (See **Appendix II** for more considerations on symbol comparison tables.)

OPTIONAL PARAMETERS

The parameters and switches below can only be set from the command line. Optional parameters that apply to all programs are described in the **Introduction** to the **Program Manual**.

/LIMit1=20 and **/LIMit2=20**

let you set "gap shift limits" for each sequence. When you already know of a long similarity between two sequences you can "zip" them together using this mode. The beginning coordinates for each sequence must be near the beginning of the alignment you want to see. The alignment continues so that gaps inserted do not require the sequences to get out of step by more than the gap shift limits. You can align very long sequences rapidly. The surface of comparison is still limited to one million. The size of a comparison can be predicted by multiplying the average length of the two sequences times the sum of the two shift limits.

/OUTfile2=SeqName1.Gap /OUTfile3=SeqName2.Gap

This program can write three different output files. The first displays the alignment of sequence one with sequence two. The second is a new sequence file for sequence one, possibly expanded by gaps to make it align with sequence two. The third, like the second, is a new sequence file for sequence two, possibly expanded by gaps to make it align with sequence one. The program writes only the first file unless there are output file options on the command line. If there are any output files named on the command line, ONLY those output files are written.

Aligned sequences (in sequence files) can be displayed with **GapShow**.

/PROtein

sets a program to expect protein sequences. The GCG Package determines whether your sequence is protein or nucleic acid by looking at its composition. If a sequence is composed entirely (98 percent or more) of IUB standard nucleotide codes (**Appendix III**), it will appear to the GCG Package to be a nucleotide sequence. This option lets you insist that your sequences are proteins. Use **/NOPROtein** to insist that your sequences are nucleic acids.

/PAIr=1.0,0.5,0.1

The paired output file from this program displays sequence similarity by putting '|'s, ':'s, and '.'s between similar sequence symbols. Normally '|'s are put between symbols that are the same, ':'s are put between symbols whose comparison value is greater than or equal to 0.50, and '.'s are put between symbols whose comparison value is greater than or equal to 0.10. You can change these "match display thresholds" from the command line. The three parameters for **/PAIr** are the display thresholds for '|', ':', and '.'. The match display criterion for a '|' changes from symbolic identity (the default) to the quantitative threshold you have set in the first parameter. '|'s will no longer be inserted between identical symbols unless their comparison values are greater than or equal to this threshold. If you still want '|'s to connect identical symbols, put an 'x' in the first parameter instead of a number. See **Appendix II** for more information about symbol comparison tables.

/PAGe[=64]

adds form feeds to the output file in order to try to keep clusters of related information together. If you do not use **/PAGe** when you print the output from this program, it may cross from one page to another in a frustrating way, especially when you print on individual sheets. You can set the number of lines per page by adding a number to the **/PAGE** qualifier.

/WIDth=50

puts 50 sequence symbols on each line of the output file. You can set the width to anything from 10 to 150 symbols.

/NOBIGGaps

Gaps that extend one sequence by more than one complete line of output are normally abbreviated with three dots arranged in a vertical line. This switch suppresses these large gap abbreviations showing all the sequence characters across from large gaps.

/LOWroad and **/HIGhroad**

The insertion of gaps is, in many cases, arbitrary, and equally "optimal" alignments can be generated by inserting gaps differently. When equally optimal alignments are possible, this program can insert the gaps differently if you select either the **/LOWroad** or the **/HIGhroad** options. Here are examples for the alignment of GACCAT with GACAT with different parameters.

```
For:        Match = 1.0        MisMatch = -0.9
       Gap weight = 1.0   Length Weight =  0.0

LowRoad:    1 GACCAT 6
              || |||       Quality = 4.0
            1 GA.CAT 5

HighRoad:   1 GACCAT 6
              ||| ||       Quality = 4.0
            1 GAC.AT 5

For:        Match = 1.0        MisMatch =  0.0
       Gap weight = 3.0   Length Weight =  0.0

HighRoad:   1 GACCAT 6
              |||            Quality = 3.0
            1 GACAT. 5

LowRoad:    1 GACCAT 6
                 |||         Quality = 3.0
            1 .GACAT 5
```

Essentially the "low road" shifts all of the arbitrary gaps in sequence two to the left and all of the arbitrary gaps in sequence one to the right. The high road does exactly the opposite. When neither "high" nor "low" road is selected, the program tries not to insert a gap whenever that is possible and uses the high road alternative for all collisions.

Printed: October 2, 1989 11:49

APPENDIX 4: FLUORESCENCE ANALYSIS OF FIRST INTERMEDIATE IN CAB REFOLDING

Association of First Intermediate: Quenching Analysis

Previous studies of bovine growth hormone (bGH) revealed that the associated state of the protein could be assessed by using fluorescence and fluorescence quenching (Havel et.al., 1988). Therefore, fluorescence studies were performed on the association of the first intermediate in the refolding pathway of CAB. Denatured CAB in 5 M GuHCl was rapidly diluted to 2.0 M GuHCl and different protein concentrations. After equilibration for six hours, the fluorescence of each sample was measured with an excitation wavelength of 296 ηm and an emission wavelength of 340 ηm. Each fluorescence reading was then normalized by the total protein concentration. The normalized fluorescence was then plotted as a function of the final protein concentration ($[CAB]_f$) as shown in Figure A4.1. These results were comparable to the phenomena observed for bGH (Havel et.al., 1988). The decreasing fluorescence per protein molecule could have been caused by intermolecular quenching of tryptophans on the surface of the first intermediate. Also, the effect of PEG on the associated protein was measured. Equilibrium refolding was performed at 2.0 M GuHCl and different protein concentrations by using PEG (8000 MW) in the dilution buffer to provide a final concentration of 30 g/l. As shown in Figure A4.1, the fluorescence per protein decreased and was higher at the nonassociating conditions (< 10 μM CAB). Since PEG (8000 MW) at 30 g/l shifted the equilibrium association toward the first intermediate as discussed in Chapter 6, the insignificant difference between the results with and without PEG shown in Figure A4.1 is likely the result of a dominate intermolecular quenching effect.

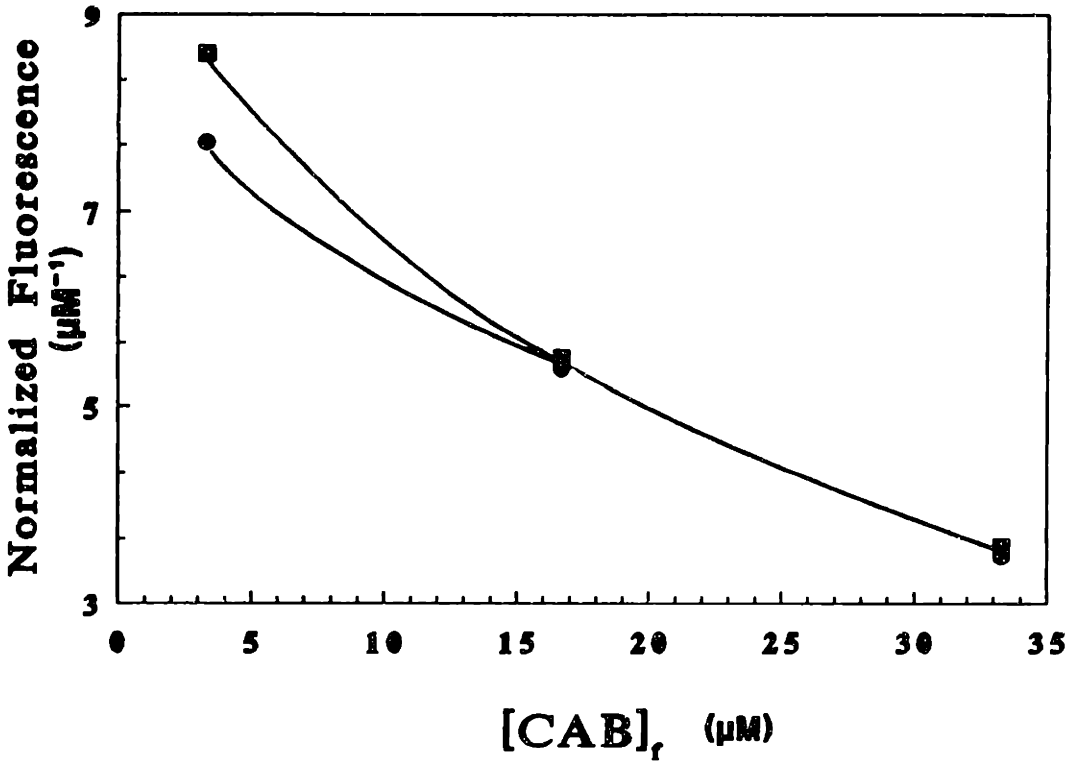


FIGURE A4.1: Normalized fluorescence ($F/[CAB]_f$) as a function of the final protein concentration, $[CAB]_f$, for equilibrium refolding in 2.0 M GuHCl. Fluorescence was measured at an excitation wavelength of 296 ηm and an emission wavelength of 340 ηm.

To further study the association of the first intermediate, fluorescence quenching studies were performed with iodide. Denatured CAB in 5 M GuHCl was diluted to 2.0 M GuHCl and different protein concentrations with and without PEG (8000 MW). After equilibration, iodide was added to provide different final concentrations ranging from 10 to 500 mM. The fluorescence was then measured for each sample by using an excitation wavelength of 296 ηm and an emission wavelength of 340 ηm and the fluorescence of a control solution at the same protein and GuHCl concentrations, F_o, was then divided by the sample fluorescence, F. Figure A4.2 displays the results of these measurements for final CAB concentrations of 6.7 μM (a), 16.7 μM (b), and 33.3 μM (c). These results were then used with Equation 5 from the Literature Review section to calculate the collisional quenching constant, K_{SV}, as shown in Figure A4.3. The collisional quenching constant, K_{SV}, was significantly lower at association conditions (> 10 μM CAB). In addition, the presence of 30 g/l PEG (8000 MW) did not have a significant effect on the quenching constant. These results clearly indicated that an association phenomena was occurring at the high protein concentrations and that PEG does not effect the quenching of the associated state.

PEG Effects on the Fluorescence Spectra

To further confirm that PEG will only interact with the first intermediate, fluorescence spectra were obtained for different final GuHCl concentrations. CAB in 5 M GuHCl was diluted to different final GuHCl concentrations and 0.10 mg/ml (3.33 μM) protein with and without PEG. After equilibration for six hours, the samples were excited at a wavelength of 296 ηm and the emission spectra was measured from 310 to 450 ηm.

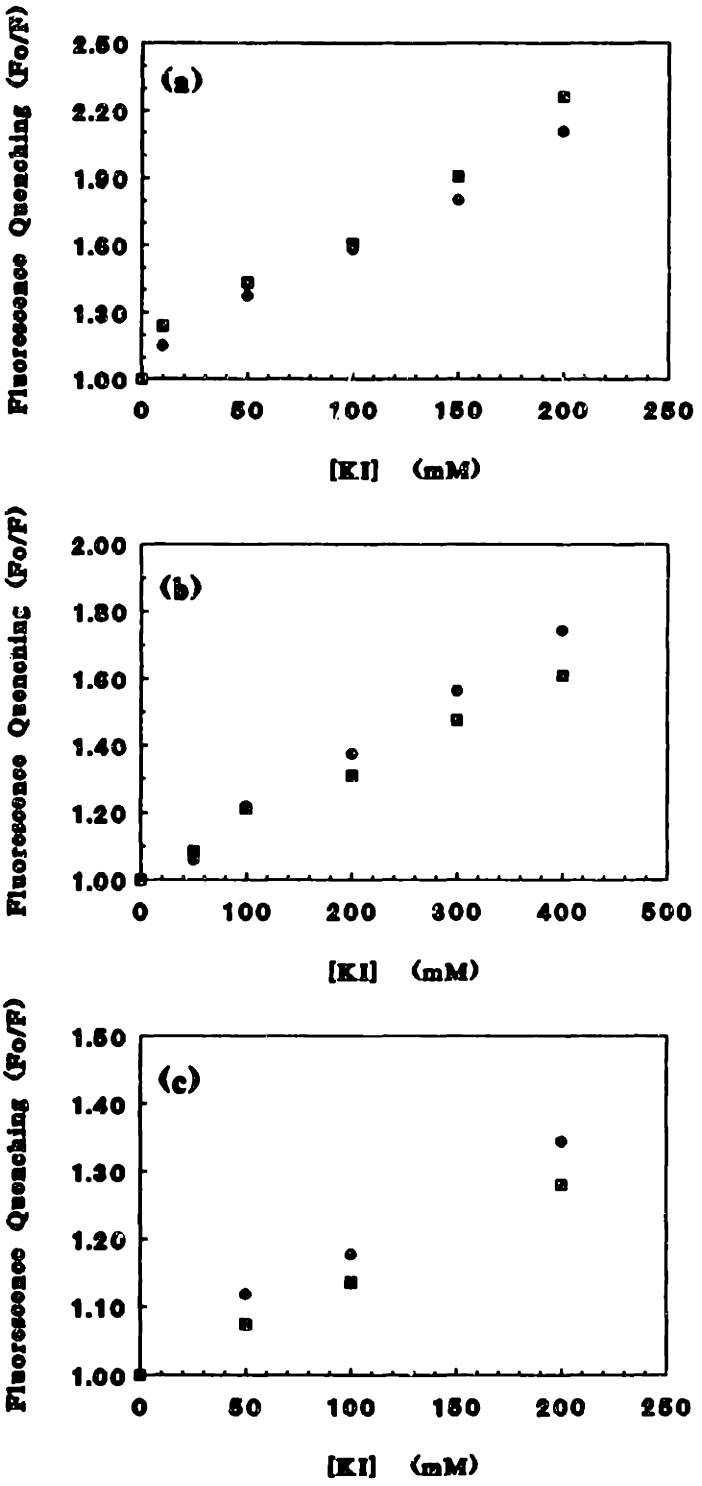


FIGURE A4.2: Fluorescence quenching results for equilibrium refolding at 2.0 M GuHCl and three final CAB concentrations: (a) 0.20 mg/ml (6.7 μM), (b) 0.50 mg/ml (16.7 μM), and (c) 1.0 mg/ml (33.3 μM). Potassium iodide, [KI], was added at different concentrations to each solution after refolding in 30 g/l PEG (8000 MW, ■) or buffer (●).

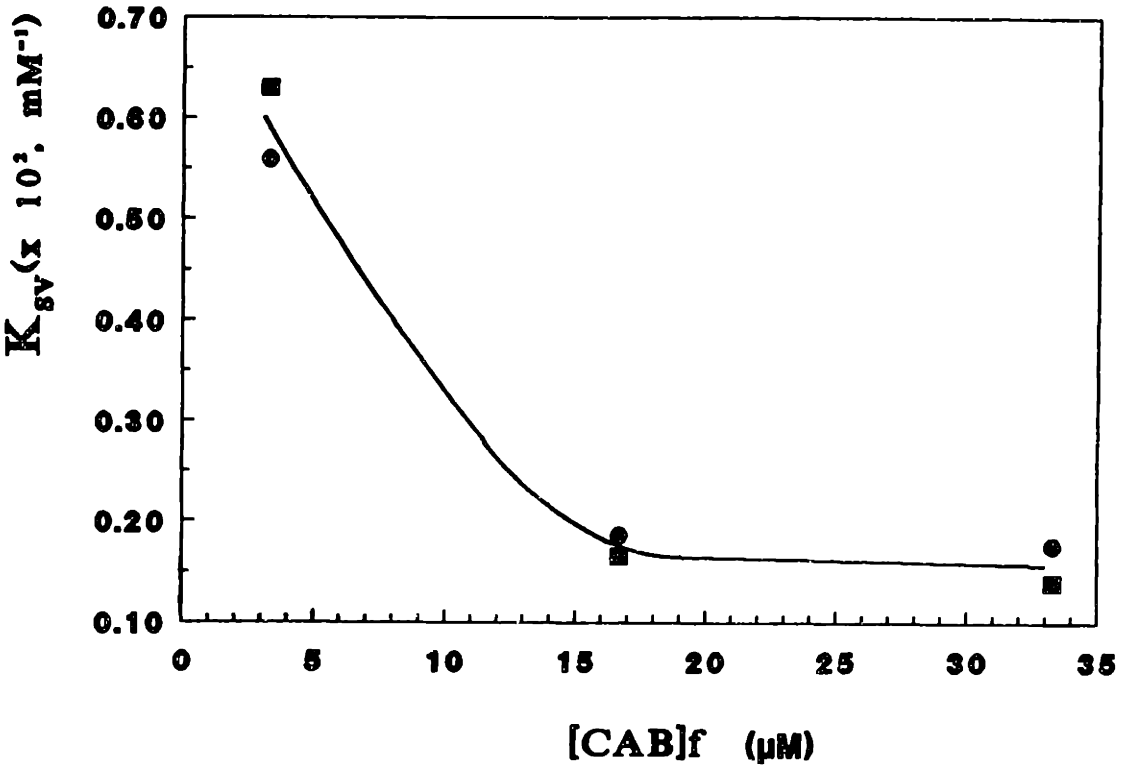


FIGURE A4.3: Collisional quenching constant, K_{SV}, calculated from the data in Figure A4.2 and Equation 5 in the Literature Review section. The data represent equilibrium refolding at different final CAB concentrations ($[CAB]_f$) in 30 g/l PEG (8000 MW, ■) or standard buffer (●).

As shown in Figure A4.4, the emission maximum increased from 340 ηm to 350 ηm with increasing GuHCl concentration. Thus, when the protein became more unfolded, the number of tryptophans exposed to the solvent environment increased and the hydrophobic core of aromatic residues dissolved. The dependence of the emission maximum on GuHCl concentration was different for refolding in 30 g/l PEG (8000 MW) as illustrated in Figure A4.5. A higher concentration of GuHCl was required to shift the emission maximum to 350 ηm for samples containing 30 g/l PEG. To analyze the effect of PEG, the difference between the two sets of spectra was calculated and plotted as shown in Figure A4.6. These results revealed that PEG primarily effected the fluorescence for the protein in 1.5 to 2.2 M GuHCl where the first intermediate is highly populated (Rodionova et.al., 1989).

Acrylamide Quenching Analysis

The fluorescence properties of CAB intermediates and their interaction with PEG was studied by using acrylamide as a quenching agent. Since acrylamide is a small molecule which can penetrate the protein matrix, it should provide additional insight into any conformational changes which result from PEG addition. Fluorescence quenching with acrylamide was first measured as a function of the final GuHCl concentration. Denatured CAB in 5 M GuHCl was diluted to 0.10 mg/ml (3.33 μM) and different final GuHCl concentrations with and without PEG (8000 MW). After equilibration for at least six hours, acrylamide was added to a final concentration of 0.20 M. The fluorescence of each sample was then measured at an excitation wavelength of 296 ηm and an emission of 340 ηm. The fluorescence of the same solutions without acrylamide, F_o, was also

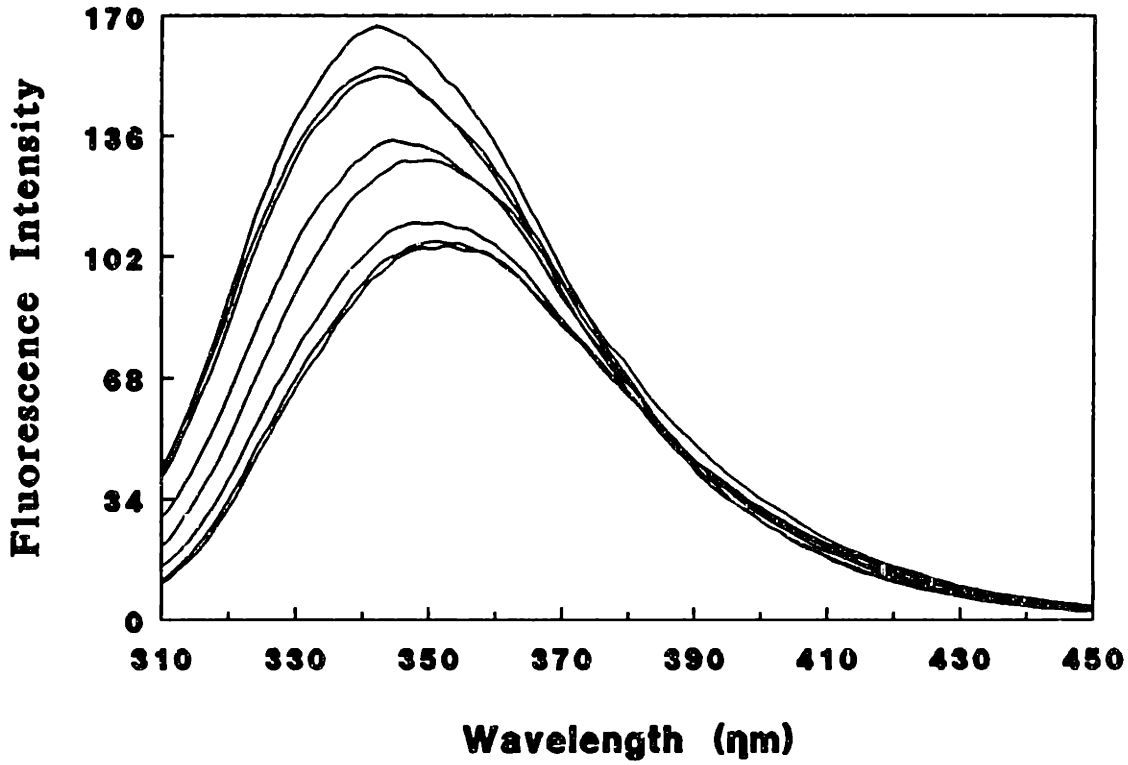


FIGURE A4.4: Fluorescence emission spectra measured at an excitation wavelength of 296 ηm. CAB in buffer (native) and CAB refolded from 5 M GuHCl to different final GuHCl concentrations are shown for a final protein concentration of 0.10 mg/ml (3.33 μM). The curve with the largest emission maximum at 340 ηm is the native protein and the maximum shifts to higher wavelengths and decreases in intensity with increasing final GuHCl concentration ([GuHCl]f= 0.50, 1.0, 1.5, 1.8, 2.0, 2.2, and 2.5 M).

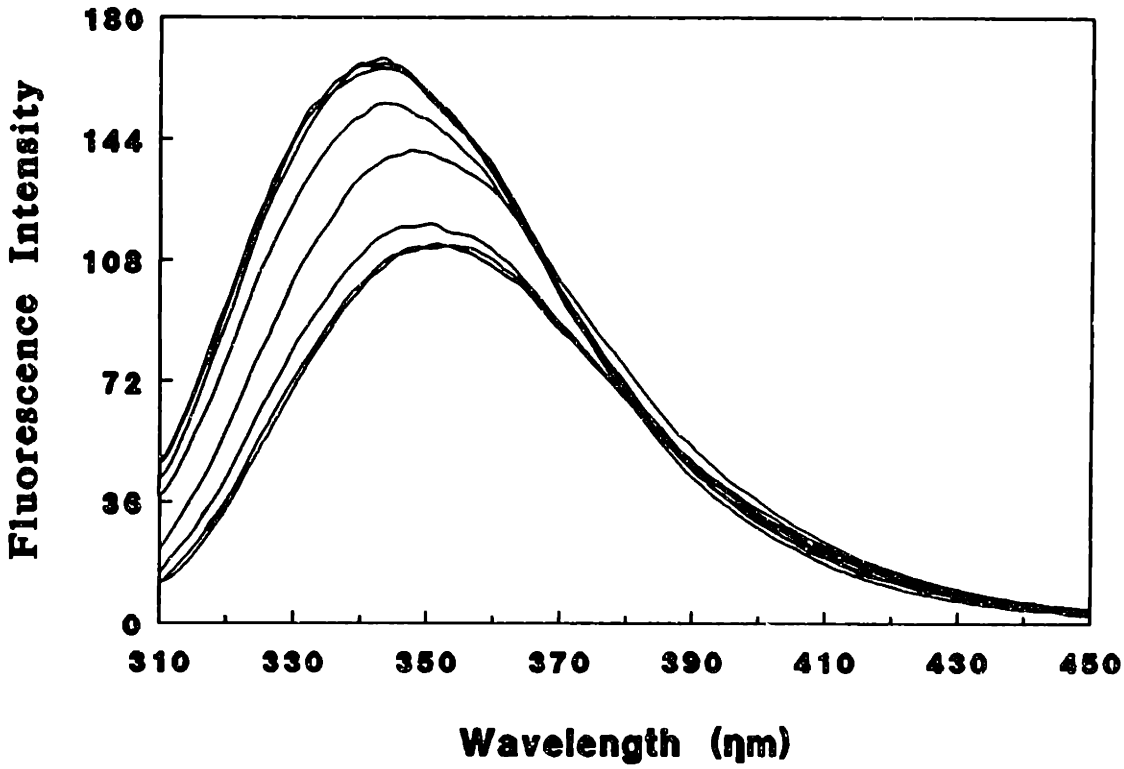


FIGURE A4.5: Fluorescence emission spectra measured at an excitation wavelength of 296 ηm. Native CAB and CAB refolded from 5 M GuHCl to different final GuHCl concentrations with 30 g/l PEG (8000 MW) are shown for a final protein concentration of 0.10 mg/ml (3.33 μM). The curve with the largest emission maximum at 340 ηm is the native protein and the maximum shifts to higher wavelengths and decreases in intensity with increasing final GuHCl concentration ([GuHCl]f= 0.50, 1.0, 1.5, 1.8, 2.0, 2.2, and 2.5 M).

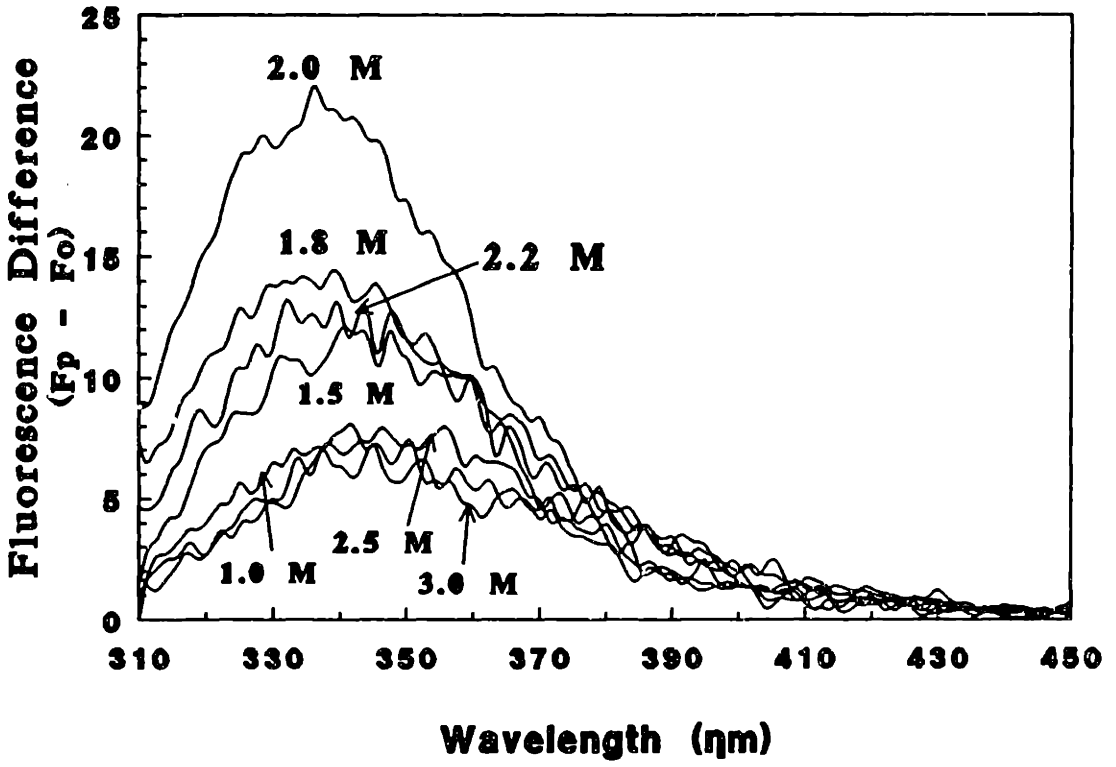


FIGURE A4.6: Difference in fluorescence emission spectra between CAB with 30 g/l PEG (8000 MW, F_p) and without PEG (F_o) measured at an excitation wavelength of 296 ηm. Native CAB and CAB refolded from 5 M GuHCl to different GuHCl concentrations are shown for a final protein concentration of 0.10 mg/ml (3.33 μM). The difference between native protein with and without PEG was negligible and not shown. The difference for refolding to each final GuHCl concentration was plotted as marked ([GuHCl]f= 0.50, 1.0, 1.5, 1.8, 2.0, 2.2, and 2.5 M).

measured and divided by the respective fluorescence of the solutions with acrylamide, F, to calculate the fluorescence quenching as shown in Figure A4.7. The fluorescence quenching reached a maximum at 1.7 M GuHCl without PEG and 1.9 M GuHCl with 30 g/l PEG (8000 MW). In addition, the overall profile for refolding in PEG was shifted to higher GuHCl concentrations. These results could be interpreted as a slight conformational change in the protein as a result of PEG bound to the surface.

To analyze the first intermediate quenching further, equilibrium refolding studies were performed at 2.0 M GuHCl and 0.10 mg/ml CAB. After equilibration, different concentrations of acrylamide were used to quench the fluorescence. The fluorescence of each sample, F, was measured at the same wavelengths and used along with the fluorescence without acrylamide, F_o, to calculate the quenching as shown in Figure A4.8. The concave upward shape of the quenching curve in Figure A4.8 has been described previously as a combination of collisional and static quenching effects (Eftink & Ghiron, 1981). The static quenching can be generally described as active quenching agent around a given fluorophore and the static quenching constant, V, can be determined from:

$$\frac{F_o / F}{\exp(V [Q])} = 1 + K_{SV} [Q]$$

By varying the quenching constants, a fit to the fluorescence quenching data was achieved as shown in Figure A4.9. The collisional quenching constants were 4.59 M^{-1} and 4.62 M^{-1} for solutions without and with 30 g/l PEG (8000 MW), respectively. The static quenching constants were also similar for two cases: 2.0 M^{-1} without PEG and 2.07 M^{-1} with 30 g/l

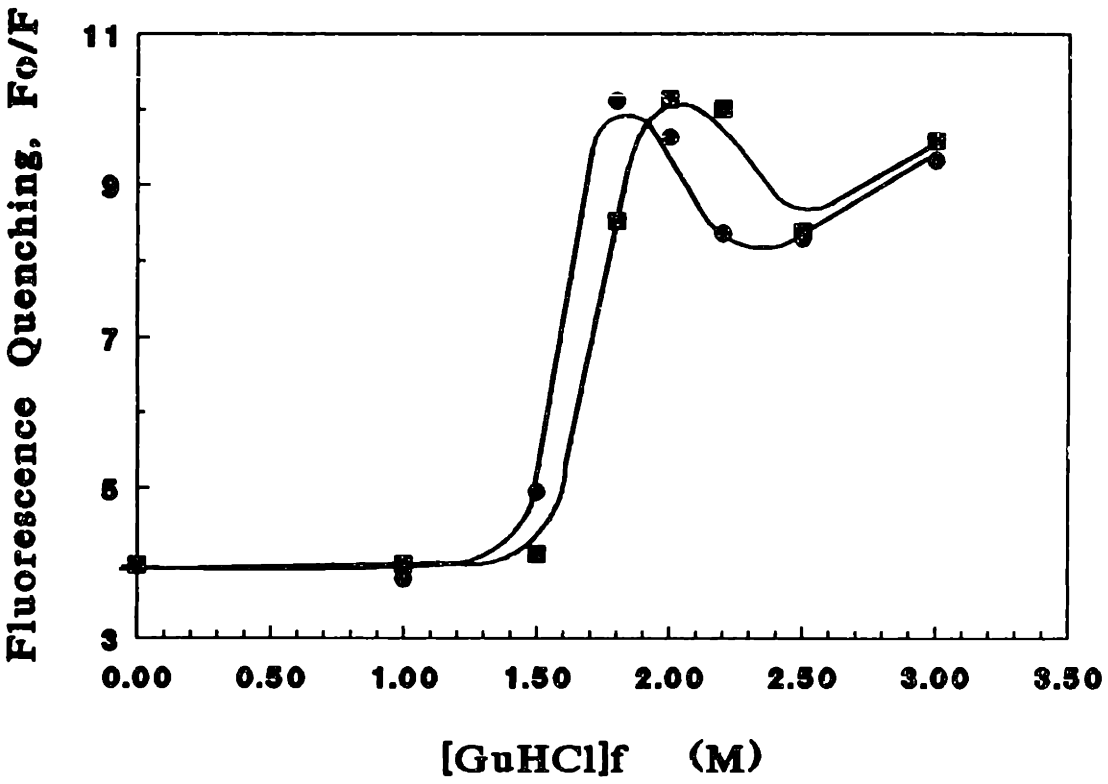


FIGURE A4.7: Fluorescence quenching by acrylamide as a function of the final GuHCl concentration at 0.10 mg/ml CAB without (●) and with 30 g/l PEG (■, 8000 MW). Refolding was performed by rapid dilution of CAB in 5 M GuHCl to different final GuHCl concentrations and 0.10 mg/ml CAB without and with 30 g/l PEG. After equilibration, acrylamide at a final concentration of 0.20 M was used to quench each solution.

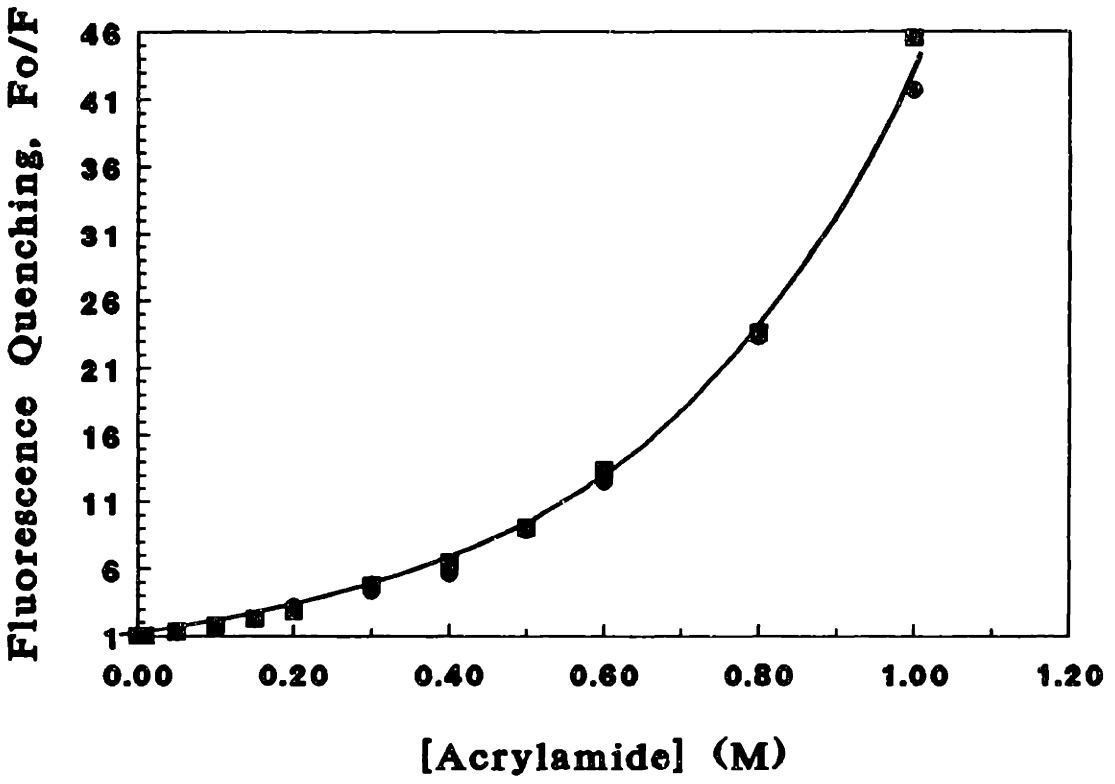


FIGURE A4.8: Fluorescence quenching of equilibrium refolded CAB (0.10 mg/ml) in 2.0 M GuHCl as a function of the acrylamide concentration without (●) and with 30 g/l PEG (■, 8000 MW). Fluorescence of native protein in buffer, F_o, and the fluorescence of the quenched solution, F, were measured at an excitation wavelength of 296 ηm and an emission wavelength of 340 ηm.

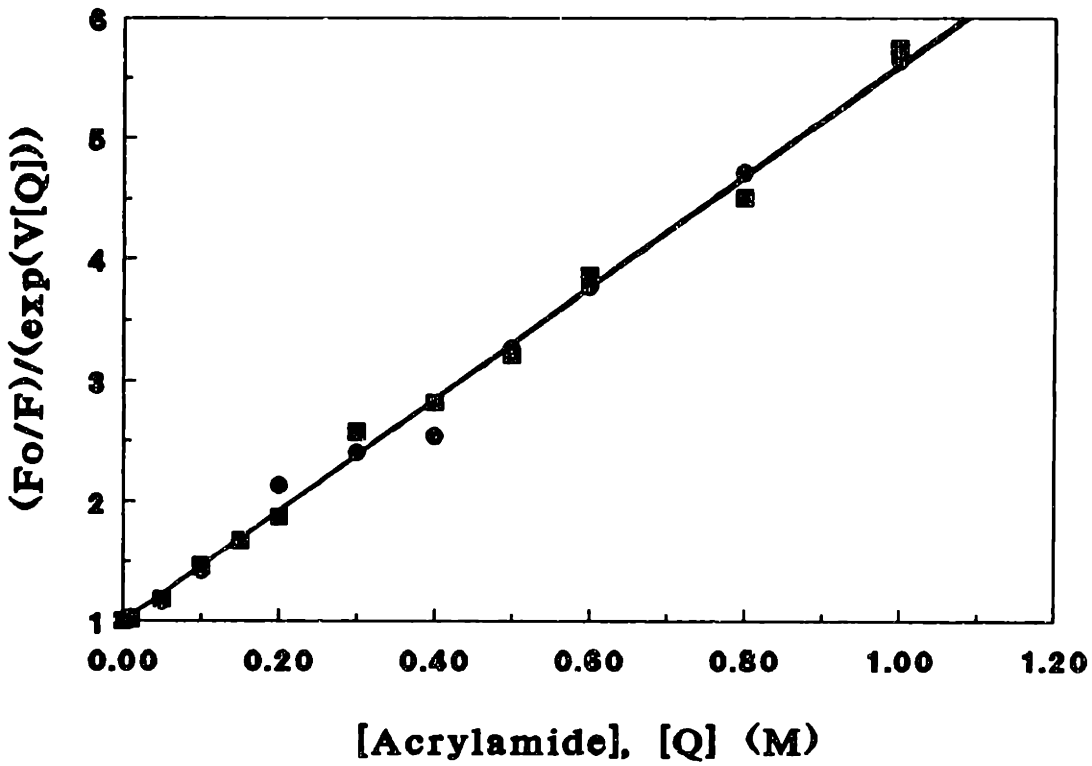


FIGURE A4.9: Calculation of quenching constants from data in Figure A4.8 and Equation A4.1 for quenching without (●) and with 30 g/l PEG (8000 MW, ■). The static (V) and collisional (K_{SV}) quenching constants were determined for each case as described in the text.

PEG. Since the constants were similar for both cases, acrylamide quenching of the first intermediate in 2.0 M GuHCl may not be affected by PEG. Acrylamide is much smaller than PEG (8000 MW) and acylamide can diffuse into the protein matrix. Therefore, acrylamide quenching would probably not be affected by PEG.

APPENDIX 5: TRANSMISSION ELECTRON MICROSCOPE STUDIES

Transmission electron microscope (TEM) studies were performed to measure the extent of aggregation and confirm the QLS analyses. The procedure which was used to perform the TEM studies has been described previously (Horne et.al., 1975; Horne, 1978). Briefly, the TEM studies were performed by placing each protein solution on a thin carbon film which covered a copper grid (400 mesh). Protein which was not adsorbed to the surface of the carbon film was rinsed away with distilled water. Uranyl acetate stain was added for a short period of time (30-45 seconds) and removed by rinsing with deionized water. The stain binds only to the carbon film and, thereby, results in a negative image. Each sample was allowed to dry overnight. After drying, the samples were analyzed by either a JEOL 100B or JEOL 200CX transmission electron microscope. Photographs taken with the transmission electron microscope show the protein as an unexposed area (white or gray) and the stained carbon as an exposed area (black or dark gray grains).

To initially test the applicability of TEM analysis, rapid dilution from 5 M GuHCl to final conditions of 0.20 mg/ml (6.7 μM) protein and 0.30 M GuHCl (aggregation conditions) was performed. At these conditions, large aggregates formed in the solution within a few minutes. Although any small multimers which were present at equilibrium could not be readily observed by QLS, the TEM photograph clearly showed the formation of both large aggregates and small multimers (Figure A5.1). As shown in Figure A5.1, the TEM analysis of this solution also revealed the presence of multimeric aggregates which were the size scale of monomers, dimers, and trimers. In addition, there appeared

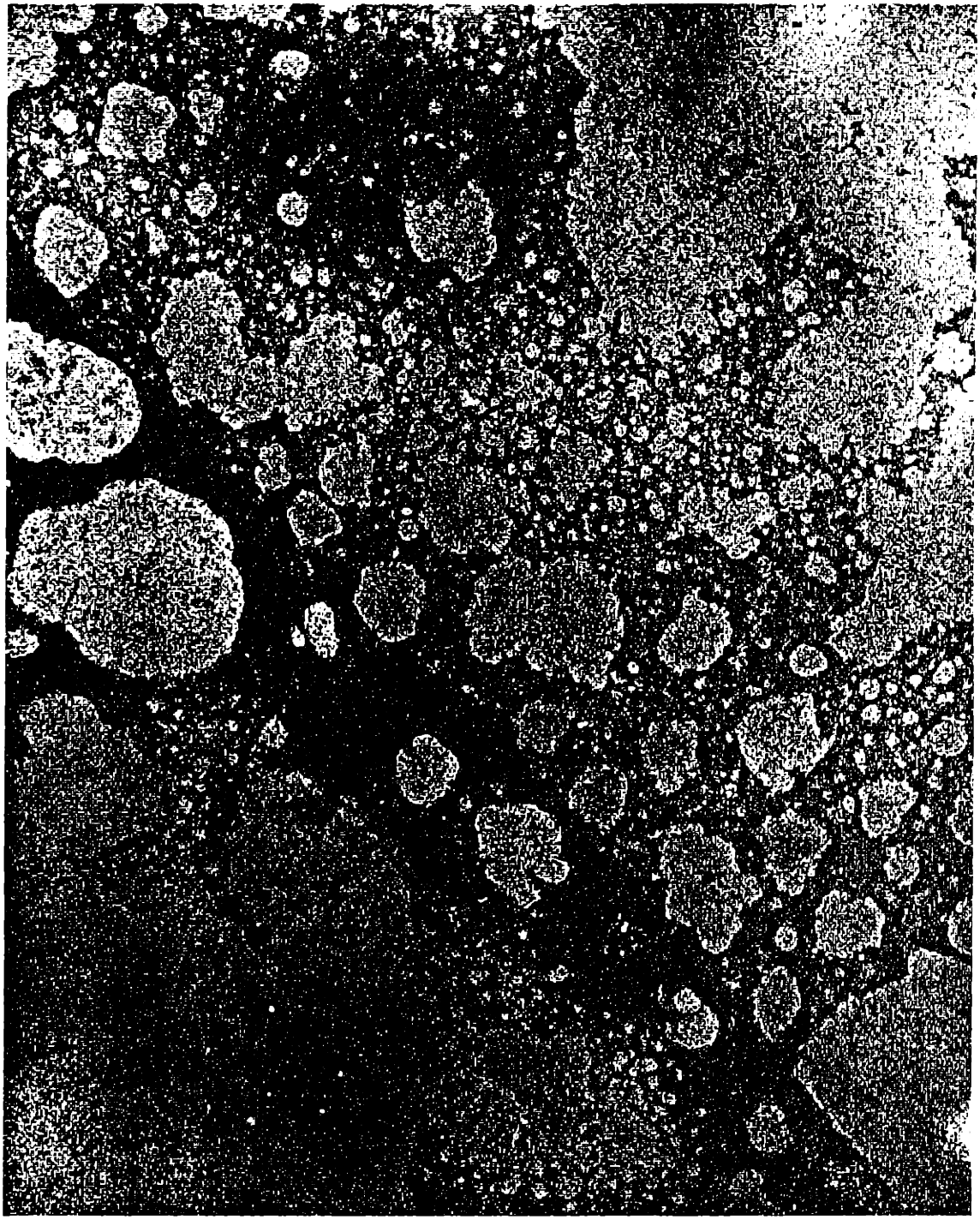


FIGURE A5.1: Transmission electron micrograph of CAB aggregates (Magnification x 150,000). CAB in 5 M GuHCl was diluted with buffer to 0.20 mg/ml CAB and 0.30 M GuHCl. After equilibration, the sample was adsorbed to the carbon film and negatively stained. The bimodal distribution of large aggregates and small multimeric species was observed on all sections of the grid. The multimer distribution based on computer enhanced measurements was 1.8 μM monomers, 1.9 μM dimers, and 0.4 μM trimers.

to be a distinct bimodal distribution of aggregates. One distribution consisted of aggregates which were small multimers and the other distribution consisted of large aggregates of greater than 50 ηm in diameter. The smaller species distribution consists of 1.8 μM monomers, 1.9 μM dimers, and 0.4 μM trimers determined from size analysis with conversion from number distribution to concentrations.

The final conditions of 0.50 mg/ml CAB and 0.60 M GuHCl also yielded large aggregates which precipitate out of solution. To confirm the existence of multimers, TEM analysis was performed on the supernatant obtained from these conditions (Figure A5.2). The supernatant consisted only of multimers and monomers of CAB. In particular, the dominant species were again monomers, dimers and trimers. At the same protein concentration (0.50 mg/ml, 16.7 μM) and higher GuHCl concentration (0.70 M), a bimodal distribution was observed (Figure A5.3). The smaller species (< 100 ηm) distribution for this case consisted of 5.9 μM monomers, 3.0 μM dimers, and 1.6 μM trimers. This distribution could be used to explain the activity level of the aggregated solution as shown in Figure 1.6 where the maximum active protein concentration is 4.2 μM. Finally, multimeric species formation at the upper limit for aggregation (Figure 1.3) of 0.50 mg/ml (16.7 μM) and 0.80 M GuHCl was confirmed as shown in Figures A5.4 and A5.5. The distribution of multimers for this case was 0.6 μM monomers, 5.53 μM dimers, and 1.7 μM trimers which was very comparable to the QLS results shown in Figure 1.2. Therefore, the TEM results provide additional evidence that dimers and trimers form during the aggregation process and the concentrations determined from TEM analysis were comparable to those obtained by QLS and activity analyses.

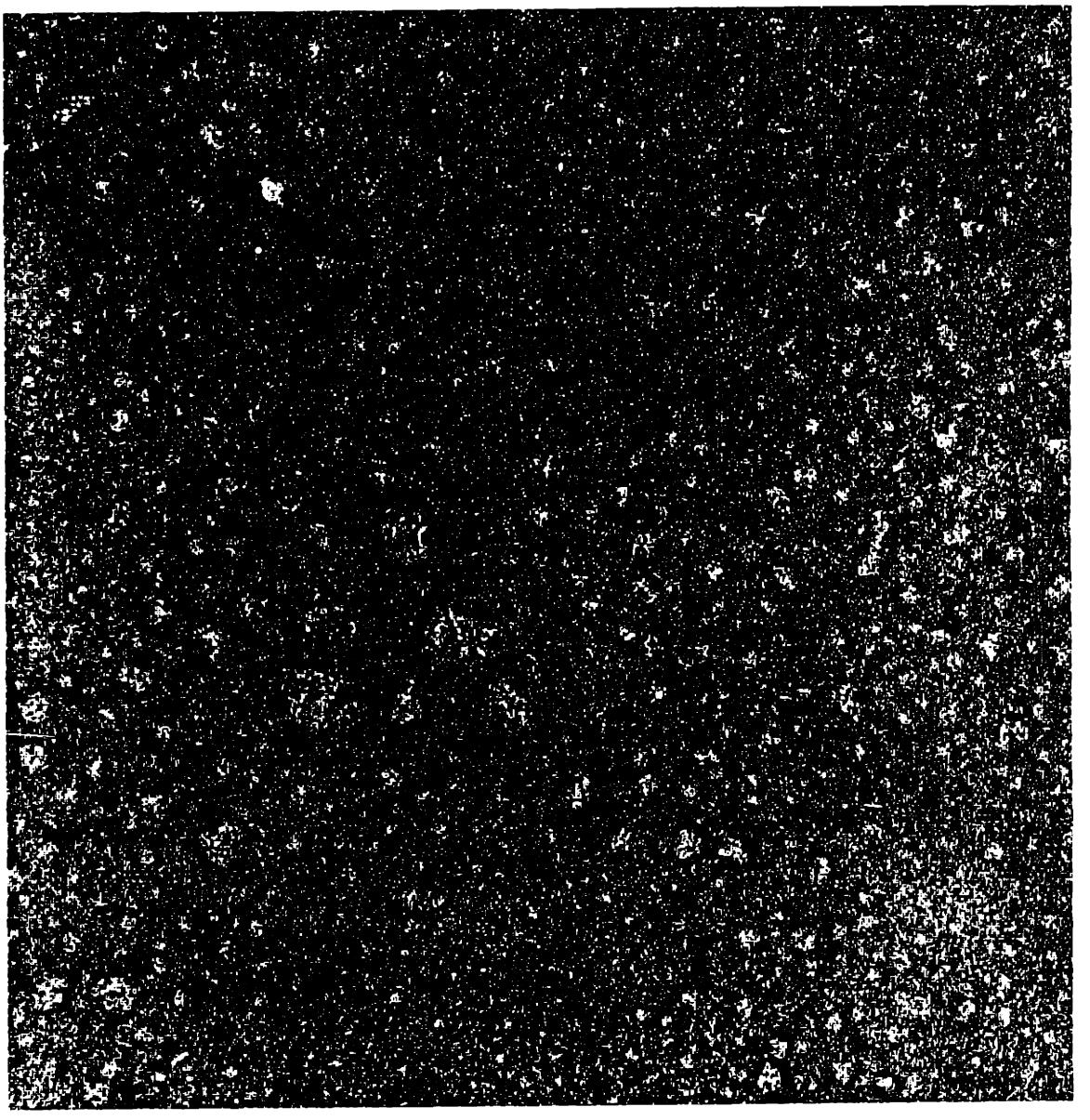


FIGURE A5.2: Rapid rate of aggregation case (0.50 mg/ml CAB, 0.60 M GuHCl) observed at equilibrium by TEM (Magnification x 150,000). The supernatant from the refolding as described in Figure 1.2.a was analyzed by TEM and revealed the presence of monomers, dimers, and trimers with few large aggregates as shown.

FIGURE A5.3: Moderate rate of aggregation case (0.50 mg/ml CAB, 0.70 M GuHCl) observed at equilibrium by TEM (Magnification x 150,000). The solution from the refolding as described in Figure 1.2.b consisted of a large aggregate distribution and a small multimer distribution. The distribution for the smaller species was 5.9 μM monomers, 3.0 μM dimers, and 1.6 μM trimers.

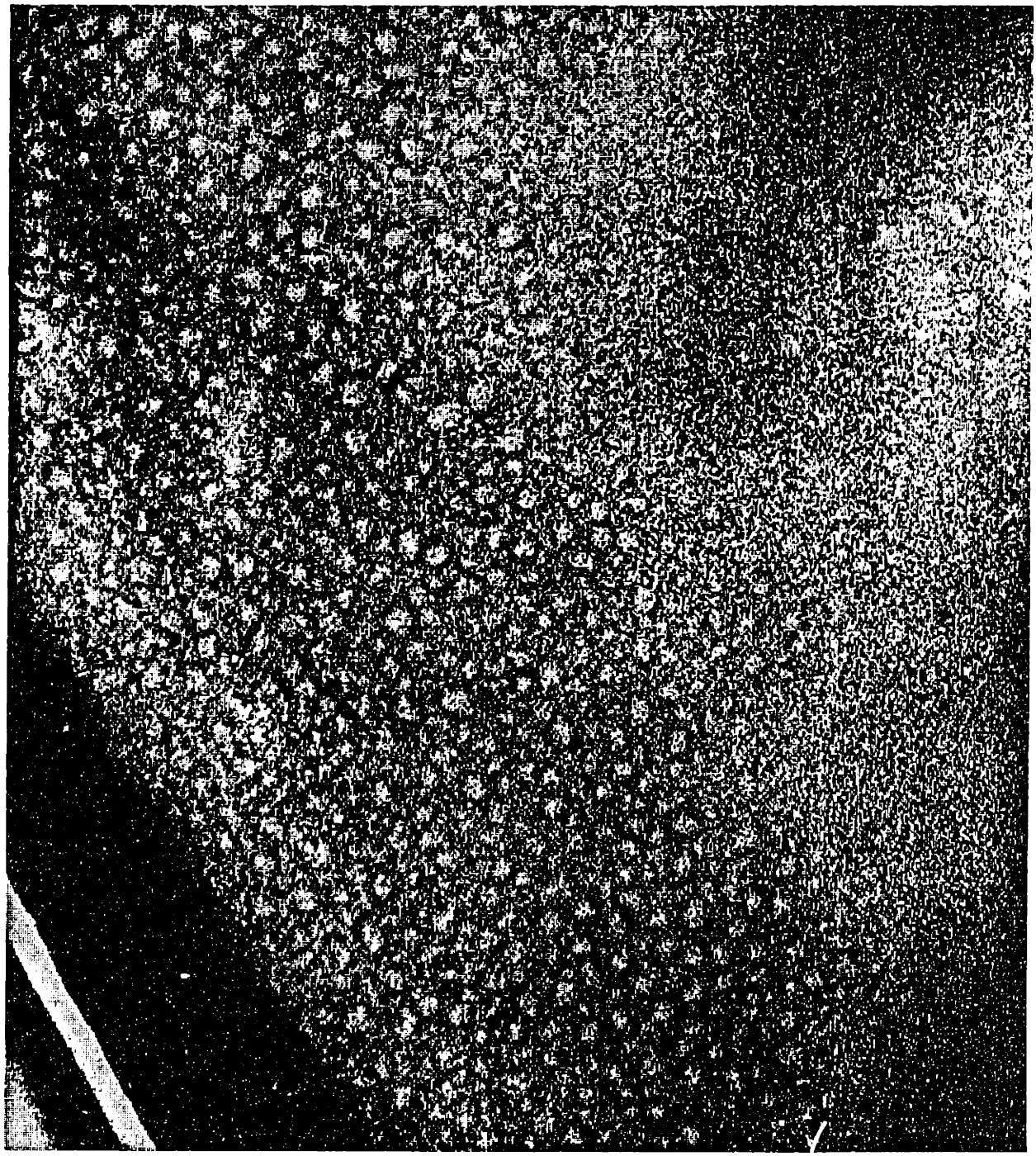


FIGURE A5.4: Slow rate of multimer formation (0.50 mg/ml CAB, 0.80 M GuHCl) observed at equilibrium by TEM (Magnification x 150,000). The solution from refolding as described in Figure 1.2.c in the Results and Discussion section was observed to form multimers in the absence of large aggregates. The concentration of each species was determined from the TEM to be 0.60 μM monomers, 5.5 μM dimers, and 1.7 μM trimers.

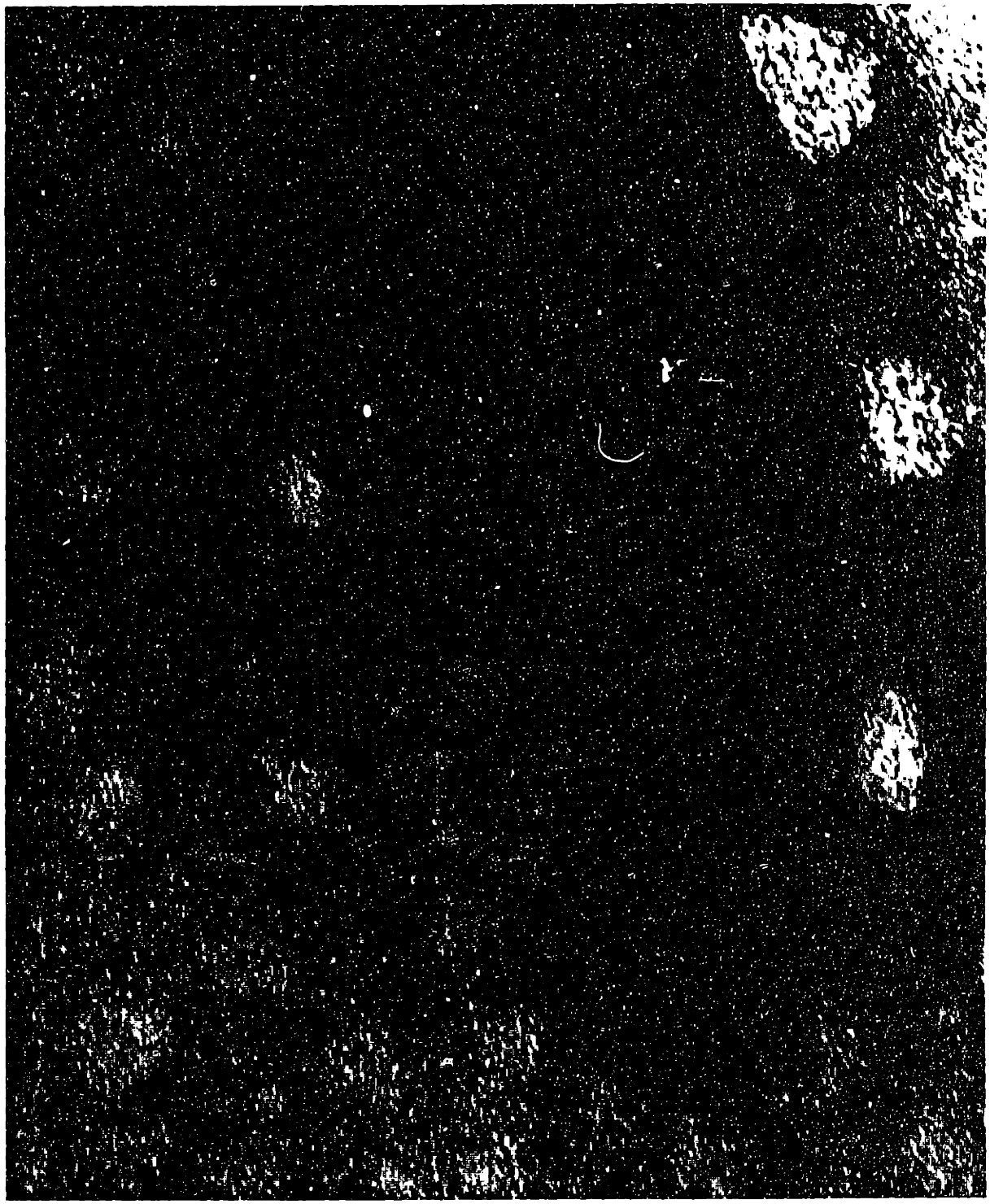


FIGURE A5.5: High Magnification (x 900,000) TEM photograph of monomer, dimer and trimer species observed at equilibrium under the conditions of slow multimer formation as shown in Figure A5.4 (see also Results and Discussion section, Figure 1.2.c).

APPENDIX 6: INHIBITOR EFFECTS ON CAB REFOLDING AND AGGREGATION

Since PEG specifically interacts with the first intermediate which does not have an active site, an inhibitor should not interfere with the action of the PEG. A noncompetitive inhibitor such as acetazolamide (AAA, see Figure A6.1) would also allow for the measurement of the formation of the active site relative to the inhibitor site on the protein. Therefore, AAA and PEGpNP were used to measure the effect of inhibitor and PEG interaction with the refolding and aggregation of CAB. AAA was used in a ten fold excess of the final protein concentration ($[CAB]_f$) since an excess concentration of inhibitor is required to completely inactivate the native protein (K_i = 0.4 μM; Pocker & Stone, 1967).

Initially, the effects of the noncompetitive inhibitor, AAA, were assessed for the final conditions which result in refolding. Dilution of CAB in 5 M GuHCl was performed by using a solution containing PEGpNP and AAA dissolved in the standard buffer. The final refolding conditions of 0.50 mg/ml (33.3 μM) CAB and 1.0 M GuHCl were achieved with a final PEGpNP concentration of 3 g/l and a final AAA concentration of 167 μM. The concentration of active protein was calculated from the hydrolysis rates as described previously (see Results and Discussion, Chapter 5). In addition, the hydrolysis rate of the native protein at the same final concentrations of protein, GuHCl, and AAA was measured as shown in Figure A6.2. The protein achieved a steady state level of activity which was the equilivent to the native protein activity at the same conditions (Figure A6.2). The rate of refolding did not appear to be altered by the inhibitor since CAB refolded without inhibitor reached a steady state at the same time as

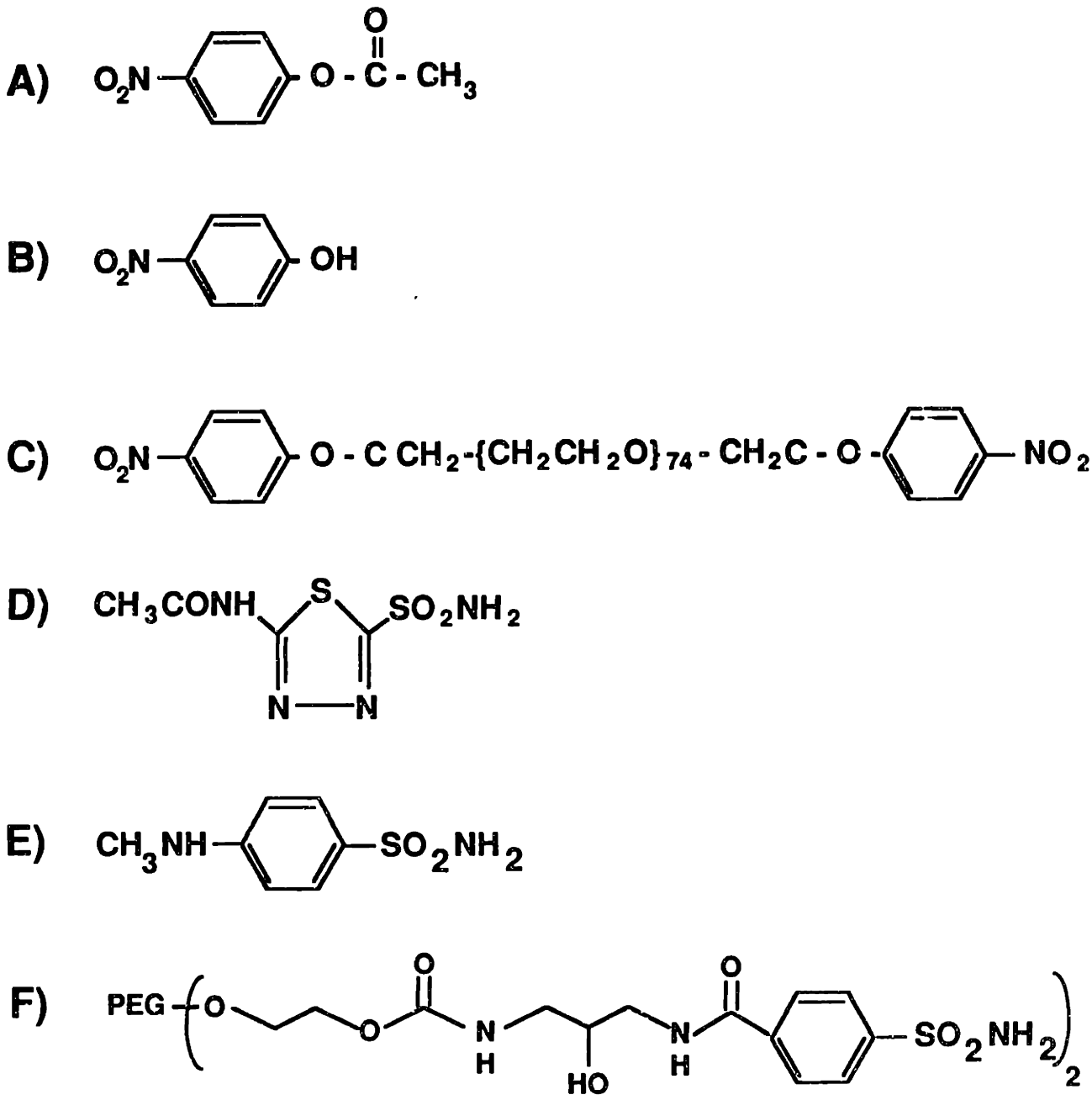


FIGURE A6.1: Chemical structures of substrates and inhibitors. A) p-nitrophenol acetate (pNPA); (B) p-nitrophenol (pNP); (C) bis(p-nitrophenol carbonate) polyoxyethylene (PEGpNP); (D) Acetazolamide (AAA); (E) p-aminomethyl benzene sulfonamide (SA); (F) bis(sulfonamide) polyethylene glycol (PEGSA)

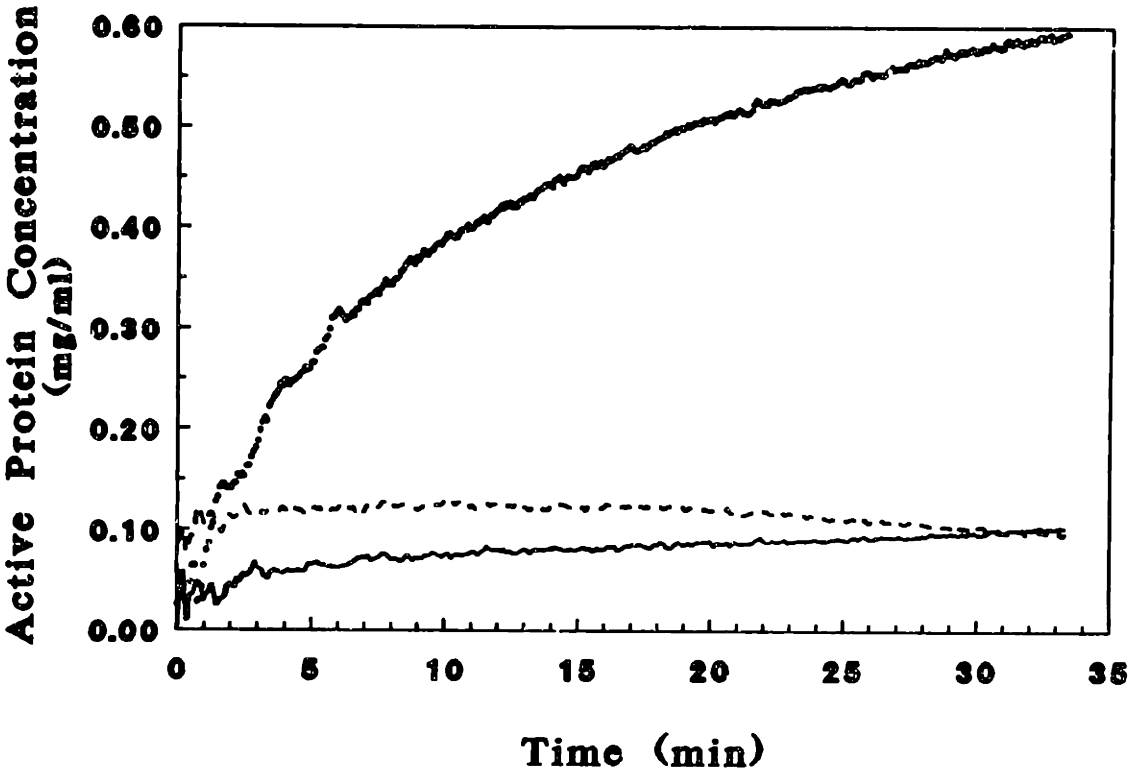


FIGURE A6.2: Effect of noncompetitive inhibitor on refolding of CAB. CAB in 5 M GuHCl was rapidly diluted to 0.50 mg/ml (16.7 μM) protein and 1.0 M GuHCl with 3 g/l PEGpNP (●). Refolding was also performed at the same conditions in the presence of 167 μM AAA (–). For comparison, the native protein at the same final conditions with 167 μM AAA is also shown (- -). The active protein concentration was determined by using hydrolysis rate of the native protein without inhibitor as described in the Materials and Methods section.

CAB refolded with inhibitor.

To determine if the noncompetive inhibitor would affect aggregation, refolding was performed by dilution of CAB in 5 M GuHCl to the aggregation conditions of 0.20 mg/ml CAB (6.7 μM) and 0.30 M GuHCl. Refolding at these conditions was performed in 67 μM AAA with and without 30 g/l PEGpNP. In addition, refolding with 30 g/l PEG (3350 MW) was also performed at these conditions as a control. As shown in Figure A6.3, refolding at aggregating conditions provided dramatically different results than those observed in Figure A6.2. For refolding with AAA only, the protein refolded to achieve 90% of its biological activity in 5 minutes and, then, its activity decreased to less than 10% after 20 minutes. As shown in Figure A6.3, these results paralleled the recovery of active protein observed for refolding in PEGpNP and AAA at the same concentration. At times greater than 7 minutes the concentration of active protein for the PEGpNP and AAA conditions was greater than that of refolding in AAA only. The difference could be explained by a higher effective concentration of substrate at the protein surface for the PEGpNP case resulting in greater rate of hydrolysis. The control case of refolding in 30 g/l PEG resulted in refolding of the protein at the same rate, but complete recovery of active protein was achieved under these conditions. The shape of the active protein concentration curve in Figure A6.3 can be explained by either the partial formation of the active site in the second intermediate which does not have the ability to bind the inhibitor or the formation of an active species which is not capable of binding the inhibitor. In either case, the inhibitor binding site must be formed after the active site to explain the results observed in Figure A6.3. In addition, the crystal structure of the AAA complex

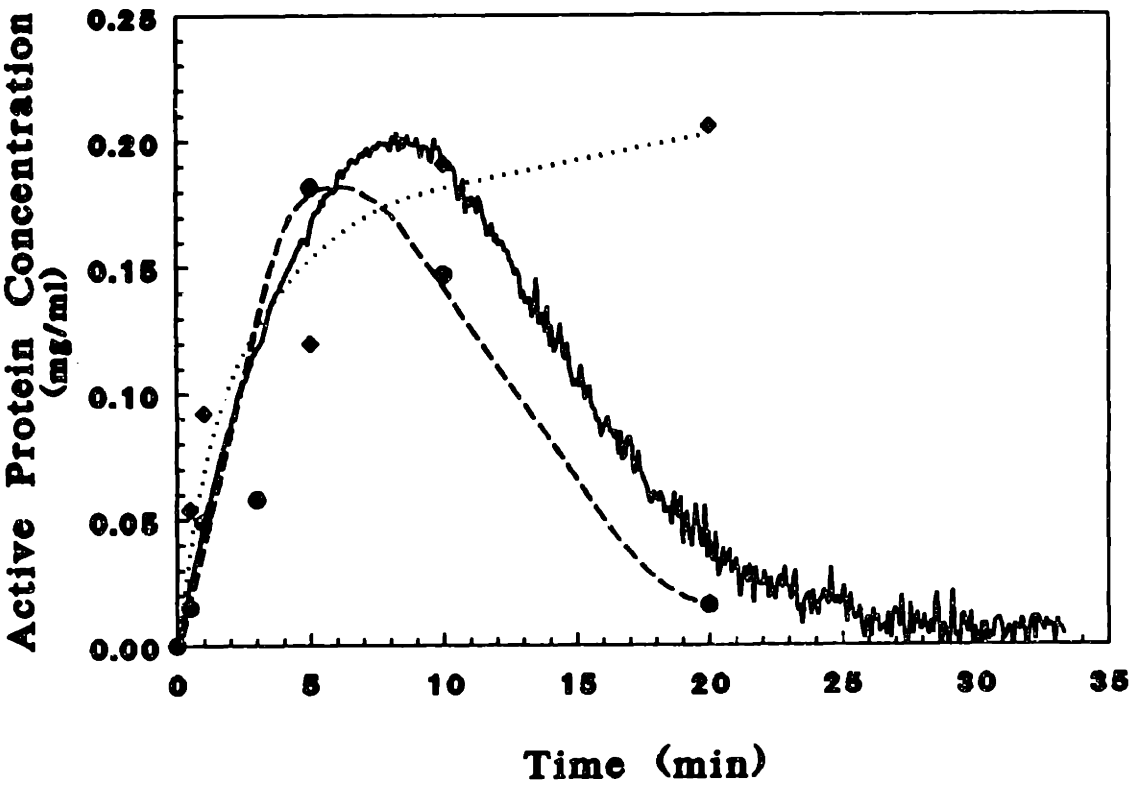


FIGURE A6.3: Noncompetitive inhibitor enhanced CAB refolding. CAB in 5 M GuHCl was diluted to the aggregation conditions of 0.20 mg/ml (6.7 μM) protein and 0.30 M GuHCl with three different solutions. Refolding in the presence of noncompetitive inhibitor at 167 μM was performed in buffer (-●-) or 30 g/l PEGpNP (–). For comparison, refolding was also performed without inhibitor in 30 g/l PEG (3350 MW, ...◆...).

with human carbonic anhydrase II revealed that the inhibitor bound to the active site in a different geometry than the substrate, p-nitrophenol acetate (pNPA) or PEGpNP (Vidgren et.al., 1990). To confirm the observed phenomena, unfolding of the native protein was performed by adding 8 M GuHCl to provide the final conditions of 0.20 mg/ml CAB and 0.30 M GuHCl with 30 g/l PEGpNP and 67 μM AAA. As shown in Figure A6.4, the active protein concentration for the unfolding mirrored the refolding results at the same final conditions. However, during unfolding, the protein became completely inactivated and, then, regained approximately 25% of its biological activity. If the second intermediate was stabilized by the PEGpNP and not inhibited by AAA, the observed active protein concentration after 25 minutes for unfolding could be explained by the activity of the second intermediate. From these results, a model of refolding with PEG and AAA was postulated as shown in Figure A6.5. Since the protein completely recovers activity at a greater rate in the presence of inhibitor and complete activity was achieved, the pathway of refolding with PEG and AAA should occur by the same pathway for PEG enhanced refolding with the inhibitor binding to the native state or partially binding to the second intermediate without inhibition. If inhibitor binds only to the native state, the rate of inhibitor binding must be slow relative to the rate of formation of the native state ($k_{AAA} < k_n$) such that the protein can recover nearly all of its acitivity before inactivation. In contrast, if the inhibitor interacts with the second intermediate without inactivation, the rate constant, k_{AAAi2}, could be slightly larger than the normal refolding rate constant which would explain the greater rate of folding observed without PEG in Figure A6.3. The pathway shown in Figure A6.5 could therefore represent the

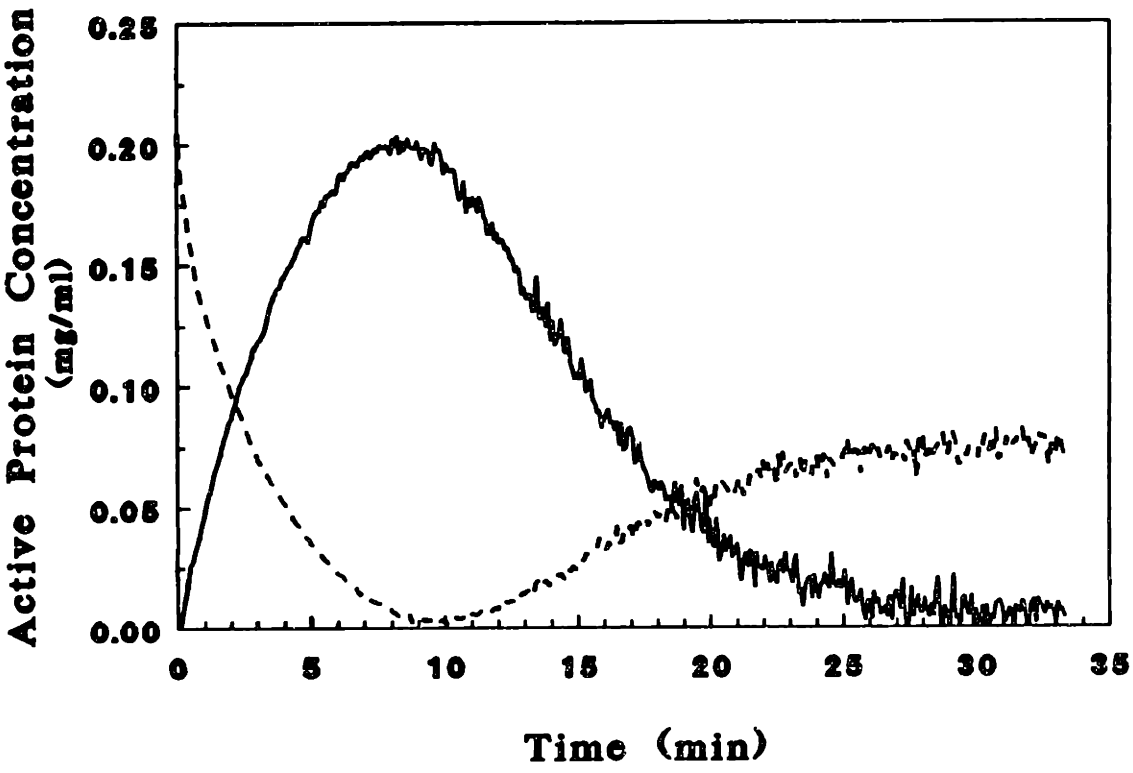


FIGURE A6.4: CAB refolding and unfolding with noncompetitive inhibitor. Denatured CAB in 5 M GuHCl was refolded by rapid dilution to 0.20 mg/ml CAB and 0.30 M GuHCl with 30 g/l PEGpNP and 167 μM AAA (–). Unfolding of CAB was performed by adding 8 M GuHCl with PEGpNP to native protein in buffer to achieve the final conditions of 0.20 mg/ml CAB, 5 M GuHCl, 30 g/l PEGpNP, and 167 μM AAA (- -).

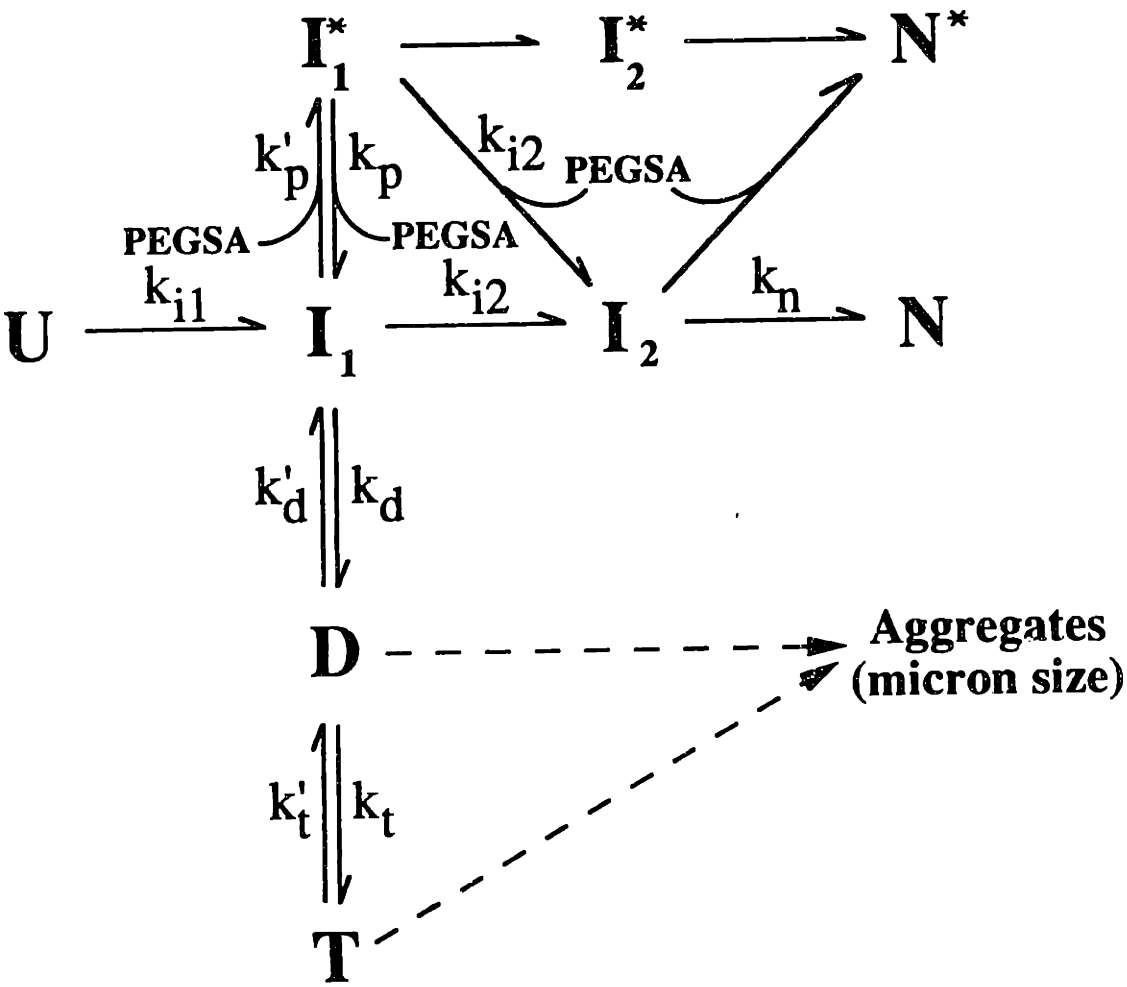


FIGURE A6.5: Proposed refolding pathway for CAB in the presence of PEG and noncompetitive inhibitor. Refolding and aggregation of CAB in PEG occurs by the pathway described previously (see Results and Discussion section, Figure 8.3). The noncompetitive inhibitor, AAA, could bind to the second intermediate and catalyze the folding to the inactive native structure (N^*) with a rate constant of k_{AAAi2}. Alternatively, AAA could bind to the native protein with a rate constant of k_{AAA} resulting in an inactive native structure (N^*).

refolding in the presence of PEG and inhibitor.

Additional studies of the combined effects of PEG and inhibitors were performed with the competitive inhibitor, sulfonamide (SA, Figure A6.1). The inhibitor was covalently attached to PEG as shown in Figure A6.1 to observe the combined effect of inhibitor and PEG. The PEGSA molecule was donated by Professor George Whitesides and synthesized by Yen-Ho Chen both from the Department of Chemistry, Harvard University. The refolding of CAB with PEGSA was initially assessed at the refolding conditions of 0.50 mg/ml (16.7 μM) CAB and 1.0 M GuHCl. Refolding was performed by rapid dilution of CAB in 5 M GuHCl to these final conditions with different concentrations of PEGSA all of which were in excess of the concentration required to completely inhibit CAB (K_i= 0.6 μM, Procter & Stone, 1967). The recovery of active protein at these conditions was then measured as a function of time as shown in Figure A6.6. For each concentration of PEGSA, the concentration of PEG was at the optimum or greater than the optimum concentration required for PEG enhanced refolding as discussed in Chapter 5 of the Results and Discussion (see Figure 5.3). The initial rate of refolding in PEGSA increased with PEGSA concentration as shown in Figure A6.6. However, the protein did not completely recover activity for these conditions.

To determine if PEGSA did provide a rate enhancement in refolding, refolding at the same final protein and GuHCl conditions was performed with 0.17 g/l PEGSA ($[PEGSA]_f/[CAB]_f$= 3) and measured after dilution by absorbance change at 280 ηm as described in Chapter 3 of the Results and Discussion section. This analysis resulted in rate constants for refolding to the second intermediate (k_{i2} = 1.48 min^{-1}, $t_{1/2}$ = 28.2 sec)

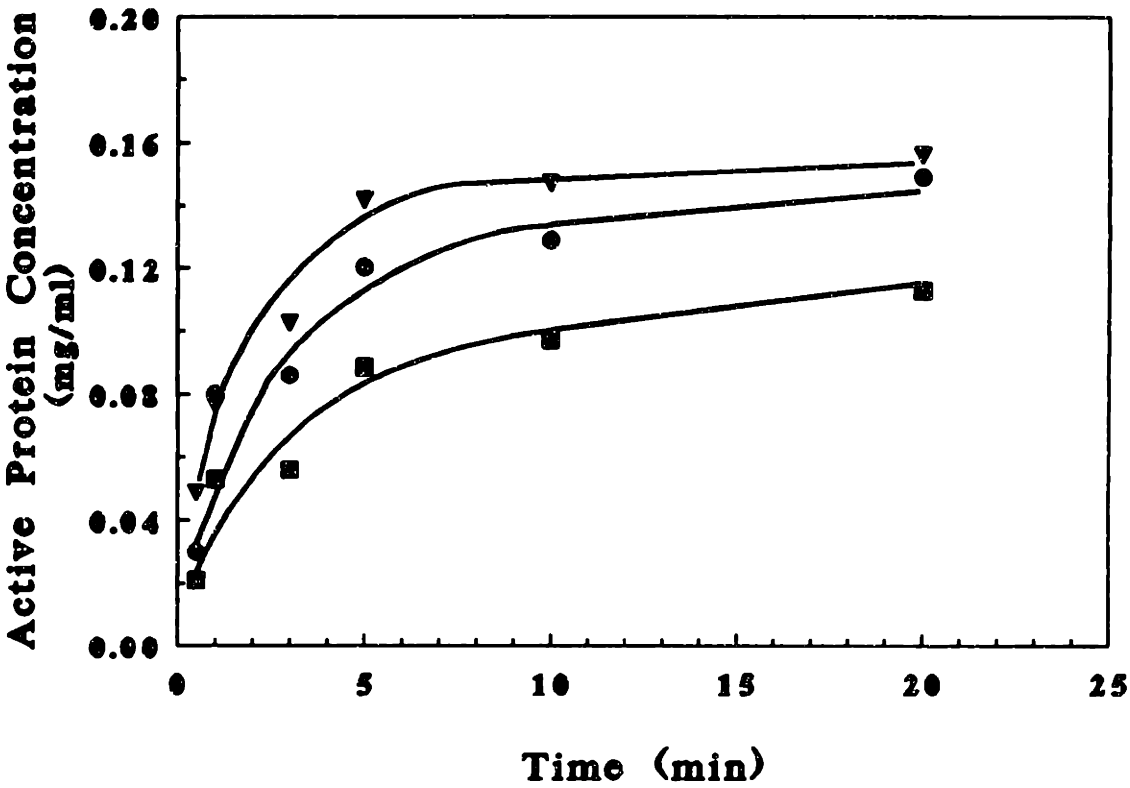


FIGURE A6.6: Refolding of CAB with PEG-inhibitor analog. CAB in 5 M GuHCl was refolded by rapid dilution to 0.50 mg/ml (16.7 μM) protein and 1.0 M GuHCl with different concentrations of PEGSA (see Figure A6.1). The final concentrations of PEGSA resulted in the following molar ratios of PEG to CAB ($[PEG]_f/[CAB]_f$) and SA to CAB ($[SA]_f/[CAB]_f$), respectively: 1) 18:1, 200:1 (▼); 2) 9:1, 100:1 (●); and 3) 2:1, 20:1 (■).

and native state (k_n = 6.04 x 10^{-2}, $t_{1/2}$ = 11.5 min) which were only slightly greater than those obtained with PEG at the same conditions (k_{i2} = 1.39 min^{-1}, k_n = 7.76 x 10^{-2} min^{-1}; Results and Discussion, Figure 8.2). To determine if PEGSA bound to a CAB folding intermediate, refolding was performed at 0.50 mg/ml (16.7 μM) CAB and 1.0 M GuHCl with 29 g/l PEG (3350 MW) and 1 g/l PEGSA ($[PEG]^f/[CAB]^f$= 18; $[SA]^f/[CAB]^f$= 200). The recovery of active protein was measured as a function of time as shown in Figure A6.7. Also, the concentration of each species was calculated from the model developed in Chapter 8 of the Results and Discussion and plotted in Figure A6.7. These results indicated that PEGSA could bind to the second intermediate and inhibit the activity. Refolding was also performed with PEG and SA to assess the effect of inhibitor at the protein surface. Denatured CAB in 5 M GuHCl was refolded by rapid dilution to 0.50 mg/ml (16.7 μM) protein and 1.0 M GuHCl with 0.17 g/l PEG (3350 MW; $[PEG]_f/[CAB]_f$ = 3) and SA at three different concentrations. As shown in Figure A6.8, the recovery of active protein was independent of the SA concentration for the conditions attempted ($[SA]_f/[CAB]_f$ > 20). Unlike refolding with PEGSA, the initial rate of refolding was approximately the same for each SA concentration. Therefore, PEGSA provided a greater effective concentration of inhibitor at the protein surface which could have facilitated a more rapid rate of refolding. From these studies, a refolding pathway with PEGSA was postulated as shown in Figure A6.9. The first intermediate interacts with the PEG portion of the polymer as in the previous model (Chapter 8, Figure 8.4, Results and Discussion). As second intermediate formed the SA portion of the polymer bound and inhibited recovery of activity. Additional experiments would be required to determine if PEGSA

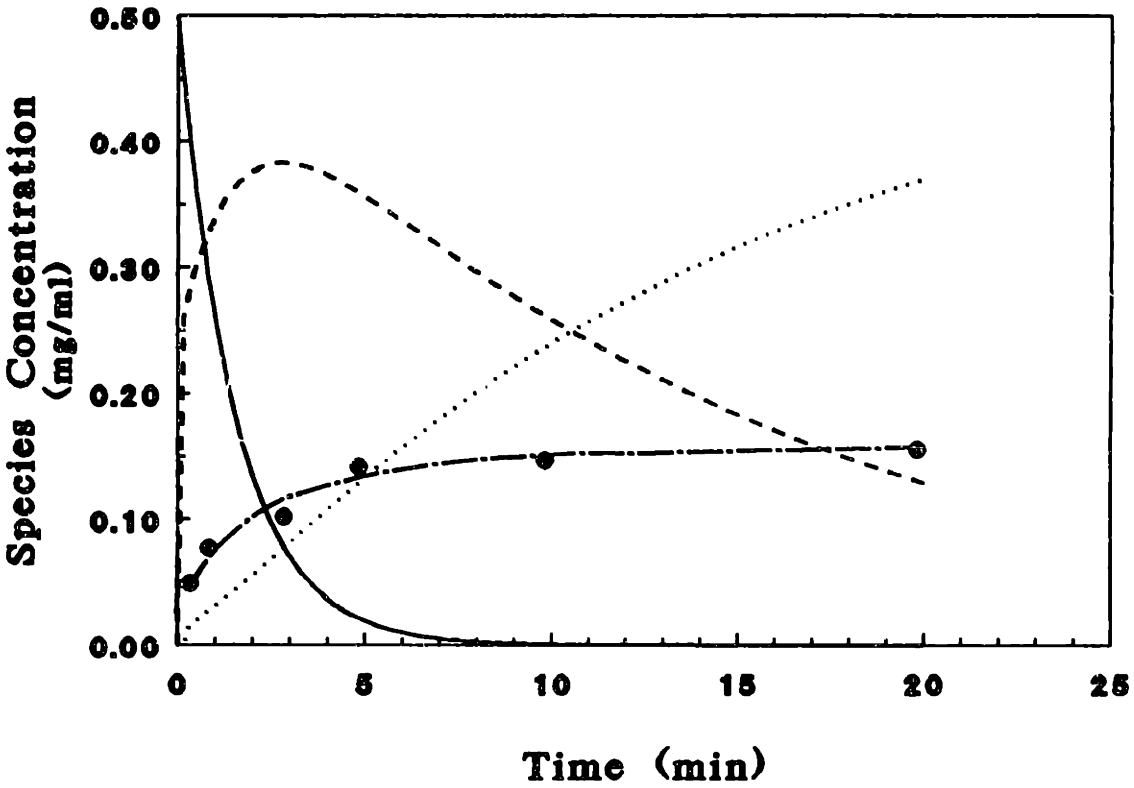


FIGURE A6.7: Determination of PEG-competitive inhibitor analog interaction with CAB refolding intermediates. Denatured CAB in 5 M GuHCl was refolded by rapid dilution to 0.50 mg/ml protein and 1.0 M GuHCl with buffer containing PEG (3350 MW) and PEGSA (- ● -). The final concentrations were 29 g/l PEG and 1 g/l PEGSA ($[SA]_f/[CAB]_f$ = 200). The concentration of the first intermediate (–), second intermediate (- -), and native protein (···) were calculated from the model developed for PEG enhanced refolding (see Results and Discussion section, Chapter 8).

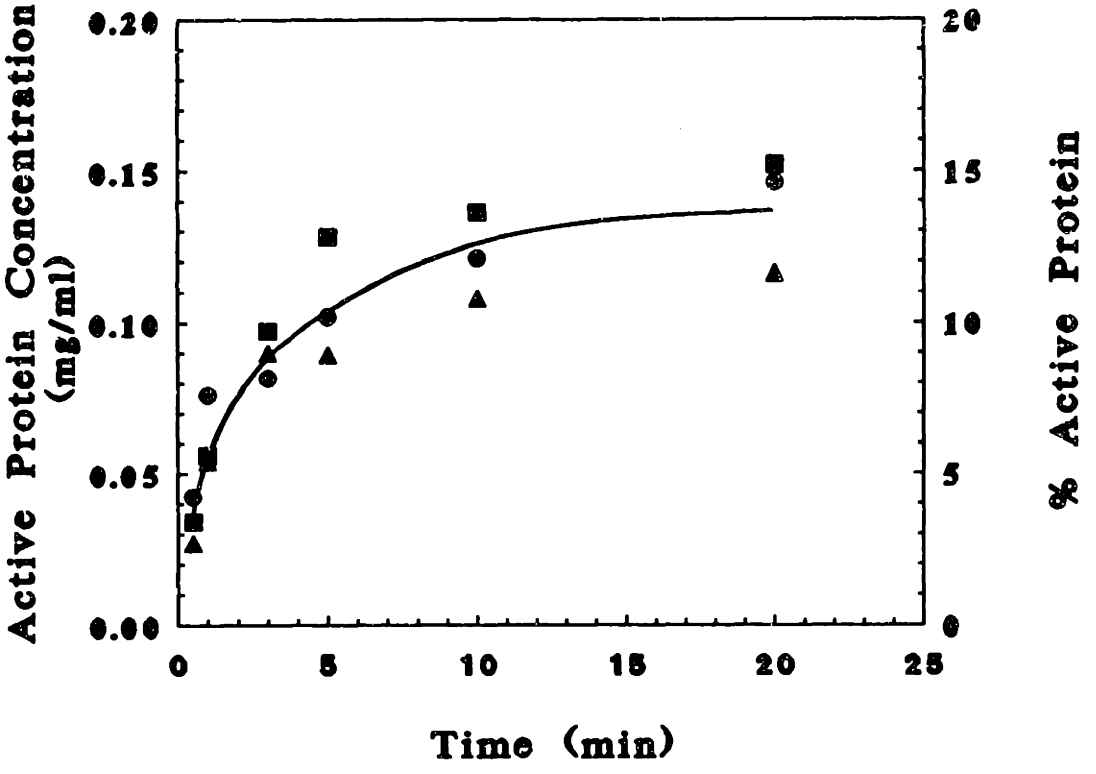


FIGURE A6.8: Refolding of CAB with PEG and competitive inhibitor. Refolding was performed at 0.50 mg/ml CAB and 1.0 M GuHCl with 30 g/l PEG (3350 MW) and different concentrations of competitive inhibitor, SA (see Figure A6.1). The final SA to CAB molar ratios ($[SA]_f/[CAB]_f$) were 20 (▲), 100 (●), and 200 (■). The solid line approximates an average of the data at each time point.

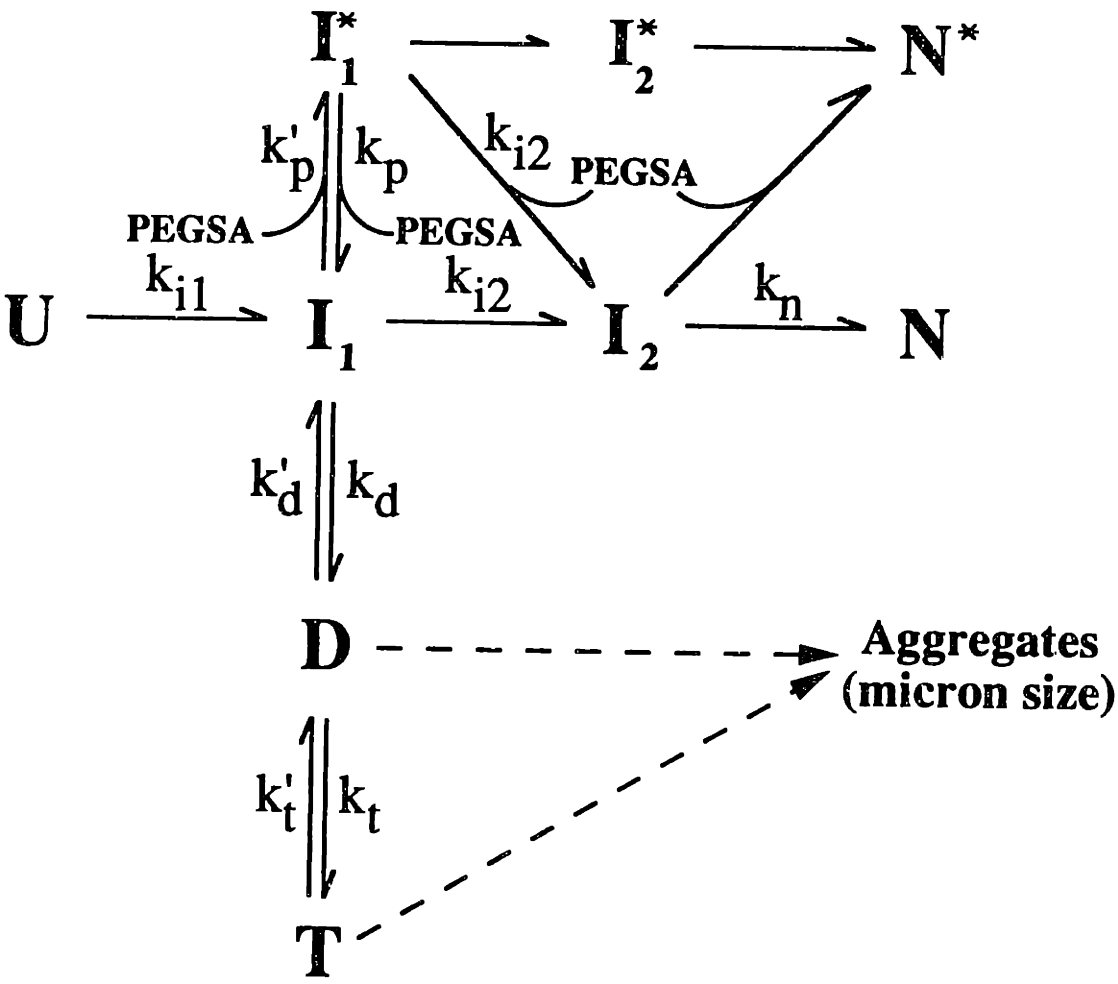


FIGURE A6.9: Proposed refolding pathway for CAB in the presence of PEG-competitive inhibitor complex (PEGSA). The PEG portion of the complex will interact with the folding intermediates as described previously (see Results and Discussion section, Figure 8.3). PEGSA will bind to the first intermediate (I_1) with the PEG portion of the molecule to form the specie, I^*_1. The PEGSA could then either remain bound while the protein folded to form the second intermediate, I^*_2, or dissociate from the first intermediate prior to the formation of the second intermediate, I_2. PEGSA could also either remain bound throughout the folding or bind at the SA moiety to form the inactive native structure, N^*.

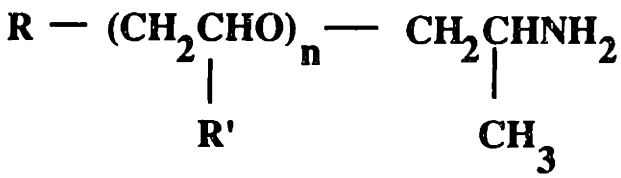
remained bound throughout the refolding process. In general, inhibitors could be useful in studying the formation of the active site during refolding and could also be used to catalyze the active site formation.

APPENDIX 7: HYDROPHOBIC POLYMERS FOR REFOLDING AND AGGREGATION OF CAB

Since PEG binds to the first intermediate in the refolding of CAB and prevent aggregation, other polymers which have similar properties could also be useful for enhancing refolding. To test this hypothesis, block copolymers of ethylene oxide (EO) and propylene oxide (PO) as shown in Figure A7.1 were used in the refolding of CAB. The protein in 5 M GuHCl was diluted to 0.50 mg/ml CAB and 1.0 M GuHCl with each polymer. The final polymer to protein molar ratio ($[Polymer]_f/[CAB]_f$) for each case was 3 which was the optimum for PEG enhanced refolding. The recovery of active protein for each polymer is shown in Figure A7.2. The complete recovery of active protein was not achieved for any of these conditions. The polymer with the lowest fraction of propylene oxide recovered the maximum concentration of active protein (0.12 mg/ml, 20% active). These results indicated that more hydrophobic molecules such as the propylene oxide portion of the polymers could bind to the first intermediate and prevent it from folding to the native state. Alternatively, these polymers could become locked in the hydrophobic active site cleft and inhibit the activity. Additional studies with these polymers could yield further insight into the interactions between PEG and the folding intermediates.

a) M Series

	PO/EO	MW
M-1000	3/19	1000
M-2070	10/32	2000



R' = H, EO (ethylene oxide)

R' = CH_3, PO (propylene oxide)

b) ED Series

	b(EO)	a+c=2.5(PO)	PO/EO
ED-6000	132	2.5	3.5/132

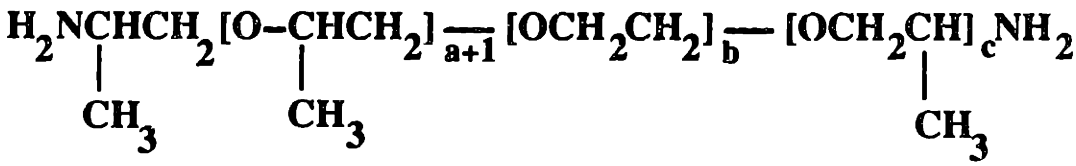


FIGURE A7.1: Chemical structure of hydrophobic polymers. The hydrophobic block copolymers were provided by Texaco Chemical Company (Houston, TX). These polymers are referred to as the JEFFAMINE polymers by Texaco. The structure and relative ratio of ethylene oxide (EO) and propylene oxide (PO) in each polymer are shown. For comparison, polyethylene glycol (PEG) is also known as polyoxyethylene (PEO).

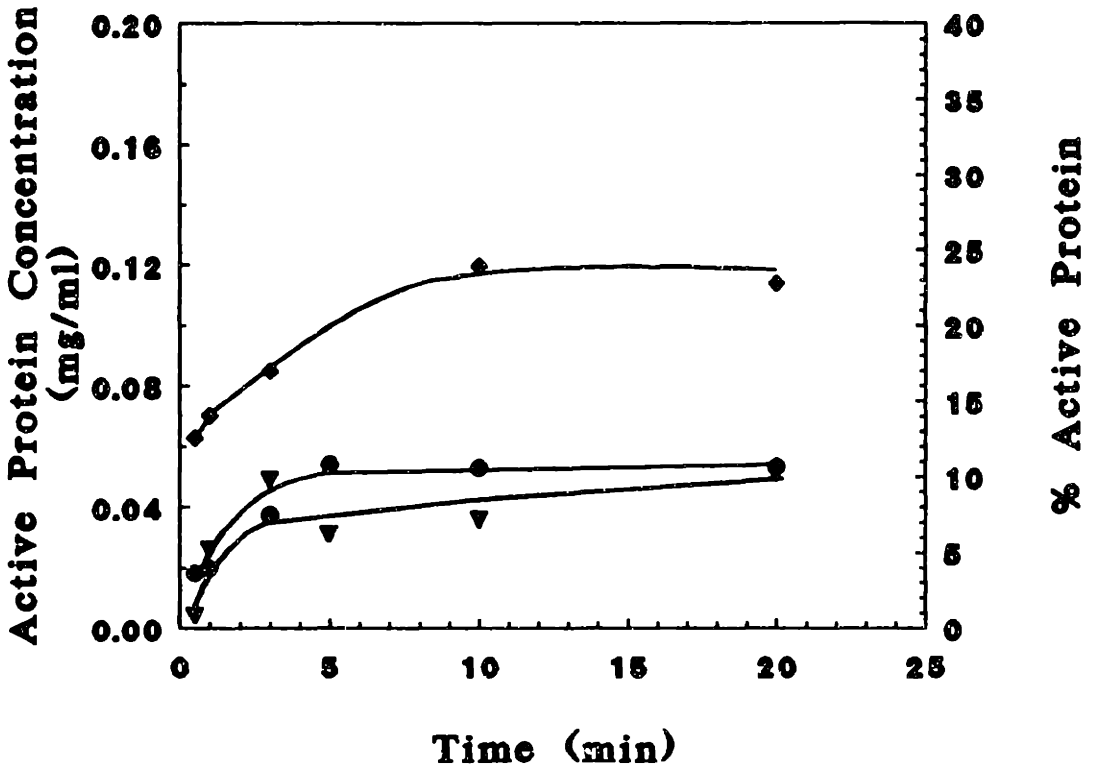


FIGURE A7.2: CAB refolding in the presence of hydrophobic polymers. CAB in 5 M GuHCl was refolded by rapid dilution to 0.50 mg/ml (16.7 μm) protein and 1.0 M GuHCl with a final molar ratio of polymer to CAB ($[Polymer]_f/[CAB]_f$) of 3 for three different polymers: M-1000 (♦), M-2000 (●) and ED-6000 (▾) (see Figure A7.1).