INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

DISSERTATION

ENERGETIC AND KINETIC FOLDING MECHANISMS OF THE PARALLEL β-HELIX PROTEIN PECTATE LYASE C

SUBMITTED BY

DOUGLAS KAMEN

DEPARTMENT OF BIOCHEMISTRY

AND MOLECULAR BIOLOGY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTORATE OF PHILOSOPHY

COLORADO STATE UNIVERSITY

FORT COLLINS, COLORADO

SUMMER 2001

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

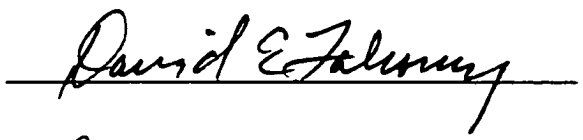
JUNE 25, 2001

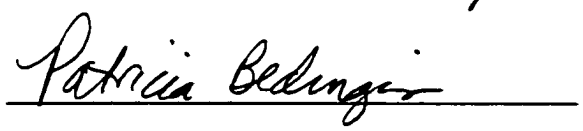
WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY DOUGLAS KAMEN ENTITLED:

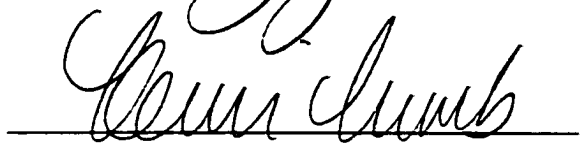
ENERGETIC AND KINETIC FOLDING MECHANISMS OF THE PARALLEL β-HELIX PROTEIN PECTATE LYASE C

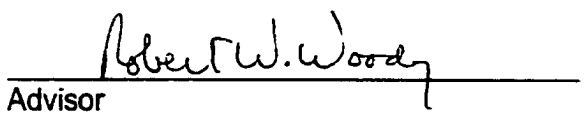
BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTORATE OF PHILOSOPHY

Committee on Graduate Work:

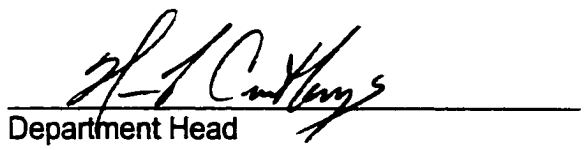








Advisor



Department Head

ABSTRACT OF DISSERTATION

ENERGETIC AND KINETIC FOLDING MECHANISMS OF THE PARALLEL β-HELIX PROTEIN PECTATE LYASE C

Pectate lyase C (pelC) was the first protein in which the parallel β-helix structure was recognized. The unique features of parallel β-helix-containing proteins – a simple topology and unusual interactions among side-chains – make pelC an interesting protein to study with respect to protein folding. In chapter 2 I discuss studies on the unfolding equilibrium of pelC. Gdn-HCl titrations monitored by far- and near-UV CD, and flourescence spectroscopies were used to monitor the isothermal denaturation of pelC. Differential scanning calorimetry and temperature dependent CD were used to observe unfolding as a function of temperature. A detailed thermodynamic analysis of the denaturation was completed.

In Chapter 3 the effects of the binding of anions to acid-unfolded pelC is presented. Addition of ANS to acid-denatured pelC leads to a large increase in the fluorescence quantum yield. This results from the induction of a partially folded protein species upon binding ANS. Thus, ANS does not act as a probe to detect partially folded species, but induces them. The mechanism of ANS binding and structure induction was studied.

A study of the folding and unfolding kinetics is presented in Chapters 4 and 5. Three kinetic folding phases are observed with far-UV CD and four are observed with near-UV CD. The two slowest phases are the result of *cis-trans*

isomerization of prolyl-peptide bonds. The fastest folding phase involves the formation of most, if not all, of the secondary structure and some tertiary structure. There is one kinetic phase observed with near-UV CD that has no corresponding far-UV CD phase. Mutagenesis results indicate that the slowest phase results from isomerization of the Leu219-Pro220 peptide bond and the intermediate slow phase is a result of the simultaneous isomerizations of multiple prolyl-peptide bonds.

Quantum-chemical calculations on the CD spectrum of pelC are described in Chapter 6. Calculations on pelC with each individual tyrosine or tryptophan mutated to alanine were carried out and the difference spectra between the wild type and mutant proteins were obtained. A more complete structural model of the folding of pelC was deduced. It is proposed that the β-helix domain is formed in the fast kinetic phase.

Douglas Kamen
Department of Biochemistry
And Molecular Biology
Colorado State University
Fort Collins, CO 80523
Summer 2001

ACKNOWLEDGEMENTS

There are so many people who have contributed in helping me in some way or another during these last five years. Thanking them all would require a separate chapter of its own. But some people stand out. I first would like to graciously acknowledge my advisory committee: Drs. Pat Bedinger, David Fahrney, Paul Laybourn and Kevin Lumb for periodically making sure that I stay humble. Everyone provided excellent comments, suggestions and general guidance.

I would like to thank all of my teachers for introducing me to this wonderful field of Biochemistry, which I am so glad to be a part of, and sharing their knowledge with me. I would like to thank Dr. Norm Curthoys for doing a great job of running the department and for being a really nice guy. Thank you to all of the people on the staff of the Biochemistry Department for doing all of the things that need to be done but rarely receive the credit that you deserve. Special thanks go to Nancy Fockler for running the whole show, Jeannine McCormack for providing a warm welcome in the early years, and Janet Shackelford for helping a "poor student" go to scientific meetings and general discussions about our "furry children". I would also like to thank all of the past members of the Woody lab for good times, good science and good lunches.

Special thanks also go to the Colorado Institute for Research in Biotechnology for a year of financial support and the National Institutes of Health for generous funding of the entire research group (USPHS GM-22994). I also

would like to thank Dr. Yuri Griko for his crucial role in the calorimetry experiments and Dr. Tom Laue for helpful discussions and suggestions regarding sedimentation experiments.

I would like to thank my mom and dad for raising me and guiding me towards who I am today, and for much needed help in some rough times. Extra special thanks go to Grandma Bea (may she rest in peace) for her inspiration, guidance and wisdom. Thank you to my entire family, including the Quinns, for their support, and their daughter. Thank you to my wonderful friends Josh Adkins, Ewa Bienkiewicz, and Conrad Schaefer and their families for the great times we shared together and for all of the support we gave to each other. It would have been almost impossible to get through this all without you.

I always save the best for last. I can not come up with the words to express my gratitude to Robert and A-Young Woody. A-Young has been a great mentor and lab partner. She has taught me so much in these last years. Robert has been not only the best science advisor possible but has been a mentor in all aspects of life. Finally I wish to thank the most wonderful family anyone can have. Calvin and Yukon have showed that you do not have to be human to love and be loved. They have also helped me keep my sanity over the past four years. To my beloved wife Kerry I thank you with all of my heart. It was you who kept me going through all of the good times and bad times. Without you I never would have made it.

Dedicated to the memory of Beatrice Kamen

Table of Contents

List of Figures

Chapter 5

Chapter 6

List of Tables

Abbreviations

AFP, anti-freeze protein
ANS, 8-anilino-1-naphthalenesulfonate
ASA, accessible surface area
CD, circular dichroism
CM, carboxy-methyl
$D_{1/2}$, denaturant concentration at the transition midpoint
DEAE, diethyl amino ethane
DSC, differential scanning calorimetry
DSS, 2,2-dimethyl-2-silapentane-5-sulfonate
gdn-HCl, guanidine-HCl
H/D, hydrogen/deuterium
HPLC, high pressure liquid chromatography
INDO/S, intermediate neglect of differential overlap/spectrosopic
IPTG, isopropyl-β-D thiogalactopyranoside
IR, infrared
LB, Luria-Bertani
MES, 2-[N-morpholino]ethanesulfonic acid
MOPS, N-morpholino propanesulfonic acid
M_r, relative molecular mass
MW, molecular weight
nm, nanometer
NMR, nuclear magnetic resonance
pelC, pectate lyase C
PPI, peptidyl-prolyl isomerase
PP_{II}, poly(proline) II
SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis
TRCD, time-resolved circular dichroism
TROSY, transverse relaxation time optimized spectroscopy
TSP, tailspike protein
UV, ultraviolet
WT, wild type

Standard abbreviations are used for the amino acids and nucleic acid bases

Chapter 1

A Perspective of Protein Folding

1.1 The Protein Folding Problem

The majority of cellular processes are controlled in some way by the action of proteins. In addition to catalytic activity as enzymes, proteins sequester small molecules and ions, provide channels through membranes, act as receptors, and determine structure in various tissues, to name but a few functions. To date there are over 13,000 protein structures in the Protein Data Bank. Clearly it is not as difficult as it once was to determine the structure of a protein to atomic resolution. One of the big holes in the knowledge of protein chemistry is how and why a seemingly random polypeptide chain becomes an active biological entity. This protein-folding problem is one that has intrigued researchers for over four decades.

The fact that enzymes can be converted from the active, native state to a physically different and inactive state has been known for almost 100 years (Anson 1945). In 1907 Hefter showed that reactive sulfhydryl groups are present in denatured, but not native egg albumin (Heffter 1907). A modern interpretation of this is that the cysteine side-chains are exposed to the solvent upon unfolding. This was perhaps the first demonstration of the different nature of native and denatured proteins (although at the time it was not even accepted that enzymes could in fact be proteins). The first paper to demonstrate a true grasp of the fundamental relation between native and denatured states of proteins was by Wu (1931). Wu discusses in his theory many principles that are considered common knowledge by today's standards. These include structural ensembles and

statistical averages of conformations, weak interactions stabilizing proteins, non-covalent changes upon denaturation, astronomical numbers of conformations for an extended polypeptide chain, and the use of the term "folding".

Complementary to his theory (which unfortunately is somewhat obscure and rarely cited) were many groundbreaking studies in the 1930's and 1940's that showed that denaturation is, in principle, reversible (Anson 1945). In the 1950's and 1960's, Anfinsen's classic experiments firmly established the reversibility of unfolding (Anfinsen 1973). More importantly, they established the fact that all information for the protein tertiary structure is encoded in the primary sequence (Anfinsen et al. 1961). The resulting "thermodynamic hypothesis" suggested that the native state is that in which, under normal physiological conditions, the Gibbs free energy of the whole system is lowest (Anfinsen 1973). Levinthal (Levinthal 1968; Levinthal 1969) made the classic counter-argument to this. He proposed that a protein in the native state might be a uniquely selected metastable state, in which the configurational energy is at a local minimum but not necessarily at an absolute minimum. He later proposed the argument that a random search for this thermodynamic minimum is impossible due to the astronomical number of possible conformations that would need to be sampled. Therefore, because protein folding is fast, it must be under kinetic control. A likely explanation is that proteins fold through specific pathways. The fact that proteins fold in a short amount of time to an apparent free energy minimum despite the enormous time required for a full conformational search has come to be known as Levinthal's paradox.

Those of us working in the protein-folding field are motivated by the preceding arguments. We seek to understand how and why a seemingly random sequence of amino acids linked together by peptide bonds adopts a unique, highly evolved and biologically active conformation. To understand this we must know what inter-residue contacts are made, and when in the course of the folding process they are made. We want to know the structural and energetic properties of the unfolded state and any intermediate states relative to the native state. Finally we want to understand what the amino-acid sequence code is. In other words, what is it about one sequence that dictates a specific structure whereas another sequence gives a totally different structure, and how can we predict the structure of a protein from the sequence alone?

Why is the protein folding problem so important? For one thing, it has been a problem that has intrigued researchers for decades. A solution to the problem will have profound effects on many areas of the life sciences. For example, there will be a more complete view of the central dogma of biology. Presently, there is at least a mechanistic understanding of DNA replication, RNA transcription from DNA, as well as protein translation. But how the translated polypeptide is transformed into the active, three-dimensionally structured protein is still a mystery. If the ultimate goal of accurately predicting structure from sequence is met, then the field of protein design will be significantly advanced. It would then be possible to produce enzymes to catalyze virtually any reaction. This would be applicable to therapeutic and industrial proteins. Enzymes are much more efficient and environmentally friendly than any industrial catalyst.

Finally, several diseases including Alzheimer's disease (Kelly 1998) and spongiform encephalopathies (Prusiner 1997) are the result of misfolded proteins. If we can understand how proteins fold, then it might be possible to understand how they misfold, and develop treatment or prevention strategies.

In this thesis I present studies on the folding of Pectate Lyase C. This protein was the first parallel β-helix protein to be observed. The work presented here represents the first biophysical study of this class of proteins. This initial step has provided some interesting observations. Future experiments based on these studies will hopefully contribute to a better understanding of protein folding in general.

1.2 Theory of Protein Folding

In response to Anfinsen's thermodynamic hypothesis (Anfinsen 1973), Levinthal (1968) proposed the question "Are there pathways for protein folding?" His argument can be formulated as follows. If a polypeptide chain of 100 amino acids was restricted to only two possible configurations for each residue, there would be about $1x10^{30}$ possible conformations. If we assume that each conformation can be sampled at the rate of 10^{13} sec^{-1} (a typical vibrational frequency), it would take over three billion years to sample all possible conformations in order to find the thermodynamic minimum in free energy. Clearly this is not the case, and more than a simple random search is involved in protein folding.

One solution that Levinthal proposed was that proteins fold through specific pathways in much the same way that the reactions of small molecules occur, or the metabolic pathways proceed. This mechanism would resolve the Levinthal paradox by restricting the conformational space that a protein explored. In support of this view, several examples of folding intermediates were observed in kinetic and equilibrium experiments over the next decade. (Creighton 1978; Kuwajima 1977; Kuwajima et al. 1976).

In the last ten years, the pathway idea has broken down. In 1991 Jackson and Fersht observed that the small protein chymotrypsin inhibitor 2 (CI2) folded without any detectable intermediates in a true two-state process (Jackson and Fersht 1991). Subsequently, several small proteins have been found to fold via the same two-state mechanism (Grantcharova et al. 2001). This means that no intermediates are needed to guide the folding process for small, single-domain proteins. These proteins spontaneously find the native state without any external influence. These observations have sparked a "new view" of protein folding based on statistical ensembles. This view stresses the importance of the reorganization of a very large number of weak interactions. This is in contrast to the reactions of small molecules in which one, or a very few, strong interactions are made or broken.

The new view of folding, the funnel model (Figure 1.1) or energy landscape theory, can be summarized as follows, as described in many reviews (Bryngelson et al. 1995; Bryngelson and Wolynes 1987; Chan and Dill 1998; Dill 1999; Dill and Chan 1997; Dinner et al. 2000; Dobson and Karplus 1999; Dobson

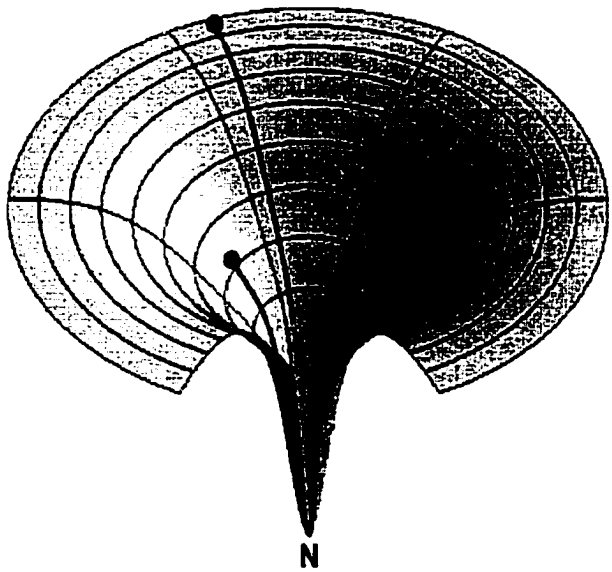


Figure 1a: Funnel representation of a hypothetical protein that folds through a two-state mechanism.

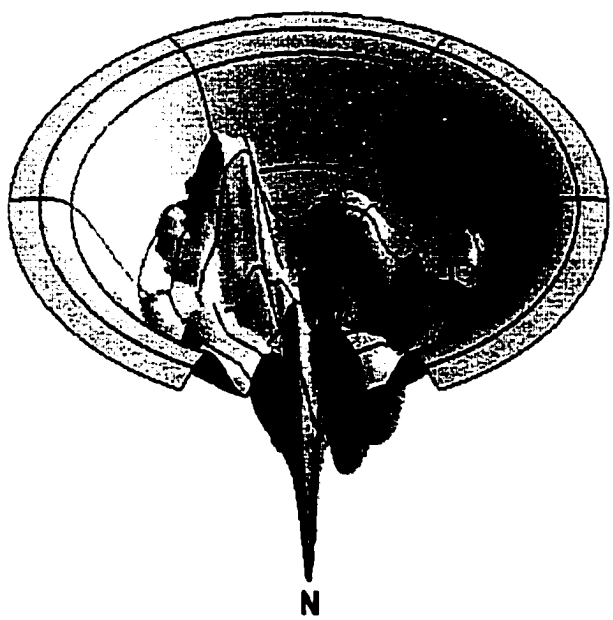


Figure 1b: Funnel representation of a hypothetical protein in which intermediates are observed in the folding process. Figures are taken from Dill & Chan (1997).

et al. 1998; Fersht 1995b; Fersht 1997; Frauenfelder et al. 1991; Harrison and Durbin 1985; Laurents and Baldwin 1998; Radford 2000; Scheraga 1996; Shakhnovich 1997; Straub and Thirumalai 1993; Tsai et al. 1999; Wolynes 1996). Protein folding is much more complex than a simple single pathway. Rather, the process involves simultaneous populations of multiple pathways. The term unfolded "state" of a protein can be misleading. This implies a specific unfolded conformation. It is more accurate to describe the unfolded state as a heterogeneous ensemble in which local interactions are dominant. Many different conformations of the protein exist in equilibrium with similar free energies. In fact, it is possible that the number of conformers in this unfolded population approaches the number of conformations proposed by Levinthal. When a folding reaction is initiated, the individual protein molecules fold from whatever specific conformation they are in at that instant. This generates the concept of multiple, parallel pathways. A popular analogy is that the unfolded protein molecules represent many skiers on the top of a mountain (Dill and Chan 1997). If all of the skiers start down the mountain at the same time, they will each go down a separate path, but all will eventually end up at the bottom. It is possible that some will go down without being hindered, but some might encounter regions of the landscape that go back uphill. This is analogous to local free energy minima that might be interpreted as folding intermediates or off-pathway traps.

Figure 1.2 illustrates a hypothetical free energy surface for folding (Dobson and Karplus 1999). The surface resembles a funnel. At the top of the

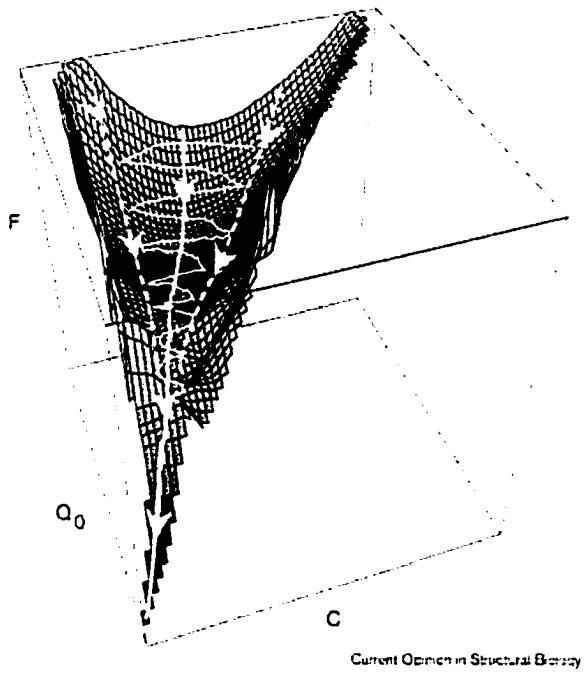


Figure 1.2: Folding free energy landscape generated for a model protein. The free energy (F) is plotted against the fraction of native contacts (Q_0) and the total number of contacts (C). The arrows and lines show the trajectories for various simulations. Taken from Dobson and Karplus (1999).

funnel is the unfolded ensemble. It is broad because there are many conformations of similar free energy. The difference in enthalpy between the folded and unfolded state is on the order of 30 – 100 kcal $mole^{-1}$ (Dobson et al. 1998). This large difference in enthalpy leads to a conformational bias towards the native state that is sufficient to avoid the need to search through an impossibly large number of conformations. As folding progresses the energy of the system decreases as more native contacts are formed. Native contacts are generally more stable than non-native ones. Thus, native contacts tend to be stable. The number of accessible conformations is further decreased as native contacts are formed and the energy difference between these and molecules without native contacts increases. Thus as the native state is approached, the entropy decreases, but the overall free energy of the system is decreased. This mechanism overcomes the Levinthal Paradox. Although the accessible conformational space is very large, an individual molecule only samples a small region of this space. Therefore, the bias in the free energy surface makes the numbers of accessible conformations much less than the full sample available by the Levinthal calculation.

One problem with this model is that it really only applies to small proteins that fold rapidly and without intermediates. There are many examples of small proteins that do populate intermediates (Barrick and Baldwin 1993; Chen et al. 1998; Cota and Clarke 2000; Houry et al. 1996; Killick et al. 1999; Kuwajima et al. 1991; Kuwajima et al. 1985; Matagne et al. 1998; Raschke and Marqusee 1997; Udgaonkar and Baldwin 1988), as well as many more examples of larger

proteins that almost exclusively form intermediates (Ayed and Duckworth 1999; Bilsel et al. 1999; Jaenicke 1987). A fundamental question is still unanswered. That is, are folding intermediates required, and if so, what are their roles? The fact that many small proteins can fold without them has led some to speculate that intermediates are experimental artifacts (Dill and Chan 1997). They are not necessary in guiding folding, rather they are nonproductive dead ends. In this view, the intermediates are compact structures with non-native contacts, and the molecules must "go uphill" in energy to break these contacts. They represent kinetic traps that, for example, involve misfolded species that hinder the folding process and slow down formation of the native state. Such intermediates appear in some lattice simulations (Dinner et al. 1996). This results in multiphasic kinetics. There are experimental results that corroborate this. Lysozyme and cytochrome c have both been shown to fold through parallel pathways with and without intermediates (Kiefhaber 1995; Sosnick et al. 1994). The alternative view is that intermediates are productive and help to direct the folding process towards the native state by acting as templates for the formation of native-like structure and restricting the conformations that need to be sampled. Interleukin 1-β has been shown to fold through a necessary intermediate (Heidary et al. 1997). It is most likely that both types of intermediates occur. It is very possible that domains (Panchenko et al. 1996) are an important feature in the folding of larger proteins, in contrast to small proteins in which the only cooperative unit may be the entire polypeptide chain (Fersht 1993). The concept of "foldons" is an

interesting one (Panchenko et al. 1996). These are discrete basic folding units that are simple and make up larger, more complex proteins.

It appears that the recent theoretical developments in protein folding have led to a greater conceptual understanding of the physical process of folding. However, this framework gives little insight into the relationship between the amino acid sequence and the final protein structure. Only through experiments will this "code" finally be broken.

1.3 Technological Advances in the Study of Protein Folding

Understanding of the physical process of protein folding has clearly advanced over the last decade. New ideas and theories have emerged as a result of experimental observations. These ideas and observations were possible because of significant developments in techniques to study protein folding. The most substantial developments have been made in the areas of kinetics and high resolution NMR, and I will limit this brief discussion to these areas.

The most popular technique for studying folding reactions is the stopped-flow technique (Pflumm et al. 1986). This is a way of mixing solutions in the millisecond time range. Usually a solution of unfolded protein in high denaturant concentration or low pH is rapidly diluted to native conditions. Spectroscopic probes such as fluorescence, circular dichroism or absorption are used to follow folding. This is a good method for looking at formation of secondary and tertiary structure, or ligand binding as a function of structure formation. Unfortunately,

this technique is limited in that many proteins have been observed to have folding events that occur in the sub-millisecond time-scale (Akiyama et al. 2000; Ballew et al. 1996a; Chan et al. 1997). This "burst phase" is often attributed to the formation of intermediates, but the true nature has been controversial. Some results suggest that the burst phase observed is merely an experimental artifact, and it is generally the unfolded protein adjusting to new solvent conditions (Sosnick et al. 1997).

In order to overcome this time constraint many new techniques have been developed that probe folding on the micro- and nanosecond time ranges. Collecting data at micro- to nanosecond resolution is not a big problem. The problem is rapidly initiating the reaction and coupling it to the method of observation. Many "ultra-rapid" experiments have utilized cytochrome c as the model protein. Protein folding has been triggered by electron transfer reactions (Chen et al. 1999; Pascher et al. 1996). Oxidized cytochrome c is unfolded in 3M guanidine-HCl (gdn-HCl) and 40° C, but the reduced protein is completely folded under the same conditions. Rapid photochemical electron injection can transfer an electron from Ru^{2+} or NADH to cytochrome c (Fe^{3+}) in about 20 nsec, thus forming reduced cytochrome c. Folding was followed by visible absorption (Pascher et al. 1996), and later by far-UV CD and visible absorption (Chen et al. 1999).

The folding of cytochrome c has also been initiated using a laser photolysis technique (Chen et al. 1998; Jones et al. 1993). In the presence of CO, cytochrome c unfolds at 4.5M gdn-HCl, but without CO, it is folded. A laser

pulse can rapidly cleave the CO from the heme iron. Once the Fe-CO bond is cleaved, the protein folds. Both fluorescence (Jones et al. 1993) and CD (Chen et al. 1998) spectroscopies have been used to observe this reaction. CD has the advantage over other optical techniques in that it can be used to observe heme structure and ligation, as well as secondary structure formation.

Laser temperature jumps have also been used extensively to initiate fast events in protein folding (Ballew et al. 1996a; Ballew et al. 1996b; Gruebele 1999; Gruebele et al. 1998; Munoz et al. 1997; Reinstadler et al. 1996; Williams et al. 1996). For these studies, small peptides and apo-myoglobin have been used. Time-resolved absorption, fluorescence and infrared spectroscopies have been used for these studies. Since water absorbs light in the infrared, a 1.54 μm IR pulse (the first OH overtone of water) can yield a temperature change of up to 30 °C in about 100 psec (Gruebele et al. 1998). Apo-myoglobin is unfolded at low temperatures (about –5° C). The temperature jump leaves the unfolded protein under thermodynamic conditions where folding is favorable.

Ultra-rapid continuous flow mixing has also been used to probe folding of cytochrome c from the micro- to millisecond time-scale (Akiyama et al. 2000; Chan et al. 1997; Takahashi et al. 1997). A protein is unfolded in denaturant. Turbulent flow created by pumping the unfolded protein solution and refolding buffer through a small gap dilutes the denaturant in microseconds. The refolded protein solution travels through the length of the cuvette. Each point along the observation cuvette corresponds to a time point. A laser is focused at each point and a spectrum is measured corresponding to the protein at that time after

folding starts. This technique has the advantage that various temperatures and conditions can be probed. This method's strongest point is that in principle any protein can be studied by this technique. In the previous techniques, samples are limited to those that can be optically triggered or cold denatured in liquid buffer.

Structural details at high resolution are necessary to characterize intermediates in folding. NMR has the potential to uncover many of these details. Hydrogen/deuterium exchange can be used to observe residue-specific information on hydrogen bonding on the millisecond time scale and protection from exchange of side-chains (Clarke and Itzhaki 1998). This translates into the formation of secondary structure and overall side-chain packing. This method can be used to determine what residues become involved in secondary structure at specific times (Raschke and Marqusee 1997). H/D exchange can also be used in conjunction with mass spectrometry (Heidary et al. 1997). This information can detect protection from exchange within populations, specifically detecting populations of intermediates.

Native-state hydrogen exchange can measure the global stability of a protein and detect metastable states (Bai et al. 1995; Mayne and Englander 2000). Under native conditions a small fraction of protein molecules in a population will occupy higher energy conformations such as partially folded states or even the fully unfolded state. These populations are described by the Boltzmann distribution. Hydrogen-exchange experiments coupled with NMR can determine the hydrogens exposed in each high-energy form, the rates of

exchange with solvent, and their sensitivity to specific perturbants. From this, the structure, free energy and surface exposure of each form can be inferred. This type of experiment has been used to deduce the intermediates along the major refolding pathway of cytochrome c (Bai et al. 1995).

One of the most powerful techniques in the study of protein folding is the protein engineering technique (Fersht 1995a; Martinez and Serrano 1999; Oliveberg 2001; Otzen et al. 1994; Riddle et al. 1999; Ternstrom et al. 1999). This method utilizes systematic mutations of every residue in the protein to probe the role of a specific contact in the stabilization of intermediates and the transition state. Suppose that a mutation destabilizes the native protein by 1 kcal / mole. Also suppose that in a kinetic experiment it is determined that the same mutation destabilizes the transition state by the same amount. It can be inferred that the interaction is the same in the transition state as in the native structure. If the transition state energy is not affected by the mutation, then it can be inferred that the interaction is not formed yet, and the interaction is the same in the unfolded protein as in the transition state. This can also be extended to probe the role of specific amino acids in intermediate conformations. This kind of information is important in determining what specific contacts occur in the rate-limiting and earliest stages of the reaction, when the proper kinetic technique is used. It indicates what residue-specific contacts are driving the initial condensation of the structure. This method can potentially reveal residue-specific rules for folding.

The latest methods for studying folding appear to be approaching the inherent speed limit for folding. These techniques have led to a profound and

comprehensive theory on folding. In the future, improvements to these experiments will combine high time resolution with high structural resolution.

1.4 Experimental Observations in Protein Folding

The new theories and ideas about protein folding have largely been developed from experiments over the last ten years. The observation of proteins that fold in a true two-state manner was perhaps one of the most important experimental advances in the entire field.

Proteins that fold through a kinetic two-state mechanism are very important to understanding folding. To date about 26 proteins have been shown to fall into this family or at least show two-state behavior under some conditions (Capaldi and Radford 1998; Grantcharova et al. 2001; Jackson and Fersht 1991; Kuhlman et al. 1998; Schonbrunner et al. 1997). Most of these proteins are small (<100 amino acids) and therefore represent the minimal protein structural unit. It is reasonable to assume that the rules of folding will be most easily understood with these proteins. Because these proteins are small, their folding processes can be compared to computational results and simulations. The size and simplicity makes these proteins applicable to study by NMR and protein engineering techniques. Thus, very detailed structural, energetic and kinetic mechanisms can be deduced.

Realistically, most proteins form some intermediates during the folding process. It is possible that characterizing the transitions to and from these

intermediates will lead to an understanding of folding. In recent years, the role of intermediates has come into question (Dill and Chan 1997). Whether intermediates facilitate folding or slow folding through kinetic and thermodynamic traps remains unclear (Baldwin 1996; Dill and Chan 1997). Nevertheless, characterization of these species' structures, and thermodynamic and kinetic properties is still an important step towards understanding the mechanism of protein folding. Intermediates can be of various character. Several cases where formation of intermediates results in productive folding have been observed (Heidary et al. 1997; Matagne et al. 1998; Udgaonkar and Baldwin 1988). These are often termed "on pathway" intermediates. Subsequently, many cases are known of intermediates that must unfold prior to the formation of the native state (Akiyama et al. 2000; Bilsel et al. 1999; Chen et al. 1999; Takahashi et al. 1997; Zitzewitz and Matthews 1999). Whether or not intermediates possess structure found in the native state, the same goal is accomplished: that is condensation of the random coil into a compact structure where long-range interactions are localized. This has the final result of reducing the conformational space available for the protein to sample. Some cases have been observed in which the protein actually forms more structure in the intermediate and must unfold to form the native state (Clarke et al. 1999).

As previously noted, the latest techniques permit folding to be studied down to the nanosecond time scale. The two systems to which this has been applied most extensively are cytochrome c and apo-myoglobin. Because folding can be optically triggered, cytochrome c has been used to look at the earliest

events in folding. Using traditional stopped-flow CD, it was shown that 44% of the secondary structure is formed in less than the 4 msec dead time of the instrument (Elöve et al. 1992). Recent time-resolved CD experiments indicate ~25% of native structure is formed in this time frame (Chen et al. 1999). There are also two "unfolding" events corresponding to misfolded species. The discrepancy in the percent of native structure formed is probably a result of the different solution conditions. The initial study used denaturant dilution, but the TRCD study used optical triggers with no change in solvent. Heme ligation was also looked at with the rapid techniques (Chen et al. 1998; Jones et al. 1993). In the earliest stages (2μsec), a native heme-histidine ligation forms concomitantly with ~8% of secondary structure. The next phase involves binding of a non-native histidine to the heme. Subsequent slower phases involve unfolding of the non-native intermediate and further progression to the native protein. Rapid mixing experiments have further characterized the secondary structure formation (Akiyama et al. 2000) and shown that the earliest event involves a nonspecific collapse of the structure followed by formation of a molten globule (Kuwajima 1996). This is consistent with the hydrophobic collapse model (Gutin et al. 1995).

Temperature-jump studies on apo-myoglobin have provided additional insights into the early formation of structure. Initial steps of folding involve formation of solvated helices followed by a step that is three orders of magnitude slower involving tertiary packing of these helices (Gilmanshin et al. 1997). This suggests that local interactions can form folding nuclei, followed by formation of

the hydrophobic core. This conclusion is more in line with the framework model of folding (Kim and Baldwin 1990). Other work has shown that a solvated, compact structure forms early (Ballew et al. 1996a). This is also consistent with stopped-flow CD results that indicate the formation of a burst-phase intermediate in less than 5msec (Jennings and Wright 1993).

Some studies on model peptides have uncovered the fundamental rates of formation of secondary structure in proteins. The time constant of propagation of α-helices has been measured to be 180 nsec (Thompson et al. 1997) whereas that for the propagation of β-hairpins is 6μsec (Munoz et al. 1997). The difference in rates is a result of the height of the free energy barrier separating folded and unfolded states. In the α-helix, the barrier is crossed after the first hydrogen bond is formed upon fixing three residues in the native geometry. Elongation is favorable because the entropy loss of fixing a residue is compensated by the formation of a hydrogen bond. For the hairpin, two peptide bonds have to be fixed in order to form one interstrand hydrogen bond until there is stabilization by side-chain interactions. In addition, α-helices can be initiated from any point within the peptide with equal probability. β-hairpins are less likely to form when the turn region is not involved (Munoz et al. 1997). The rate of helix initiation has also been measured (Clarke et al. 1999). This was done from completely random coil peptides. It turns out that the rate-limiting step is formation of the first loop (50 msec). Helix propagation from this nucleus is then very fast (180 nsec) (Munoz et al. 1997). Because many proteins fold faster that 50 msec, it is possible that transient structure in unfolded proteins can serve as

nucleation sites for folding. There is much controversy over this result, however. Statistical mechanical models suggest that the helix formation rate agrees with the rates measured in temperature jump experiments described above (Eaton et al. 2000).

Recent data indicate that the fundamental physics underlying the protein folding process might be simpler than current thinking would lead one to believe (Baker 2000; Plaxco et al. 1998). Protein folding rates and mechanisms may be largely determined by native state topology rather than interatomic interactions. Mutations that do not alter the overall protein fold usually have less than a ten-fold effect on the folding rate. Sequence differences in homologous proteins usually do not have a large effect on the transition state structures. Finally, the folding rates of many proteins correlate with the contact order. This is a measurement of the average sequence separation between residues that make contact in the three-dimensional structure. These ideas suggest that topology may be the main factor in determining folding mechanisms.

Although a lot of attention has been focused on fast events in folding, most proteins exhibit slow phases that occur from tens of seconds even up to hours. Most often the slow folding phases are the result of the slow isomerization of x-proline peptide bonds (x = any amino acid) (Brandts et al. 1975; Cook et al. 1979; Schmid and Baldwin 1978). Peptide bonds formed by non-proline residues are thermodynamically preferred in the planar *trans* form. The planarity is due to the partial double bond character of the peptide bond, and the *trans* isomer preference is due to more favorable interactions and larger

conformational entropy relative to the *cis* isomer. X-pro peptide bonds can exist as a mixture of *cis* and *trans* isomers because of steric constraints associated with linkage of the proline side-chain with the amide nitrogen. Thus, in an unordered polypeptide chain there can be a *cis:trans* ratio of about 30:70 (Brandts et al. 1975). Isomerization occurs with an enthalpy of activation of about 19 kcal/mol (Brandts et al. 1975), and therefore can be very slow at room temperature . For many years there has been debate over whether the isomerization is an important event in folding or just an experimental artifact. The discovery of peptidyl-prolyl *cis-trans* isomerase (PPI) might lead one to think that the former is correct (Harding et al. 1986; Lang et al. 1987). Also, some enzyme function is regulated by *cis-trans* isomerization of prolyl-peptide bonds (Schiene and Fischer 2000). Proline 191 of mannose-binding protein is *trans* in the absence of Ca^{++}, and the *cis* form binds Ca^{++} and is active (Ng et al. 1998; Ng and Weis 1998). Therefore, the binding of Ca^{++}, and hence biological activity, is regulated by isomerization of prolyl-peptide bonds (Ng et al. 1998; Ng and Weis 1998). Characterizing slow steps in folding and identifying essential prolines is important in understanding the mechanism of protein folding.

1.5 Pectate Lyase C and the Parallel β-Helix Proteins

The model system that was chosen for these protein folding studies was pectate lyase C (pelC) from *Erwinia chrysanthemi* (Yoder et al. 1993). PelC was chosen because it possesses a rare, and at one time unique, backbone topology.

PelC was the first protein to be observed to be in a parallel β-helix conformation (Jurnak et al. 1994; Kobe and Kajava 2000; Yoder and Jurnak 1995). The structure is interesting because it forms three parallel β-sheets in a unique way. As shown in figure 1.3, a single strand from one sheet is connected to another strand of the next sheet by a loop. This strand is then connected to another strand from the third sheet by another loop. The strands are labeled PB1, PB2 and PB3, and the connecting turns are labeled T1, T2 and T3. T1 connects PB1 with PB2 and so forth. This motif occurs six to ten times in this protein. The result is a large coil, or helix, composed of parallel β-sheets, hence the name parallel β-helix. This protein family was chosen to study folding primarily because of this relatively simple topology. It seemed logical to assume that a very simple fold would yield a relatively simple folding mechanism. PelC was chosen because at the time that this project was initiated only ten β-helix proteins were known to exist, and four of these were trimers. PelC appeared to be one of the smaller and simpler of the remaining monomers. The fact that it was the first to be observed gave a nice historical significance to the project.

PelC is a virulence factor involved in soft rot disease of plants (Tamaki et al. 1988). It catalyzes the cleavage of the pectate component of the cell wall (Barras et al. 1994; Rao et al. 1996; Rexova-Benkova and Markovic 1976; Sutherland 1995). The actual reaction consists of the cleavage of the α-1,4-glycosidic linkages in pectate. Cleavage proceeds via a β-elimination mechanism, resulting primarily in the formation of trisaccharides. Other pectic and catabolic enzymes further break down the trimers to monomers for use

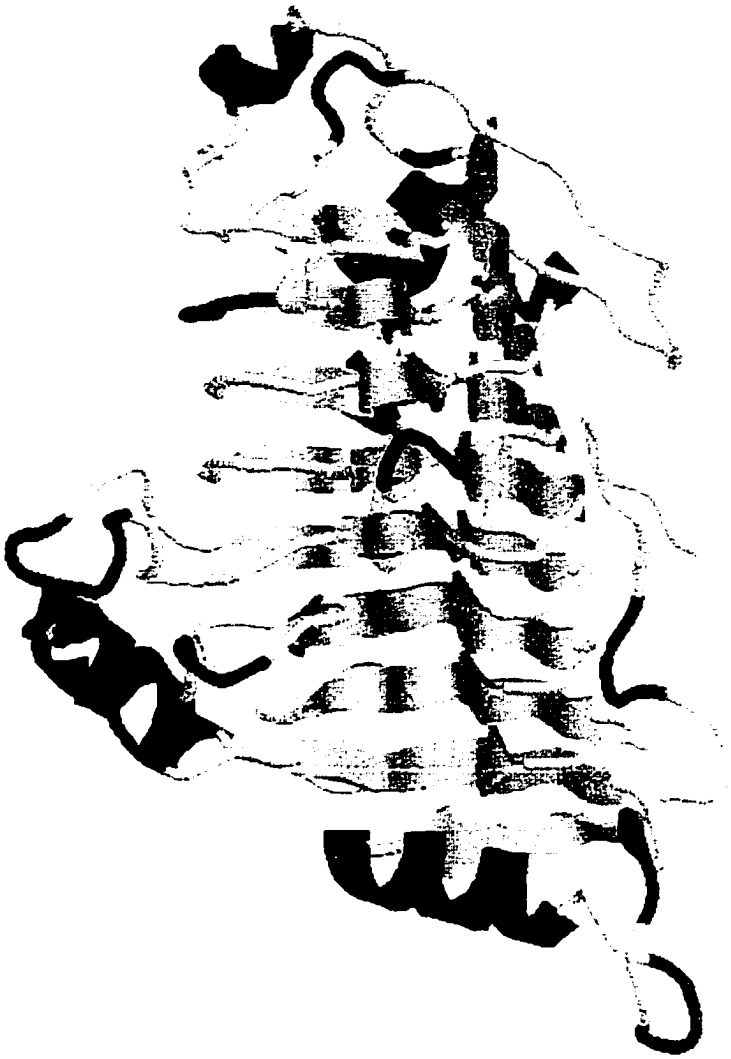


Figure 1.3: A model of the structure of pelC from x-ray crystallography data.

as nutrients (Barras et al. 1994). PelC is synthesized with a 22-amino acid residue signal peptide that labels the protein for secretion. The "Out" secretion pathway in *E. crysanthemi* functions in *E. coli*, resulting in efficient secretion to the periplasmic space in *E. coli*. Because of this, and the fact that pelC is cloned into *E. coli*, purification is simplified. Calcium is required for activity of pelC, and the calcium is thought to act as a general acid, with an active site lysine acting as a general base (Rao et al. 1996). PelC, like most polysaccharide lyases, is optimally active at alkaline pH (for pelC optimal activity is at pH 8.8-9.5).

The structure of pelC has been determined by x-ray crystallography and refined to a resolution of 2.2Å (Yoder et al. 1993) (Figure 1.3). As mentioned, an important feature of this protein is the uniform and relatively simple topology of the β-helix domain. The β-helix structure differs significantly from the α-helix secondary structure. The typical α-helix is characterized as right-handed with a rise per residue of 1.5Å, a period of 3.6 residues per turn and backbone dihedral angles (ϕ,ψ) within specified limits (ϕ= –57, ψ= –47). We also observe a significant dipole associated with the alignment of the peptide bonds. The β-helix of pelC is characterized as right-handed with a rise per residue of 0.22Å and an average period of 22 residues per turn. Some parallel β-helix proteins have a left-handed sense. No dipole is observed in the β-helix. The ϕ,ψ angles for the parallel β-strands of pelC range from -104° to -140° and 109° to 150°, respectively. Another structural feature of pelC that makes it an interesting folding model is the side-chain interactions (Figure 1.4). We observe several aliphatic stacks comprised mainly of valine and isoleucine, and to a lesser

degree, leucine and alanine. Inside the T2 turn there is a stack of six asparagine residues. In the T1 turn there is a stack of three serine residues. We also observe several stacks of aromatic residues inside and outside the β-helix (Figure 1.4). It is very likely that these side-chain stacks play a significant role in the folding of pelC, particularly in the structural specificity and integrity. Especially noteworthy are the polar stacks found within the hydrophobic core of pelC. Two stacks, one of six asparagines and one of three serines are observed. The extensive hydrogen bond networks of these stacks must be satisfied upon formation of the hydrophobic core. Formation of the hydrophobic cores of proteins is a very rapid event, thus formation of these interaction will be a very rapid and cooperative process.

Since the first parallel β-helix structure was observed in 1993 with pelC, twenty more protein structures have been solved that contain a parallel β-helix fold or domain. Table 1.1 (references therein) shows a list of all the known β-helix proteins to date. It is interesting to note the broad range of species, kingdoms and functions of these proteins. Parallel β-helix proteins are found in organisms ranging from bacteria and archea all the way to humans and other eukaryotes. These proteins also have a wide range of functions. It appears that the majority of these proteins are involved in catalyzing the breakdown of polysaccharides as either lyases or hydrolases. One of the more advanced and intricate functions is in signaling or cell-cell recognition, such as the cases of P.69 pertactin and the insulin-like growth factor receptor. Some of the recent structures to be published were of two antifreeze proteins from insects.

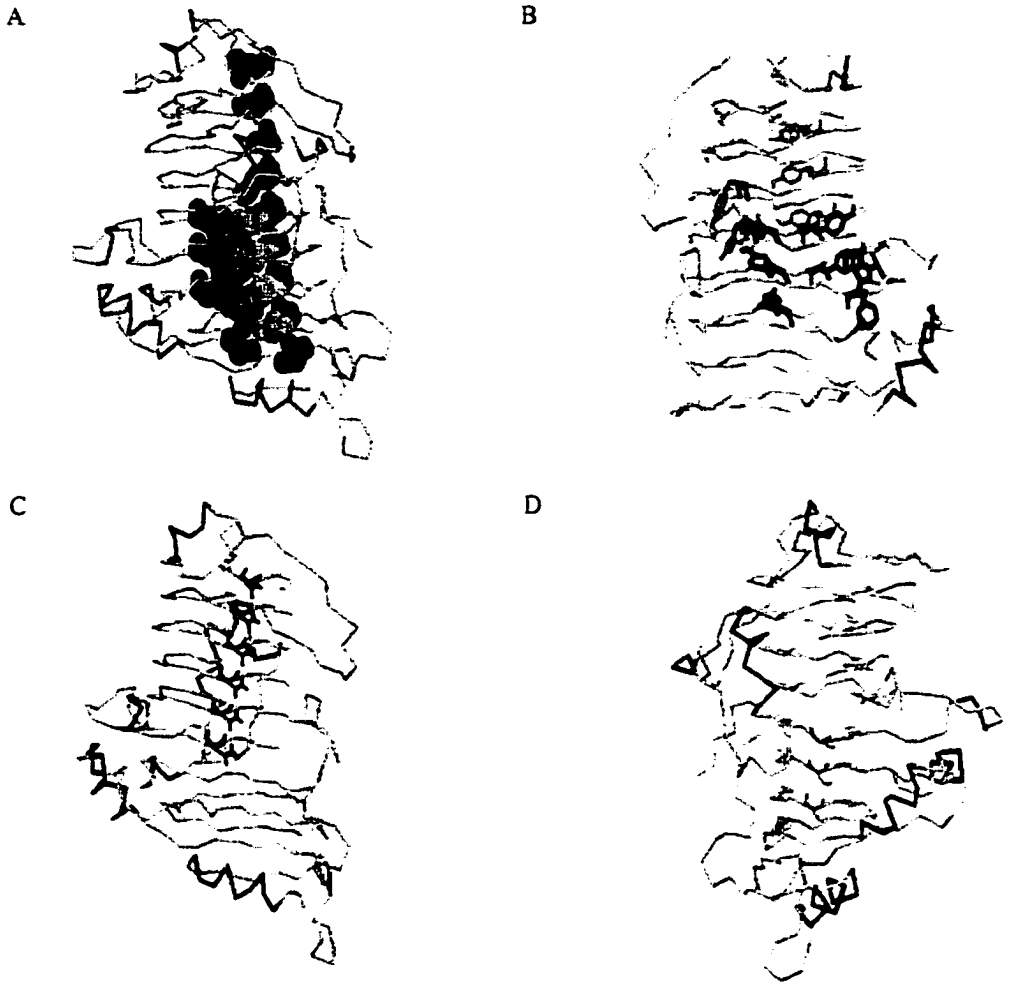


Figure 1.4: Side-chain stacks in pelC. *a.* Stacks of branched-chain amino acids make up the hydrophobic core of pelC. *b.* Aromatic stacks are found on the inside and outside of the pelC core. *c.* The asparagine ladder is located inside the core of pelC, as is the serine stack (*d*). Both stacks are colored purple.

Table 1.1: All of the parallel β-helix proteins that have been observed to date arranged in order of first observation. Species highlighted in bold are eukaryotes. Asterisks indicate that these proteins are trimers.

	Protein	Organism	Function	Helix sense	Reference
1	Pectate Lyase C	*Erwinia chrysanthemi*	Polysaccharide lyase	Right	(Yoder et al. 1993)
2	P22 tailspike	*Bacteriophage P22*	Cell adhesion	Right*	(Steinbacher et al. 1994)
3	Pectate lyase E	*Erwinia Chrysamthemi*	Polysaccharide lyase	Right	(Lietzke et al. 1994)
4	Pectate lyase	*Bacillus subtilis*	Polysaccharide lyase	Right	(Pickersgill et al. 1994)
5	UDP-*N*-Acetylglucosamine Acyltransferase	*Escherichia coli*	Acyl transferase	Left*	(Raetz and Roderick 1995)
6	P.69 pertactin	*Bordetella pertussis*	Cell Adhesion	Right	(Emsley et al. 1996)
7	Carbonic anhydrase	*Methanosarcina thermophila*	Proton transfer	Left*	(Kisker et al. 1996)
8	Tetrahydrodipicolinate *N*-Succinyltransferase	*Mycobacterium bovis*	Succinyltrasferase	Left*	(Beaman et al. 1997)
9	Rhamnogalacturonase A	*Aspergillus aculeatus*	Polysaccharide hydrolase	Right	(Petersen et al. 1997)
10	Pectin lyase A	*Aspergillus niger*	Polysaccharide lyase	Right	(Mayans et al. 1997)
11	Xenobiotic acetyltransferase	*Pseudomonas aeruginosa*	Acetyltransferase	Left*	(Beaman et al. 1998)
12	Type-1 insulin-like growth factor receptor	***Homo sapiens***	Receptor	Right	(Garrett et al. 1998)
13	Polygalacturonase	*Erwinia carotovora*	Glycosyl hydrolase	Right	(Pickersgill et al. 1998)
14	Pectin lyase B	*Aspergillus niger*	Polysaccharide lyase	Right	(Vitali et al. 1998)
15	*N*-acetyl-glucosamine-1-phosphate uridyltransferase	*Escherichia coli*	Acetyltransferase Uridyltransferase	Left*	(Brown et al. 1999)
16	Endopolygalacturonase II	*Aspergillus niger*	Polysaccharide hydrolase	Right	(van Santen et al. 1999)
17	Chondroitinase B	*Flavobacterium heparinum*	Polysaccharide lyase	Right	(Huang et al. 1999)
18	TmAFP	***Tenebrio molitor***	Antifreeze protein	Right	(Liou et al. 2000)
19	AFP	***Choristoneura fumiferana***	Antifreeze protein	Left	(Graether et al. 2000)
20	Pectin Methylesterase	*Erwinia chrysanthemi*	Ester hydrolase	Right	(Jenkins et al. 2001)
21	*N*-acetyl-glucosamine-1-phosphate uridyltransferase	*Streptococcus pneumoniae*	Acetyltransferase Uridyltransferase	Left*	(Sulzenbacher et al. 2001)

In recent years, parallel β-helix proteins have been the focus of some interest. A paper published in 1997 suggested that amyloid type fibrils formed from peptides that were in a β-helical conformation (Lazo and Downing 1997). A subsequent study suggested that amyloid fibrils might be assembled *in vivo* from β-helical precursors (Lazo and Downing 1998). This conclusion was based on modeling studies.

It has long been suspected that parallel and anti-parallel β-sheets have different infrared absorption spectra. This is due to the fact that the hydrogen bond pattern is different resulting in different vibrational modes. It has also been suspected that the parallel β-helix proteins might have a unique feature in the circular dichroism spectrum. Some spectroscopic studies have been carried out to determine if there are unique circular dichroism or infrared spectra (Khurana and Fink 2000; Sieber et al. 1995). Unfortunately, neither study detected any unique spectral features.

It has been noted that parallel β-helix proteins of similar function have significant sequence homology, but the proteins of divergent functions show no obvious sequence homology (Jenkins et al. 1998). Structural features conserved in parallel β-helix proteins have led Jenkins *et al.* to propose that these proteins evolved from a common ancestor rather than converged towards a common fold (Jenkins et al. 1998). It would be important to reevaluate this conjecture. Several new protein structures have been observed since this paper, and many of them, specifically the insect antifreeze proteins, have unique features or little structural homology. A recent article described a computer program that can

identify parallel β-helix proteins by identifying the sequence of the T2 turn (Mackenzie 2001). This program indicated that several hundred proteins in the SWISS-PROT database might potentially form parallel β-helices.

There have been some folding studies of this class of proteins. With the exception of the biophysical studies on pelC presented here, two groups working with the tailspike protein from bacteriophage P22 (TSP) have carried out all other folding studies of parallel β-helix proteins. Jonathan King's group at Massachusetts Institute of Technology has carried out the majority of these studies. This group has made significant contributions towards understanding the *in vivo* folding mechanism and trimer assembly of TSP (King et al. 1996; Robinson and King 1997). Another focus of their studies has been the mechanism of misassembly and the incorrect folding pathways of TSP (Betts and King 1999; Betts and King 1998; Haase-Pettingell and King 1997). Some additional studies have been carried out in the laboratory of Robert Seckler at the Universität Potsdam. This group has published some work on the *in vitro* folding of TSP (Fuchs et al. 1991). Their most recent work has been with a fragment of the TSP corresponding to the β-helix domain (Miller et al. 1998; Schuler and Seckler 1998). These studies did not, however, go into great detail regarding the thermodynamic and kinetic folding mechanism.

It appears that parallel β-helix proteins represent an exciting new class of proteins. Although new to the scene of known folding motifs, parallel β-helix proteins have developed quite a strong following in the last eight years. The very simple, repetitive topology has intrigued researchers and led to some interesting

new insights and technical developments. With the discovery of new parallel β-helix proteins, even more will be learned about protein structure, function, evolution, and physical properties.

Chapter 2

The Stability, Structural Organization, and Denaturation of Pectate Lyase C

The work in this chapter has been published in the journal *Biochemistry*. Some modifications from the published manuscript have been made. The introduction has been edited to avoid redundancies from Chapter 1. Figures listed in the publication as "data not shown" have been included. This work was done in collaboration with Dr. Yuri Griko who was essential in performing the calorimetry experiments and discussing those results. The literature citation is:

Kamen, D. E., Griko, Y., and Woody, R. W. (2000). "The stability, structural organization, and denaturation of pectate lyase C, a parallel beta-helix protein." *Biochemistry*, 39(51), 15932-43.

2.1 Introduction

A fundamental aim in biological sciences is to understand the mechanism by which a protein adopts its functional three-dimensional structure. Studies on small model proteins such as lysozyme, myoglobin, and α-lactalbumin have led to significant advances toward this aim. A type of protein fold known as the parallel β-helix, first reported in 1993, has been observed in several proteins (Emsley et al. 1996; Lietzke et al. 1994; Mayans et al. 1997; Pickersgill et ai. 1994; Yoder et al. 1993a; Yoder et al. 1993b). The first to be observed, and the focus of this study, was pectate lyase C from *Erwinia chrysanthemi* (Yoder et al. 1993a; Yoder et al. 1993b).

PelC is a single polypeptide chain of 37676 Da that apparently consists of a single domain. The structure of pelC, which has been determined by x-ray crystallography at 2.2Å resolution, is shown in Figure 1.3 (Yoder et al. 1993b). A notable feature of this structure is the relatively high parallel β-sheet content making up the dominant fraction of the secondary structure (Table 2.1). The most interesting characteristic of the polypeptide backbone is that these parallel β-sheets wind up to form a large right-handed coil, called a parallel β-helix. This makes a very simple and highly ordered topology that went unobserved and unpredicted in more than 30 years of protein crystallography and nuclear magnetic resonance spectroscopy. PelC contains about 30% parallel β-sheet (Table 2.1), thus making it a good system to study the folding and structural organization of this type of secondary structure.

Table 2.1: Fractions of secondary structure in pelC determined by x-ray diffraction and circular dichroism.

	α-helix	β-sheet	turns	Other
X-ray*	10.7	26.6	4.0	58.7
DSSP+	8.8	32.0	7.9	51.3
CDPro#	24.3	21.3	22.0	32.5

*Estimates from the x-ray diffraction structure (Yoder et al. 1993a).

+Objective analysis of the structure from x-ray diffraction using the method of Kabsch and Sander (Kabsch and Sander 1983).

#Estimation of secondary structure from circular dichroism (Johnson 1999; Provencher and Glockner 1981; Sreerama and Woody 1993).

The goal of the experiments in this chapter is to investigate the physical properties and the folding of wild-type pelC. In order to understand the folding mechanism and structural organization, we have monitored the equilibrium of chemical and thermal denaturation by circular dichroism (CD), fluorescence, and differential scanning microcalorimetry (DSC). From these experiments, the free energy of folding ($\Delta G°_{H_2O}$) was determined. We have determined that there are two structural components of pelC that differ in thermal stability, and each one unfolds in a simple two-state manner. These structural units are capable of displaying some degree of independence. Although recently a fragment of the tailspike protein from bacteriophage P22 containing the parallel β-helix domain was isolated and studied along with several mutants (Miller et al. 1998; Schuler and Seckler 1998), our study represents the first biophysical characterization of a complete parallel β-helix protein.

2.2 MATERIALS AND METHODS

Purification of pelC. PelC was purified from the overexpressing *Escherichia coli* strain HB101 carrying a plasmid bearing the gene for wild-type pelC (pPEL410), using published methods (Kita et al. 1996). The cells were grown for ~48 hours at 22-25° C in Luria-Bertani (LB) media supplemented with 50 μg/ml ampicillin and 1mM isopropyl-β-D-thiogalactopyranoside (IPTG) at culture initiation. Cells were harvested by centrifugation at 8 K x g . Spheroplasts were prepared by the method of Witholt et al. (1976).

Cells from a 2 L culture were suspended in 200 ml of 0.2 M Tris-HCl (pH 8.0), then centrifuged at 8 K x g. The pellet was resuspended in 32 ml of 0.2 M Tris-HCl (pH 8.0). 40 ml of 0.2 M Tris-HCl/1 M sucrose and 800 μl of 50 μM EDTA were added. The suspension was mixed well and allowed to stand for 5 min. 800 μl of fresh hen egg-white lysozyme solution (6 mg/ml), then 80 ml distilled water were added, mixed well and allowed to stand for twenty minutes at room temperature. The resulting spheroplasts were pelleted by centrifugation at 15 K x g and the supernatant periplasmic fraction was retained. The supernatant was dialyzed against 5 mM Tris-HCl (pH 8.0) at 4°C for ~48 hours using tubing with an M_r cutoff of 6-8 kDa.

The dialyzed supernatant was applied to a Bio-Rad CM Bio-Gel A cation exchange column in 5 mM Tris-HCl (pH 8.0) at room temperature. The column was washed and pelC was eluted with a gradient of 0 M-0.25 M NaCl. Fractions were collected and analyzed for absorbance at 280 nm and pectinolytic activity. Peak pelC fractions were pooled, dialyzed against distilled water, lyophilized and stored at -20°C.

Pectinolytic activity assay. The assay was performed by the method of Kita et al. (1996). In a 1-cm path length quartz cuvette, 5 μl of diluted pelC were added to 125 μl of 1% sodium polypectate and 870 μl of 50 mM Bis-Tris-Propane, pH 9.5, containing 0.5 mM $CaCl_2$. The increase in absorbance at 232 nm was monitored, corresponding to the cleavage of polypectate. One unit of pectate lyase activity

is defined as 1 μmole of product formed per minute (1.73 absorbance units min^{-1}) (Kita et al. 1996).

SDS-PAGE and HPLC were used to assay for protein purity. Mass spectrometry and N-terminal sequencing confirmed the presence of pelC.

Circular dichroism spectroscopy. CD spectra of pelC were obtained with a Jasco J-720 spectropolarimeter (Tokyo, Japan) in cylindrical, water-jacketed quartz cuvettes. The temperature was regulated by circulating water. Protein concentrations were 10 - 20 μM. Far-UV measurements were made in a 0.01- or 0.02-cm path length cell, and near-UV CD spectra were recorded in a 1-cm path length cell. All concentration measurements were converted to residue concentration by multiplying the protein concentration by the number of amino acid residues. Protein concentrations were determined in 6 M guanidine-HCl by UV absorption using a Cary 118 spectrophotometer. A molar extinction coefficient of 61590 $M^{-1}cm^{-1}$ was calculated using the method of Edelhoch (1967). All CD spectra were obtained by averaging 20 scans, subtracting a baseline and converting the raw data to units of molar ellipticity.

Fluorescence spectroscopy. All fluorescence spectra were collected on an Aviv Instruments (Lakewood, NJ) model ATF-105 differential ratio spectrofluorometer using 1-cm quartz cuvettes and thermoelectric temperature control. PelC was excited at 282 nm, which is the excitation maximum, with excitation bandwidth of 2 nm and emission bandwidth of 4 nm. Raw photomultiplier tube signal was

converted to sample fluorescence, and a rhodamine-B quantum counter corrected for lamp fluctuations. Data were collected at 1 nm steps, and averaged for 1 second per datum. 8-anilino-1-naphthalenesulfonate (ANS) was used in a molar excess of 200 relative to the protein concentration.

Equilibrium unfolding experiments. Individual guanidine-HCl (gdn-HCl) solutions were prepared volumetrically as described by Pace (1988) using ultrapure gdn-HCl purchased from ICN Biomedicals (Costa Mesa, CA). The concentration of a stock solution was determined by the difference in refractive index between the guanidine solution and the buffer. A 6M stock solution was diluted in 0.1M sodium phosphate, pH 7.0 or 5.0, 150 mM sodium chloride, to the required final concentration. The necessary volume of a 100 μM protein stock solution was added to the individual gdn-HCl solutions and allowed to equilibrate for 36 – 48 hours with the temperature regulated at 25°C. Changes in far-UV CD were monitored at 218 nm at 25°C using a final protein concentration of 10 μM in a 0.1-cm path length cell. Changes in near-UV CD were followed at 277 nm at 25°C using a final protein concentration of 10 μM in a 1-cm path length cell. For denaturant titrations, changes in tryptophan fluorescence were followed by exciting at 290 nm and monitoring emission at 326 nm. For fluorescence measurements, protein concentrations were 1 μM.

Chemical denaturation data analysis. Data from gdn-HCl unfolding curves were analyzed assuming a two-state transition, using the linear extrapolation method

as described by Pace (1988). An entire unfolding experiment can be fit to equation 1.

$$y = b_f + (m_f * [\mathrm{Gdn}]) + (b_u + (m_u * [\mathrm{Gdn}])) \frac{\exp((-\Delta G^\circ_{H_2O} + m * [\mathrm{Gdn}]) / RT)}{1 + \exp((-\Delta G^\circ_{H_2O} + m * [\mathrm{Gdn}]) / RT)} \quad (1)$$

The quantities R and T represent the gas constant and the absolute temperature, respectively; y is the value of any observable parameter (CD, fluorescence, etc.) under specific conditions; b_f and b_u are the y-intercepts of the folded and unfolded baseline, respectively; m_f and m_u are the slopes of the folded and unfolded baselines, respectively; $\Delta G^\circ_{H_2O}$ is the standard free energy of unfolding in water in the absence of denaturant; and m is the change in free energy with denaturant concentration. The constant m has been proposed to be proportional to the change in solvent-exposed surface area with increasing denaturant concentration. It is an indication of the cooperativity of the system. The nonlinear regression function in the program Axum (MathSoft, Inc., Cambridge, Mass.) was used to fit the experimental data to equation 1. Initial parameter estimates were obtained from linear fits to the folded and unfolded baselines, and a linear fit to the free energy change. Estimated errors are those calculated from the nonlinear regression and represent the standard deviation of the mean.

Calorimetry. Calorimetric measurements were performed with the prototype DASM-4 model microcalorimeter built at the Johns Hopkins University. All

calorimetric measurements were conducted at a heating rate of 1 K/min and excess pressure of 1.5 atm. All experiments were performed in buffer solutions with an ionic strength of 20 mM. The protein concentrations in these experiments were varied from 25 to 44 μM depending on solvent conditions and tendencies to aggregate (pH 10, 25 μM; pH 8, 34 μM; pH 7, 37 μM; pH 6, 37 μM; pH 5, 33 μM; pH 4, 28 μM; pH 3, 44 μM). The partial specific heat capacity has been determined according to the procedure described elsewhere (Privalov and Potekhin 1986), assuming that the molecular weight of pelC is 37.7 kDa and the partial specific volume is 0.726 cm^3/g, calculated from the amino acid composition (Durchschlag 1986). The partial heat capacity of the unfolded polypeptide chain was calculated according to Privalov & Makhatadze (1990), using the known heat capacity values of amino acid residues.

The calorimetric transition enthalpy, $\Delta H (T_{tr})$ was determined from the area of the heat absorption peak by extrapolating the heat capacity of initial and final states to the mid-transition temperature, T_{tr}. This is in accordance with the previously determined temperature dependencies of these functions: linear for the native and polynomial for the unfolded state. The entropy of transition was determined at the transition temperature, T_{tr}, as: $\Delta S (T_{tr}) = \Delta H (T_{tr})/T_{tr}$. A deconvolution analysis of the protein excess heat capacity function was performed using the program based on the sequential procedure of Freire and Biltonen (1978). The DSC data were fitted assuming a multistate mechanism comprising one or more irreversible steps using software described elsewhere (Freire et al. 1990).

Thermal Denaturation Monitored by CD. Thermal denaturation was carried out, and monitored by far- and near-UV CD on a Jasco J-720 spectropolarimeter equipped with a NESLAB RTE-110 water bath. The rate of change of the temperature was 30° C / hour, controlled by circulating water. Data were collected at 0.5° C intervals. The midpoint of the denaturation curves was determined by taking the second derivative of the curve and using the temperature where the second derivative intersected the zero line.

Analytical Ultracentrifugation. All measurements were made using a Beckman XL-I analytical ultracentrifuge. Samples were prepared identically to those used for the chemical denaturation experiments. Data were collected using concentrations of 5, 10, and 15 μM at 18 K, 25 K, and 33 K revolutions per minute. The path length of the observation cell was 1.2-cm. Samples were spun for 12 hours. Three scans were taken at two-hour intervals. Identical results ensured that equilibrium was achieved. Origin software (Microcal Software Inc, Northhampton, Mass.) was used to fit the data to a single-species model of association using a partial specific volume of 0.726 cm^3/gm.

2.3 RESULTS

Sedimentation analysis. Seven of the known parallel β-helix proteins are trimeric (Beaman et al. 1997; Beaman et al. 1998; Kisker et al. 1996; Raetz and Roderick

1995; Steinbacher et al. 1994). Association of the monomeric pelC could lead to deviations from simple two-state behavior and thus to anomalies or complications in a study of protein denaturation. Sedimentation equilibrium experiments were carried out at pH 5 and 7. The results indicate that, under all conditions studied, pelC behaves as an ideal monomer with a relative molecular mass of 38019 ± 268 Da (Figure 2.1, and Table 2.2).

Reversibility of chemical unfolding. A requirement for thermodynamic analysis of any process is that it is reversible. A logical first step is to show that after unfolding in strongly denaturing conditions the native structure can be completely regained (Figure 2.2). The far-UV CD spectrum of native pelC is typical of a protein with a high fraction of β-sheet secondary structure. The secondary structure content of pelC (Table 2.1) determined from the x-ray structure (Sieber et al. 1995) analyzed by the DSSP method (Kabsch and Sander 1983), and from CD using the CDPro secondary structure prediction package (Johnson 1999; Provencher and Glockner 1981; Sreerama et al. 1999; Sreerama and Woody 1993) differ substantially more than the rms errors normally encountered in this type of analysis. The CD secondary structure analysis uses a reference set of proteins for which the 3D-structures and CD spectra are known. The program compares the sample CD spectrum with the reference spectra in the basis set. The secondary structure content is estimated from the known crystal structures in the basis set. Discrepancies arise in the analysis of pelC because this type of protein is not well represented in the reference set used for the secondary

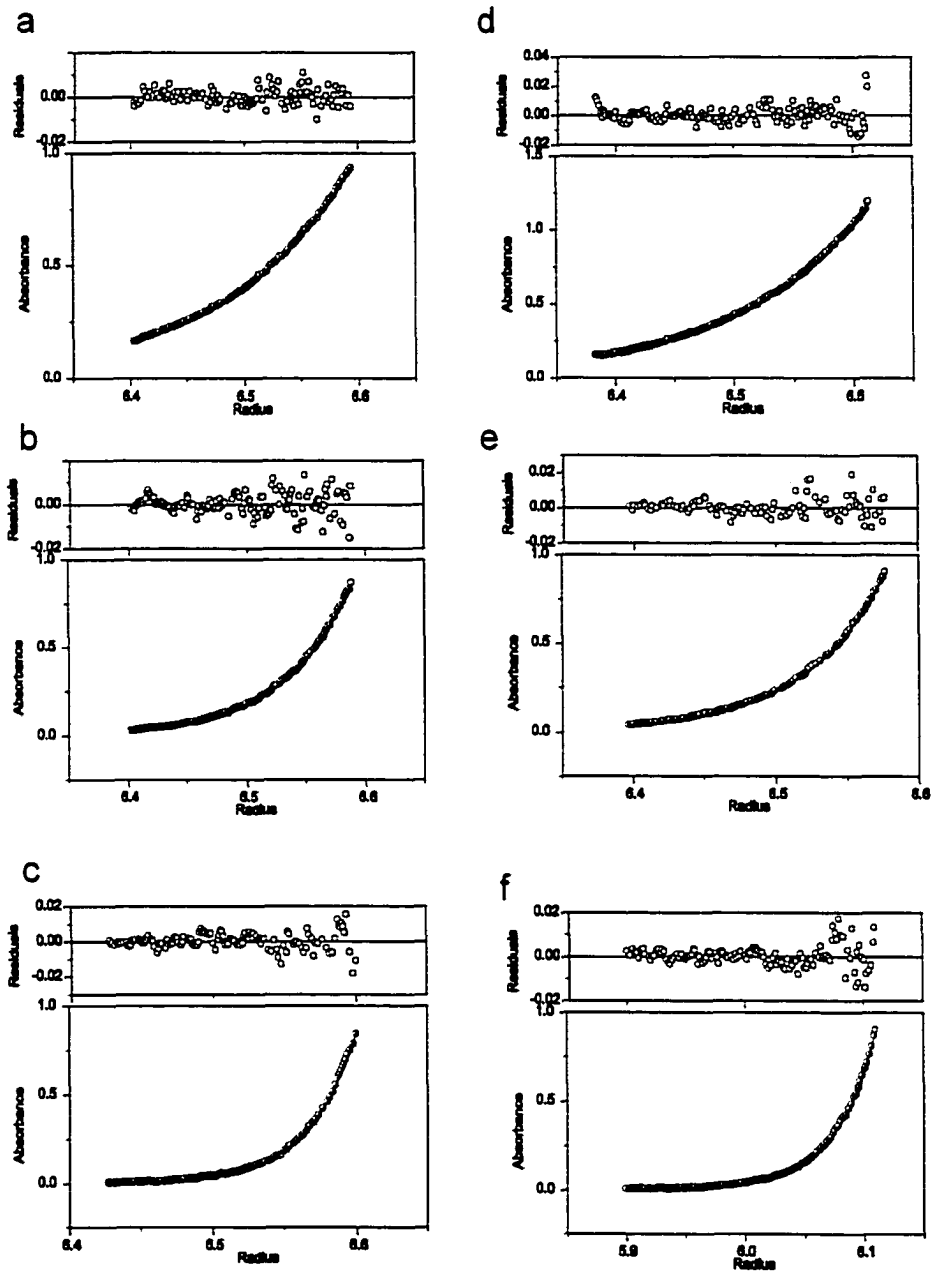


Figure 2.1: Sedimentation equilibrium of pelC at pH 5 (*a, b, c*) and pH 7 (*d, e, f*) fit to a single species model. Solution conditions were the same as the gdn-HCl denaturation experiments. Speeds are 18 K (*a, d*), 25 K (*b, e*), and 33 K rpm (*c, f*).

Table 2.2: Analytical ultracentrifugation sedimentation equilibrium results for pelC at pH 5 and 7.

	Starting Conc. (μM)	Molecular mass determined at speed x 1000 RPM 18	25	33
pH 7	5	39529	40122	38688
	10	37023	38548	37027
	15	38402	38006	38185
pH 5	5	39069	37879	37707
	10	37317	35310	36286
	15	39006	38497	38043

The average molecular mass from all runs is 38109 ± 268.

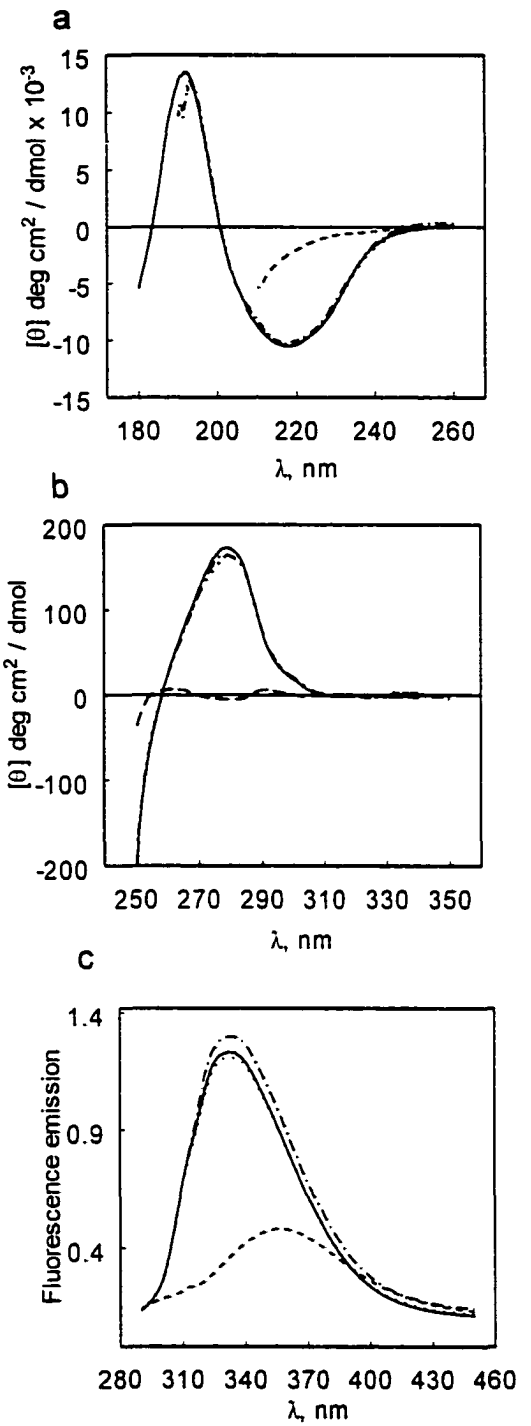


Figure 2.2: Far-UV (*a*), near-UV CD (*b*), and fluorescence spectra (*c*) of pelC at pH 7 (—), pH 5 (–·–·–), pH 7 and 6M guanidine-HCl (---), and refolded from 6M guanidine to 0.6M (···). For the far-UV CD the final gdn-HCl concentration is 0.44M.

structure analysis by CD. In addition, a small shoulder is present at about 225-230 nm (Figure 2.2a). This shoulder probably arises from aromatic contributions, and it may be responsible for the overestimation of the α-helix content. Figure 2.2 shows far- and near-UV CD spectra and fluorescence spectra of native, unfolded, and refolded protein. It is clear from this evidence that native-like secondary structure, tertiary structure, and tryptophan environment are regained upon dilution of gdn-HCl from 6M to less than 0.6M. The far-UV CD spectrum shows the broad minimum centered at 218nm, with an intensity of about –10500 deg cm^2 / dmol (Figure 2.2a). The CD spectrum of pelC in 6 M gdn-HCl clearly indicates a change in structure. The characteristic broad β-sheet band centered at 218nm has been lost, and the decreased intensity and altered band shape is that of a predominantly unordered polypeptide. At 25°C and 6M gdn-HCl there might be a contribution from the polyproline II (PP_{II}) conformation to the CD spectrum. The CD spectrum, which is weakly negative and devoid of features in the 220-230nm region, suggests such a contribution (Sreerama and Woody 1994). β-sheet secondary structure is clearly regained upon dilution of 6 M to 0.44M gdn-HCl.

In the near-UV CD region, the ordered environment of aromatic chromophores and disulfide bonds are responsible for circular dichroism (Woody and Dunker 1996). The native state is characterized by a broad, positive band centered at about 280nm which is due to the aromatic side-chains, with a small shoulder at about 295nm possibly arising from disulfide contributions (Figure 2.2b). The near-UV CD of the fully unfolded state in 6M gdn-HCl is featureless,

and even becomes slightly negative in the 280nm region. Upon dilution of the denaturant, the native spectral features reappear quantitatively. This indicates that the native arrangement of aromatic side-chains of pelC is completely lost at the elevated denaturant concentration and is fully regained upon refolding from the chemically denatured state.

Intrinsic tryptophan and tyrosine fluorescence are good probes of changes in protein conformations. They are sensitive to the degree of solvent exposure, i.e. the polar environment of the aromatic side-chains. Native pelC, upon excitation at 282nm, shows a broad emission band with a maximum at 333nm (Figure 2.2c). In 6M gdn-HCl, a loss of 70% of the intensity is observed, along with a shift in the emission maximum 21nm to the red to 354nm. In addition, there is a weak shoulder at about 315nm. The latter feature is typical of tyrosine fluorescence, and indicates a loss of energy transfer from tyrosine to tryptophan. These changes are consistent with the process of protein unfolding and exposure of buried tryptophan residues to the solvent, as well as an increase in the average distance between tyrosine and tryptophan side-chains. The loss of intensity and red shift in tryptophan emission results from solvent quenching (Chen and Barkley 1998; McMahon et al. 1992). After a ten-fold dilution from 6M to 0.6M gdn-HCl, pelC completely regains the native emission spectrum, indicating that the aromatic side-chains are in an environment identical to that of the native conformation.

One of the most convincing arguments for reversibility is the reappearance of enzymatic activity after unfolding. After complete unfolding in 6M gdn-HCl,

and removal of the denaturant by dialysis, the enzymatic activity of pelC is regained and is essentially identical to that of freshly purified pelC, within experimental error. The specific activity of pelC is 2.2×10^6 u/g. The activity of a sample of pelC unfolded, then refolded, was measured at 2.3×10^6 u/g.

These results indicate that the structure of pelC can be completely disrupted by a chaotropic denaturant and can then be fully recovered upon dilution of the denaturant.

Guanidine-HCl unfolding of pelC at pH 5 and 7. PelC can be reversibly unfolded by titration with gdn-HCl. Figure 2.3 shows the changes in the optical properties of pelC at pH 7 upon increasing the denaturant concentration. The molar ellipticity at 218nm is an indication of the amount of intact β-sheet secondary structure within pelC (Figure 2.3a). A two-state model of unfolding is assumed for this analysis. A smooth transition occurs upon going from 0.0-3.0M gdn-HCl with no indication of intermediate conformations. In the transition region, there are no plateaus or inconsistencies between different physical properties that might imply the existence of intermediates or structural domains. Another important result to note is that an identical curve results from dilution of the denaturant with buffer. The equilibrium value for the CD signal obtained by starting the reaction with either folded or unfolded protein is identical. This indicates that chemical denaturation is thermodynamically reversible and equilibrium is achieved. Another essential feature to notice is the single, highly cooperative transition from the folded to the unfolded state. This allows one to

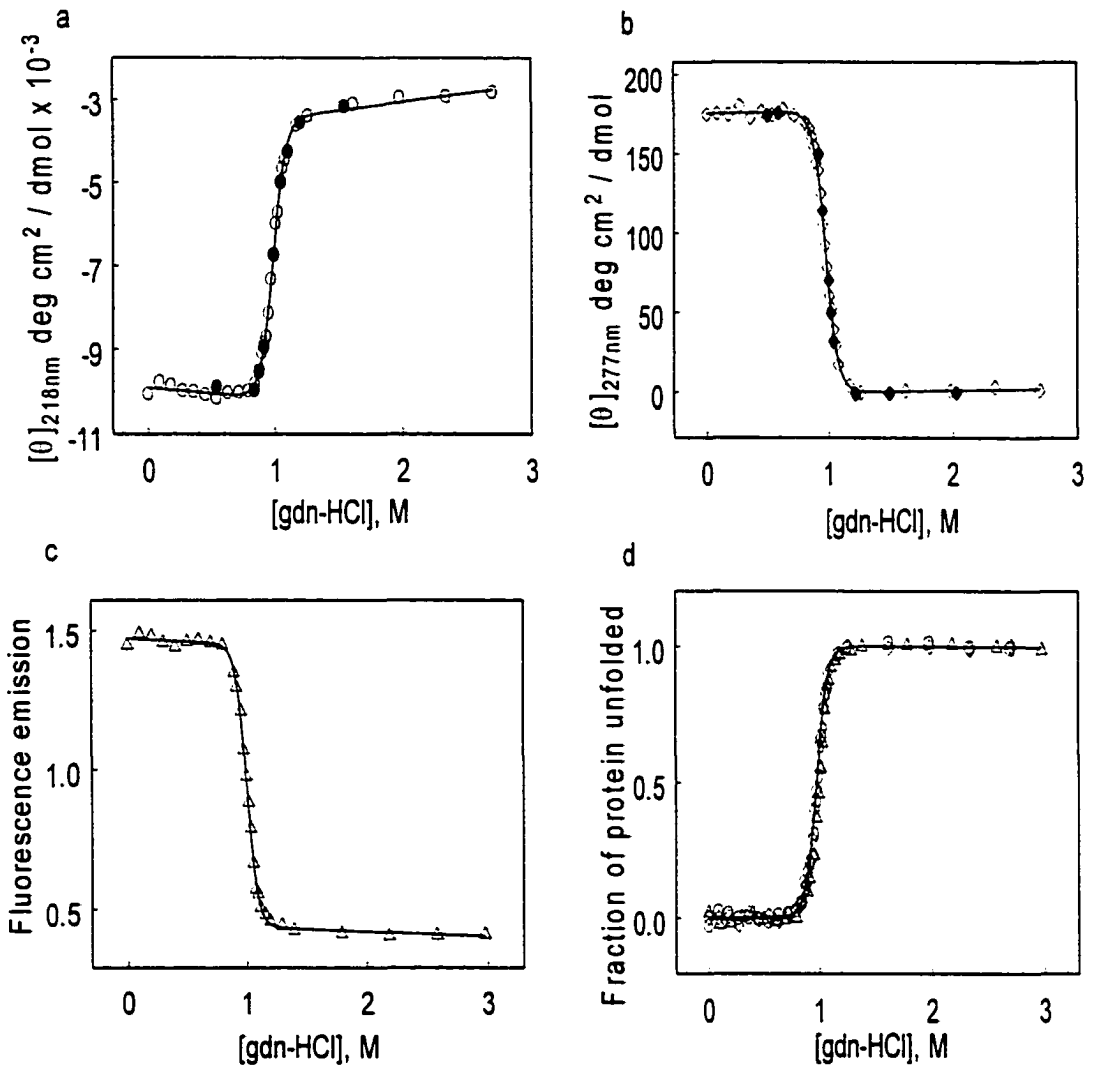


Figure 2.3: Denaturation at pH 7 followed by (*a*) CD at 218 nm and (*b*) CD at 277nm and (*c*) fluorescence emission at 326 nm with excitation at 290 nm. Open symbols represent the addition of denaturant to a solution of folded protein, and closed symbols are for the dilution of denaturant by buffer. The transitions shown in (*a-c*) have been converted to fraction of protein unfolded and shown in (*d*).

define the equilibrium constant for points in the transition region and to calculate the free energy of unfolding using the linear extrapolation method as described in Materials and Methods (Table 2.3). These three facts allow us to be confident that thermodynamic parameters extracted from the curves are accurate. This analysis yields: $\Delta G^{\circ}_{H_2O}$ = 12.06 ± 0.59 kcal / mol, *m* = 12.19 ± 0.59 kcal / mol M, and $D_{1/2}$ = 0.99 ± 0.1M. The native baseline region is linear but shows a slight negative slope. This is probably a result of the molecular structure adjusting to changing solvent conditions resulting from increased ionic strength from the gdn-HCl. The denatured baseline also is linear but has a significant positive slope. This results from increasing amounts of PP_{II} structure within the denatured protein ensemble as the concentration of gdn-HCl is increased at constant temperature.

Figure 2.3b shows the near-UV CD as a function of gdn-HCl concentration at pH 7. Aromatic sidechains are primarily responsible for the near-UV CD. Again there is no concentration range within the transition region where any intermediate conformations are detectable, and the transition is completely reversible. Figure 2.3b shows the molar ellipticity at 277nm fit to a two-state model. This unfolding transition is similar to that observed in the far-UV. It is a single smooth transition from the native to the unfolded conformation. In this case the native and denatured baselines show a negligible slope. Therefore, the only significant change to the aromatic structure occurs in the transition region. A nonlinear least-squares analysis of the data yields results identical, within experimental error, to those obtained from far-UV CD experiments (Table 2.3).

The same experiment was repeated with fluorescence emission spectroscopy. Figure 2.3c shows the denaturation process monitored with excitation at 290nm and emission at 326nm, fit to the same two-state model as the CD experiments. This transition is qualitatively similar to the CD-monitored transitions, with a single sharp transition and no intermediate conformations stabilized. In this case, the slope of the native and denatured baselines results from changing solvent conditions causing an increase in collisional quenching. A nonlinear least-squares analysis of these data also yields results identical to those determined by CD (Table 2.3).

An important observation is that the transition region appears to occur at similar denaturant concentrations as determined by the three separate probes. Figure 2.3d shows the three spectral plots converted into fraction of protein unfolded, and fit to the two-state model used in all three unfolding experiments. Indeed, the transitions monitored by the three separate probes are coincident, and to a first approximation, this system can be treated as a two-state process.

PelC has its maximum thermal stability at pH 5 (Table 2.4, Figure 2.8b). This is four pH units below the isoelectric point (pI = 9). Many proteins exhibit a maximum in thermal stability at or about their pI. We therefore decided to investigate the behavior of pelC at pH 5. In some respects, the results are similar to those obtained at pH 7 but, surprisingly, there were some interesting features that arose when the pH was reduced to 5. Figure 2.2 shows that at pH 5 and 7, pelC shows similar spectroscopic properties. The far- and near-UV CD

Table 2.3: Thermodynamic results from gdn-HCl denaturation at pH 7 and 5*.

		218nm CD	277nm CD	Fluorescence emission
pH 7	ΔG° (kcal / mol)	12.06 ± 0.59	11.56 ± 0.40	13.03 ± 0.67
	m (kcal / mol M)	12.19 ± 0.59	11.81 ± 0.40	13.04 ± 0.67
	$D_{1/2}$ (M)	0.99 ± 0.1	0.98 ± 0.06	1.00 ± 0.1
pH 5*	ΔG° (kcal / mol)	12.10 ± 0.38	14.92 ± 0.50	14.02 ± 0.67
	m (kcal / mol M)	12.51 ± 0.39	15.70 ± 0.52	14.51 ± 0.70
	$D_{1/2}$ (M)	0.97 ± 0.06	0.95 ± 0.06	0.96 ± 0.1

*Because the two-state model on which the analysis is based does not hold at pH 5, these results do not accurately describe the denaturation process at this pH, as indicated by the significant differences between the values obtained from the three spectroscopic probes.

and fluorescence spectra are identical when measured for the two pH values. We can conclude that there are no significant perturbations in the structure of native pelC in going from pH 5 to 7. However, there is a change in the denaturation behavior. When denaturation was monitored by far-UV CD (backbone unfolding), there appeared to be virtually no change in the behavior (Figure 2.4a). However, there was an increase in the overall free energy change and the *m* value of unfolding for the transition when following side-chain properties, such as near-UV CD or fluorescence (Table 2.3, and Figures 2.4b and c), but only a small change in the unfolding midpoint, which is the same as for pH 7 within experimental error. The unfolding transition monitored by three probes still appears coincident under these conditions (Figure 2.4d) despite the differences in the $\Delta G^{\circ}_{H_2O}$ and *m* values as measured by the three probes. At pH 7, $\Delta G^{\circ}_{H_2O}$, *m*, and $D_{1/2}$ are almost identical when measured by CD and fluorescence. At pH 5 the $\Delta G^{\circ}_{H_2O}$ value measured by CD at 218nm is lower than that calculated from the parameters sensitive to side-chain conformation. One still observes what appears to be global unfolding of the protein due to a high degree of cooperation within the protein, as indicated by the apparent coincidence of the unfolding transitions. However, it appears that 277nm CD and fluorescence are measuring a different transition than 218nm CD is measuring. The discrepancy between the $\Delta G^{\circ}_{H_2O}$ and *m* values from the different spectral probes indicates that at pH 5 there is a decrease in the cooperativity within pelC, and therefore these data do not support a two-state model of denaturation.

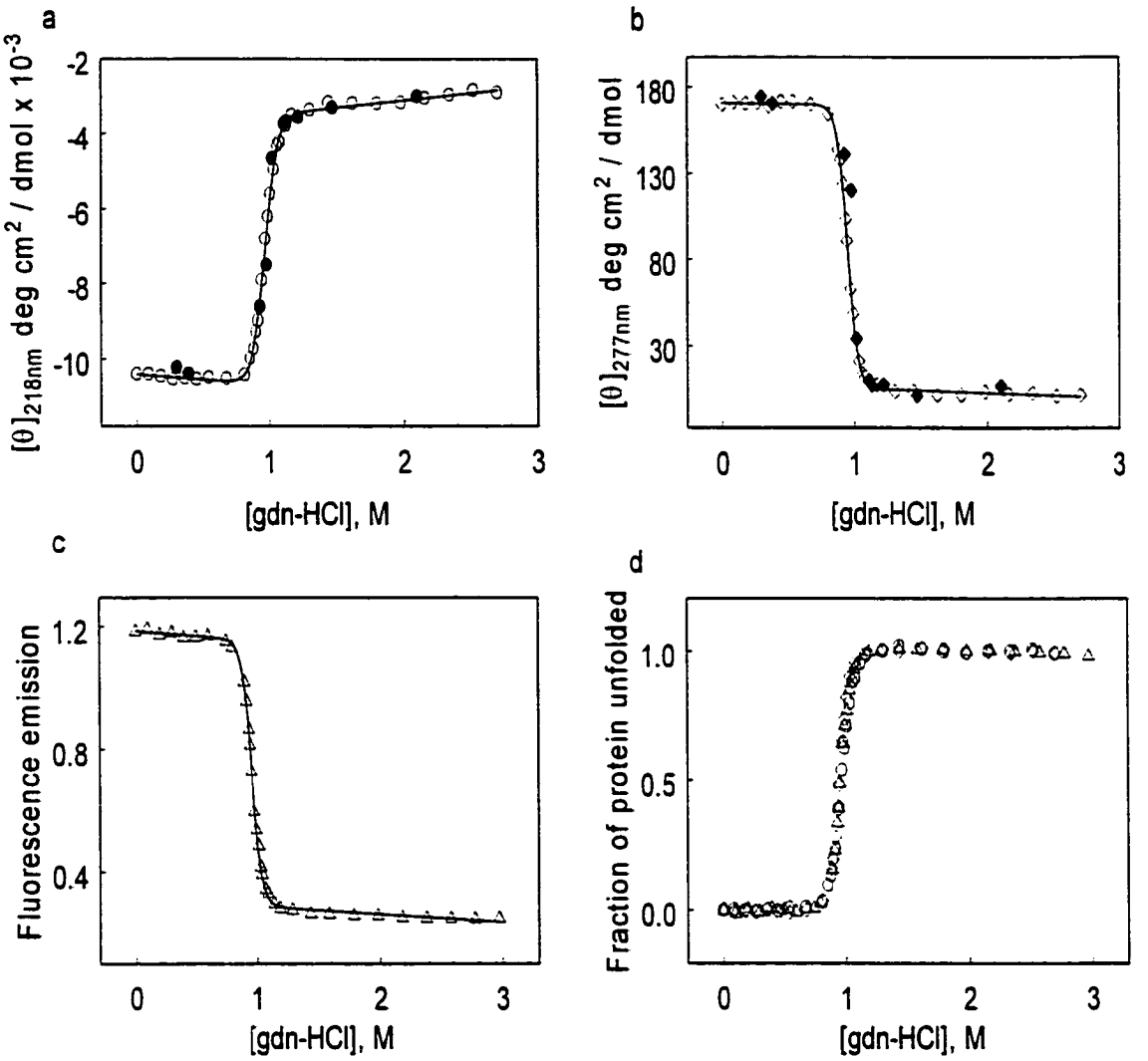


Figure 2.4: Denaturation at pH 5 followed by (*a*) CD at 218 nm and (*b*) CD at 277nm and (*c*) fluorescence emission at 326 nm with excitation at 290 nm. Open symbols represent the addition of denaturant to a solution of folded protein, and closed symbols are for the dilution of denaturant by buffer. The transitions shown in (*a-c*) have been converted to fraction of protein unfolded and shown in (*d*).

We tested the idea that a molten globule intermediate may be responsible for the deviations from apparent two-state behavior at pH 5. When ANS is added to a solution of pelC in acidic buffer (pH < 3), a large increase in the fluorescence emission at 500nm is observed (see chapter 3, Figure 3.1). This indicates that ANS is bound in a hydrophobic environment. What is actually happening is that ANS is inducing the formation of ordered structure via electrostatic interactions (see chapter 3). Several other lines of evidence indicate that the protein is completely unfolded at pH 2. Calorimetry indicates that there is no cooperative increase in heat absorption with an increase in temperature (Figure 2.5a), implying the absence of structure at pH 2. Isoelliptic points in the far-UV CD (209 nm, Figure 2.5b) and the near-UV (257 nm, Figure 2.5c) indicate that pelC exhibits essentially two-state denaturation behavior over the pH range of 2.2-7 at 25° C. At pH 2.2, the protein is completely unfolded, whereas at pH 3, folded and unfolded protein are present to a comparable extent. Furthermore the temperature dependence of the far-UV CD signal indicates that from 0-80°C there is only a modest change in the equilibrium distribution of unordered conformations at pH 2.2, and no cooperative transition is observed (Figure 2.6). The temperature dependence of the CD spectrum at pH 5 and 2M gdn-HCl also indicates a change from PP_{II} conformation to a more random distribution of structures as the temperature is increased (Figure 2.7), suggesting that there is no residual secondary structure in denatured pelC.

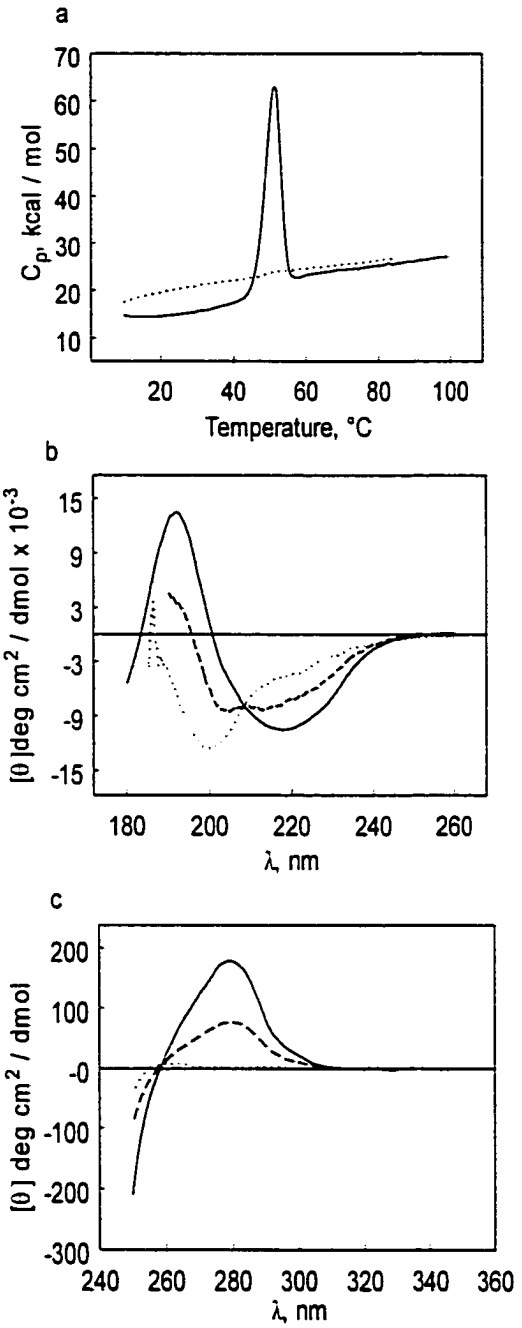


Figure 2.5: Native and acid-denatured pelC. Calorimetric unfolding curves (*a*) at pH 2.2 (···) and pH 7 (—), (*b*) far-UV CD spectra at pH 2.2 (···), 3 (---), and 7 (—), and (*c*) near-UV CD spectra at pH 2.2 (···), 3 (---), and 7 (—). The CD spectra were collected at 25°C.

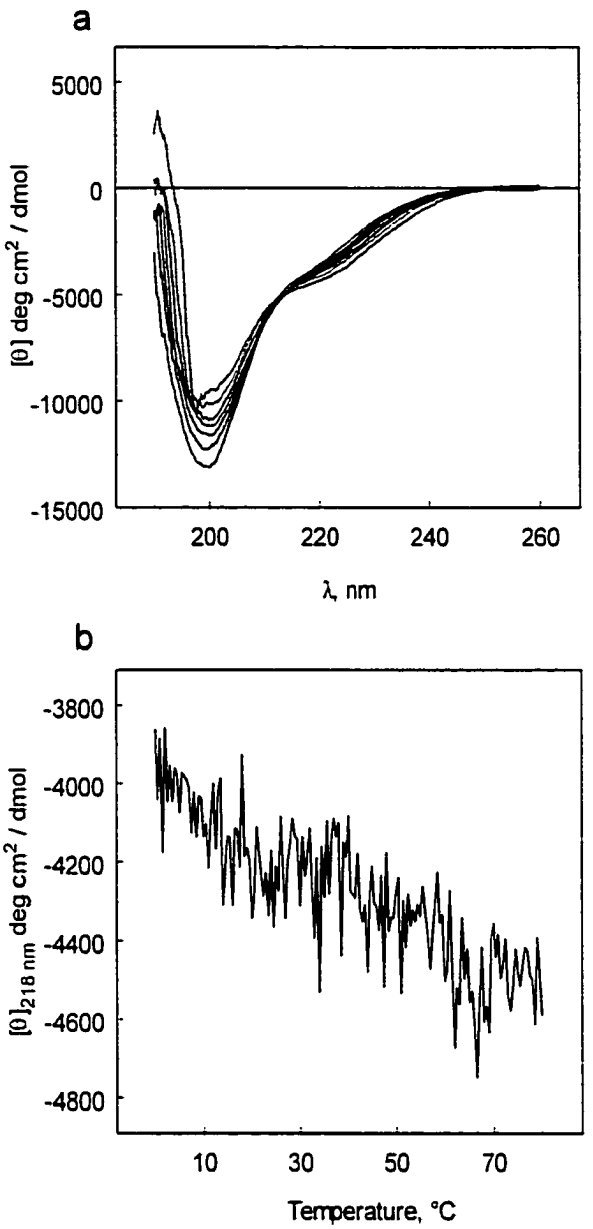


Figure 2.6: (*a*) Temperature dependence of the CD spectrum of pelC at pH 2.2 at 0, 10, 20, 30, 40, 60, and 80 °C (reading from bottom to top at 200 nm). (*b*) Temperature dependence of the 218 nm CD signal.

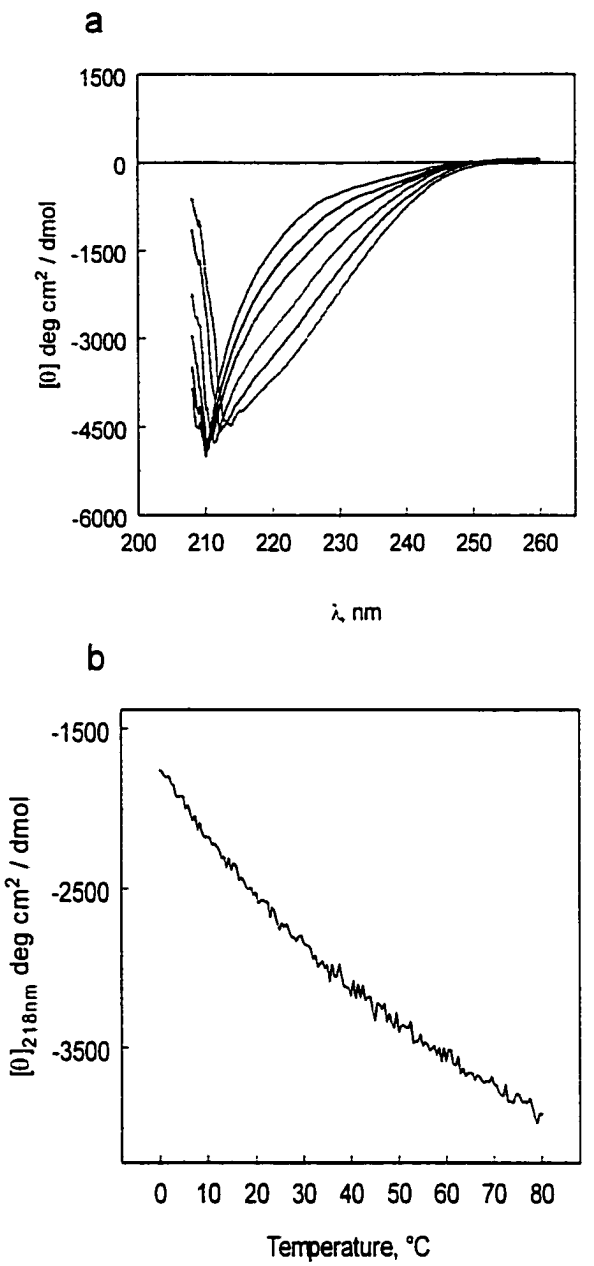


Figure 2.7: (*a*) Temperature dependence of the CD spectrum of pelC at pH 5 and 2 M gdn-HCl at 0, 10, 20, 30, 40, 60, and 80 °C (top to bottom). (*b*) Temperature dependence of the 218 nm CD signal.

Calorimetric unfolding of pelC. Figure 2.8 shows the heat capacity (C_p) as a function of temperature for the irreversible thermal denaturation of pelC (see discussion). The protein undergoes a cooperative transition accompanied by an increase in heat capacity and extensive heat absorption (ΔH). The denaturation transition appears as a single peak (Figures 2.8a and b), and upon baseline correction appears to be symmetric (Figure 2.9a). To a first approximation, denaturation may be treated as a two-state transition. However, the calorimetrically measured enthalpy exceeds the van't Hoff enthalpy, whereas they should be equal for a two-state transition (Table 2.4). A deconvolution analysis of the curves has been carried out. Under all solvent conditions studied, the denaturation transition can be resolved into two simple two-state transitions. Figure 2.9a shows the deconvolution of the transition at pH 6, and the two transitions involved. This might mean that pelC contains two structural blocks that strongly cooperate with each other. This cooperativity is so strong that chemical and thermal denaturation appear to be a single cooperative transition at pH 7. Similar observations have been made for the proteins pepsinogen and phosphoglycerate kinase (Griko et al. 1989; Mateo and Privalov 1981).

At first glance, there appears to be an unusual temperature dependence of the heat capacity peaks. The height of the heat absorption peak shows an initial increase, then decrease with increasing temperature (Figure 2.8b) that might be interpreted as a decrease of the enthalpy of denaturation with increasing temperature. At the same time, the positive ΔC_p observed upon pelC denaturation requires that the enthalpy change should always increase with

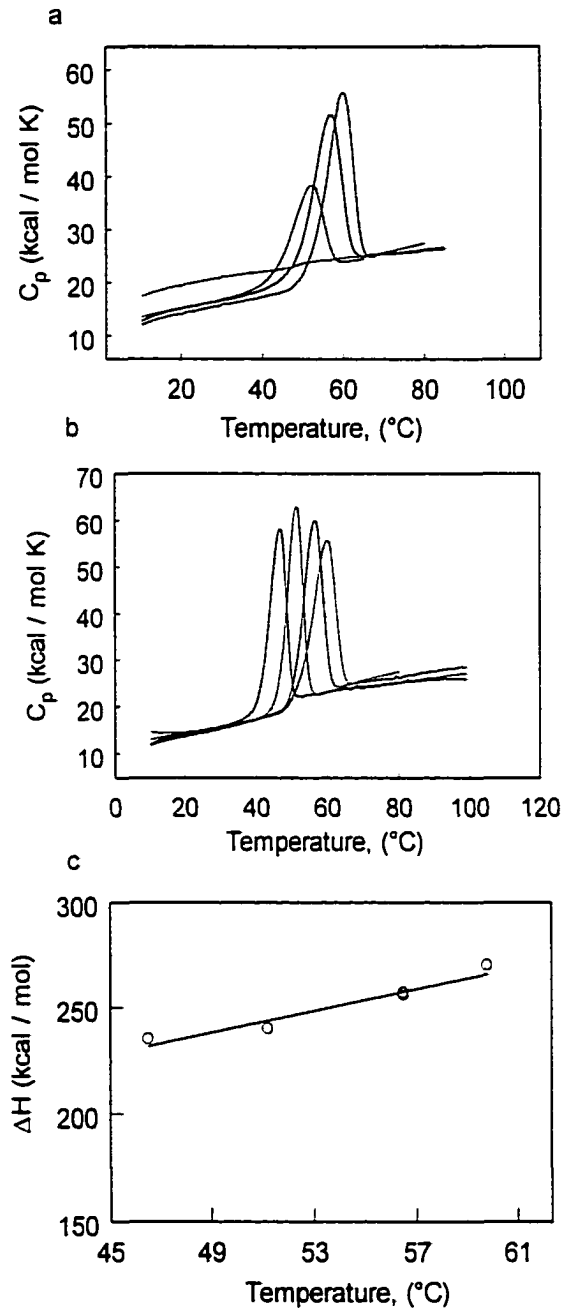


Figure 2.8: Calorimetric unfolding curves for pelC in order of increasing peak height for (*a*) pH 2.2, 3, 4, and 5 and from left to right for (*b*) pH 8, 7, 6 and 5. See Table 2.4 for a summary of the thermodynamic data obtained from these curves. (*c*) ΔH_{cal} vs. temperature indicates that the transition enthalpy increases with increasing temperature.

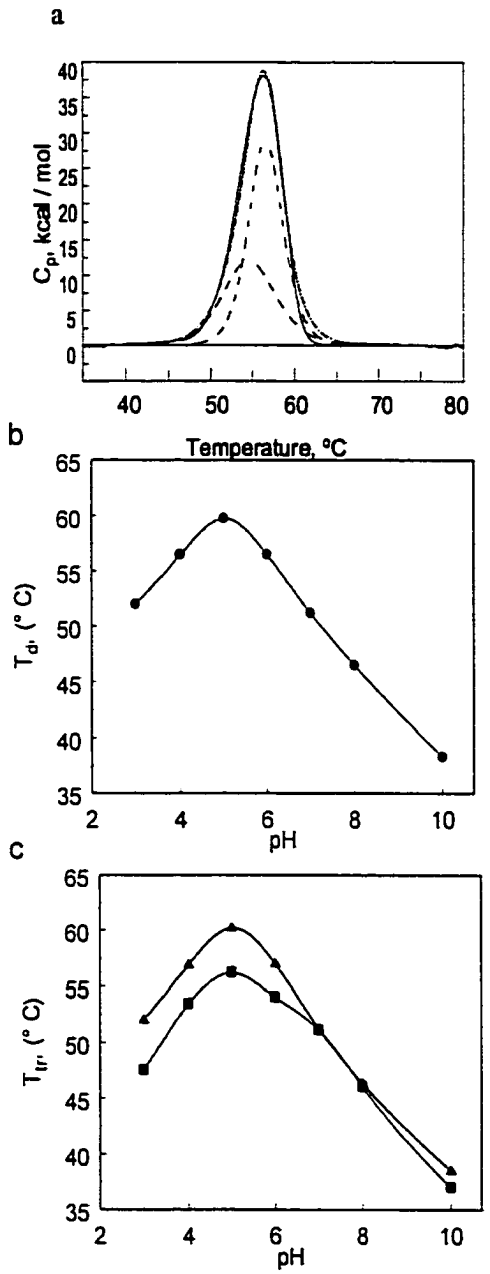


Figure 2.9: (a) Deconvolution analysis of pelC denaturation at pH 6 indicating the experimental curve (—) and deconvoluted component curves (---) and the sum of the calculated components (-----). (b) Observed calorimetric peak maximum as a function of pH. (c) Transition temperatures for the deconvoluted component curves T_1 (■) and T_2 (▲).

Table 2.4: Thermodynamic properties of pelC determined by differential scanning microcalorimetry.

pH	T_m °C	ΔH_u Kcal /mol	ΔH_{vH} Kcal/ mol	$\Delta H_{vH}/\Delta H_u$	ΔS_u Kcal / mol K	$\Delta H(25°C)$ Kcal/mol	$\Delta S(25°C)$ Kcal/ mol K	$\Delta G(25°C)$ Kcal/mol	$\Delta C_p(T_{max})$ Kcal / mol K
8	46.5	235.30	201.09	0.85	0.74	149.30	0.46	12.87	4.00
7	51.2	240.20	219.64	0.91	0.74	150.80	0.45	15.74	3.40
6	56.5	255.30	203.07	0.80	0.77	132.50	0.38	18.33	3.90
5	59.8	270.20	181.67	0.67	0.81	134.50	0.38	20.88	3.90
4	56.5	256.10	174.09	0.68	0.78	130.10	0.38	18.25	4.00
3	51.8	162.00[a]	178.30[a]	1.11[a]	0.50	89.60	0.27	9.67	2.70

T_m is the midpoint temperature, ΔH_u is the calorimetric enthalpy determined by the area under the heat capacity curve, ΔH_{vH} is the van't Hoff enthalpy, ΔS_u is the change in entropy at the transition temperature, ΔG is the free energy of unfolding, and ΔC_p (T_{max}) is the change in heat capacity at the maximum temperature.

[a] ΔH_{vH} is greater than ΔH_u at pH 3 because of the large population of unfolded molecules in equilibrium with the folded protein.

increasing temperature. The explanation for the unusual evolution of the heat absorption peaks is in the decrease of the cooperation between the two structural blocks in pelC. Because of this decrease in cooperation within the protein, the integral heat absorption peak of the heat capacity will be broader, and the peak area (calorimetric enthalpy) increases (Table 2.4, Figure 2.8c). Figure 2.9b shows the apparent transition temperature, or more appropriately, the peak maximum, for pelC. This indicates that the protein's maximum thermal stability is at pH 5. Figure 2.9c shows that the stability of the structural units is very similar from pH 7-8. Above pH 8 and between pH 7 and 5, the difference in thermal stability increases. This means that the structural blocks respond differently to changes in pH, implying some degree of independence. Below pH 5 the difference in thermal stability remains constant with temperature, and the observed decrease in the T_{tr} is a result of the titration of acidic side chains and stabilization of the unfolded conformation. Tables 2.4 and 2.5 give a summary of the thermodynamic data for the overall transition and the two components from the deconvolution analysis, respectively.

CD-monitored thermal denaturation. The results obtained from calorimetry suggest that two cooperative structural blocks are present. The temperature-dependence of the far-UV CD signal was therefore examined at pH 5 and 7 in gdn-HCl concentrations corresponding to those in the equilibrium unfolding transition (Figure 2.10). At high temperatures, pelC solutions less than 0.6M in gdn-HCl aggregate irreversibly and precipitate. If the time spent at high

Table 2.5: Thermodynamic results for the analysis assuming two two-state transitions.

pH	Tm_1 °C	Tm_2 °C	T_{max} °C	ΔH_1 Kcal/mol	ΔH_2 Kcal/mol	$\Delta H_{cal.}(T_{max})$ Kcal/mol
8	46	46.3	46.5	81.8	165.1	235.3
7	51.1	51.2	51.2	131.0	135.3	240.2
6	54.0	57.0	56.5	121.0	158.8	255.3
5	56.2	60.2	59.8	111.4	159.5	270.2
4	53.4	56.9	56.5	102.3	136.4	256.1
3	47.5	52.0	51.8	66.5	105.4	162

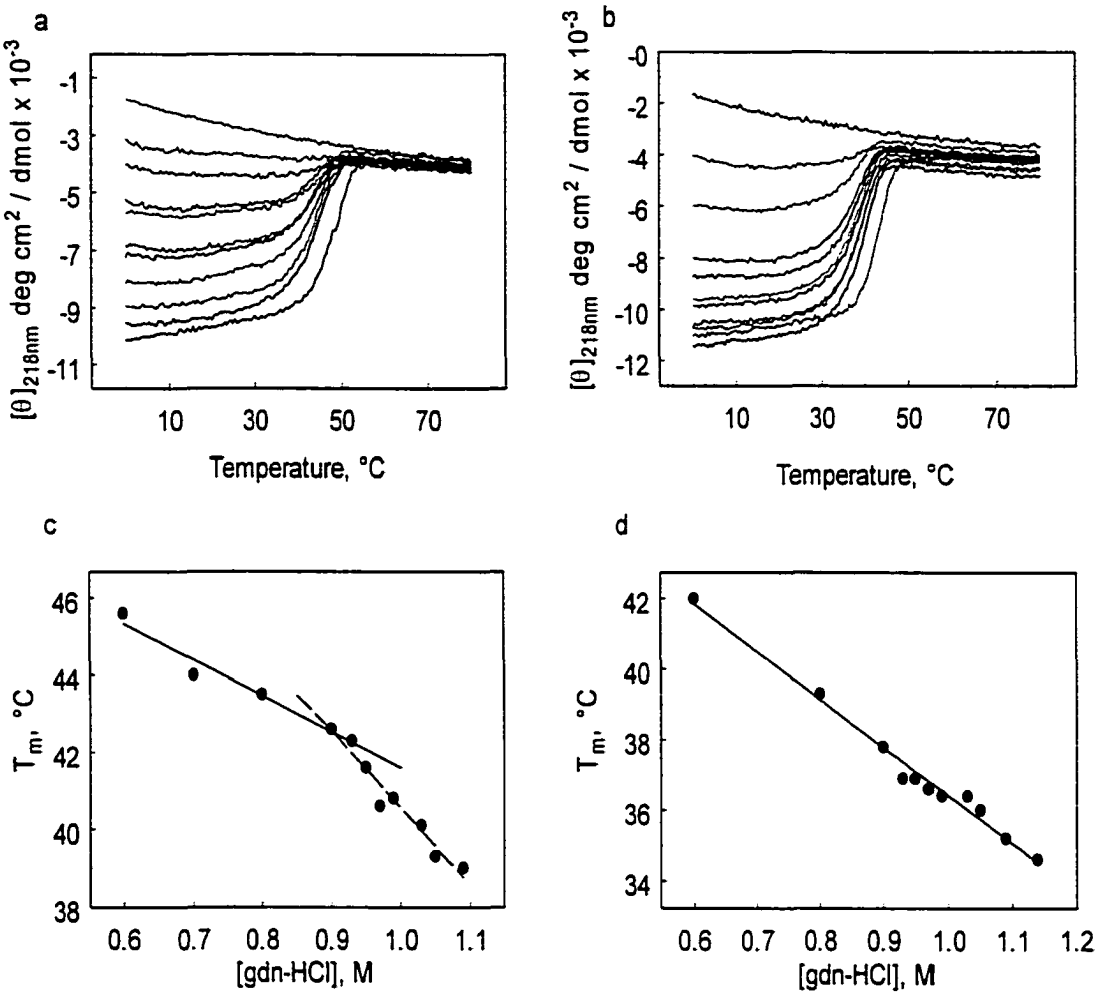


Figure 2-10: Thermal denaturation of pelC monitored by far-UV CD at 218nm in (a) pH 5 phosphate/NaCl, and (b) pH 7 phosphate/NaCl. Midpoint temperature vs. gdn-HCl concentrations is shown for (c) pH 5 and (d) pH 7. The curves correspond to gdn-HCl concentrations of 0.6, 0.9, 0.93, 0.95, 0.97, 0.99, 1.03, 1.05, 1.09, 1.14 and 2.0M reading from bottom to top in the low temperature region.

temperatures is short, the thermal transition of pelC is not accompanied by large effects of irreversibility in gdn-HCl solutions of 0.6M or greater. PelC undergoes a cooperative transition in all gdn-HCl concentrations in the transition region. It is only at gdn-HCl concentrations exceeding 1.15M that a transition is not observed. At both pHs and at gdn-HCl concentrations greater than 0.99M, the initial 218nm ellipticity at 0° C is less negative than that at 20° C. Under these conditions at pH 5, the unordered regions of pelC have a substantial fraction of the PP_{II} conformation. Near-UV CD (Figure 2.11) indicates no slope in the baseline from 0° to 20° C. This indicates that there is no detectable change in the tertiary structure. However, it appears that at pH 7 and temperatures below 25° C there is a significant positive slope in the near-UV CD baseline (Figure 2.11) indicating that the unfolded state becomes populated at low temperatures. The decrease in magnitude of the observed CD signal as the temperature is decreased and the unfolded form is favored may result from cold denaturation of pelC at low temperatures. This can lead to decreased cooperation of the protein's internal interactions and deviations from two-state behavior (Griko et al. 1989).

A plot of the T_m versus denaturant concentration shows yet another difference in behavior at pH 5 and pH 7 (Figure 2.10). At pH 7, the midpoint temperature is a linear function of denaturant concentration, but at pH 5, there is a break in the plot at about 0.9M gdn-HCl. The T_m is a linear function of gdn-HCl concentrations above and below 0.9M gdn-HCl, the start of the equilibrium unfolding transition. This evidence supports the notion that there are two

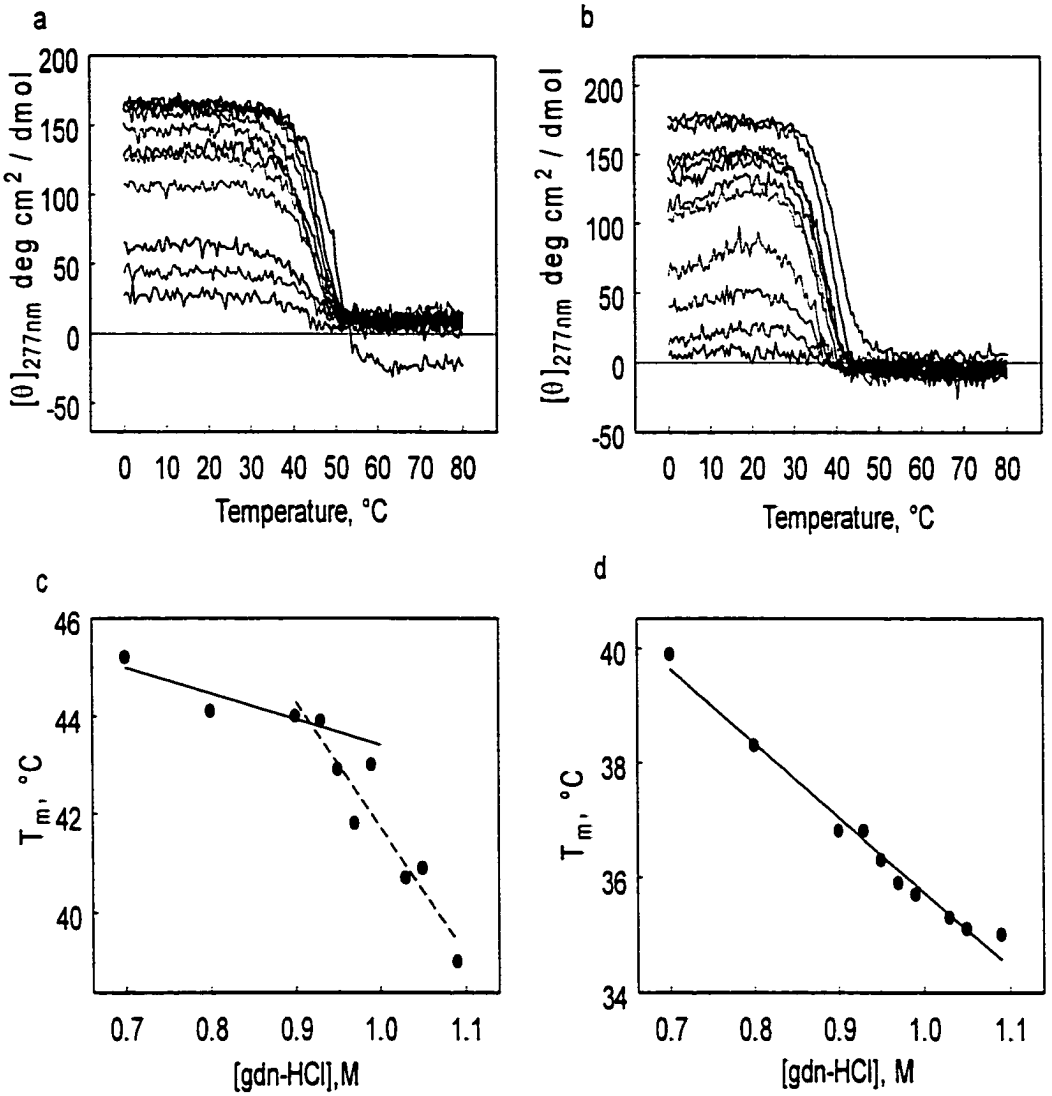


Figure 2-11: Thermal denaturation of pelC monitored by near-UV CD at 218nm in (a) pH 5 phosphate/NaCl, and (b) pH 7 phosphate/NaCl. Midpoint temperature vs. gdn-HCl concentrations is shown for (c) pH 5 and (d) pH 7. The curves correspond to gdn-HCl concentrations of 0.7, 0.8, 0.9, 0.93, 0.95, 0.97, 0.99, 1.03, 1.05, 1.09 and 1.15M reading from top to bottom in the low temperature region.

structural units and these units behave differently under different pH conditions. At pH 7, these structural units show similar thermal properties and, therefore, their denaturation processes are not resolved. However, at pH 5, where the structural units show some independence in the transition region, we observe two separate behaviors. In more native conditions (≤ 0.8M gdn-HCl), the T_m is linear, but in the transition region there is a loss of cooperation within the molecule causing a change in the behavior of the midpoint temperature. The slope of the T_m dependence is steeper, indicating a loss of stability due to domain independence. In other words, as the gdn-HCl concentration increases, the independence of the structural units also increases. We can conclude that this molecule is very strongly dependent on intramolecular cooperativity for its stability, and therefore, any loss of this cooperativity results in a concomitant decrease in stability. The same experiments monitored in the near-UV give results that are virtually identical to those from the far-UV (Figure 2.11). This supports the observation of two simple two-state transitions.

2.4 DISCUSSION

Denaturation of pelC. To a reasonable first approximation, gdn-HCl unfolding of pelC at pH 7 can be treated as a two-state process at equilibrium. A single, sharp, and fully reversible denaturation transition is observed, which is coincident when measured by three different probes. Regardless of the conformational property followed during denaturation, the transition profile is the same. This

implies that no intermediate conformations are present that can be detected by the techniques employed. The techniques used are very sensitive to conformations and environments of the backbone and side-chains. Quite often, in protein unfolding studies, one observes that the near-UV CD shows that unfolding occurs at a lower denaturant concentration or temperature than that observed by far-UV CD. This is often interpreted as evidence for an intermediate state, but this is not always the case. In the case of pelC, what is observed is somewhat unusual. At pH 7 the transitions induced by chemical and thermal denaturation and monitored by spectroscopic probes suggest that a two-state process accurately describes the transition. Three independent spectroscopic probes yield a coincident chemical denaturation transition as well as thermodynamic parameters that are identical within experimental error. When thermal denaturation in the presence of gdn-HCl is followed by two spectroscopic probes at pH 7, there also appears to be a two-state transition. Although thermal denaturation is only partially reversible, and therefore not a strict equilibrium process, one can still observe the denaturation and approximate the transition midpoint. The dependence of the midpoint temperature on solution conditions is related to structural changes. Under conditions that might stabilize an intermediate or favor a non-two-state process, such as 0.9-1.2M gdn-HCl, one still observes only a single transition. A plot of the midpoint temperature vs. denaturant concentration yields a straight line. This implies that the decrease in midpoint temperature is a result of an overall destabilization of the protein, and it is not due to an increasing population of intermediate structures. The midpoint

temperatures at each gdn-HCl concentration are also approximately equal when measured by far- and near-UV CD.

The spectroscopically based observations are contradicted by calorimetry, which indicates that the overall unfolding process is not a two-state transition, but can be resolved into two separate two-state transitions. A true test of a two-state process can be made using calorimetry. The ratio of the van't Hoff enthalpy to the true calorimetric enthalpy (ΔH_{vH} / ΔH_{cal}) for a two-state process is one. Typically an experimental error of ± 5% is allowed (Privalov 1979). For pelC at pH 7, this ratio is 0.91. This value is close to one, but it is sufficiently outside of the margin of error such that one is unable to assume a two-state transition. Under all other solvent conditions examined with calorimetry, the two-state model fails even more conclusively (Table 2.4). The existence of an equilibrium intermediate is not supported by our spectroscopic data. However, the spectroscopic and calorimetric data support the presence of two structural blocks or cooperative units. The molecule's overall stability suggests an unusually high degree of cooperativity between these structural blocks, which makes the denaturation appear as a two-state process. The deconvolution of the calorimetric denaturation data yields the two transitions implied by the ratio of van't Hoff and calorimetric enthalpies (Figure 2.9).

In spite of the fact that thermodynamic analysis can generally be applied only to reversible transitions, it can also be employed for partially reversible processes, if some calorimetric criteria are satisfied (Privalov and Potekhin 1986). The formalism of equilibrium thermodynamics can be used to analyze

scanning calorimetry data if the irreversible part of the transition does not exhibit measurable exothermic effects during and after the transition. It was noted that changes in the heat capacity function in this case do not affect the shape of the calorimetric curve, which exhibits a Gaussian shape similar to reversibly denaturing proteins under similar solvent conditions. The heat capacity increment of denaturation often appears unaffected if exothermic effects due to aggregation are absent. Usually these characteristics apply to those proteins that have denaturation transitions that are only partially reversible judging by the calorimetric criteria. In some cases the degree of the reversibility of the denaturation transition strongly depends on the temperature to which the protein has been heated. With a decrease in the maximum temperature to which the protein has been exposed, many proteins exhibit calorimetric profiles like those expected in the absence of an exothermic effect of aggregation. This allowed us to apply thermodynamic analysis, with some restrictions, to obtain denaturation parameters of pelC. In all cases, we tested for sources of irreversibility that might affect the structural and thermodynamic characteristics of its thermal unfolding. According to established criteria, based on Lumry-Eyring models (Lumry and Eyring 1954), pelC can be classified as a type-A irreversible system (Sanchez-Ruiz 1992). Figure 2.12 shows the dependence of the transition temperature and the denaturational enthalpy on scan rate. Both situations show negligible effects, therefore, equilibrium thermodynamic analysis is permissible.

Denaturation was also studied at pH 5 where deviations from the two-state model are most prominent. Under these conditions, differences in the apparent

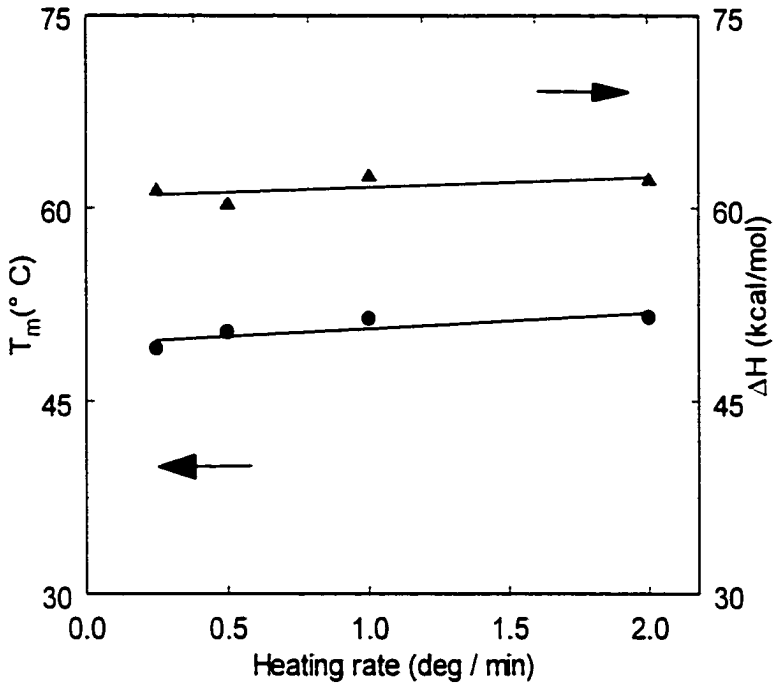


Figure 2.12: Dependence of the midpoint temperature (●) and denaturational enthalpy (▲) on heating rate for pelC thermal denaturation at pH 7.

ΔG° and *m* values that cannot be accounted for by experimental error are obtained using different spectroscopic parameters. The denaturation transitions still appear to be nearly coincident but the analysis of the thermodynamic data obtained indicates that the two-state model fails (Table 2.3). These results would suggest that the free energy change measured from the backbone denaturation is less than that measured from the side-chains, which is not permitted under the constraints of the two-state model. Therefore, the two-state model cannot accurately be applied at pH 5, and the thermodynamic parameters derived from the spectroscopic analysis are not accurate.

Calorimetry suggests that, under these conditions, two separate two-state transitions occur. At pH 5 the individual transitions show the most independence and the individual transition temperatures have the greatest difference. If we follow the thermal denaturation by CD at various gdn-HCl concentrations, we see that there is a nonlinear dependence of the midpoint temperature. At denaturant concentrations within the transition region, the slope of the T_m vs. denaturant concentration becomes more negative, suggesting an enhanced dependence of stability on denaturant concentration. This decrease in stability results from a loss of intramolecular cooperativity, which supports the idea that there are two individual two-state transitions. This is further supported by the data from the near-UV CD that shows a similar break in the midpoint dependence (Figures 2.10 and 2.11).

Several lines of evidence suggest that these anomalies are not due to a partially structured intermediate. These include the lack of secondary and

tertiary structure at pH 2.2 over a wide range of temperatures. Although ANS binds to pelC at pH 2.2, as indicated by a large increase in fluorescence emission, this is due to the formation of ordered secondary structure induced by ANS binding ((Ali et al. 1999), and Chapter 3). If a structured intermediate state can exist at extreme pH conditions, this would imply that a similar structured state could also exist in an equilibrium denaturation situation. The absence of a thermal transition in 2M gdn-HCl also indicates that there is no ordered structure remaining in the unfolded ensemble. In addition, no ANS binding occurs in the gdn-HCl-induced unfolding transition. Lastly, far- and near-UV CD indicate that the loss of secondary structure due to acid denaturation parallels the loss of tertiary structure under identical conditions.

There are three lines of evidence that indisputably preclude the possibility of a two-state process. The ratio of the van't Hoff to calorimetric enthalpy is less than one. Gdn-HCl denaturation at pH 5 using three independent spectroscopic probes yields inconsistent values for $\Delta G°$. Finally, thermal denaturation at pH 5 exhibits a nonlinear dependence of the midpoint temperature on denaturant concentration.

The results discussed can be applied to the protein folding phenomena in a more general sense. The existence of two cooperative units that are not at all obvious in the crystal structure, and are also rather elusive to spectroscopic probes, implies that the results obtained in this study may not be unique. It would be rather surprising if a relatively large protein, such as pelC, folded via a simple two-state mechanism. Consequently, larger proteins that do exhibit

apparent two-state behavior may very well be composed of closely interrelated structural blocks whose independence and existence require a detailed characterization in order to be elucidated.

Stability of pelC. PelC contains three relatively large parallel β-sheets, consisting of six, eight and ten strands per sheet. This protein has a free energy of unfolding that places it among the most stable globular proteins characterized (Creighton 1993). It has been reported that β-sheets attain maximum stability with five or more strands (Sheridan et al. 1979). This may be due to cooperative effects. Strands on the edge of the sheet are only hydrogen-bonded on one side. By increasing the number of strands in a β-sheet, it is possible to increase the overall stability by increasing the cooperative stability of the internal strands. Side-chain interactions probably play a role in the high stability of pelC as well. Stacks of aliphatic amino acids are arranged in a linear fashion, with C_α distances spaced at an optimal 4.5-5.5 Å (Yoder et al. 1993a). These residues form a highly ordered, compact hydrophobic core with many van der Waals contacts. In addition, stacks of aromatic residues are observed in pelC. These rings are stacked in a face-to-face fashion with a slight offset. This allows for optimum orientation of ring hydrogens and π orbitals. Betts and co-workers have shown that aromatic stacks in P22 tailspike protein (TSP) are important for stability and specificity (Betts et al. 1999). It has been suggested that parallel β-sheets are less stable than antiparallel sheets (Richardson 1977). This is because parallel sheets are often found buried and it has been inferred that they

can not tolerate solvent access to their hydrogen bonds. Theoretical calculations indicate that this is in fact due to less effective interchain packing and less favorable interchain electrostatic interactions (Chou et al. 1982). However, in the case of pelC the high stability may be a result of efficient side-chain packing resulting from the β-helix fold. Calorimetric measurements imply the existence of two structural units, each of which gives rise to a simple two-state transition. The high degree of cooperation between these units is a major contributor to the high stability of the protein. It should be noted that the free energy reported ($\Delta G°_{H_2O}$) is the free energy of unfolding and represents the work necessary to unfold the entire protein. The stability of the native state would be equal to the stability of the least stable structural component. In pelC, this value is difficult to measure because there is no apparent symmetry that might indicate two energetically similar units, and the structural blocks have not been identified.

The most striking result from chemical denaturation experiments is the large *m* value. In a survey of 16 proteins for which *m* values have been measured, the largest being a 47 kDa fragment of the P22 tailspike protein (Miller et al. 1998), a range of 1.2-4.4 kcal / mol M has been observed (Ahmad and Bigelow 1982; Ahmad and Bigelow 1986; From and Bowler 1998; Haezebrouck et al. 1998; Hu et al. 1992; Kuwajima et al. 1991; Reid et al. 1998). PelC exhibits an *m* value of 12 kcal / mol M at pH 7, a value 3-fold greater than any observed in the survey. A similarly structured and much larger protein, TSP from phage P22, yields an *m* value of 3 kcal / mol M (Miller et al. 1998). This m value was measured from urea denaturation, which typically yields a lower m value than

gdn-HCl denaturation (Myers et al. 1995). Empirical relationships have been developed that can be used to estimate the m value for gdn-HCl denaturation from experimentally determined m values by urea denaturation (Myers et al. 1995). Thus, for TSP, $m_{gdn\text{-}HCl}$ is expected to be approximately 6.8 kcal / mol M. There are several ways to explain the large value for pelC. From the x-ray structure, there are no significantly exposed loops or elements of secondary structure that would enhance solvent exposure in the folded form. The number of β-strands per sheet may also be responsible for the large *m* value, as the number of strands can influence the cooperativity of the system (Sheridan et al. 1979). Because of the high $\Delta G°$ value and the low midpoint of unfolding (about 1M gdn-HCl), a strong driving force is required to favor the folded conformation. Perhaps the largest contributor to the high *m* value is the strong cooperation between structural units. It should be noted that all the proteins in this survey are less than half as large as pelC, except TSP, which is about twice as large and represents only a fragment of the entire protein. Additionally, due to the relatively large size of pelC, many of its residues are buried when folded. Upon unfolding, there is a large change in solvent-accessible surface area. Using the DSSP program (Kabsch and Sander 1983), the solvent-accessible surface area (ASA) was calculated for both the folded and unfolded forms. Coordinates for an unfolded polypeptide chain with the pelC sequence were generated using β-sheet ϕ and ψ angles with Insight II software (Molecular Simulations Inc., San Diego, CA). In order to include the effects of disulfide bonds, a simple energy minimization was performed with the two native disulfide bonds of pelC intact.

The difference in accessible surface area between the folded and unfolded chain (ΔASA) is 35189Å^2 without disulfides and 30701Å^2 with disulfides intact. It has been shown that ΔASA, *m*, and ΔC_p correlate well with each other for small proteins that exhibit two-state folding (Myers et al. 1995). In the case of pelC, ΔASA is about what one would expect for unfolding of a protein of its size. However, the experimental value for *m* is higher than what one would expect (calculated value = 8.0 kcal / mol M) and ΔC_p (calculated value = 6.0 kcal / mol K) is lower (Myers et al. 1995). Although this correlation has been established for small, two-state proteins (Myers et al. 1995), it is still useful as a guideline in our situation. A recent study of the denaturation of maltose binding protein reported a comparable *m* value of 12 kcal / mol M (Sheshadri et al. 1999). The protein denatures via a two-state mechanism, and therefore, *m* was correlated with its large ΔC_p. In our case, a large *m* is accompanied by a ΔC_p lower than the expected value for a protein the size of pelC. Our results indicate that for proteins such as pelC, which unfold via a mechanism more complex than a simple two-state process, the *m* value may show a different dependence on ΔC_p and ΔASA and/or depend on additional properties.

When the x-ray structure of pelC was solved, it yielded a novel conformation. Denaturation studies thus far imply the existence of a novel folding mechanism. It appears that two structural units are present in pelC and the cooperativity between them is so strong that unfolding at first appears as a two-state process. This high cooperativity may contribute to the remarkable stability of this protein. Kinetic experiments on pelC folding and unfolding are

described in Chapters 4 and 5 and have added to our understanding of the unique folding properties of parallel β-helix proteins.

Chapter 3

A Partially Folded Intermediate Conformation is Induced in Pectate Lyase C by the Addition of 8-Anilino-1-Naphthalene Sulfonate

The work presented in this chapter has been submitted for publication to the journal *Protein Science*. This chapter is identical to the submitted manuscript, except all figures listed as "data not shown" have been included.

3.1 Introduction

8-anilino-1-naphthalene sulfonate (ANS) is a small hydrophobic, charged fluorescent dye. It is frequently used for the investigation of equilibrium and kinetic protein-folding intermediates (Ptitsyn et al. 1990; Semisotnov et al. 1987; Semisotnov et al. 1991). When ANS is bound to a protein in a nonpolar environment, there is a large increase in the fluorescence quantum yield (Stryer 1965; Turner and Brand 1968). ANS is a good probe for protein-folding intermediates, especially molten globules, because their partially structured nature provides access for ANS to bind exposed hydrophobic regions (Ptitsyn et al. 1990), whereas ANS has a very weak affinity for fully unfolded or folded proteins. ANS has also been widely used as a probe for kinetic intermediates in protein folding (Ptitsyn et al. 1990; Semisotnov et al. 1987). Partially folded transient species will bind ANS, and the time dependence of its emission is indicative of the lifetime of the species.

Acid denaturation is a widely used method of inducing protein conformational transitions. The major cause of acid denaturation of proteins is the charge-charge repulsion of basic amino acid side chains that have uncompensated positive charges resulting from the neutralization of carboxylate groups. It has been shown that addition of excess acid to an acid-denatured protein can induce a significant fraction of secondary structure (Goto et al. 1990a). A similar result is obtained upon the addition of neutral salts (Goto et al. 1990b). This has been interpreted as the formation of a molten globule state

stabilized by the binding of anions (Goto et al. 1990a; Goto and Fink 1989; Goto et al. 1990b). It has recently been demonstrated that the addition of ANS to acid-denatured cytochrome *c* also induces the formation of a molten globule (Ali et al. 1999). This occurs because the sulfonate group of ANS behaves as a counterion to the positive charges, reducing the charge density, thus permitting refolding. It has also been reported that ANS acts as a conformational tightening agent to acid-expanded bovine serum albumin and γ-globulin (Matulis et al. 1999). A similar phenomenon occurs in the acid-unfolded form of pectate lyase C (pelC) from *Erwinia chrysanthemi*.

PelC contains a relatively large number of amino acids with basic side chains (Jurnak et al. 1994; Yoder and Jurnak 1995). Of the 353 amino acids, 37 are lysine, arginine or histidine; only 32 are glutamate or aspartate; and the pI is 9. We propose that ANS acts as a counterion that neutralizes some of the positive charges in acid-unfolded pelC, causing subsequent burial of the dye and increased ANS fluorescence. The relatively high affinity of ANS for pelC results from a combination of hydrophobic and ionic interactions. The results we present suggest that caution should be exercised in inferring the existence of folding intermediates in acid denaturation, based upon ANS binding alone.

3.2 Materials and Methods

Materials. PelC was prepared as previously described (Chapter 2). The samples were buffered at pH 2 with 25mM glycine-HCl. ANS was purchased

from Sigma (St. Louis, MO) and recrystallized several times from water (Weber and Young 1964). Nile Red was purchased from Sigma. 2,2-dimethyl-2-silapentane-5-sulfonate was purchased from Cambridge Isotopes (Andover, MA). 2-naphthalenesulfonate was from ACROS Organics (Fairlawn, NJ). Other reagents were of high quality and used without further purification.

Fluorescence spectroscopy. Fluorescence measurements were obtained on an Aviv Instruments model ATF-105 fluorometer (Lakewood, NJ). The excitation wavelength was 350 nm with a 2 nm bandwidth. Emission was monitored from 600-400nm with an 8 nm bandwidth. The temperature was regulated at 25° C by a thermoelectric cell holder. The pelC concentration was 1μM. ANS was used in 200-fold molar excess. A reference blank containing only buffer and ANS was subtracted from each spectrum.

Circular dichroism spectroscopy. All circular dichroism spectra were obtained on a Jasco J-720 spectropolarimeter (Tokyo, Japan). A 0.02-cm path-length cell was used to collect data in the far UV, and a 1.0-cm cell was used to collect data in the near UV. The temperature was regulated at 25° C with circulating water. PelC concentrations were 5μM. Spectra were collected with a scan rate of 100 nm/sec, 1 nm bandwidth, and a 4-sec response time. A blank containing buffer and other reagents was subtracted from all spectra.

Analytical ultracentrifugation. Analytical ultracentrifugation experiments were performed on a Beckman XL-I (Palo Alto, CA). For sedimentation equilibrium experiments, pelC concentrations were 2, 3, and 5μM. The ANS concentration was 100μM. Because ANS has a strong absorbance in the same region as proteins, interference optics were used. For sedimentation velocity experiments, pelC concentrations were 1, 2.5, 5, and 7.5 μM and the ANS concentration was 100μM. The samples were spun at a speed of 60K rpm and data were collected every thirty seconds for two hours. PelC has a partial specific volume of 0.726 cm^3/g. Samples were dialyzed against buffer overnight prior to experiments. Data were analyzed with Origin software (Microcal Software Inc, Northhampton, MA).

3.3 Results

CD and fluorescence of pelC/ANS. PelC, like many proteins, is susceptible to denaturation by acid. ANS fluorescence emission was used to probe the structure of pelC at pH 2. CD spectroscopy was further used to assess any structural changes in acid-unfolded pelC upon addition of ANS.

The fluorescence quantum yield of ANS is dramatically increased when the dye is in a nonpolar environment, as compared to an aqueous environment. Hydrophobic regions in a partially structured protein provide such an environment. Partially unfolded proteins may be present in a protein solution at low pH, in high salt, or in the presence of moderate concentrations of chaotropic

denaturants. Figure 3.1 shows that as the pH of a solution of pelC is lowered below 4, the fluorescence emission of added ANS increases. The emission is 50-fold more intense at pH 2 than at pH 7. A result such as this suggests the presence of a stable hydrophobic core or binding pocket. This was surprising because evidence from CD experiments indicated that pelC was completely unfolded at pH 2 (Kamen et al. 2000). Further CD experiments were conducted to determine if ANS interacts with acid-unfolded pelC.

Figure 3.2 shows the far-UV CD spectra of pelC alone and in complex with ANS at pH 2. The CD spectrum of pelC alone at pH 2 shows the characteristics of an unfolded protein conformation. There is a slight negative shoulder at 220 nm and a very prominent negative band at 200 nm. This is very typical of an unordered protein, which is an ensemble of many conformations. Titration with ANS results in a dramatic change in the shape of the CD spectrum as a complex between ANS and pelC forms. The shoulder at 220 nm evolves into a very distinct band with a broad minimum at 216 nm. The negative band at 200 nm in the unordered protein spectrum, in the absence of ANS, evolves into a positive band with a maximum at about 192 nm upon addition of 50-fold excess of ANS. A CD spectrum of this type is characteristic of β-sheet, which is the dominant fraction of secondary structure in pelC (Figure 3.2b). It appears that the binding of ANS to pelC is responsible for the induction of secondary structure in acid-unfolded pelC. This is a very interesting result because it raises further questions about the usefulness of ANS as a probe for partially folded protein

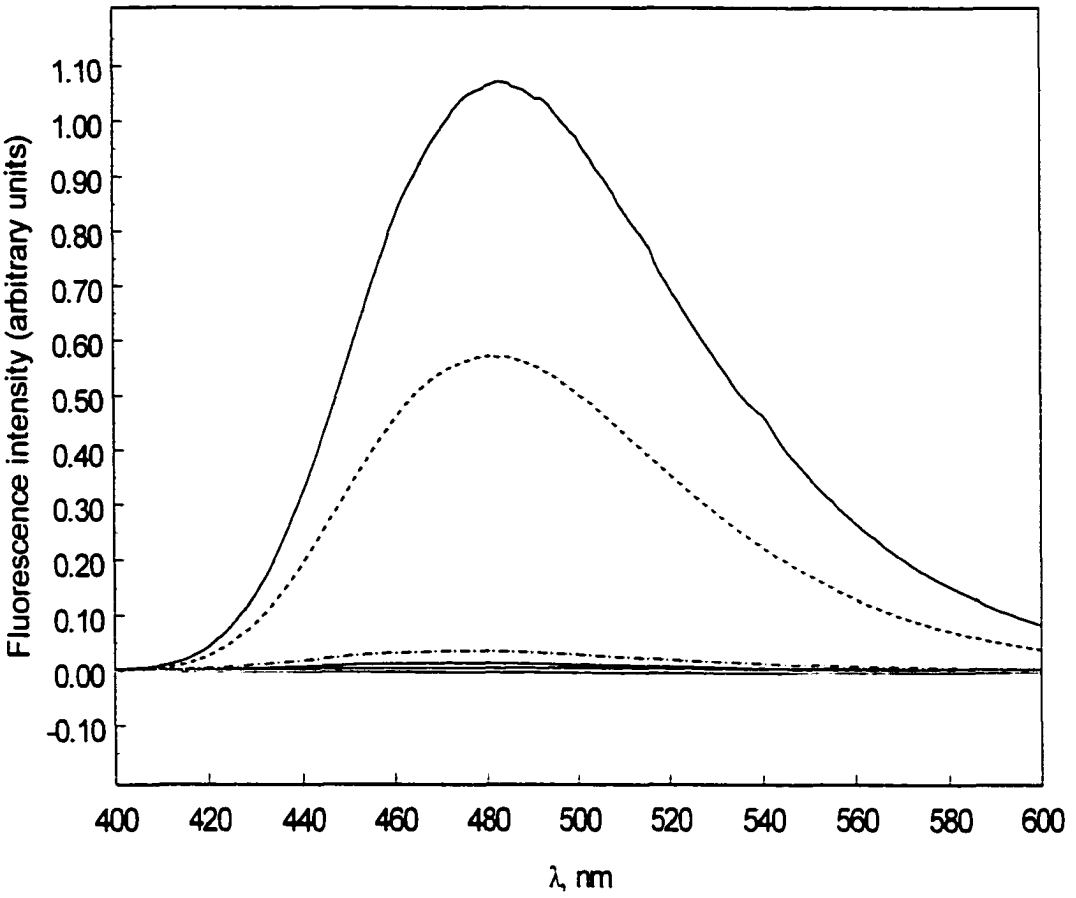


Figure 3.1: Fluorescence emission spectra of pelC/ANS complex at pH 2-9. The cluster of solid lines at approximately zero intensity is pH 4-8, pH 9 (–·–·–), pH 3 (- - -), and pH 2 (——). Protein:ANS ratios were 1:200.

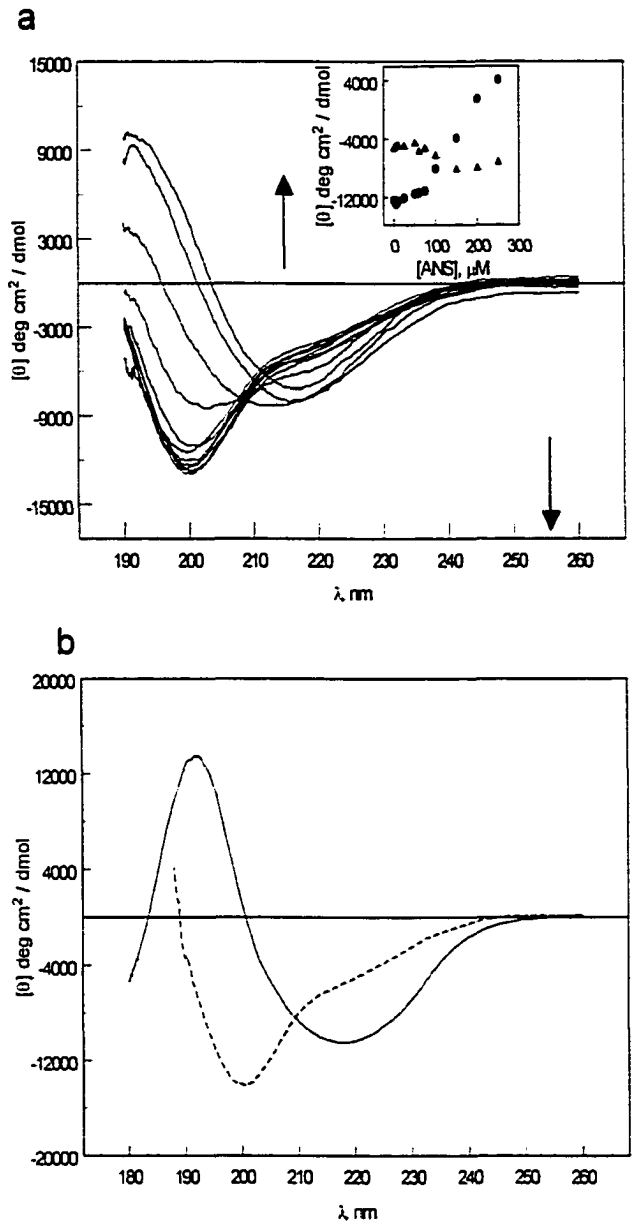


Figure 3.2: *(a)*. Far-UV CD spectra of pelC (5μM) with increasing concentration of ANS. The inset to *A* shows the ellipticity value at 200 nm (•) and 215 nm (▲). PelC:ANS ratios were 1:1, 2, 5, 10, 12, 15, 20, 30, 40, and 50. Arrows indicate the direction of increasing ANS concentration. *(b)*. Far-UV CD spectra of pelC at pH 2 (---) and pH 7 (——) without ANS.

conformations. If in fact the probe that is used to observe these structures is actually inducing an intermediate conformation, other independent probes are required to substantiate the existence of partially folded intermediates.

The far-UV CD spectrum of a protein can be deconvoluted to estimate the fractions of secondary structure components (Greenfield and Fasman 1969). An analysis of the secondary structure of pelC under several conditions is shown in Table 3.1. The far-UV CD spectrum of pelC in the native state in water was compared with the acid-unfolded state at pH 2 and the acid-unfolded state at pH 2 with 50-fold excess of ANS. An analysis of the CD data indicates that β-sheet and α-helix are stabilized (their fractions increase) and a significant fraction of unordered structure (~50%) is lost upon the addition of ANS. Compared to native pelC, the ANS-induced intermediate has about 60% higher β-sheet content as estimated by CD. In fact, the CD-derived estimates of α-helix and β-sheet fractions for the ANS-induced form agree better with the x-ray data for the native structure than do the CD-derived data for the native protein. The poor estimates from CD for the native structure may be attributable to the absence of proteins with a similar β-sheet architecture in the reference proteins used in the CD analysis. Alternatively, contributions of aromatic side-chains in the native protein that are absent in the ANS-induced conformation may be responsible for the poor analysis of the native structure by CD. In any case, the analysis is consistent with our proposed model.

Near-ultraviolet CD was also used to study the pelC/ANS complex. It has been shown that a significant induced CD occurs in the near-UV spectrum when

Table 3.1: Fraction of secondary structure determined by CD or x-ray crystallography for pelC*

	α-helix	β-sheet	Turns	Unordered
Native (CD)	0.243	0.213	0.220	0.325
Native (DSSP)	0.088	0.320	0.079	0.513
pH 2/no ANS	0.077	0.158	0.125	0.639
pH 2/1:50 ANS	0.095	0.360	0.234	0.311

* Secondary structures calculated by CD are an average from three methods (Provencher & Glöckner, 1981; Sreerama & Woody, 1993; Johnson, 1999)

Secondary structures determined from x-ray crystallography are from the DSSP method (Kabsch & Sander, 1983)

ANS is bound to model β-sheet peptides at alkaline pH (Sato and Woody 1980). ANS also displays an induced CD when bound to bovine serum albumin (Anderson 1969; Daniel and Yang 1973) and NAD-dependent dehydrogenases (Ivanov and Nagradova 1974; Vanek 1975). Induced CD is also observed in the near-UV CD of the pelC/ANS complex (Figure 3.3). The CD spectrum of the native conformation of pelC displays a maximum at 280 nm (Figure 3.3b). In the complex with ANS, the maximum at 280 nm is lost, but new maxima at 260 and 340 nm appear. These features are absent in proteins such as pelC that lack extrinsic chromophores. However, ANS has absorption bands at the same wavelengths (inset to Figure 3.3a). Therefore, these bands arise from induced CD of ANS. ANS is a molecule that is optically inactive in solution, but upon binding to a protein can give rise to CD bands. This is interesting for several reasons. Because ANS is bound to the protein matrix, the electronic transitions can couple with protein chromophores, thus allowing the transitions to be observed. It demonstrates that ANS is bound to the protein in a chiral environment in such a way that its conformation is well defined. For example, the aromatic rings are buried in a hydrophobic region. It is also interesting from a more fundamental perspective in that optically inactive probes can display optical activity when bound to a protein matrix. It suggests that the binding site favors a particular conformation of the ligand. This can potentially provide information regarding the structure of the binding site. It can also provide clues as to how the ligand is distorted upon binding. This could lead to the more efficient design of synthetic ligands or inhibitors.

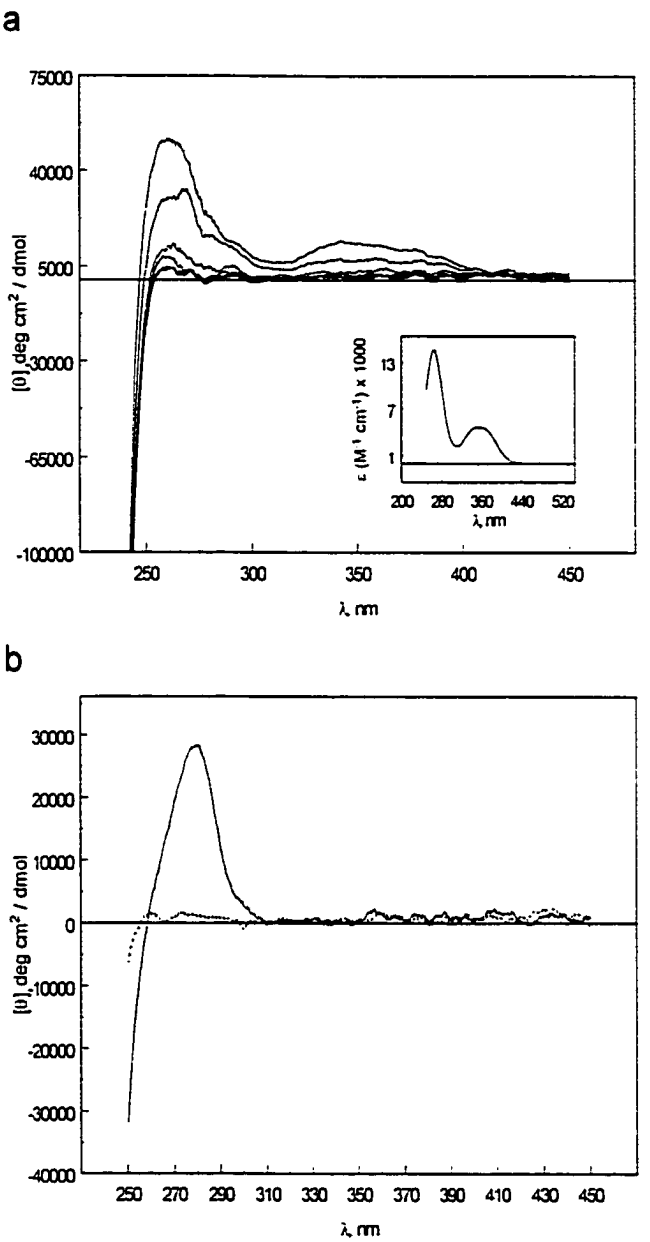


Figure 3.3: *(a)*. Near-UV CD spectra of pelC (5μM) with increasing concentration of ANS. PelC:ANS ratios are 1:1, 5, 10, 15, 25, and 40. The inset to *A* shows the absorption spectrum of ANS in water. *(b)*. Near-UV CD spectra of pelC at pH 2 (----) and pH 7 (——) without ANS. In both figures, the molar ellipticity is based on the protein concentration.

Neutral and anionic reagents, and pelC. Some experiments were done to probe the nature of the interaction of ANS with pelC. Nile Red is an uncharged dye similar in structure to ANS and has previously been used as a neutral control for the induction of structure in cytochrome *c* (Ali et al. 1999). Nile Red shows no effect on the CD spectrum of pelC at pH 2 (Figure 3.4a). The pH dependence of the fluorescence spectrum of Nile Red from pH 8-2 shows two regions where the fluorescence changes sharply (Figure 3.4b). This suggests that at pH 2, the dye is not neutral but has a charge of +2. This is consistent with the chemical structure of Nile Red (Figure 3.4c) which indicates that there are two nitrogens capable of being protonated at low pH. This fact, along with the low solubility of the dye in aqueous buffer (Sackett and Wolff 1987) prevented any conclusions from being made regarding the structure-inducing ability of uncharged ANS analogs on pelC.

It has been demonstrated that ion pair formation is responsible for the interaction of ANS with some proteins (Matulis and Lovrien 1998). NaCl is capable of inducing structure in pelC, as demonstrated by the far-UV CD spectra (Figure 3.5a). However, it requires about 100 mM salt, whereas only 100μM ANS is required to achieve a comparable CD spectrum. We tested whether or not binding of ANS is enhanced as a result of combining hydrophobic interactions with electrostatic interactions. 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) is a very simple molecule containing a sulfonate group, as in ANS, and a short aliphatic chain (Figure 3.6a). This reagent is capable of inducing structure in pelC, but it requires about 8mM DSS. This suggests that ionic interactions play a

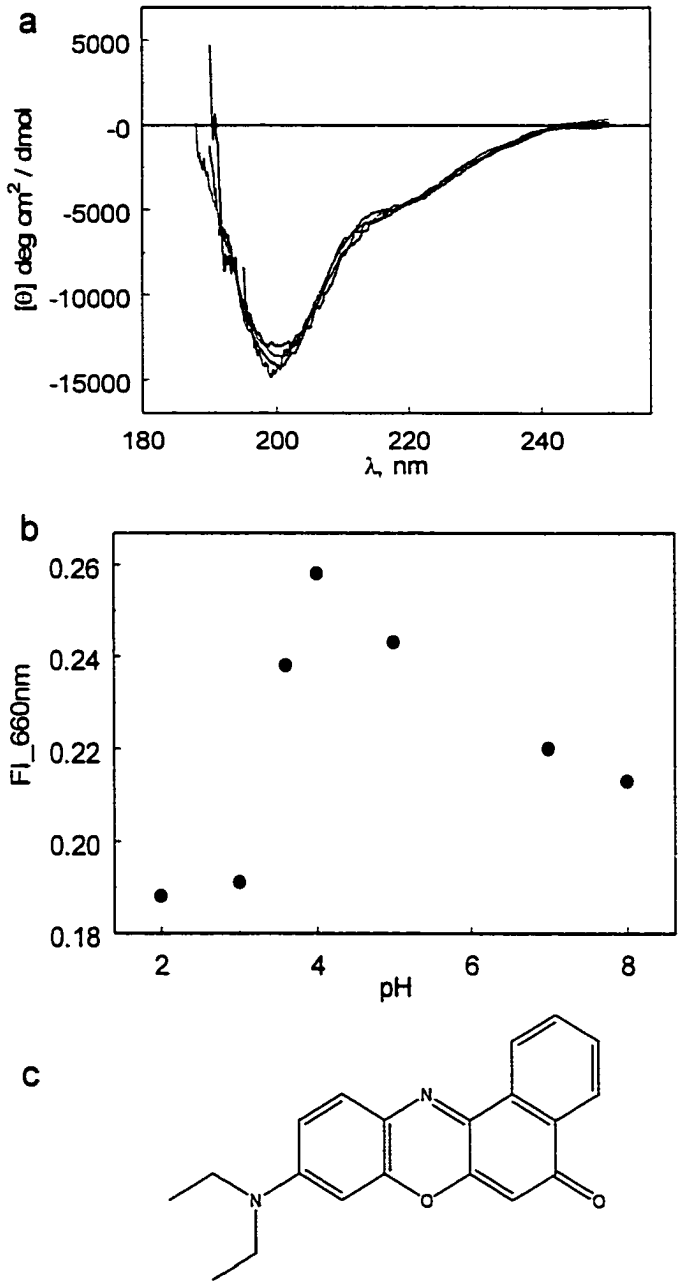


Figure 3.4: Interaction of Nile Red with pelC at pH 2. (*a*) No secondary structure is induced in pelC with Nile Red concentrations of 0, 50, 100, and 200 μM. (*b*) Nile Red fluorescence emission as a function of pH shows a change due to titration of nitrogens. (c) The chemical structure of Nile Red.

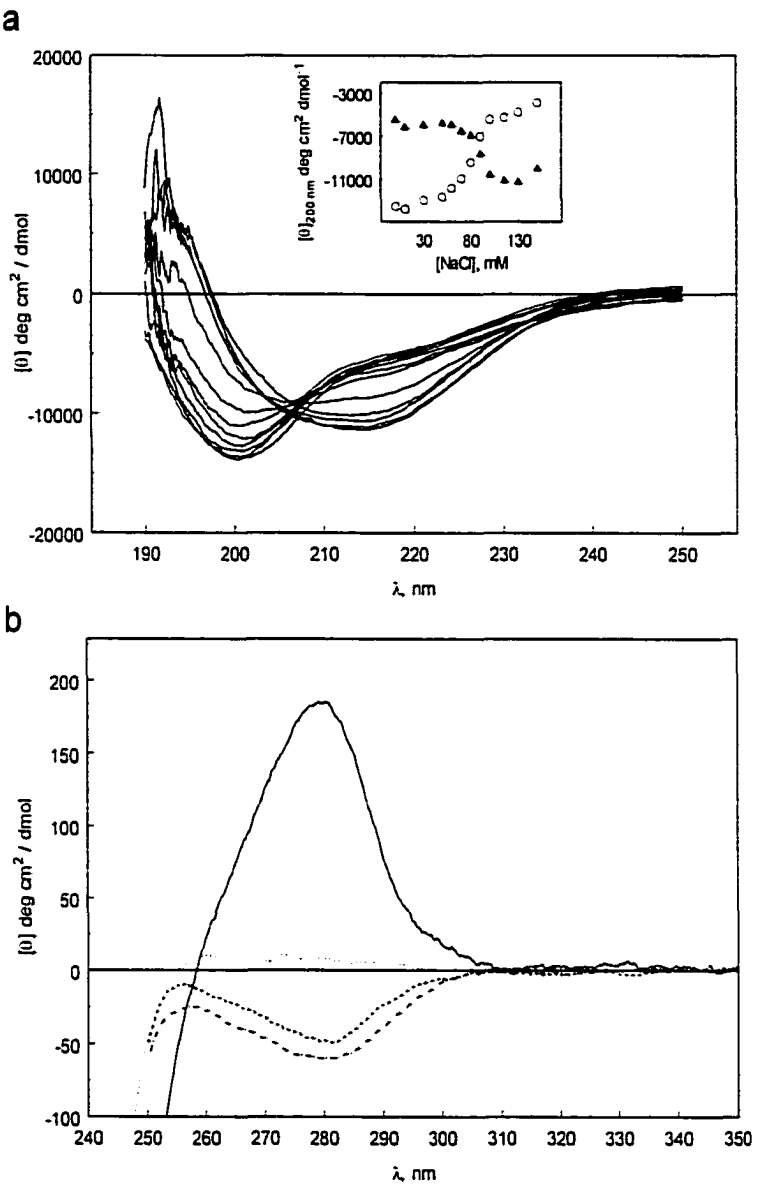
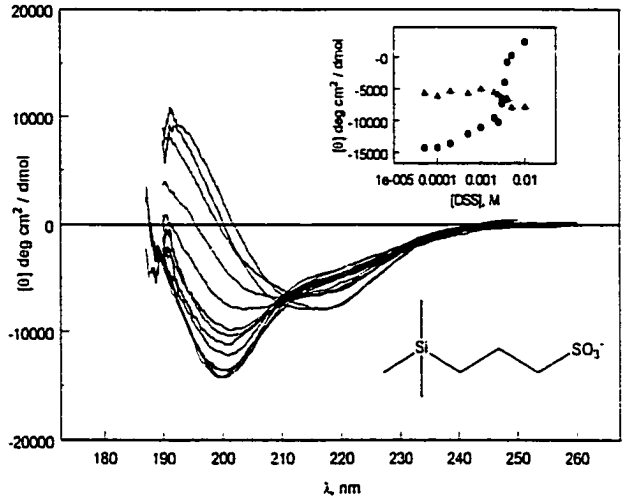


Figure 3.5: *(a)*. Far-UV CD spectra of pelC at pH 2 in the presence of NaCl (0, 10, 30, 50, 60, 70, 80, 90, 100, 115, 130, 150 mM looking from bottom to top in the short wavelength region). The inset shows the ellipticity values at 200 nm (o) and 215 nm (▲). *(b)*. Near-UV CD spectra of pelC at pH 7(——) and pH 2 (····) without NaCl, and pH 2 with 100 (----) and 150mM (—·—) NaCl.

a



b

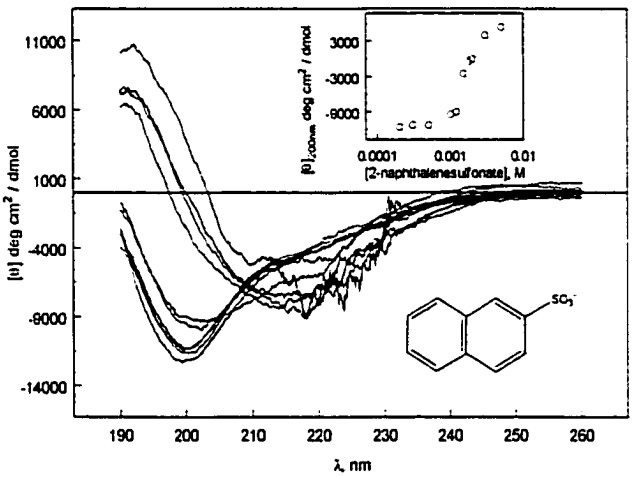


Figure 3.6: *(a)*. Far-UV CD spectra of pelC at pH 2 in the presence of increasing amounts of DSS (0.05, 0.1, 0.2, 0.5, 1, 2, 2.5, 3, 3.5, 4, 5, and 10 mM) (the structure is shown as an inset). The inset to *A* shows the ellipticity value at 200 nm (•) and 215 nm (▲). *(b)*. Far-UV CD spectrum of pelC (5μM) at pH 2 with increasing amounts of 2-naphthalenesulfonate (0.05, 0.2, 0.3, 0.5, 1, 1.2, 1.5, 1.8, 2, 3, and 5 mM). The inset shows the ellipticity at 200nm (○). The structure of 2-naphthalenesulfonate is also shown as an inset. The poor signal to noise ratio near 220 nm is a result of strong absorption by the naphthalene.

strong role in the structure formation in pelC but also that they are not the whole driving force. The ability of 2-naphthalenesulfonate to induce structure in pelC was investigated. This molecule contains a sulfonate group but the amount of hydrophobic surface area is reduced compared to ANS. The hypothesis was that a molecule with a negative charge, but reduced hydrophobic surface area, would require higher concentrations to induce structure in pelC compared to ANS, but not so high as using NaCl alone. Induction of secondary structure appears at about 1mM 2-naphthalenesulfonate concentration (Figure 3.6b). This supports the notion that a combination of hydrophobic and electrostatic interactions are responsible for the formation of structure in pelC.

8-toluido-1-naphthalenesulfonate is structurally very similar to ANS, differing only in the addition of a methyl group to the phenyl ring (inset to figure 3.7a). This molecule has an identical structure-inducing affect as ANS does on pelC. Interestingly enough, so does the 6,2 isomer of TNS, which is quite different in its arrangement of aromatic rings, compared to ANS (Figure 3.7b). These experiments suggest that there is nothing unique to the structure of ANS that is responsible for the induction of a conformational change in pelC.

It is important to point out that the induced intermediate conformation is not a molten globule. Molten globules lack tertiary structure. This is demonstrated by a near-UV CD of essentially zero. Figure 3.5b indicates a significant near-UV CD intensity when pelC is at pH 2 in 100 and 150mM NaCl. If the induction of structure occurs through a similar mechanism with NaCl as it

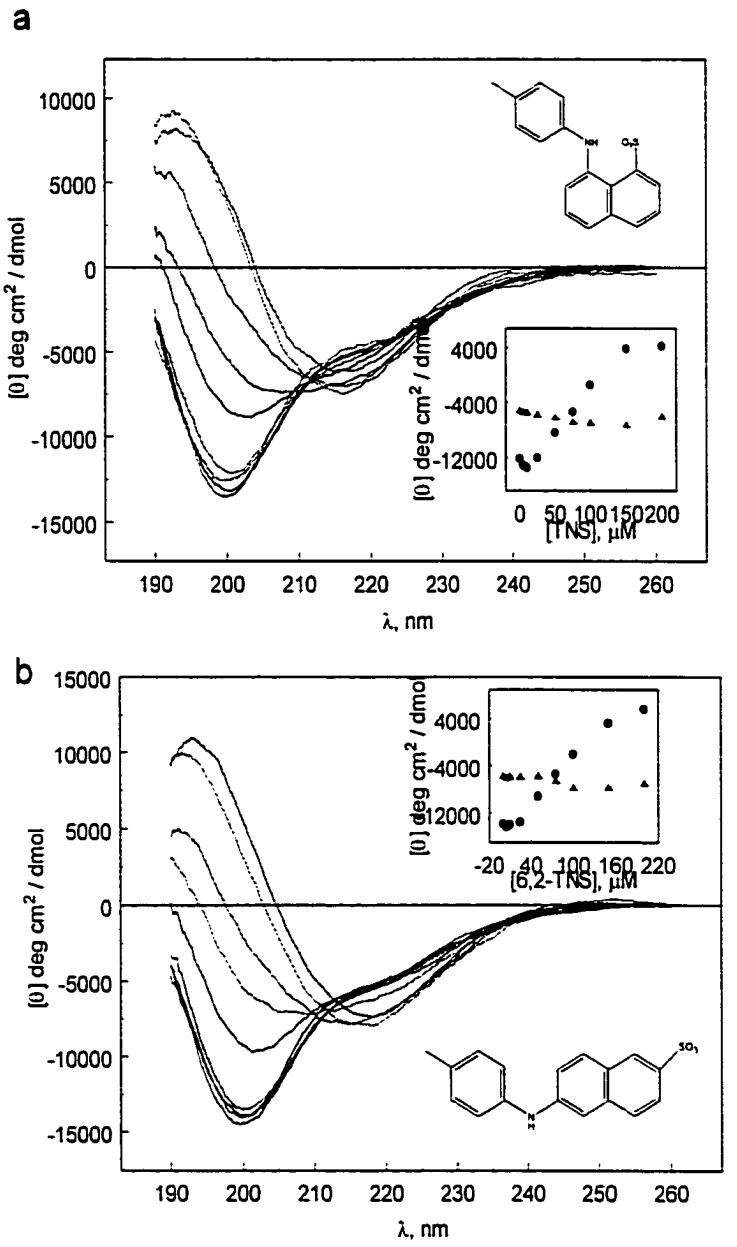


Figure 3.7:CD spectra of pelC in the presence of increasing amounts of 8,1-TNS (*a*), and 6,2-TNS (*b*). In both cases, the dye concentrations are 0, 5, 10, 25, 50, 75, 100, 150, and 200 mM and the pelC concentration is 5 μM. The insets show the ellipticity at 200 (●) and 215 nm (▲).

does with ANS, then we can assume that the intermediates are similar in nature. Native pelC shows a positive CD amplitude at 280nm but the chloride-induced intermediate is negative at 280nm. It is possible that the ANS-induced structure also has a negative CD from aromatic side-chains but it is not observed due to the stronger CD intensity of bound ANS. It is also not clear if the intermediate tertiary structure is native or not. It is possible that the negative CD band observed in high salt is due to aromatic side chains in a non-native conformation. It is also possible that these aromatic side-chains that are structured in the intermediate are native-like and make a negative contribution to the overall positive near-UV CD of the native protein.

Sedimentation analysis. Spectral features such as those observed for pelC in the presence of ANS might be the result of intermolecular association. Sedimentation equilibrium experiments were therefore performed to assure that pelC is still monomeric under these solution conditions. The results indicated that pelC is indeed behaving as a monomer. However, when ANS is in the solution the apparent relative molecular mass ($M_{r\,app}$) is lower than, but within the estimated error of, the expected monomer $M_{r\,app}$ of 37.7 kDa. The data suggest that the low apparent molecular mass is due to non-ideality in the solution. The observed molecular mass decreases with increasing rotor speed and concentration (Table 3.2). For pelC at pH 2, $M_{r\,app} = 38.0 \pm 2.6$ kDa. For pelC at pH 2 with ANS, $M_{r\,app} = 34.8 \pm 7.9$ kDa (reported errors are 95% confidence intervals). The low $M_{r\,app}$ results because charged protein molecules repel each

Table 3.2: Relative molecular mass of pelC complexed with ANS determined by analytical ultracentrifugation

		Rotor speed x 1000		
		18	25	33
[PelC], (μM)	2	44.2 (37.8,-2.9e-6)	30.7 (37.6, 1.0e-5)	25.2 (37.7, 3.6e-5)
	3	50.1 (N/D)	30.1 (37.6, 2.2e-5)	23.8 (37.2, 1.7e-5)
	5	49.5 (N/D)	30.6 (37.7, 4.2e-6)	29.3 (32.1, 2.1e-5)

Numbers in parentheses indicate the apparent molecular mass determined using a second virial coefficient, and the coefficient used. A mass was not determined for 3 and 5 μM at 18000 RPM using a second virial coefficient.

other, which counters the tendency of the protein to sediment towards the outside of the cell. The result is an $M_{r\,app}$ that is lower than expected. Although the mechanism of structure reformation in pelC involves neutralization of charges, the residual charge on the protein is likely to be sizeable. The data fit well to a single-species model (Figure 3.8). Addition of a second virial coefficient to the analysis yields the expected molecular mass (Table 3.2). It was not possible to fit the data to a single-species model using the dimer or tetramer molecular mass. It was also not possible to fit the data to a monomer-dimer, or monomer-tetramer equilibrium model. These observations suggest that aggregation is not likely to be a problem. To decrease or eliminate non-ideality, the solution conditions are usually adjusted to have an ionic strength of about 150mM. This was not possible in this case because addition of salts such as 50 mM NaCl to the ANS-containing solutions resulted in precipitation of the protein.

Sedimentation velocity experiments were also carried out on native pelC (50 mM $NaPO_4$, 100 mM NaCl, pH 7), pelC at pH 2 (50 mM Gly-HCl), and pelC at pH 2 in the presence of 100μM ANS (Figure 3.9). A g(s*) analysis (Stafford 1992) of the sedimentation data for native pelC indicated the presence of a single species with $M_{r\,app}$ of about 33 kDa and s* (the apparent Svedberg coefficient) of 3.5 S. For pelC at pH 2 a single species is also observed with $M_{r\,app}$ of ~40 kDa, and s* of 1.8 S. When ANS is present, two species are observed. One has a rather expanded conformation with $M_{r\,app}$ of 41 kDa and s* of 2.3 S. The other is somewhat more compact with $M_{r\,app}$ of 28 kDa and s* of about 6 S. The two species are assigned to: (1) pelC unfolded without ANS bound and (2) pelC

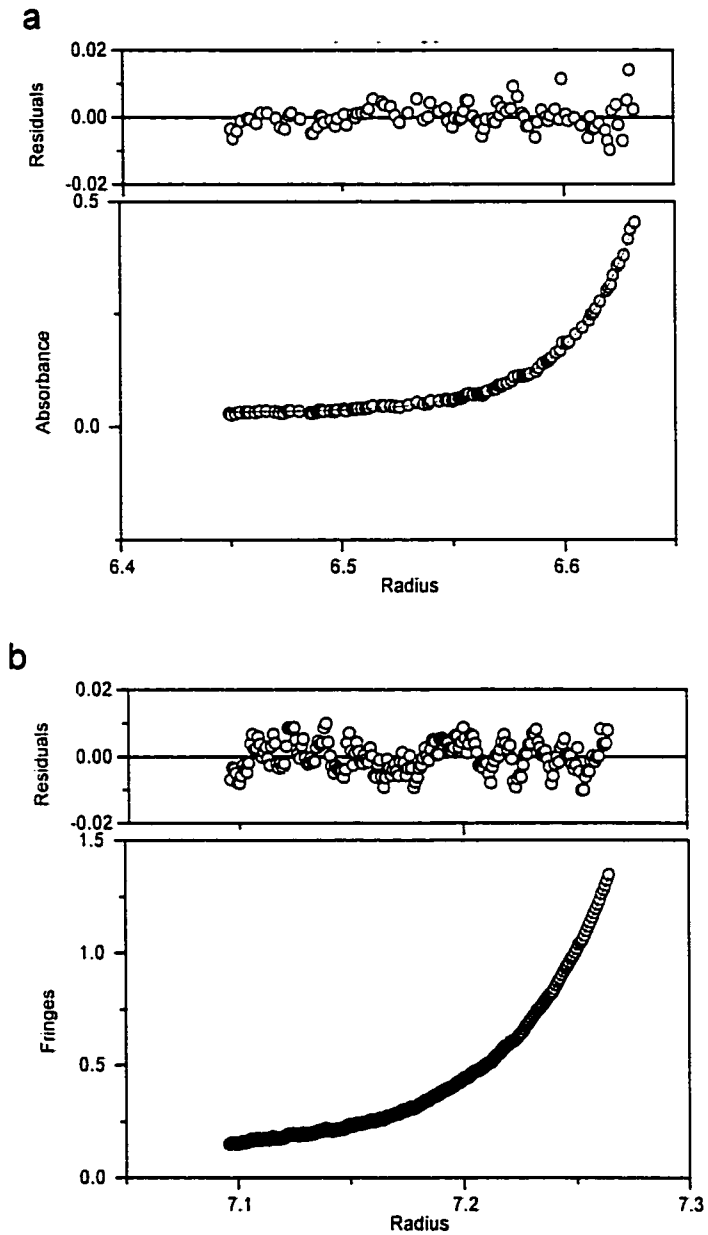


Figure 3.8: Sedimentation equilibrium data for pelC at pH 2 without ANS (*a*) and with 100μM ANS (*b*). Both sets of data are fit to a single-species model. Absorption optics were used for samples without ANS, and interference optics were used for samples containing ANS. Symbols represent the measured signal and the lines are a fit to a single-species model. The rotor speed in these plots was 25000 rpm.

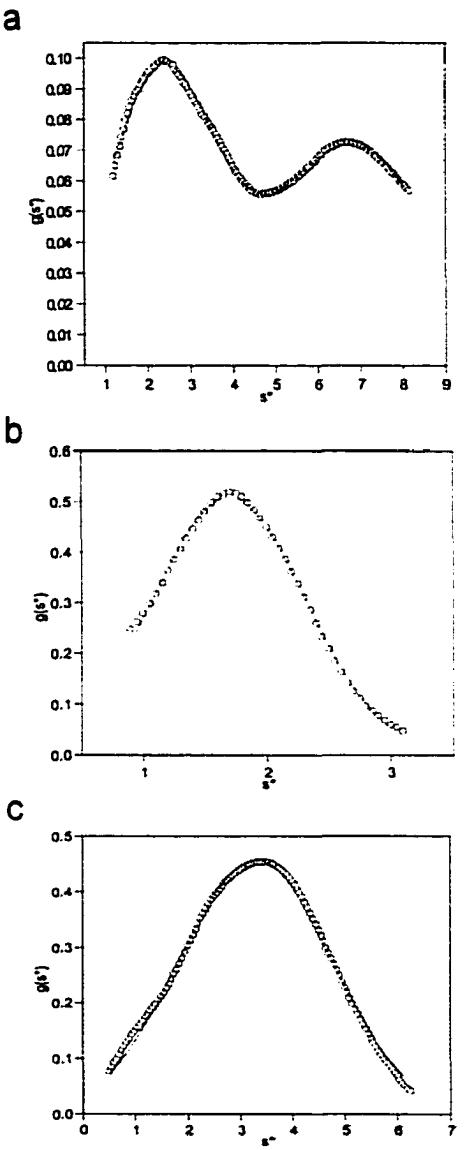


Figure 3.9: Sedimentation velocity data for pelC at pH 2 with 100μM ANS (*a*). The symbols are the measured g(s*) values, (---) are the gaussian fits to each species, and (——) is the sum of the two gaussian peaks. (*b*) Data for pelC at pH 2 and (*c*) pH 7. The lines and symbols are the same as above.

refolded to the intermediate with ANS bound. These data indicate direct observation of two distinct species of similar mass and different shapes. This strongly supports the idea of a conformational change in pelC upon addition of ANS.

Overall, the sedimentation data support the idea that pelC is monomeric when refolded by ANS. The observed spectral features are not the result of aggregation.

Discussion

A significant fraction of β-sheet structure is induced, along with a small increase in the fraction of α-helix, when acid-denatured pelC is complexed with ANS. Extension of the polypeptide chain in the process of acid denaturation is the result of the repulsion of excess positive charges from basic amino acid side-chains. For pelC at pH 2, the net charge on the protein is +37. The negative sulfonate group of ANS partially neutralizes the basic groups, which results in stabilization of native-like secondary structure. The increased fluorescence emission of ANS results from the burial of this dye in the hydrophobic regions of the protein upon induction of folding. The large fraction of native-like secondary structure and the apparent lack of native-like aromatic structure in the pelC-ANS complex indicate that a partially structured intermediate is stabilized in the complex. Our results indicate that this phenomenon can also occur with the addition of salt. However, it requires almost 1000-fold higher concentrations of

NaCl than ANS to observe similar results. We conclude that the interaction between ANS and acid-unfolded pelC is both electrostatic and hydrophobic in nature and this accounts for the relatively high affinity of ANS for acid-unfolded pelC.

These CD results are consistent with the findings of Ali *et al.* (1999), who showed that ANS binding to the acid-denatured form of oxidized cytochrome *c* leads to refolding of the protein. Taken together, the two studies strongly suggest that ANS is a non-innocuous probe that can significantly perturb the unfolded state of a protein, particularly if the protein has a sizeable positive charge as in the cases of cytochrome *c* and pelC. Thus, reports of molten globule formation detected by ANS fluorescence enhancement must be considered suspect unless corroborated by other types of evidence.

Chapter 4

The Folding Kinetics of the Protein Pectate Lyase C Reveals Fast-Forming Intermediates and Slow Proline Isomerism

4.1 Introduction

In recent years, significant advances in the field of protein folding have been made. These advances have come in the form of insightful theoretical models (Dill and Chan 1997; Dinner et al. 2000; Dobson et al. 1998) and information-rich experimental advances (Brockwell et al. 2000; Clarke and Itzhaki 1998; Dobson and Karplus 1999). Protein folding is currently viewed from the perspective of free energy surfaces in which the delicate balance between enthalpy and entropy contribute to guide an unfolded polypeptide chain along a free energy surface towards the native, active biological entity. Developments and advances in instrumentation and techniques have led to observation of folding reactions in the milli-, micro-, and nano-second time scales (Akiyama et al. 2000; Chan et al. 1997; Chen et al. 1998; Gruebele et al. 1998; Takahashi et al. 1997; Yeh et al. 1997). These achievements over the last several years have led to a shift in the way one thinks about protein folding (Dill and Chan 1997) and a deeper understanding of the physical phenomenon of folding.

Small, single-domain proteins that fold via a two-state mechanism have been the focus of many recent studies (Capaldi and Radford 1998; Jackson and Fersht 1991a; Kuhlman et al. 1998a; Kuhlman et al. 1998b). These studies have provided much of the basis for understanding the fundamental properties of protein folding. In this chapter an initial characterization of the folding kinetics of pectate lyase C (pelC) from *Erwinia chrysanthemi* is presented. PelC is a member of the class of proteins containing a parallel β-helix folding motif (Yoder

et al. 1993a; Yoder et al. 1993b). PelC was the first protein observed to possess such a backbone topology (Yoder et al. 1993a). The majority of the regular secondary structure is composed of parallel β-sheets. The architecture of the sheets results in the backbone winding up to form a large helix composed of β-sheets. There are two disulfide bonds in pelC that remain intact throughout all of the experiments reported here. In addition, there are 12 proline residues. One of these prolines, Pro220, is involved in a *cis* peptide bond (Yoder et al. 1993a).

The *in vitro* folding of this class of proteins has been largely untouched. Several studies on the *in vivo* folding, misfolding and assembly of the tailspike protein from bacteriophage P22 have been published (Betts and King 1999; Haase-Pettingell and King 1997; King et al. 1996; Robinson and King 1997). In addition, some *in vitro* work has been done on the tailspike protein (Fuchs et al. 1991; Miller et al. 1998). We propose that because of the relatively simple topology of the parallel β-helix, these proteins will be excellent folding models, yielding relatively simple and easily understood mechanisms. It is also important to understand the nature of the folding of these proteins because amyloid fibrils may be assembled from β-helical precursors (Lazo and Downing 1998).

The study presented represents a detailed characterization of the *in vitro* kinetic folding mechanism of a parallel β-helix protein. Evidence is presented that indicates that the folding of pelC is dominated by the slow *cis-trans* isomerization around X-pro peptide bonds. Two slow kinetic phases have been assigned to proline isomerization. One fast folding phase is also observed that corresponds to the rapid and highly cooperative formation of the secondary

structure. There are two kinetic phases that have been assigned to the formation of ordered side-chain structure. A folding model of pelC is presented. With the exception of complications arising from proline isomerization, the model presented is relatively simple, consistent with the simple topology of the parallel β-helix.

4.2 Material and Methods

Materials. All chemicals purchased were of the highest quality and purity. MES (2-[N-morpholino]ethanesulfonic acid) was >99.5% pure and purchased from Sigma Chemicals (St. Louis, MO). Ultra-pure guanidine-HCl (gdn-HCl) was purchased from ICN Biomedicals (Aurora,OH). Sodium chloride (99.9%) was purchased from Fischer Chemicals (Fair Lawn, NJ). PelC was prepared as described in Chapter 2.

Unfolding kinetics. Unfolding of pelC was initiated by manually diluting a stock solution of pelC into a concentrated solution of gdn-HCl in a cuvette thermostated at 25° C. The dead-time for this procedure was between 15 and 20 seconds. Unfolding was monitored by CD at 218nm using a Jasco J-720 CD spectropolarimeter (Tokyo, Japan). A 0.1-cm path-length cell was employed for all measurements. The temperature of the cell was maintained by circulating water from a NES-LAB RTE 110 water bath. PelC concentrations were 5-10 μM. Kinetic traces utilized the following parameters: 0.5 – 1 second time resolution,

0.5 – 1 second response time, and a spectral band width of 2.0 nm. Solution conditions were 25mM MES, 15mM NaCl, pH 6, and 1.0 – 5.0M gdn-HCl. All kinetic experiments were fit to the following model:

$$\theta = \sum_{i=1}^{n} a_i * \exp(-k_i * t) + c \qquad (1)$$

θ is an observable parameter (CD, fluorescence,...), a_i is the amplitude of the ith phase, k_i is the rate constant of the ith phase, t is the time, and c is the equilibrium value of the observed property.

Folding kinetics. The folding of pelC was followed by either manual mixing or the stopped-flow technique. For manual mixing experiments, a mixing device similar to that used by Kuwajima et al. (1985) was employed. A rectangular cuvette holder was made. The holder was designed so that coolant could be circulated to regulate the temperature. In addition, a magnetic stirring device was installed (Starna Cells, Atascadero, CA). The result was a thermostated cuvette holder capable of holding a 1-cm path-length cell with a rotating stir bar inside. Dead times for this device were determined by measuring the time to mix a solution of 0.1% (+)-pantoyl lactone in 5 M gdn-HCl, 25mM MES into 25mM MES. The CD at 220 nm was followed as a function of time. The time it took for the CD signal to reach a constant value was taken as the mixing time. In all cases, additional time was allowed for the instrument response. The resulting dead-time was four seconds. In cases of high denaturant concentration or low temperature when the

folding is slow, a resolution of 5 seconds with a 4-second response time was used. PelC concentrations were 1 μM in 5mM MES, 50mM NaCl for data collected at 218nm. CD in the near UV was observed at 277nm using a pelC concentration of 5μM.

To observe the earliest phases of the folding process, stopped-flow mixing was used. A Bio-Logic SFM-3 stopped flow module (Claix, France) was interfaced with a Jasco J-720 CD spectropolarimeter. The stopped-flow device was equipped with a high-density solution mixer and a small-volume syringe for the sample. Solutions of pelC in 5M gdn-HCl, 25mM MES, 50mM NaCl, pH 6 were equilibrated to 25° C in the reservoir of the stopped-flow device. Dilutions were made into identically buffered solutions containing the necessary concentration of gdn-HCl in order to give the desired final gdn-HCl concentrations by a 50-fold dilution. A 0.2 or 1-cm flow cell was employed. Dead-times were 12 msec. Data were collected at 20 msec intervals with a response time of 16 msec and a bandwidth of 2.0nm. Multiple transients were averaged (25-30). For all CD measurements, pelC concentrations of 5μM were used.

Temperature dependence of folding. The temperature dependence of the rate constant was determined for the slow folding phases. Activation enthalpies ($\Delta H^{\ddagger}$) were calculated using Eyring's equation from absolute rate theory (Eyring 1935):

$$\ln (k_f / T) = \ln (k_B / h) + \Delta S_f^{\ddagger} / R - \Delta H_f^{\ddagger} / RT \qquad (2)$$

In the above equation k_f is the folding rate constant, T is the absolute temperature in Kelvin, k_B is Boltzmann's constant, *h* is Planck's constant, $\Delta S_f^{\ddagger}$ and $\Delta H_f^{\ddagger}$ are the transition state entropy and enthalpy changes for folding, respectively, and R is the gas constant. The folding rate constants for the slowest phase were measured with far-UV CD. Folding was measured in gdn-HCl concentrations of 0.1 – 0.4 M at temperatures from 5-25° C. The activation enthalpy was determined for this phase using the extrapolated rates in 0M gdn-HCl at each temperature. These rates were fit to equation 2, yielding activation parameters for folding. The activation enthalpy for the intermediate slow phase was obtained using the P220A mutant as described in Chapter 5. Fluorescence emission spectroscopy was used to monitor folding from a jump in the pH from 2-6.

Double-jump folding assay. Double-jump assays to test for the presence of slowly unfolding populations were performed for the slow phases of pelC folding. For the slowest phase, CD and fluorescence spectroscopies were used. PelC was unfolded in 5M gdn-HCl in 25mM pH 9.5 Gly-NaOH at 0° C. Refolding was initiated after various delay times by 50-fold dilution of the unfolded pelC solution into 50mM pH6 MES, 50mM NaCl at 25° C. The final concentration of pelC was 1μM. A 1-cm path-length cell was used. Refolding amplitudes were determined by fits of the data to equation 1. For the intermediate phase, only fluorescence was used because this method yielded better signal-to-noise ratios.

Folding catalyzed by peptidyl-prolyl isomerase. Human peptidyl-prolyl isomerase (also known as PPI or cyclophilin) was purchased from Sigma Chemical Co. (St. Louis, MO). Because the activity of PPI is significantly reduced in solutions containing gdn-HCl, folding was carried out by dilution of pelC unfolded in 9M urea at pH 8.0 (50 mM Tris) to 0.3 M urea. Folding was followed by fluorescence emission spectroscopy. A final PPI concentration of 1 μM was used and included in the refolding buffer.

4.3 Results

Unfolding kinetics of pelC. Unfolding of pelC was monitored by far-UV CD. In all cases examined, the unfolding kinetics fit to a single-exponential model (Figure 4.1). This suggests that unfolding is composed of a single process. Previous studies of the denaturation of pelC at equilibrium suggested that two energetic blocks are present (see Chapter 2). This is not observed in the kinetic unfolding experiments, perhaps because the thermodynamic stabilities of the structural blocks are similar, or because the thermodynamic driving force for unfolding is so great that the two phases are not resolved. One interesting observation is made, however. The dependence of ln k_u on denaturant is nonlinear (Figure 4.2). Similar observations have previously been interpreted as reflecting a change in the structure of the transition state with a change in denaturant concentration (Matouschek et al. 1995; Oliveberg 1998; Otzen et al. 1999; Silow and Oliveberg 1997).

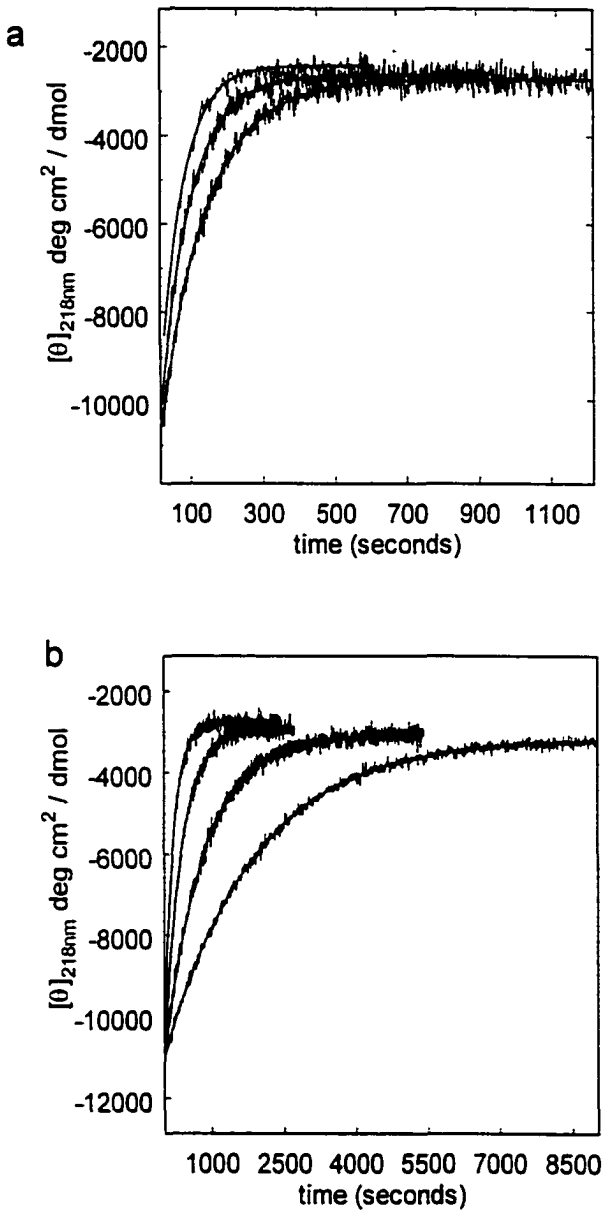


Figure 4.1. Unfolding kinetics of pelC monitored by CD at 218nm at pH 6. *a.* Unfolding (from right to left) in 4, 4.5, and 5 M gdn-HCl. *b.* Unfolding of pelC (from right to left) in 2, 2.5, 3, and 3.5 M gdn-HCl.

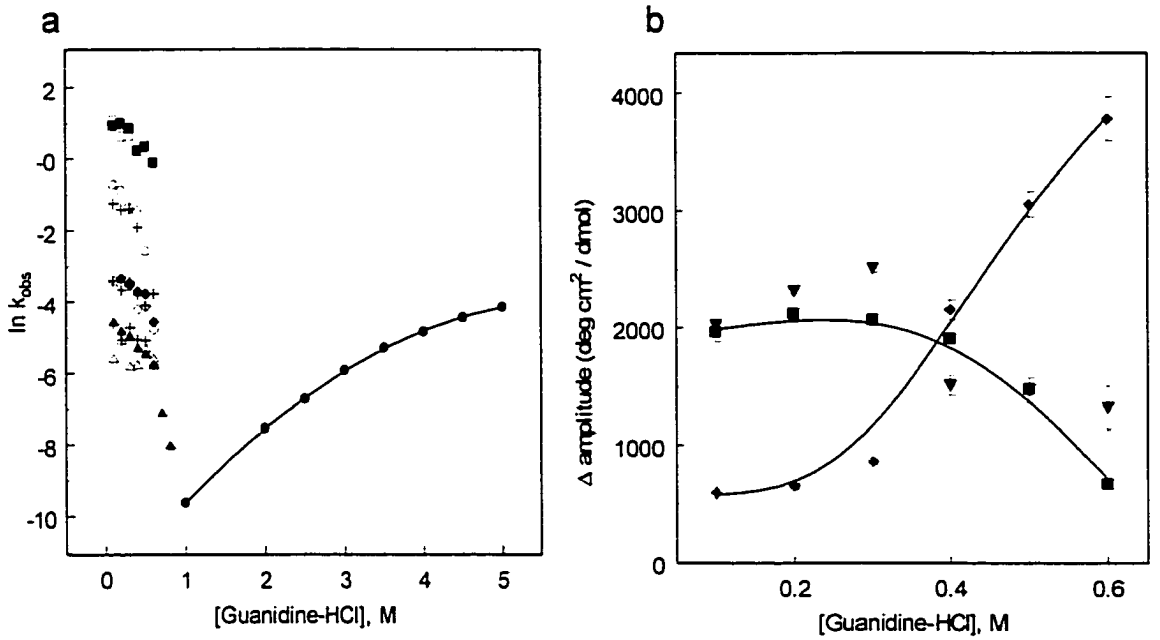


Figure 4.2. *a.* The natural log of the observed folding and unfolding rate constants as a function of final denaturant concentration. The filled circles are from unfolding experiments and are fit to a second order polynomial (*ln* k_u = -12.2 + 2.8([gdn-HCl]) – 0.2([gdn-HCl])^2). The filled squares, diamonds and triangles represent the rate constants (k_{f1}, k_{f2}, and k_{f3} respectively) as determined by far-UV CD. The open symbols represent the respective rate constants as determined by near-UV CD. The open circles represent the phase observed by near-UV CD and not by far-UV CD. The plus signs represent the rate constants as determined by fluorescence. *b.* The change in the folding amplitude determined by far-UV CD. The symbols correspond to those in *a*. The lines are a weighted least-squares fit of the data points.

Folding kinetics of pelC. The folding kinetics of pelC can be broken down into two parts, which include fast and slow phases. Stopped-flow CD was used to observe the fast-folding events and manual mixing techniques were used to observe the slow phases. The CD at 218nm (far-UV) is an excellent indication of the formation of β-sheet secondary structure. Figure 4.3a shows a representative folding experiment of pelC monitored by CD at 218nm. Unfolded pelC in 5 M gdn-HCl was diluted to 0.2 M to initiate refolding. The folding process consists of three phases. The fastest phase has a relaxation time in buffer of 0.25 seconds (Figure 4.3 inset) that comprises about 50% of the total amplitude. There are also two slow phases with relaxation times of 21 and 46 seconds. Figure 4.2a shows the dependence of the rate constant on final denaturant concentration for folding and unfolding. The initial phase is much faster than the other two phases. Figure 4.2b shows the change in amplitude of each phase as a function of denaturant concentration. The amplitude is proportional to the number of molecules that are actually populated in that phase. As the denaturant concentration is increased, the number of molecules populating the fast phase decreases, while the number of molecules in the intermediate phase increases. The number of molecules in the slowest phase shows little dependence on denaturant concentration. This suggests that an intermediate form in the fast phase allows efficient folding. As this is destabilized, more molecules must populate other intermediates that have higher activation barriers, leading to slower folding. Figure 4.3b shows the time-

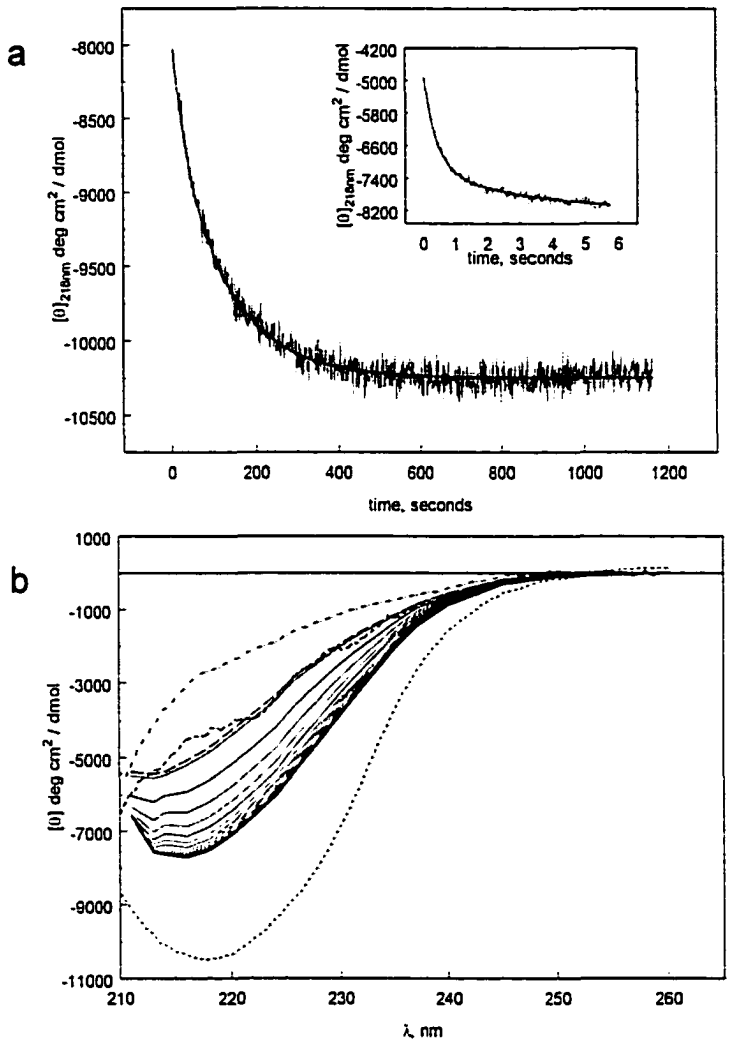


Figure 4.3. *a.* A representative folding curve for pelC measured by far-UV CD. Conditions are 5 mM MES pH 6, 50 mM NaCl, 0.2 M gdn-HCl, 25° C. The main figure shows the results from manual mixing and the inset is from stopped-flow mixing. *b.* Time-resolved CD spectra of the fast events in pelC folding. Final conditions were the same as *a*. Shown are the spectra of pelC in 5 M gdn-HCl (——), pelC at pH 2 (- - -) and native pelC (---). The solid lines are pelC folding from 0.24-2.2 seconds at 200msec intervals. Also shown is the spectrum at 0 seconds determined by extrapolation (——).

resolved CD spectrum of the fast phase. This suggests that the intermediate has a very native-like far-UV CD spectrum.

The folding of pelC was also examined in the near UV. CD at 277nm is sensitive to the chiral environment of aromatic side-chains. It is a good probe of the formation of tertiary structure. The folding process monitored at 277nm has some interesting similarities and differences relative to that followed by 218nm CD. A representative kinetic trace is shown in figure 4.4a. One feature is the negative amplitude of the fastest phase (Figure 4.4a inset). This is interesting for two reasons. The rate of this phase is identical to the fast phase in the far-UV. It is also opposite in sign to the native CD. There are two possible explanations for this observation. An intermediate is formed in the fast phase with native-like secondary structure and non-native tertiary structure. Or, the intermediate is native-like and the aromatic residues that are structured have a negative contribution to the native CD spectrum, which overall is positive. The latter seems plausible due to the fact that at low pH, a salt-induced intermediate also has a negative sign for the CD in the near-UV (Figure 4.4b). It is also interesting in that there are four kinetic phases resolved in the near-UV (Figure 4.2a). Three of these phases have identical rates as the rates determined by far-UV CD, and one phase is unique (relaxation time = 1.1 seconds). This indicates that the secondary structure forms very rapidly along with some tertiary structure, followed by the formation of the remainder of the tertiary structure, or an off-pathway intermediate is formed rapidly. The 1.1-second phase may involve the destruction of this intermediate, as the native state is formed.

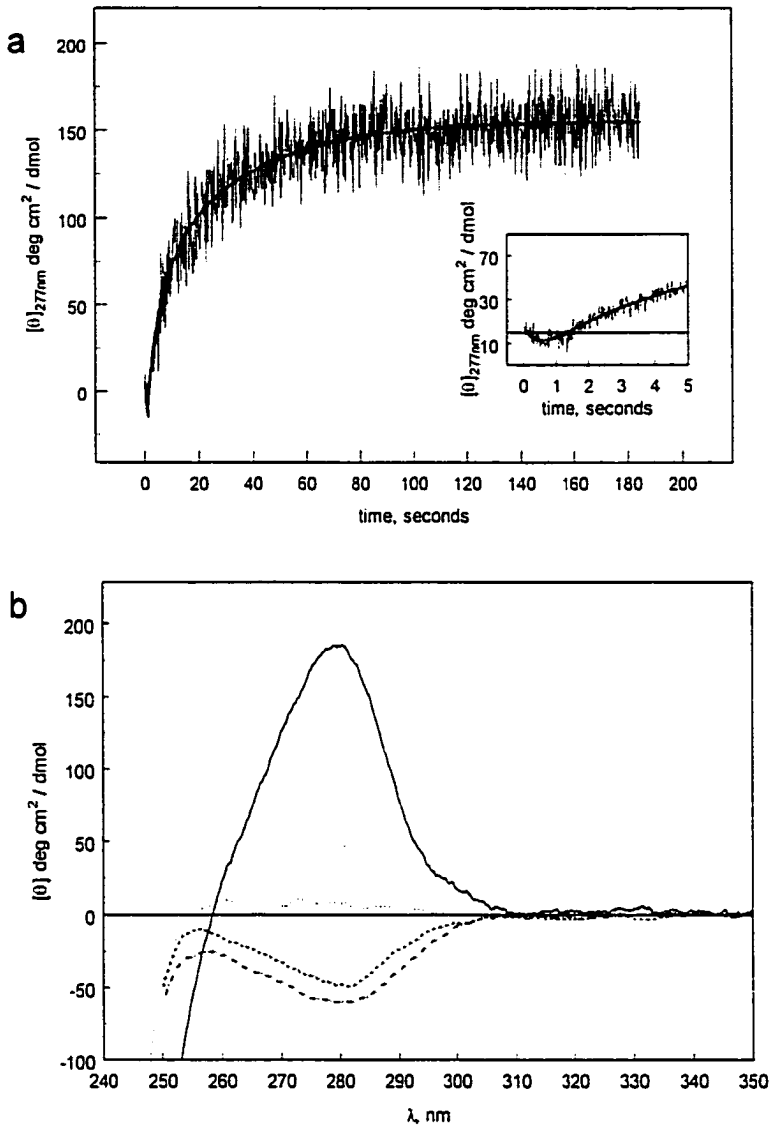


Figure 4.4. *a*. A representative folding curve for pelC measured by near-UV CD at 277 nm. Conditions were 25 mM MES, 50 mM NaCl, 0.3 M gdn-HCl, 25° C. The data are cut for clarity, but equilibrate to the native CD value. The inset shows the fast phase and the change in sign. *b*. The equilibrium intermediate CD spectrum of pelC. Conditions are 25 mM gly-HCl pH 2, 25° C, with no salt (···), 100 mM NaCl (---), 150 mM NaCl (-·-·-), and native pelC, pH 7 (—).

At high temperatures, and especially in gdn-HCl concentrations above 0.6M, single-exponential kinetics are observed (Figure 4.5a and b). This might suggest that at these extreme conditions, all the intermediates in the folding process are destabilized and no longer populated, thus yielding two-state kinetics. Intermediates are still present under these conditions but are in rapid equilibrium with the unfolded state. Evidence for the presence of intermediates can be seen from the temperature dependence of the folding rate constant of the slowest phase (Figure 4.5c). For simple two-state processes, the rate constant must monotonically increase with temperature. In all cases examined, the slow folding rate constant does not increase monotonically with temperature, but is maximal at a certain temperature. According to the steady-state approximation, an on-pathway intermediate in rapid equilibrium with the unfolded state would behave in this manner (see appendix A4). This suggests that the intermediate is necessary for folding and is populated to some degree under all conditions examined.

The two slow steps of pelC folding are due to proline isomerization. For many proteins, the slow kinetic phases have been shown to be the result of *cis-trans* isomerization about prolyl-peptide bonds (Brandts et al. 1975; Hurle and Matthews 1987; Kiefhaber et al. 1990; Schmid and Baldwin 1978). In model peptides this isomerization occurs with a relaxation time of 10-100 seconds (Brandts et al. 1975). In the unfolded ensemble, the *trans* isomer of proline is more favorable than the *cis*, but in a folded protein, structural constraints can

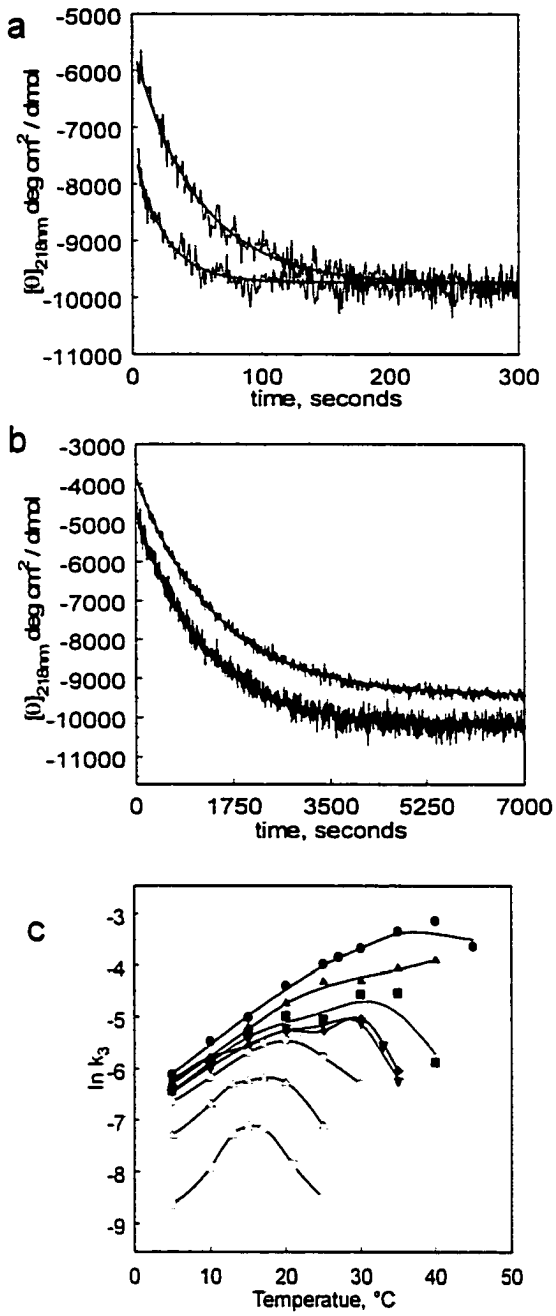


Figure 4.5. Folding kinetics measured by far-UV CD (218 nm) by manual mixing. *a.* 0.1 and 0.2 M gdn-HCl at 40° C and (*b*) 0.7 M gdn-HCl at 25° C and 0.8 M gdn-HCl at 15° C (bottom to top in both). Conditions were the same as from Figure 4.3. *c.* The natural log of the rate constant for the slow folding phase at various denaturant concentrations (top to bottom: 0.1-0.8 M gdn-HCl) and temperatures (5°-45° C).

result in the stabilization of the *cis* isomer. This is somewhat common with X-pro peptide bonds because the relative stability of the *trans* isomer is decreased. The equilibration of the isomers occurs relatively slowly due to a high enthalpic barrier. In pelC, the two slow phases have relaxation times in denaturant-free buffer of 21 and 46 seconds. It was therefore hypothesized that proline isomerization is occurring.

One way to test for the presence of proline isomerization is to use the double-jump technique. This procedure takes advantage of the fact that there is a high enthalpic barrier to proline isomerization relative to protein folding or unfolding. If the slow phases are due to a slow isomerization, such as prolyl-peptide bond isomerization, then the amplitudes of these phases will increase with increasing delay time. This is because at short delay times relatively few prolines will have undergone isomerization from the native conformation. As time goes by, equilibrium is reached and a stable population of incorrect isomers will be present. Figure 4.6 shows double-jump experiments using CD and fluorescence. Using fluorescence spectroscopy, three kinetic folding phases are observed with rates identical to those observed by CD (Figure 4.2a). This technique yielded much better signal-to-noise ratios and was therefore used for very precise determination of the relative amplitudes for the kinetic traces obtained at various delay times. It is demonstrated in Figure 4.6a that the relative change in amplitude for the slowest phase, as measured by CD and fluorescence, is identical. The overall results indicate behavior typical of proline isomerization. With increasing delay time, the relative amplitude of the initial

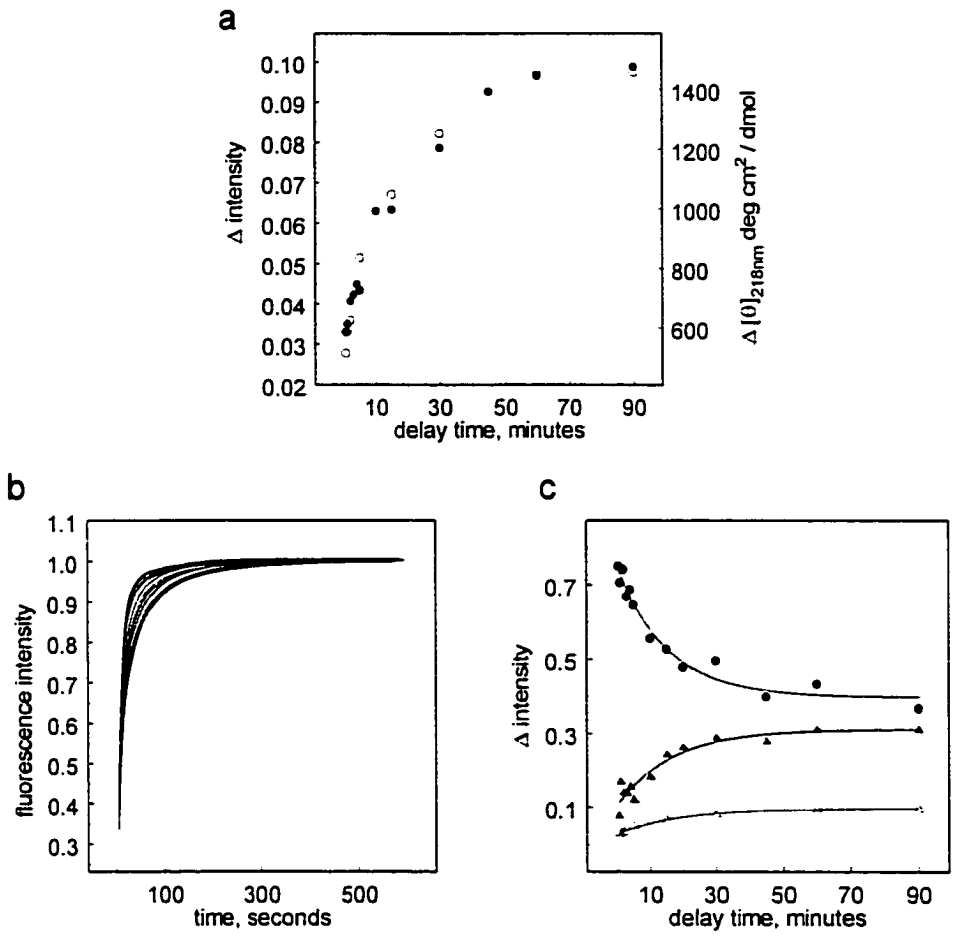


Figure 4.6. Double-jump refolding experiments for pelC. *a*. The change in amplitude for the slowest phase of folding at various refolding delay times measured by fluorescence (o) or far-UV CD (•). *b*. Refolding curves measured by fluorescence at increasing refolding delay time. *c*. The change in amplitude of each observed kinetic phase at increasing refolding delay time determined by fluorescence for the 1.1 sec (•), 21 sec (▲) and 46 sec phase (△).

phase decreases exponentially (Figure 4.6c). However, the relative amplitudes of the two slower phases both increase exponentially with delay time. This is very strong evidence that there are multiple populations of unfolded protein molecules that interconvert relatively slowly. Unfolding of pelC occurs very rapidly at 0° C, pH 9.5, and in 5 M gdn-HCl (~25 seconds). Because of the high activation enthalpy, *cis-trans* isomerization of prolyl-peptide bonds will be dramatically slower. Thus, the equilibration of the *cis-trans* isomers will be very slow compared to unfolding, effectively decoupling these processes. This experiment allows one to observe the isomerization directly. The increase in amplitude as the protein is left unfolded for longer times before refolding results from more non-native prolyl peptide bonds forming, and more molecules populating the slow kinetic phases.

Proline isomerization occurs with a characteristically high activation enthalpy (Brandts et al. 1975). In model peptide systems this is 19 ± 3 kcal / mol (Brandts et al. 1975; Cheng and Bovey 1977). Using equation (2) the activation enthalpy can be calculated from the rate dependence on temperature. $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are calculated from the Eyring plots shown in figure 4.7. The data are derived from the rates in denaturant extrapolated to 0M gdn-HCl. The slowest phase has $\Delta H^{\ddagger}$ = 21.2 ± 0.9 kcal / mol and $\Delta S^{\ddagger}$ = 5.8 ± 0.2 cal / mol K. This suggests that a typical proline isomerization event is responsible for the slow folding of this phase. For the intermediate slow phase $\Delta H^{\ddagger}$ =16.0 ± 0.7 kcal / mol and $\Delta S^{\ddagger}$ = -11.2 ± 2.5 cal / mol K (determined by experiments described in Chapter 5). These activation parameters are also typical of proline isomerization.

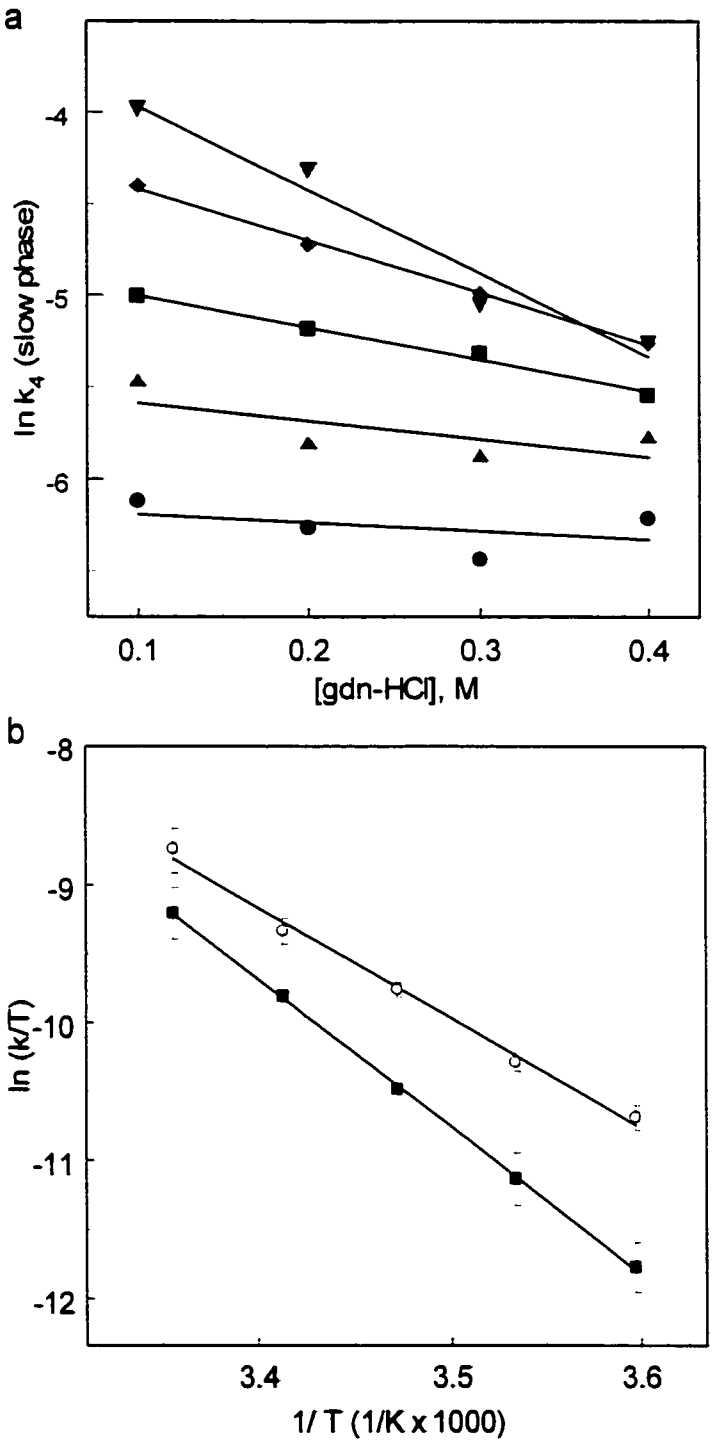


Figure 4.7. (a) ln k_f for pelC in 0.1 - 0.4 M gdn-HCl at temperatures from (bottom to top) 5 - 25° C. (b) Eyring plots for the intermediate slow phase from P220A (○) and the slowest phase (■) from wild-type pelC.

4.4 Discussion

The folding of small proteins is generally a relatively fast process, especially the formation of secondary structure (Dill and Chan 1997; Dinner et al. 2000; Dobson et al. 1998). Secondary structure formation for many larger proteins occurs on a relatively fast time scale as well (Jaenicke 1987). In small and large proteins alike, the folding steps that occur on time scales of the order of tens of seconds or more are not the result of the formation of ordered secondary structure. In many cases, along with the added complexity of intramolecular domain association and subunit association, the slow steps of folding involve slow *cis-trans* isomerization of X-pro peptide bonds (Brandts et al. 1975; Hurle and Matthews 1987; Kiefhaber et al. 1990; Schmid and Baldwin 1978). In larger proteins, retardation of protein folding by this isomerization is much more common. One reason is that large proteins contain more amino acids and the statistical probability of prolines being present is greater. The presence and location of prolines are certainly not random occurrences. One of the specific roles of proline is to form a tight turn or kink in the backbone. Larger proteins often have highly evolved and optimized functions requiring finely tuned architecture. Proline side chains are very important in fine tuning backbone architecture, and *cis* prolines are often functionally important.

In studying the folding mechanism of RNAase A, Baldwin and co-workers observed multiple unfolded populations (Garel and Baldwin 1973; Garel and

Baldwin 1975; Garel et al. 1976) corresponding to fast and slow folding kinetics. Brandts *et al.* (1975) first proposed the proline isomerization hypothesis for RNAase A, which stated that the slow- and fast-folding forms of the protein differ in the *cis-trans* configuration of X-pro peptide bonds. The original proposal suggested that the slow and fast folding forms of the protein exist in equilibrium and the slow form must convert to the fast form prior to folding to the native protein. It was subsequently shown that in RNAase A, an intermediate is formed with some degree of compact structure but with some molecules having the incorrect proline conformation (Nall et al. 1978; Schmid and Baldwin 1979a). This cleared up some problems with the original ideas. It seems unlikely that an entire protein could remain completely unfolded until one proline obtains its correct isomeric configuration. This leads to the view that a sizeable fraction of the protein can acquire native structure while the local region about the incorrect proline isomer is somewhat unordered.

There are 12 prolines in pelC, and one of these, Pro 220, is in the *cis* conformation in the native enzyme. In model peptides and unordered proteins, the ratio of *trans:cis* isomers of proline is about 70:30 (Brandts et al. 1975). The slow kinetics result from the slow interconversion of these isomers. The slowest kinetic folding phase for pelC exhibits properties that are very typical of proline isomerization, including a relaxation time of 46 seconds, a high activation enthalpy, and an increase in the amplitude of the slow phase with increasing time of unfolding during a double-jump assay. The intermediate slow phase shows similar properties. The relaxation time of 21 seconds and the increase in

amplitude in the double jump assay are consistent with proline isomerization, as is the 16.0 kcal / mole activation enthalpy. One of the prolines that is responsible for the slow phases is Pro220. This residue is in the *cis* conformation in the native structure (Yoder et al. 1993a). It is located within the central β-helix region at the end of a strand. Results from site-directed mutagenesis indicate that isomerization of the Leu219-Pro220 peptide bond results in the slowest kinetic phase observed (see Chapter 5). In the mutant P220A, the 46-second kinetic phase is eliminated. It is also possible that prolines that are *trans* in the native protein demonstrate slow kinetics because, in the unfolded protein, some small population is *cis*. In pelC, the 21-second phase is not the result of isomerization of a native *cis* proline. Folding kinetics of P220A show a slow phase that is identical to the 21-second phase observed in wild-type pelC. Experiments with other proline→alanine or valine mutants indicate that no single prolyl-peptide bond is responsible for this phase. It is more likely that this phase results from an ensemble of prolines in the incorrect isomeric configuration. It is also possible that this phase results from isomerization of non-prolyl peptide bonds. This phenomenon has recently been observed by Pappenberger *et al* (2001).

A proposed kinetic model is presented in figure 4.8, and can be described as follows for the case of folding with all of the prolines in the correct native configuration. The major fast phase results in the complete formation of the secondary structure. This occurs in a single cooperative step. At the same time, certain aromatic side chains of the molecule are forming a defined tertiary structure. The possibility of a population of molecules forming rapidly with

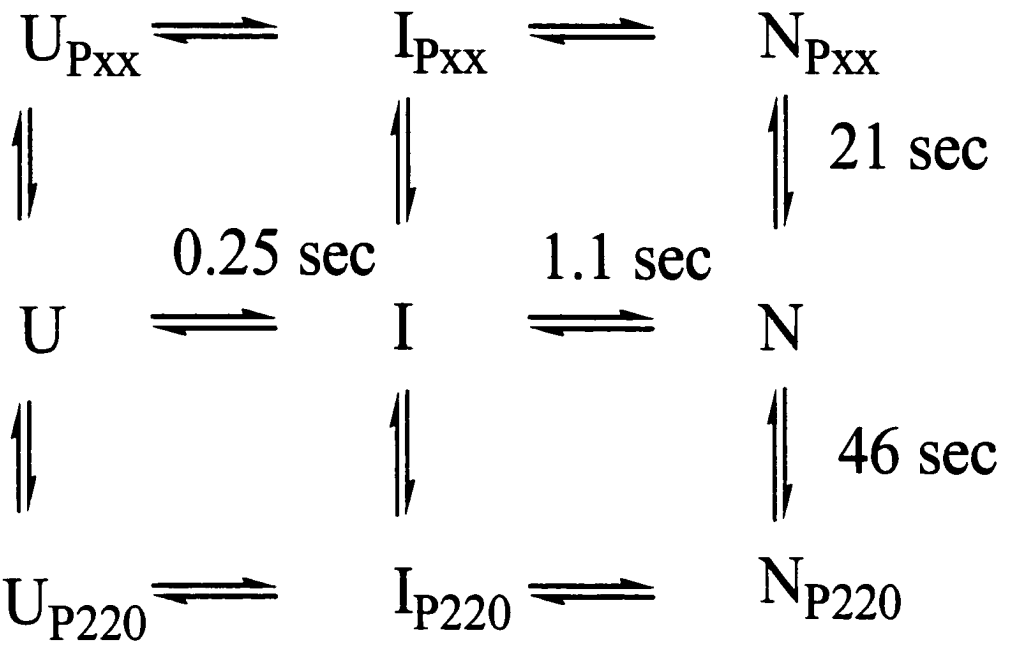


Figure 4.8: A model for the folding mechanism of pelC. U is the unfolded protein, I is the intermediate formed in the fast phase, and N is the native protein. N with a subscript represents a pseudo-native state in which the only difference from the native protein is the configuration of a peptide bond. U, I and N represent the major pathway taken when all peptide bonds are in the correct conformation. U_{P220}, I_{P220}, and N_{P220} represent the slowest pathway when Pro220 is in the *trans* conformation in the unfolded state. U_{Pxx}, I_{Pxx} and N_{Pxx} is the pathway taken when other prolines are in the *cis* conformation in the unfolded state.

completely native side-chain structure was ruled out by the negative amplitude in the fast phase of folding observed by near-UV CD. It is very possible that the side-chains that are structured at the end of this phase are in a native conformation. Calculations suggest that the aromatic side chains in the β-helix region make an overall negative contribution to the near-UV CD spectrum (see Chapter 6). Also, at low pH and high salt an intermediate with a negative CD amplitude in the near-UV is stable in pelC (Figure 4.4b). It is possible that these intermediate states are structurally related. It might only be a coincidence, but it is quite an interesting one. The next step in the folding process involves the ordering of the remainder of the side-chains.

A kinetic model for folding with prolines in the non-native configuration is also presented in Figure 4.8. When all of the prolines are in the correct configuration, there is rapid folding to the intermediate containing much or all of the native secondary structure, followed by the slightly slower formation of the remaining native tertiary structure. The case of folding with one or more prolines in the incorrect configuration is more complicated. It seems unlikely that the proline isomerization is occurring in a completely unordered protein. It is more likely that the majority of the protein is folded correctly but the region of the protein near the incorrect proline isomer is unstructured or improperly folded. When the unfolded protein solution is transferred to conditions stabilizing the native fold, there is a rapid condensation of the structure via at least three pathways. For the 46-second phase, the region around the incorrect proline is highly disordered because of the inability to form some native contacts. This

mimics the behavior of a complete random coil. The 21-second phase occurs slightly faster, probably because structural constraints put strain on the incorrect prolyl peptide bond facilitating its conversion to the correct configuration. That is to say that the local environment of the polypeptide chain lowers the activation enthalpy thus resulting in a slightly faster isomerization rate. This results in an activation enthalpy that approaches the lower limit observed for proline isomerization. It is also possible that an intermediate helps stabilize the correct backbone geometry. One commonly observed property of proline isomerization is the independence of the rate on denaturant concentration (Schmid and Baldwin 1979b). This is not observed for pelC, but can be explained by our model. The presence of an intermediate strongly influences the rate of folding for pelC, and the intermediate's stability is strongly dependent on denaturant concentration. As the intermediate is destabilized, the rate dramatically slows, and subsequently, only one phase is observed, corresponding to a rate-limiting proline isomerization.

The enzyme peptidyl-prolyl-isomerase has been shown to catalyze the *cis-trans* isomerization of X-pro peptide bonds in small polypeptides and some proteins (Lang et al. 1987). The catalytic ability of this enzyme was tested on pelC (Figure 4.9). In a fluorescence experiment monitoring folding in 0.3 M urea, it was found that for the slow phases a slight rate enhancement occurs (approximately two-fold). For the slowest phase the rate increases from 0.0070 sec^{-1} to 0.0147 sec^{-1}, and for the intermediate slow phase the rate goes from 0.0301 sec^{-1} to 0.0855 sec^{-1} in the presence of 1μM human PPI. PPI does not

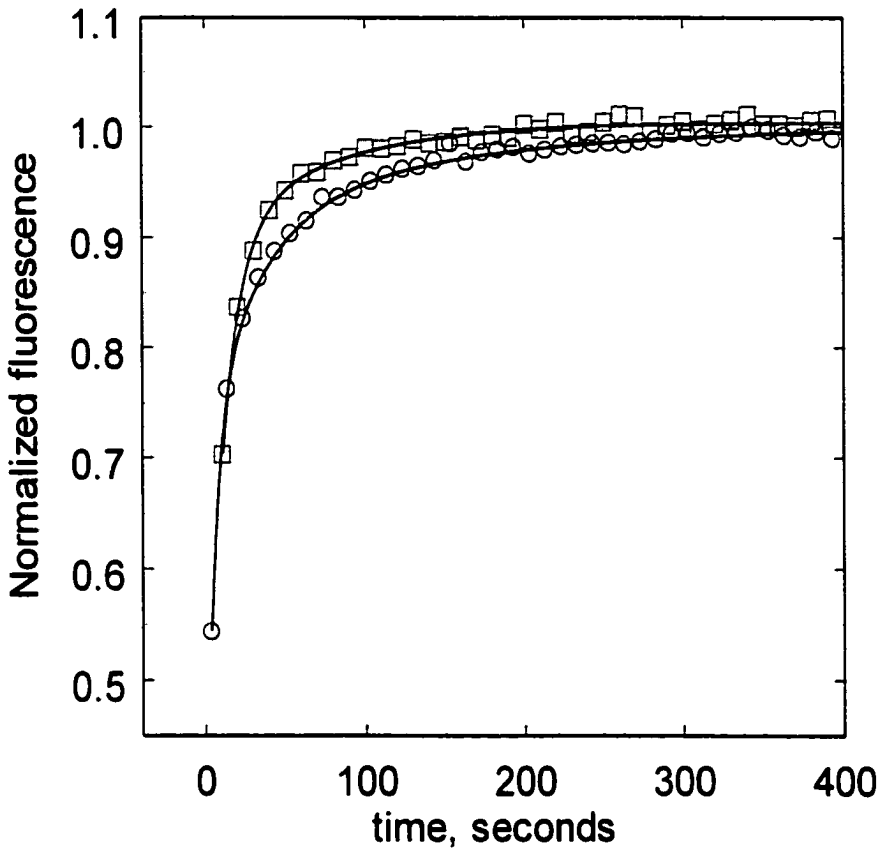


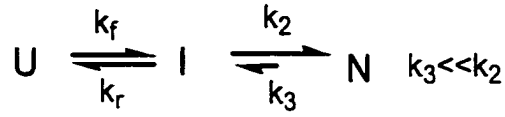
Figure 4.9: Folding of wild-type pelC in 0.3 M urea at pH 8.0 (○) and the same process in the presence of 1μM PPI (□).

necessarily show dramatic rate enhancements or catalyze all prolyl isomerization events (Jackson and Fersht 1991b; Lang et al. 1987). There must be access for the enzyme active site to the suspect prolines. In this case access to the incorrect prolines is limited. This observation is consistent with the proposed folding model. The formation of essentially all of the secondary structure occurs in about 0.25 seconds. This occurs in parallel with the formation of some degree of tertiary structure. It is therefore not surprising that the folded regions of pelC protect the incorrectly folded proline residues from PPI.

Near-UV CD is a powerful and sensitive tool for observing the formation of side-chain structure during protein folding, but it has gone widely underutilized. This is possibly because far-UV CD overshadows the usefulness of near-UV CD. Different far-UV CD spectra are very characteristic of specific secondary structures and combinations thereof, and the rules governing this are quite well understood. But the only rules for CD in the near-UV are that aromatic side-chains and disulfide bonds are the chromophores. There is no *a priori* way to tell from CD what the specific conformation of the side-chain is. This study demonstrates the qualitative utility of near-UV CD in understanding a kinetic protein folding mechanism. It is not too surprising that there is a change of the sign of the folding amplitude. Although to our knowledge, this behavior has not been observed with near-UV CD, similar observations have been made with far-UV CD (Clarke et al. 1999; Matagne et al. 1998; Matagne et al. 2000; Radford et al. 1992). These, however were overshoots of the native amplitude and not opposite in sign to the native CD.

The results presented in this chapter demonstrate that the kinetic folding mechanism of pelC is overall rather simple. The formation of secondary structure occurs in a single cooperative step, along with some ordering of side-chains. The remainder of the side-chains fold in a second step. The most complex part of the folding mechanism is well characterized and involves two proline isomerization steps. One of these slow phases is due to isomerization of Pro220 and the other is a result of the parallel isomerization of several other prolines. Systematic mutations of all of the proline residues in pelC have been carried out in order to understand more completely the slow folding kinetics of pelC. This study is presented in Chapter 5.

4.5 Appendix A4



U and I are in rapid equilibrium. We want to derive k_{obs} (the rate of formation of N from U).

$dI/dt = k_fU - k_rI - k_2I + k_3N \approx 0$ The steady-state approximation assumes that the concentration of I changes very slowly.

$k_fU = I(k_r + k_2) - k_3N$

$\approx I(k_r + k_2)$ Neglecting k_3N in comparison to $I(k_r + k_2)$.

$k_fU / (k_r + k_2) = I$ Now we have an expression for I in terms of U and rate constants. We can get an expression for the formation of N from I.

$dN / dt = k_2I$ Neglecting k_3N in comparison to k_2I

$dN / dt = (k_2k_f)U / (k_r + k_2) = k_{obs}U$

therefore $k_{obs} = (k_2k_f) / (k_r + k_2)$

For temperature dependence:

Consider the limiting behavior of k_{obs} at high and low temperature.

$$k_{obs} = k_2k_f / (k_r + k_2)$$

Assume that $\Delta H^{\ddagger}_r >> \Delta H^{\ddagger}_2$, and that at low T, $k_r << k_2$. Then at low T,

$$k_{obs} \approx k_2 k_f / k_2 = k_f \text{ and } \Delta H^{\ddagger}_{obs} = \Delta H^{\ddagger}_f$$

At high T, $k_r >> k_2$ and

$$k_{obs} \approx k_2 k_f / k_r = k_2 K_{eq}$$

$$\Delta H^{\ddagger}_{obs} = \Delta H^{\ddagger}_2 + \Delta H^{\ddagger}_f - \Delta H^{\ddagger}_r = \Delta H^{\ddagger}_2 + \Delta H$$

$\Delta H^{\ddagger}_r > \Delta H^{\ddagger}_f + \Delta H^{\ddagger}_2$, then $\Delta H^{\ddagger}_{obs} < 0$ at high temperature and we have the observed anomalous temperature dependence.

Chapter 5

Folding kinetics and stability of pelC proline mutants

5.1 Introduction

It is well established that slow phases in protein folding exist. One of the most important developments in understanding slow kinetics of protein folding came in 1975. Brandts *et al.* (1975) proposed that the slow step in folding is due to the *cis-trans* isomerization of prolyl-peptide bonds in the unfolded state. X-pro peptide bonds (X is any amino acid) can populate the *cis* conformation much more readily than any other peptide bond. The highly unfavorable *cis:trans* ratio in non-prolyl peptide bonds is due in part to the *trans* isomer having greater conformational entropy relative to the *cis* isomer, and steric clashing of the side-chain with the preceding amino acid. The presence of a C_δ bonded to the backbone nitrogen in proline reduces the advantages of *trans* over *cis*. In an unordered polypeptide the *trans:cis* ratio of x-pro peptide bonds is about 70:30 (Brandts et al. 1975). Many proteins contain prolyl-peptide bonds that are *cis* in the native conformation. Conformational constraints in the folded protein can overcome the less favorable interactions and stabilize the *cis* bond. This is often important in forming the correct architecture of the enzyme active site for binding substrate or properly positioning a catalytically active group.

In the folding kinetics of pelC, a total of four phases are observed. Two of these events occur on the second or faster time scale and are attributed to the formation of secondary structure and ordering of side-chains. Two phases show kinetics occurring on the order of tens of seconds. Proline isomerization in model peptides occurs in the time range of ten to 100 seconds (Brandts et al. 1975).

Evidence was presented in Chapter 4 that indicates that both of these slow phases are due to proline isomerization. However, although these experiments are conclusive, they give no clue as to which specific prolines are responsible for the observed slow kinetics. The Leu219-Pro220 peptide bond is in a *cis* conformation as indicated by the x-ray structure of pelC (Yoder et al. 1993b). This proline is conserved among several pectate lyases (Heffron et al. 1995). It serves a critical role in that it properly positions Arg218, an important group in catalysis. In addition to Pro220, pelC contains 11 more prolines, at positions 51, 79, 124, 139, 159, 263, 286, 302, 316, 324, and 335. All of these prolines are involved in *trans* peptide bonds (Figure 5.1).

In order to have a more complete model for the folding mechanism of pelC, one needs to characterize the slow kinetics and determine which prolines are responsible. Three methods have been used to identify "essential" prolines. Lin and Brandts developed the technique of isomer-specific proteolysis (Lin and Brandts 1983; Lin and Brandts 1984). This utilizes the fact that certain proteases only recognize *trans* peptide bonds and therefore can not hydrolyze prolyl-peptide bonds in the *cis* configuration. Peptide fragments are analyzed by peptide sequencing and slowly isomerizing bonds can be identified. A review of the literature has concluded that this method is not consistently reliable (Adler and Scheraga 1990; Schultz et al. 1992), and would be very complicated to use with pelC considering that there are twelve prolines. NMR has also been used to measure the *cis:trans* ratio in heat-unfolded RNAase A (Adler and Scheraga 1990). Prolines that had a significant population of the incorrect native isomer

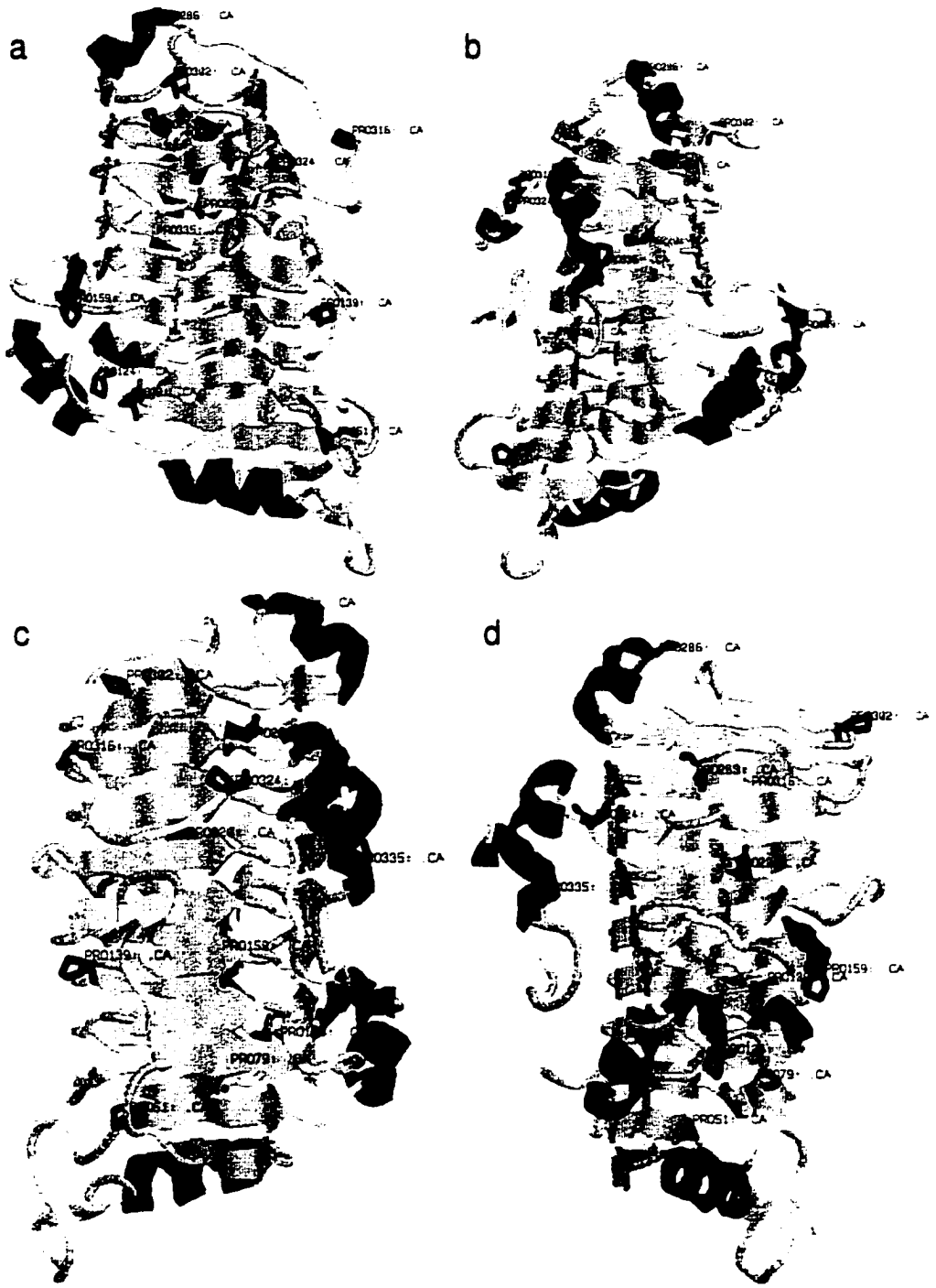


Figure 5.1: The structure of pelC rotated 90 degrees with respect to (*a*) in each frame showing all of the prolines in the protein. Pro220 is colored purple and all others are blue.

were identified. The percentage measured by NMR was in good agreement with the percentage of incorrect isomer measured by kinetic experiments. The two prolines identified were proposed to be the ones responsible for the observed slow kinetics. This method was effective, but would be very difficult to apply to a protein the size of pelC because it requires assignment of resonances in the NMR spectrum. Advances in the field of protein engineering in recent years have led to numerous insights into protein structure and function. Several groups have used site-directed mutagenesis to study slow kinetics in folding. Schultz *et al* (1992) has demonstrated conclusively that the slow kinetics in RNAase A are due to the two *cis* prolines, 93 and 114. Kelley and Richards (1987) showed that replacement of Pro76 with alanine in thioredoxin eliminates the slow phase. In both of these cases the suspect prolines were *cis* in the native structure. Schindler *et al* (1996) have investigated the role of *trans* prolines in the folding of RNAase T1. They concluded that *trans* prolines do not noticeably affect folding. More recently, Eyles and Gierasch (2000) investigated the *trans* proline contribution to the folding of CRABP 1. They found that mutation of Pro85 to Ala resulted in elimination of the slowest folding phase.

The mutagenesis approach seemed like the best way to characterize the slow folding kinetics in pelC. In all, ten Pro→Ala and two Pro→Val mutants were constructed, including *cis* P220A, and the folding of these mutants was characterized. P79A and P124A were constructed but several attempts to express these mutants failed. Instead, Pro79 and Pro124 were mutated to valine. P220A is responsible for the slowest kinetic phase of pelC folding. The

intermediate slow phase may result from multiple proline isomerizations with similar activation energies. Finally, the intermediate slow phase isomerization is strongly coupled to the formation of an intermediate on the folding pathway.

5.2 Materials and Methods

Site-directed mutagenesis. The Quickchange Site-Directed Mutagenesis kit (Stratagene Inc, La Jolla, CA) was used to construct all mutants. Plasmid pPEL410 containing the gene for wild type pelC was used as the template in all cases except for P139A (this utilized the P220A plasmid and therefore was a double mutation). The mutagenic primers used are shown in Table 5.1. Primers were always PAGE-purified. In most cases, the prolines were mutated to alanine, and where possible, the optimum codon preferences were used. In two cases, Pro79 and 124, the prolines were mutated to valine.

Mutated plasmids were transformed into *Escherichia coli* XL1-Blue super competent cells (Strategene Inc, La Jolla, CA) and grown for about 16 hours at 37° C on LB agar plates supplemented with 50 μg / ml ampicillin. Plasmid DNA was isolated and purified using a Qiagen plasmid mini-prep kit (Valencia, CA). Mutations were verified by DNA sequencing. Mutant plasmids were then transformed into *E. coli* strain HB101 for protein expression. Transformation-competent cells were prepared by the $CaCl_2$ method as described by Maniatis (1982). Transformations were carried out by a method described by Maniatis (1982). Transformants were grown at 37° C on LB agar plates supplemented with 50 μg / ml ampicillin for 13 hours to limit the number of background colonies.

Table 5.1: Mutagenesis primer sequences. Bold sequences indicate the altered bases.

Mutation	Coding Strand Primer
P51A	5'-GGCGGCGCCTAT**G**CGCTGGTAATCACC-3'
P79V	5'-GTGGAGCAAAGAC**GTT**CGTGGCGTAG-3'
P124V	5'-CGTATCGGCTACCTG**GTT**GGCGCTAAAGATGGC-3'
P139A	5'-CCGCGTCGACGATTCG**G**CGAATGTCTGGGTTGACC-3'
P159A	5'-GTGCGACGGCACA**G**CGGACAACGACACC-3'
P220A	5'-CGTTAACGCCCGTCTG**G**CGTTGCAACGTGGTGG-3'
P263A	5'-GGTTCGAGAAGGCGATAAAC**G**CGGTAACGTCCCGTTATGACGGC-3'
P286A	5'-GCAACAACATCACCAAA**G**CGGCCGACTTCTCTACC-3'
P302A	5'-CGGCCGACACCAAG**G**CTTATGTGAATGCCG-3'
P316A	5'-CGGCCACCTTC**G**CGACCGTGGCCTAC-3'
P324A	5'-GCCTACAACTACA**G**CGCGGTCAGCGCACAATGC-3'
P335A	5'-GAAGGACAAACTG**G**CTGGCTATGCCGGC-3'

Purification of mutant pelC proteins. The purification of all mutant proteins was similar to the wild-type pelC purification, with some modification. Growth of cells and formation of spheroplasts were identical. It was found that some mutants tended to aggregate on the CM Biogel column and caused the column to clog. Because pelC is the only protein in the *E. coli* periplasm that appears to bind to the negatively charged carboxymethyl resin of the CM Biogel, it was determined whether it is the only protein to pass through the corresponding DEAE Biogel. In fact, this procedure showed results comparable to the wild-type preparation. The main modification to purify the mutant proteins was to pass the dialyzed periplasmic fraction over a DEAE Biogel-A column equilibrated to pH 8. The column flow-through that showed absorbance at 280 nm was retained. Proteins were assayed for purity by SDS-PAGE. Proteins that migrated on the gel at the same rate as wild-type pelC were assumed to be the mutants. In general, this procedure provided protein that was of sufficient purity for kinetic experiments. In very rare cases, the flow-through was passed through a CM column and eluted with the procedure for the wild-type protein. With the majority of contaminating proteins removed, mutant pelC proteins did not appear to aggregate. This resulted in sufficient purity.

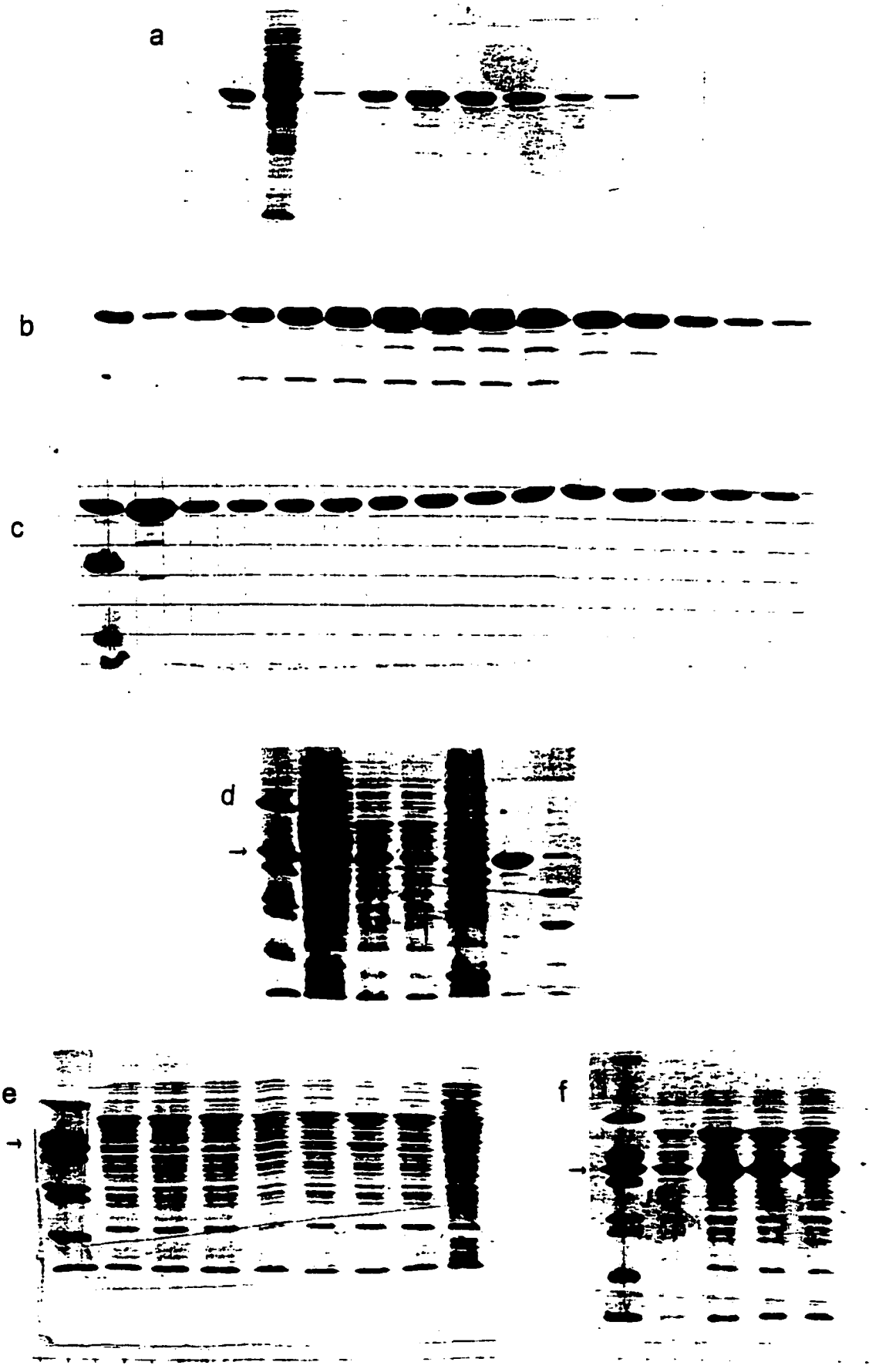
Characterization of proline mutants. All experiments on pelC mutants were carried out identically to those of the wild-type pelC (see chapters 2 and 4), with the following exceptions. Some refolding experiments were carried out by pH-jump. P220A rapidly unfolds at pH 2 in 25 mM glycine-HCl buffer. Refolding was

initiated by dilution (generally ten-fold) into 100 mM MES / 50 mM NaCl at pH 6.1. The resulting solution had a pH of 6.0-6.1. The activation enthalpy of P220A was measured in a manner similar to wild-type pelC, except the pH-jump method was used to initiate refolding. This avoided performing the experiment at several gdn-HCl concentrations and extrapolating to 0 M denaturant. For equilibrium denaturation experiments, 25 mM *N*-morpholino propanesulfonic acid (MOPS) buffer was used at pH 7.0.

5.3 Results

Expression and purification of pelC mutants. All twelve mutant constructs were successfully expressed and purified. In all cases, a two-liter culture of bacteria yielded between 8 and 115 mg of pure protein. For expression tests, 100 ml of LB cultures were grown and lysed according to the standard protocol. For P220A, P139A/P220A, and P263A this method yielded good expression and sufficient quantities for some initial experiments. For the rest of the mutants, expression in 100-ml cultures was very low and sometimes hardly detectable (Figure 5.2). A very interesting result occurred when the culture volume was increased to two liters. The level of mutant protein expression compared to other *E. coli* proteins was dramatically increased. The reason for this increase may be that the bacteria are getting less aeration in the larger cultures. This may cause them to get to a stationary phase quicker. At this point they can turn their metabolic energy towards protein production rather than cell division.

Figure 5.2: Several SDS-PAGE results for the purification of pelC mutants. *a.* Small-scale preparation of P220A from a DEAE Biogel A column. Lane 1: WT pelC, lane 2: column load, lane 3-9: flow-through fractions. *b.* Large-scale preparation of P220A. Lane 1: WT pelC, lane 2-15: flow-through fractions. *c.* P220A from *b* after elution from a CM Biogel A column. Lane 1: MW markers supplemented with WT pelC, lane 2: column load, lane 3-15: peak P220A fractions. *d.* Preparation of P139A/P220A. Lane 1: MW markers, lane 2: spheroplast suspension, lane 3: spheroplast supernatant, lane 4: column load, lane 5: precipitated proteins after dialysis, lanes 6 and 7: fractions 30 and 17. *e.* PelC mutants from 100ml culture. Lane 1: MW markers, lanes 2-9: P51A, P79A, P124A, P286A, P302A, P316A, P324A, P335A. *f.* PelC mutants from 2L culture. Lane 1: MW markers, lanes 2-5: P51A, P286A, P324A, P335A spheroplast supernatants. Arrows indicate the position of WT pelC.



Regardless of the reason, the important result is that there was sufficient yield of pelC mutant proteins.

Structure and activity of P220A. P220A is the most drastic of the Pro→Ala mutations. A proline involved in a *cis* peptide bond is changed to alanine that favors the *trans* configuration. One might expect significant perturbations in the structure and activity of this mutant. The activity of P220A was measured in a manner identical to that for wild type pelC (see chapter 2). The spectrophotometric assay that monitors the cleavage of polypectate over time indicated that the specific activity of wild-type pelC is 1.9×10^6 units / gram. The same assay under identical conditions indicated that the specific activity of P220A was only 3.0×10^4 units / gram. The mutant is about 65-fold less active than the wild-type enzyme (Figure 5.3). Such a decrease in activity could suggest a large alteration in the overall structure of pelC due to the mutation. CD and fluorescence experiments were carried out to determine the extent of structural perturbations.

The fluorescence emission spectrum of P220A at pH 7 indicates an intense band with an emission maximum at 331 nm (Figure 5.4). The spectrum of wild type pelC is also shown. The intensities of the two spectra are similar but the mutant maximum is shifted two nanometers to the blue. The intensity is arbitrary but both are typical of folded proteins. However, the shift in the maximum suggests that there may be a slight decrease in the solvent exposure of tryptophan side-chains.

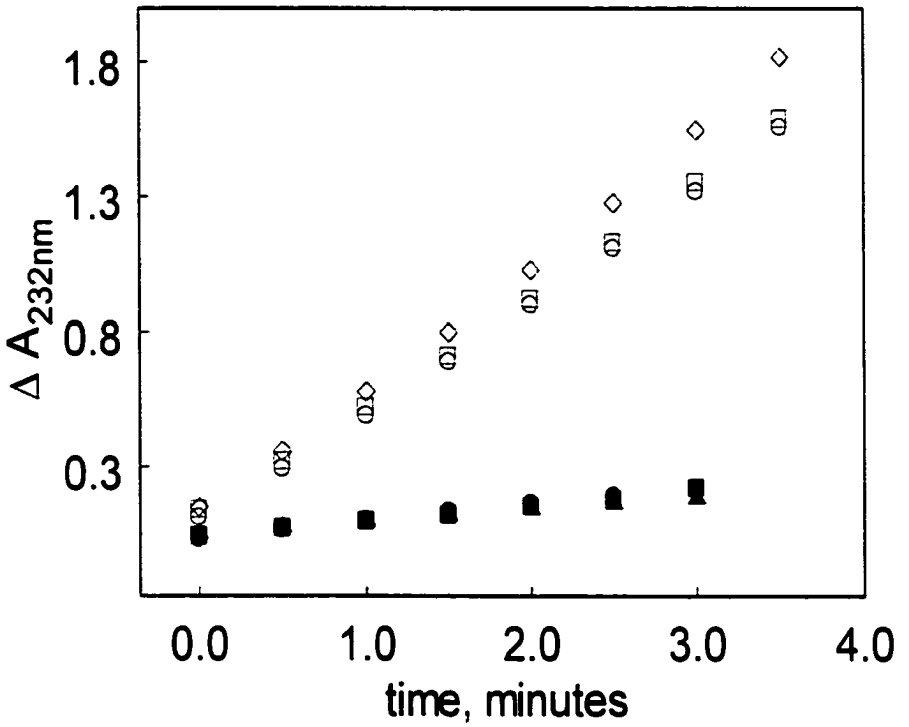


Figure 5.3: Activity assay of wild type pelC (open symbols) and P220A (filled symbols). The different symbols represent data from multiple experiments.

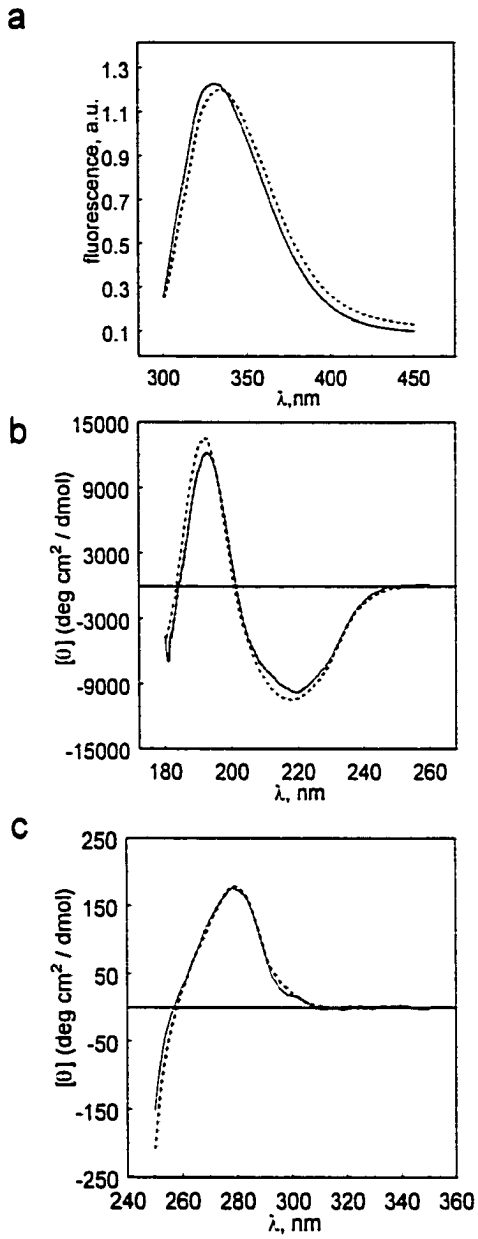


Figure 5.4: Spectroscopic analysis of P220A (——) compared to wild type pelC (----). Fluorescence emission, *a*; far-UV CD, *b*; and near-UV CD,*c*. Conditions were identical for both proteins.

Circular dichroism experiments were performed to assess the degree of change in the secondary structure and ordering of aromatic side-chains. The far-UV CD spectrum indicates a slight loss of intensity in both the long and short wavelength regions. This is consistent with a minor perturbation in some elements of secondary structure. An analysis of the secondary structure using CDPro is shown in table 5.2. This analysis indicates that the β-sheet fraction actually increases by about 3%. The fractions of α-helix and unordered structure decrease (4% and 2% respectively) and the fraction of turns increases (3%). The mutation probably alters the structure in a way that makes the CD spectrum appear more similar to other β-sheet proteins in the database.

A comparison of the near-UV CD of the wild-type and P220A mutant shows no change. This is unusual, considering the noticeable changes in the far-UV CD and fluorescence spectra. This argues against any overall changes in the geometries of the aromatic side-chains or in the coupling of the side-chain transitions.

Equilibrium denaturation of P220A. Elimination of a proline involved in a *cis* peptide bond can result in large changes to the overall stability of the protein. The free energy of unfolding for P220A in water ($\Delta G^{\circ}_{H_2O}$) was measured using gdn-HCl titration in a manner identical to that for the wild-type protein. Figure 5.5 shows the CD spectra of P220A at increasing concentrations of gdn-HCl along with the transition at 218 nm. A comparison of the denaturation of the mutant

Table 5.2: Fractions of secondary structure in pelC and P220A determined by x-ray diffraction and circular dichroism.

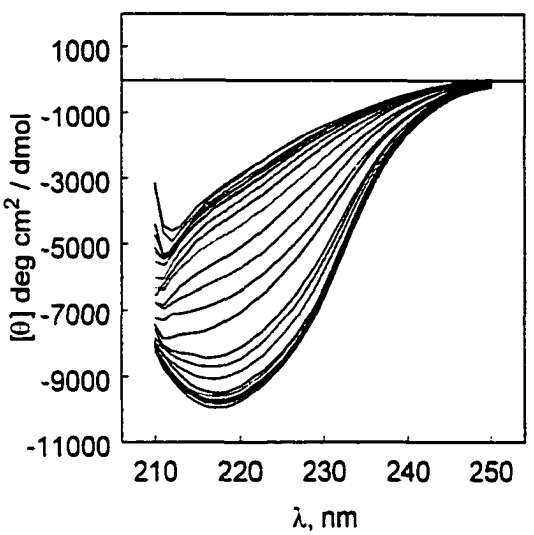
	α-helix	β-sheet	turns	Other
X-ray* WT pelC	10.7	26.6	4.0	58.7
DSSP+ WT pelC	8.8	32.0	7.9	51.3
CDPro# WT pelC	24.3	21.3	22.0	32.5
CDPro# P220A	20.2	24.6	24.7	30.2

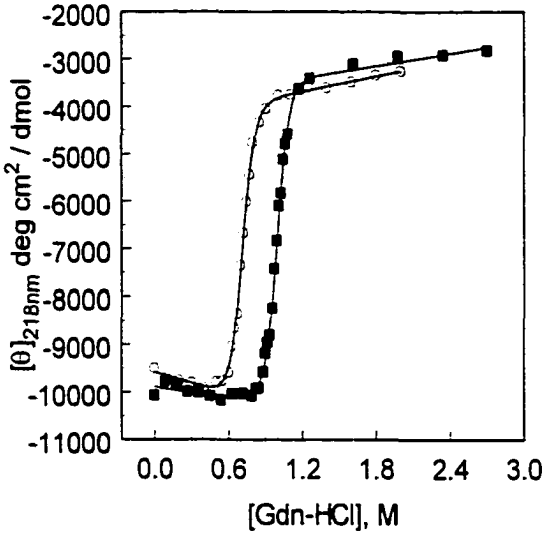
*Estimates from the x-ray diffraction structure (Yoder et al. 1993a).

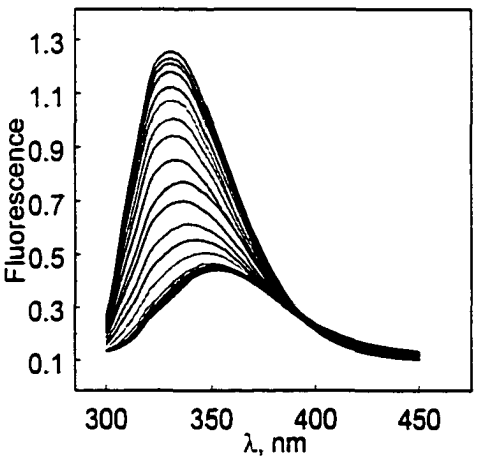
+Objective analysis of the structure from x-ray diffraction using the method of Kabsch and Sander (Kabsch and Sander 1983).

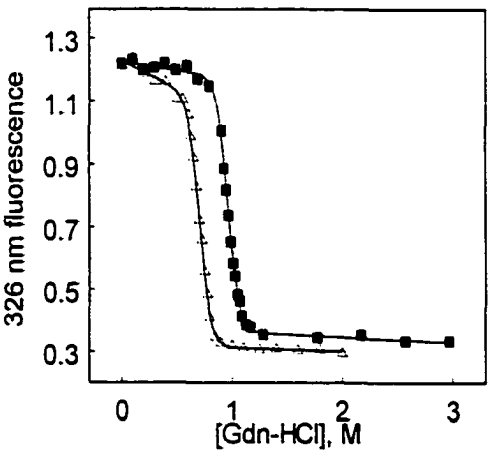
#Estimation of secondary structure from circular dichroism (Johnson 1999; Provencher and Glockner 1981; Sreerama and Woody 1993).

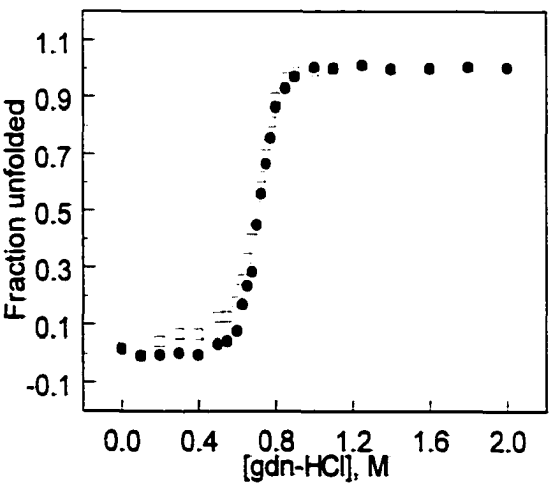
Figure 5.5: Gdn-HCl titration of P220A. *a*. Far-UV CD spectra of P220A from 0 to 2.2 M gdn-HCl. *b*. CD values at 218 nm for P220A (○) and WT pelC (■) fit to a two-state model. *c*. Fluorescence emission spectra of P220A from 0 to 2.2 M gdn-HCl. *d*. Fluorescence values at 326 nm for P220A (△) and WT pelC (■) fit to a two-state model. *e*. The fraction of P220A unfolded vs. gdn-HCl concentration. Filled circles (●) are from far-UV CD and open squares (□) are from fluorescence.











with pelC indicates that the midpoint of denaturation for the mutant is significantly lower than that of the wild type. However, the overall shape of the transition curve appears to be unchanged. Table 5.3 shows the results of the thermodynamic analysis. $\Delta\Delta G^{\circ}_{H_2O}$ is about -4 kcal / mol but the *m* value is essentially unchanged. This indicates that although there is a large loss of stability, the change in accessible surface area is not altered and hence, the overall denaturation mechanism is unchanged.

In order to demonstrate that the two-state assumption is valid, denaturation was monitored by fluorescence emission. The results are in excellent agreement with the CD results (Table 5.3). The spectroscopic data were converted to fraction of protein unfolded. The curves do not overlay perfectly, however. This is because of the significant slope of the folded baseline for the fluorescence titration, which could result from solvation effects in the unordered region around the mutation.

Kinetic unfolding of P220A. Unfolding of P220A was monitored by far-UV CD. In all cases the unfolding fits well to a single exponential function (Figure 5.6a). Also, the rate of unfolding measured at all denaturant concentrations is much faster than that for the wild type (Figure 5.6b). The lower limit for the rate of unfolding at 5 M gdn-HCl was deduced from the dead time of mixing. This seems like a good estimate because it falls on a straight line extrapolated from the other unfolding data. It is not surprising that the unfolding rate increases. The stability of the mutant is 4 kcal / mol less than that of the wild type. This

Table 5.3: Thermodynamic results from gdn-HCl denaturation at pH 7 for pelC and P220A.

	218nm CD WT pelC	218nm CD P220A	Fluorescence WT pelC	Fluorescence P220A	ΔΔG CD	ΔΔG Fl
ΔG° (kcal / mol)	12.06 ± 0.59	8.28 ± 0.36	13.03 ± 0.67	8.35 ± 0.41	3.78	4.68
m (kcal / mol M)	12.19 ± 0.59	11.59 ± 0.49	13.04 ± 0.67	11.80 ± 0.56		
$D_{1/2}$ (M)	0.99	0.71	1.00	0.71		

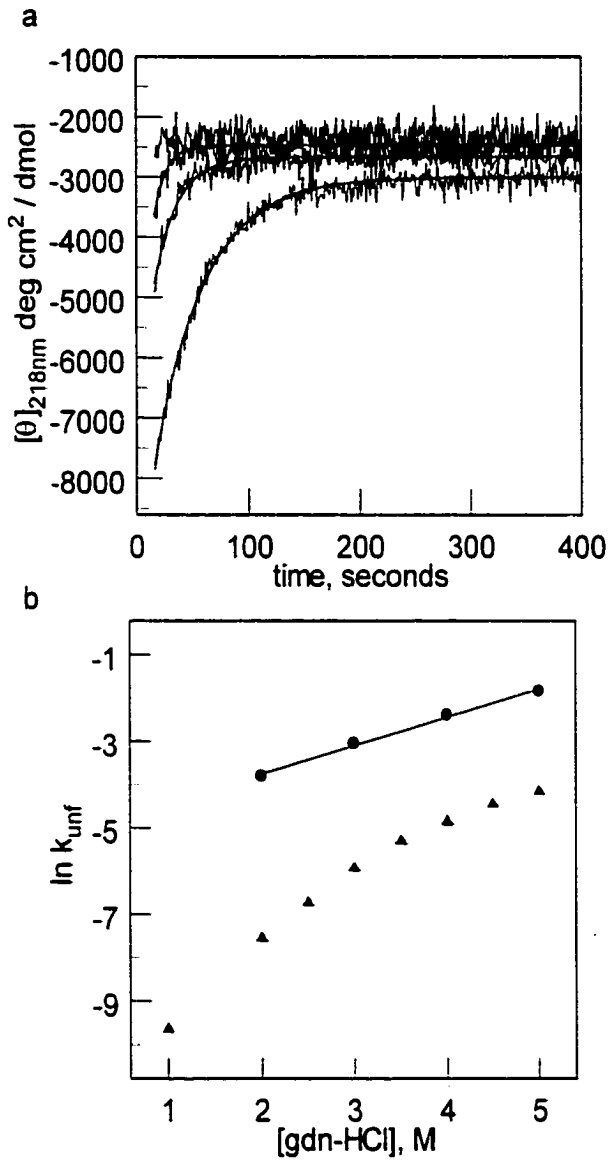


Figure 5.6: *a*. Unfolding kinetics of P220A monitored by far-UV CD at 218 nm in 2, 3, 4, and 5 M gdn-HCl (bottom to top). Each curve is fit to a single exponential model. *b*. The natural log of the observed unfolding rate constant at various denaturant concentrations for P220A (●) and wild-type pelC (▲).

can result in a noticeable change in the unfolding kinetics.

Refolding kinetics of P220A. Many examples have been reported indicating the disappearance of slow kinetic folding phases upon mutation of prolines involved in *cis* peptide bonds (Kelley and Richards 1987; Schultz et al. 1992; Wood et al. 1988). The folding kinetics of P220A were studied by far-UV CD and fluorescence. In all cases examined, the manual mixing refolding monitored by CD showed a single slow phase with a relaxation time in denaturant-free buffer of $\tau = 21$ seconds. The wild-type protein shows two slow phases with relaxation times $\tau_1 = 21$ sec and $\tau_2 = 46$ sec. Figure 5.7 shows a representative refolding experiment at 0.3 M gdn-HCl. Stopped-flow CD experiments were also performed (inset to Figure 5.7). These data indicate the presence of a fast kinetic phase very similar to that observed in wild type pelC. Closer examination of the kinetic data reveals that the refolding rates for both kinetic events are essentially the same in wild type pelC and P220A. Figure 5.8 shows the dependence of the folding rates and amplitudes on gdn-HCl concentration. This indicates that the folding mechanism is identical in both proteins, except the slow isomerization of the Leu219-Pro220 peptide bond is eliminated. The amplitudes of each respective phase show excellent agreement between wild-type pelC and P220A. There is a slight increase in the amplitude of the fast phase in P220A. One might expect this increase to be larger. In the mutant the slow phase is eliminated. The molecules that would have populated that phase should now be

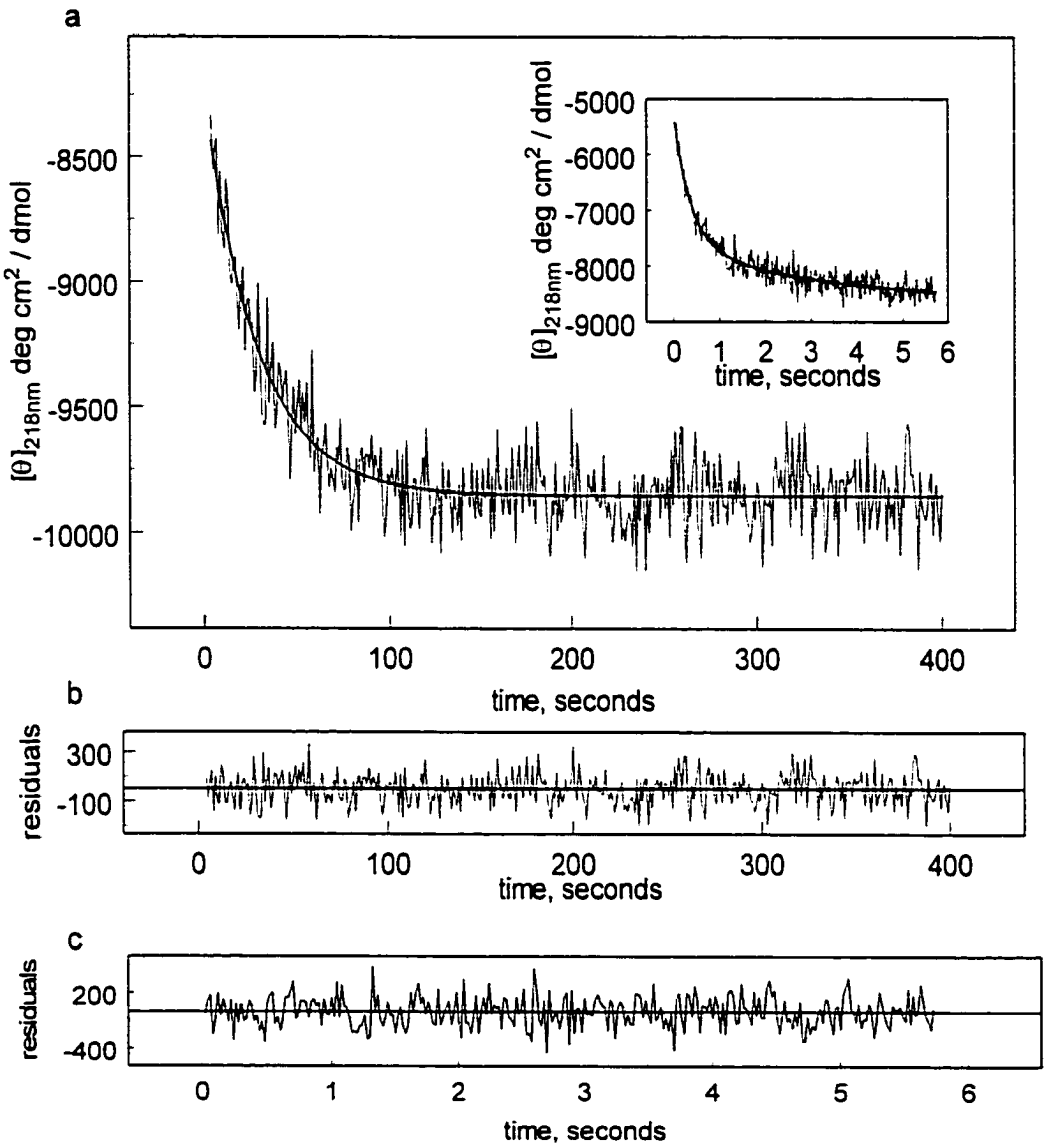


Figure 5.7: *a*. Refolding kinetics of P220A in 0.3M gdn-HCl monitored by far-UV CD at 218 nm. The solid line through the data is a single exponential fit. The inset to *a* shows the first six seconds of the folding process measured by stopped-flow CD. *b* and *c* show the residuals to the fits for the slow and fast phases.

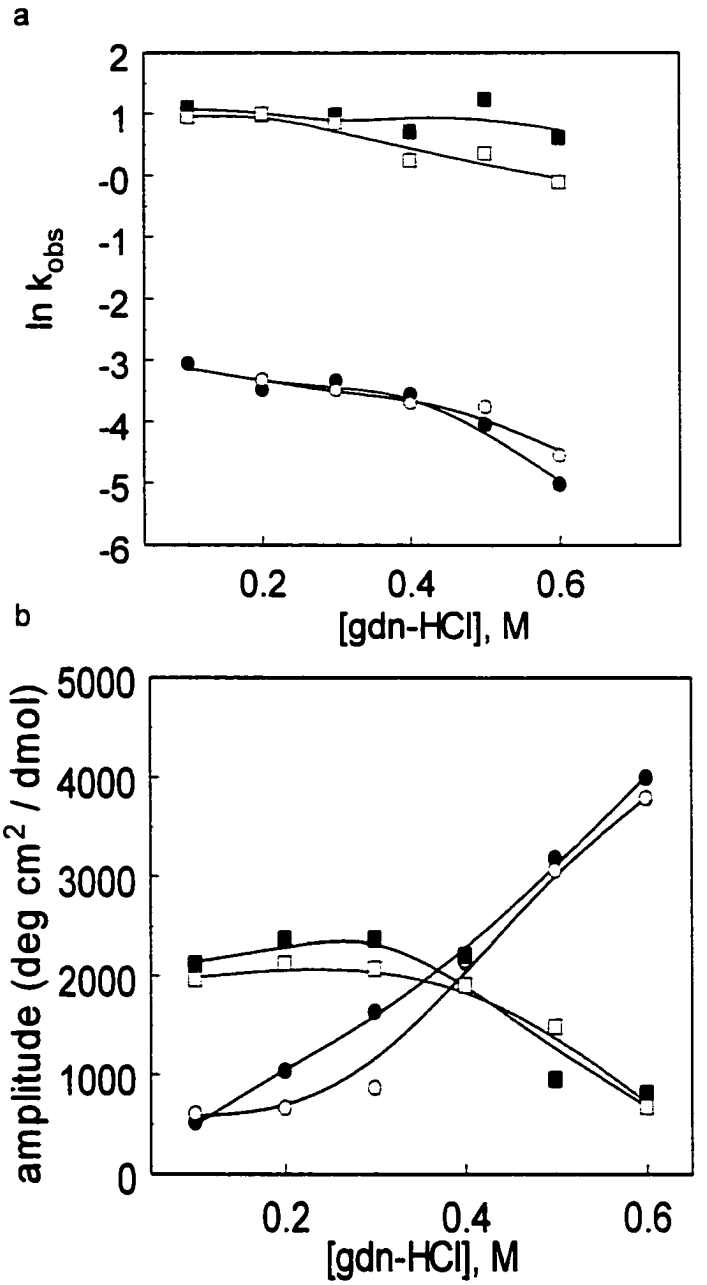


Figure 5.8: *a*. The natural log of the observed rate constants for P220A (filled symbols) and wild type pelC (open symbols); for the fast phase (squares) and 21-second phase (circles). *b*. Folding amplitude for each phase in P220A and wild type pelC folding. Solid lines are a weighted least-squares fit of the data.

able to populate the fast phase. The increase is not large because the equilibrium amplitude is lower in the mutant, resulting in a smaller total change.

It was observed that a pH jump from 6 to 2 resulted in extremely rapid and reversible unfolding of P220A. This change occurs in the dead time of manual mixing. pH jumps from 2 to 6 were carried out to determine the folding rate of the observed slow phase in denaturant-free buffer. This experiment is a good control to verify that extrapolating to 0 M denaturant is in fact accurate. CD was not used because the folding amplitude was less than the noise level of the instrument under the conditions employed. Fluorescence emission measurements gave good signals and low noise levels. Figure 5.9 shows a pH jump from 2 to 6 at 25° C. These data fit to two exponentials, in agreement with near-UV CD and fluorescence results from wild-type pelC. The slow phase has a relaxation time of τ = 21.0 seconds. This is in quantitative agreement with the rate extrapolated from the CD results in gdn-HCl for wild-type pelC.

Analysis of the 21-second phase in P220A. In wild-type pelC it was demonstrated by double-jump assays that the 21-second phase was also due to proline isomerization in the unfolded protein. Double-jump assays were carried out on P220A to further assess if this phase is the same in P220A as in wild-type pelC. Both CD and fluorescence were used for this experiment. In both cases, an increase in the folding amplitude is observed with increasing unfolding delay time (Figure 5.10). Previously, with wild-type pelC, CD could not be used to study the 21-second phase, but the mutant shows a positive result for this

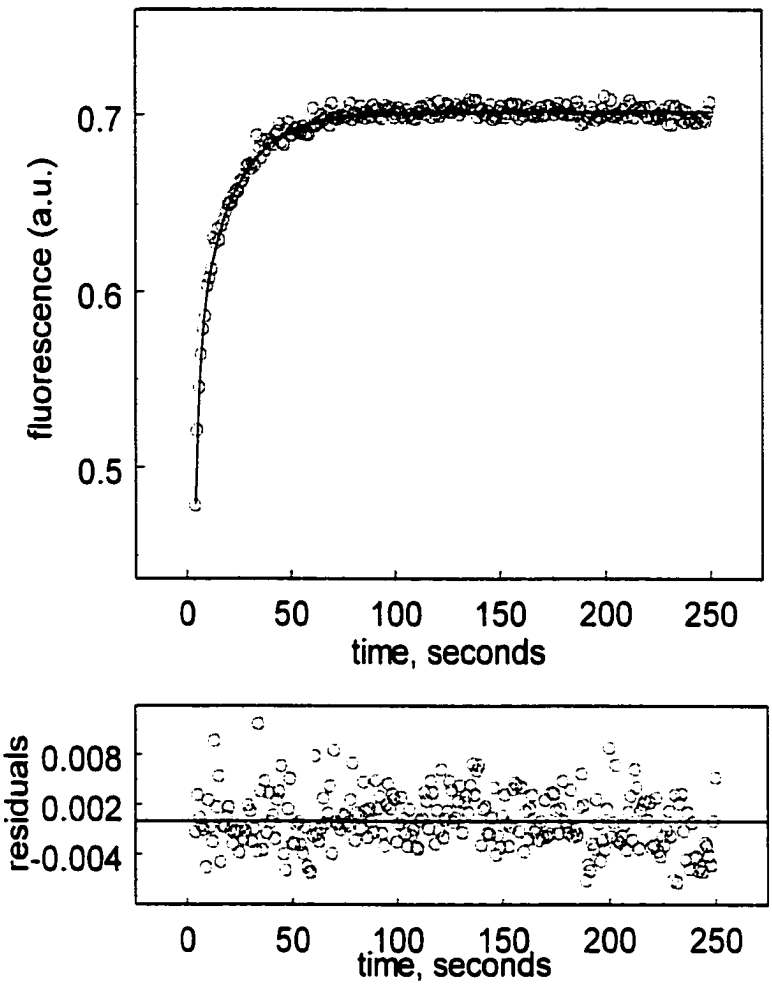


Figure 5.9: Folding of P220A monitored by fluorescence emission at 326 nm. Excitation was at 290 nm. Unfolding was initiated by lowering the pH to 2. Refolding was then initiated by a jump in the pH from 2 to 6. Residuals for the two-exponential fit are shown.

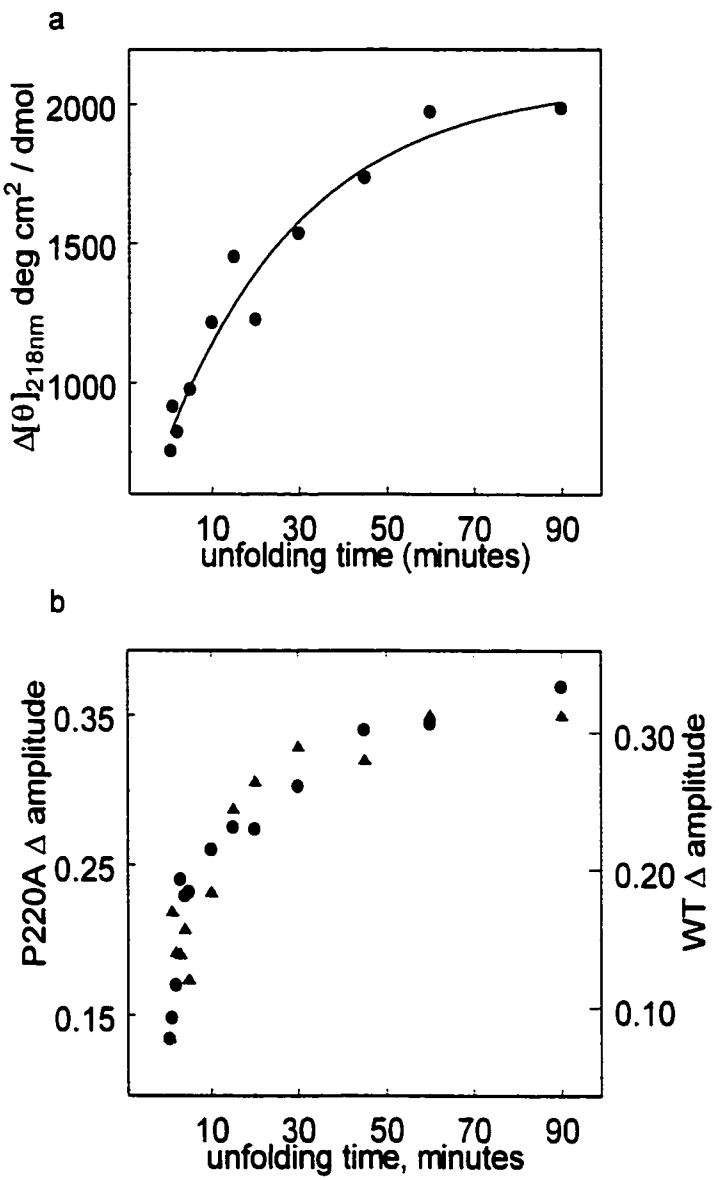


Figure 5.10: Double-jump refolding experiments for P220A monitored by CD at 218 nm (*a*), and by fluorescence emission (*b*). In (*b*) P220A is represented by (●) and wild-type pelC is represented by (▲). Each is scaled to their respective axes.

experiment (Figure 5.10a). The results obtained for the mutant by fluorescence are in excellent agreement with those for wild-type pelC. The relaxation time for isomerization for wild type pelC at 0° C and pH 9.5 is 16.5 minutes and P220A is 14.7 minutes. These data, combined with the refolding kinetics, suggest that the same proline causes the slow kinetics in wild type pelC and P220A.

Activation enthalpy of the 21-second phase in P220A. In wild-type pelC it was difficult to obtain an accurate measurement of the activation enthalpy ($\Delta H^{\ddagger}$) for the 21-second phase. This was because the amplitude was low and an extrapolation to 0 M denaturant concentration was necessary. An accurate measurement of $\Delta H^{\ddagger}$ was obtained for P220A. Because the same event is going on in wild-type pelC as in the mutant this can serve as an accurate measurement for the wild-type. Refolding was initiated by a pH jump from 2-6. The refolding kinetics were monitored by fluorescence emission at temperatures from 5° to 25°C. Figure 5.11 shows the *ln* k vs. temperature and the Eyring plots to determine $\Delta H^{\ddagger}$. This method yielded excellent data with very little scatter in the individual points. The results were $\Delta H^{\ddagger} = 16.0 \pm 0.7$ kcal / mol, and $\Delta S^{\ddagger} = -11.2 \pm 2.6$ cal / mol K. This value for $\Delta H^{\ddagger}$ is in agreement with model peptides and other known proteins (Brandts et al. 1975). This again provides further support to the claim that this phase is also due to proline isomerization.

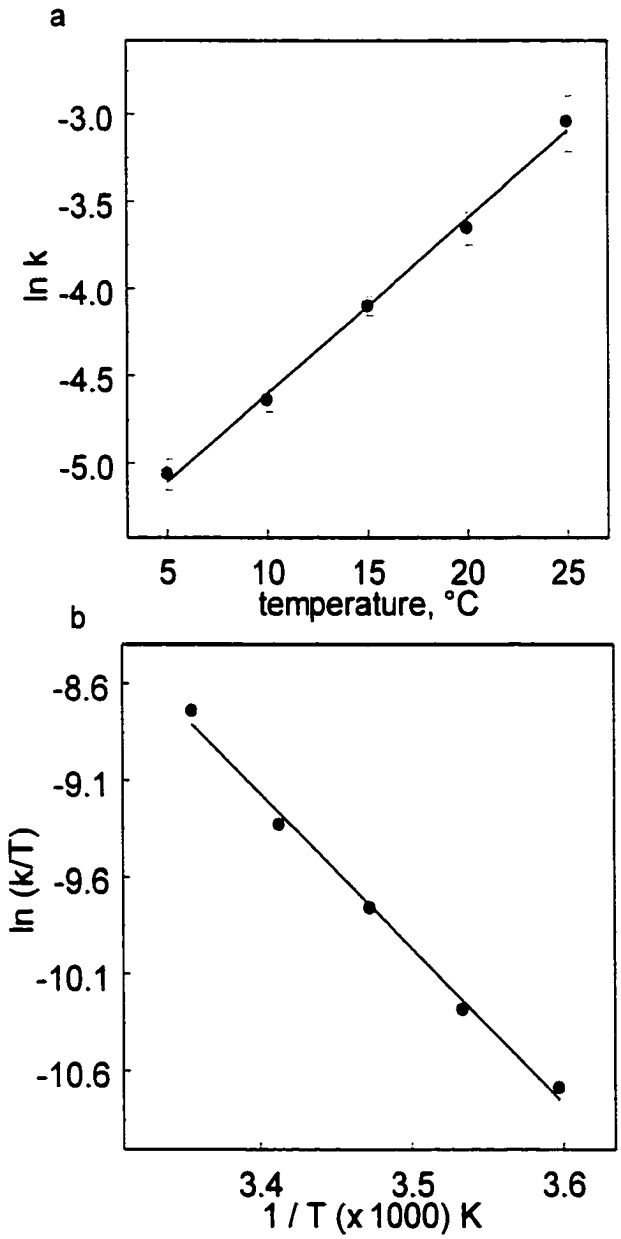


Figure 5.11: *a*. Temperature dependence of the observed rate constant for the slow folding phase of P220A monitored by fluorescence emission. The error bars represent the standard deviation for the average of three runs for each temperature and the line is a straight-line fit. *b*. Eyring plot for P220A. The data were taken from *a*.

Structure of trans proline to alanine mutants. In order to identify the proline responsible for the 21-second folding phase, 9 of the remaining 11 prolines were mutated to alanine. The mutagenesis and purification procedure used for P220A worked well for all of these other mutants. DNA sequencing demonstrated that the mutation was present in all cases.

A possible result of any mutation is an observable change in the overall secondary or tertiary structure of the protein. CD experiments were carried out on all Pro→Ala mutants of pelC. Figure 5.12 shows the far-UV CD spectrum of each mutant compared to that of the wild-type pelC. In most cases the mutant CD spectrum is almost identical to the wild-type spectrum. P139A appears to be slightly less intense in the long-wavelength region and even more so in the short-wavelength region. This is partly due to the fact that P139A is a double mutation with P220A. This may explain the loss of intensity in the 218nm region but not the 190nm region. Pro139 is located within the T1 turn. This is coincidentally next to Ser138 of the serine stack. It is possible that this turn is disrupted, thus resulting in the loss of some interactions. The CD spectrum of P263A appears to have some difference compared to that of wild-type pelC. Pro263 is also located in a tight turn (T3) at the end of a β-strand. Replacement by alanine could introduce some flexibility normally restricted by the proline, thus causing some loss of interactions. The far-UV CD spectra of all of the other mutants (Pro 51, 159, 286, 302, 316, 324, and 335) show no significant changes. Most of these mutants are located in regions of the protein that have been assigned as unordered. Only Pro335 is involved in some secondary structure, the middle

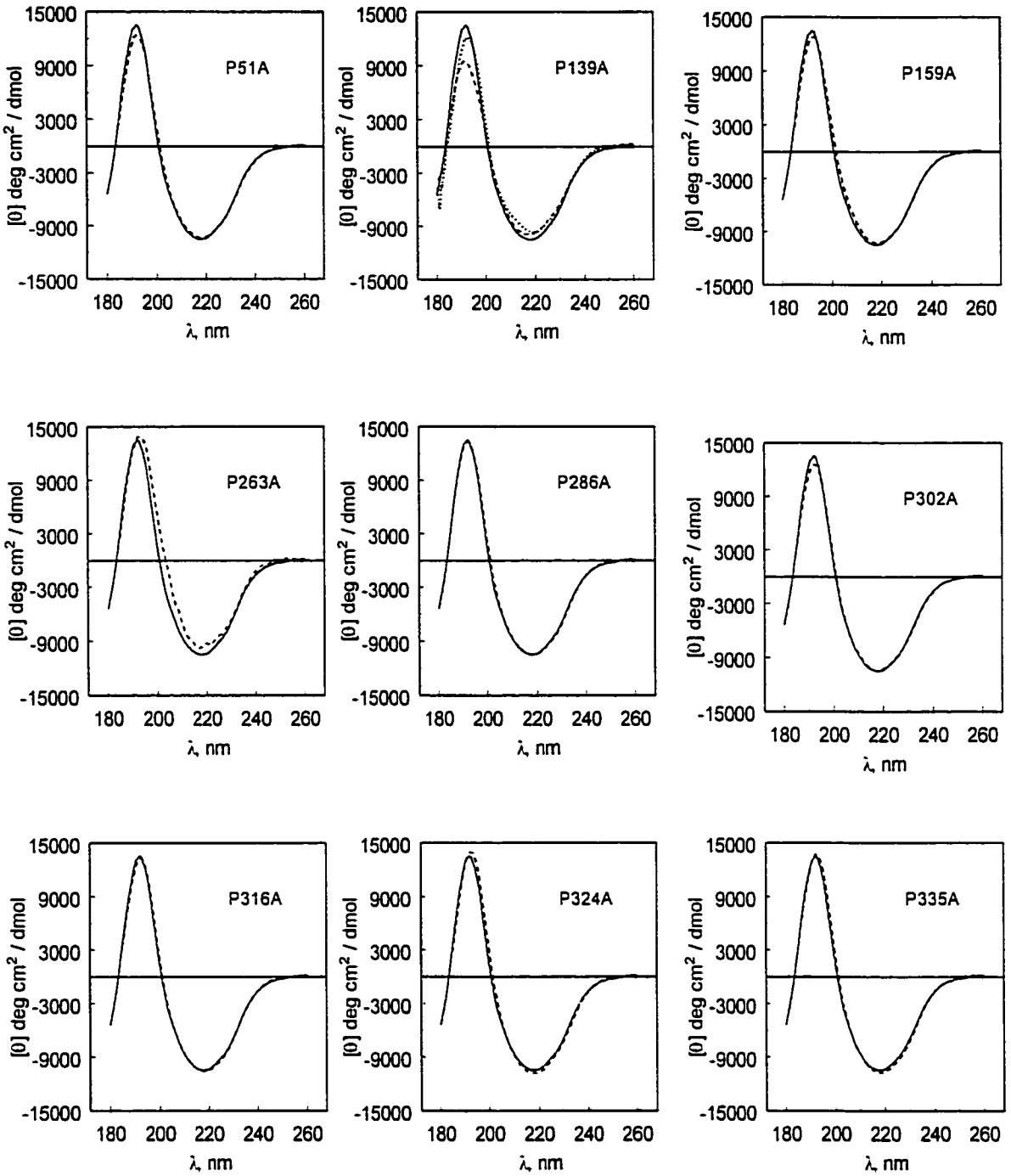


Figure 5.12: Far-UV CD spectra for the proline-alanine mutants of pelC involved in *trans* peptide bonds. In all cases the mutant spectrum is the dashed line and wild type pelC is shown as a solid line. For P139A, the spectrum of P220A is shown as a dotted line.

residue of a 3_{10} helix. Pro286 is located at the N-terminus of a short α-helix, and Pro51 is located at the N-terminus of a β-strand. Mutation of these prolines to alanine appears to have no noticeable affect on the secondary structure.

The P220A mutation had no effect on the near-UV CD spectrum. However, several of the *trans* proline mutants showed small effects (Figure 5.13). P139A showed the largest of these differences. This is not due to the double mutation with P220A because the latter mutation results in no change to the near-UV CD. Trp142 is close in sequence to Pro139. Distortion of the local sequence could change the geometry of Trp142, resulting in a loss of CD intensity or a change in sign. Pro159 also shows a lower near-UV CD intensity. Trp75 is only about 4Å away. Pro159 caps a short segment of α-helix. Mutation to alanine might result in a slightly longer helix. This might result in altered excited-state coupling between the tryptophan indole ring and the peptide groups in the helical geometry. The CD spectrum of P316A also shows some loss of intensity. Phe315 might have its geometry distorted. This residue is within 4Å of several tyrosines involved in an aromatic residue cluster between the N and C termini. Distortion of this geometry might affect coupling within this cluster.

Refolding of trans mutants, Pro→Ala. Mutation of Pro220 to Ala resulted in the elimination of the slowest phase of folding. The intermediate slow phase is also due to proline isomerization. The Pro→Ala mutants that were made were subjected to kinetic experiments in order to identify the essential proline responsible for this phase. Kinetic refolding experiments were performed on

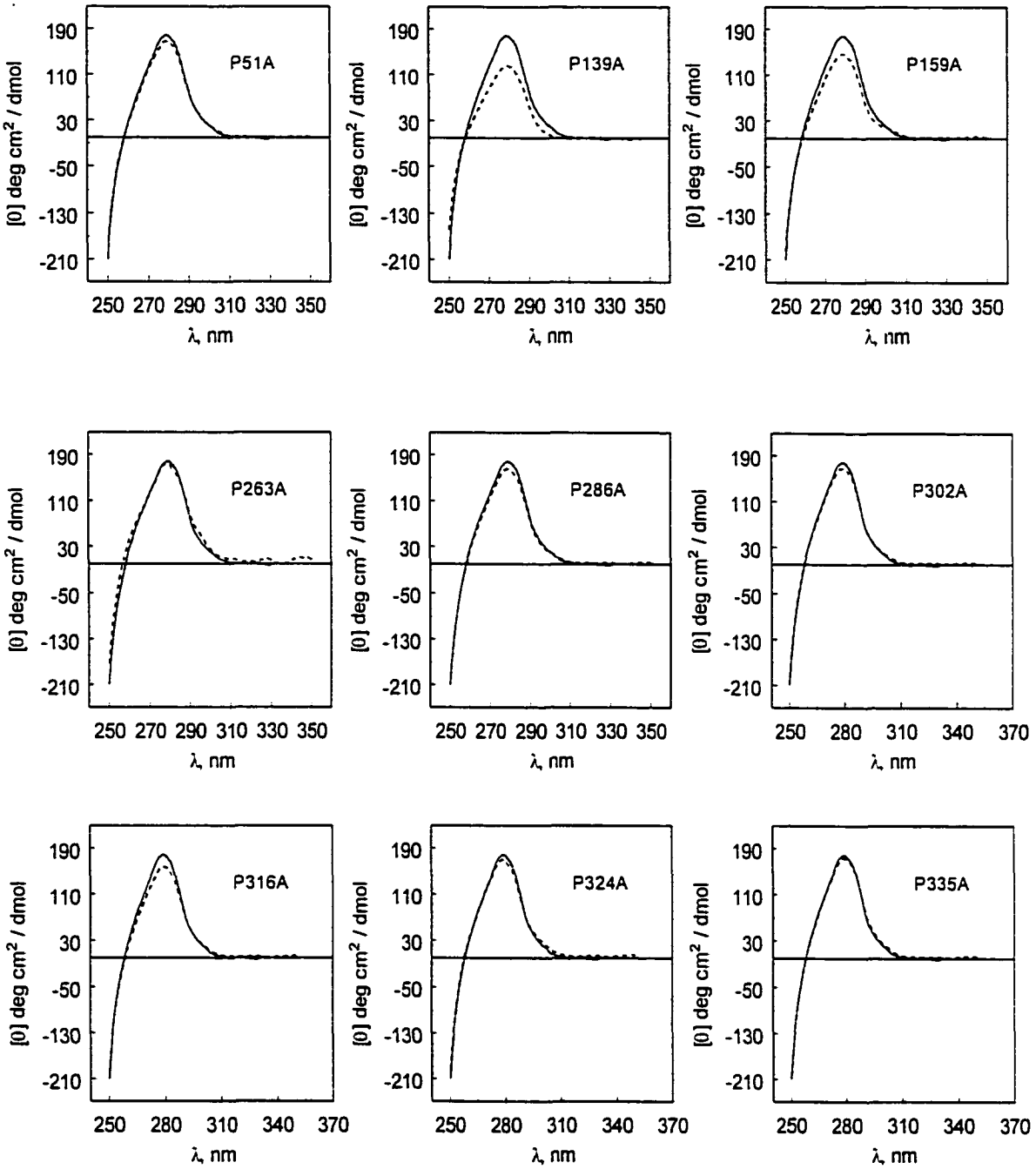


Figure 5.13: Near-UV CD spectra for the Pro→Ala mutants of pelC involved in *trans* peptide bonds. In all cases the mutant spectrum is the dashed line and wild type pelC is shown as a solid line.

these mutants in a manner identical to those for the wild-type protein. Figure 5.14 shows the kinetic refolding experiments for all of the Pro→Ala mutations. In all cases the data show two kinetic phases with rates and amplitudes comparable to wild type pelC. P139A/P220A shows a single phase with kinetic parameters comparable to the respective intermediate phase in wild type pelC. Table 5.4 lists the kinetic parameters from these experiments. It appears that none of these Pro residues are solely responsible for the observed 21-second phase.

CD spectra of Pro→Val mutants. In two cases, the Pro→Ala mutant could not be expressed at significant levels. These cases were Pro79 and Pro124. It is interesting that these residues, although 45 residues apart in sequence, are within about 6Å of each other in the tertiary structure. Both of these amino acids have ϕ and ψ angles that correspond to the poly-proline II structure. Alanine does not preferentially populate this conformation. However, valine and other β-branched amino acids do have some preference for this structure (O'Connell et al. 1999). These two prolines were mutated to valine and experiments similar to those for the alanine mutants were carried out. Figure 5.15 shows the CD spectra of P79V and P124V compared to the spectra of wild-type pelC. The far-UV CD spectra of both mutants are comparable to those for the wild-type. The only notable difference is the intensity of the 190 nm band. In both mutants, this is decreased slightly. Both spectra differ in the intensity of the near-UV band also. There is probably some distortion in the structure of near-by aromatic side chains, but it is not substantial. Overall the mutations do not significantly perturb

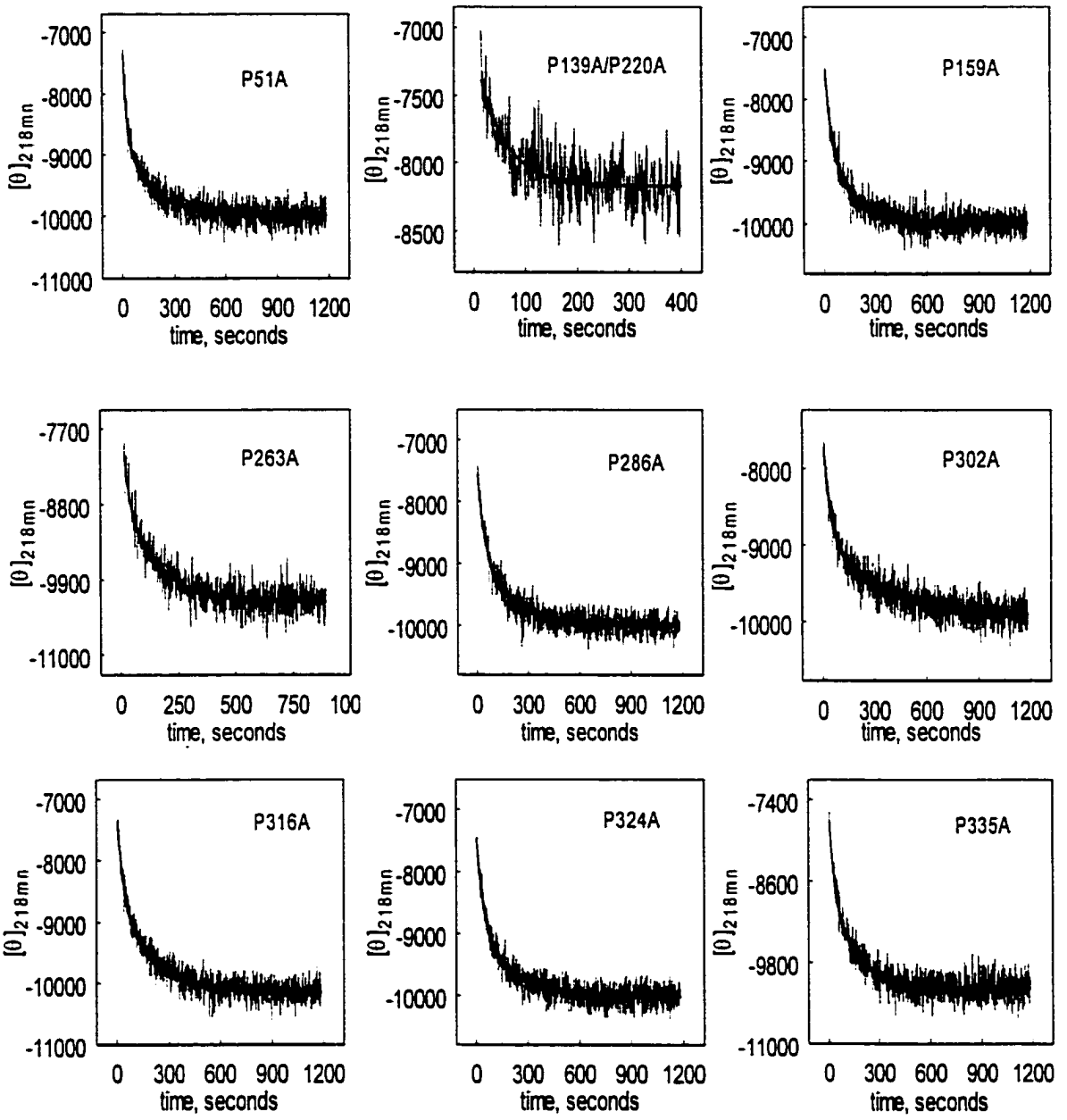


Figure 5.14: Refolding kinetic traces for pelC Pro→Ala mutants studied by manual mixing and CD at 218 nm. All curves are fit to double-exponential models except P139A/P220A, which is fit to a single exponential. Final conditions for all experiments are 0.3 M gdn-HCl, pH 6 MES 5 mM, NaCl 50 mM and 25° C. The units for the ordinate are deg cm^2 / dmol.

Table 5.4: Rates and amplitudes for WT and Pro→Ala or Val pelC mutants

Mutant	k_2 sec^{-1}	k_3 sec^{-1}	a_2 deg cm^2 / dmol	a_3 deg cm^2 / dmol
P51A	0.0365	0.0065	1580	1353
P139A	0.0187		1040	
P159A	0.0370	0.0079	1338	1292
P263A	0.0365	0.0068	1403	1570
P286A	0.0251	0.0057	1349	1244
P302A	0.0218	0.0030	1317	1003
P316A	0.0263	0.0051	1385	1562
P324A	0.0286	0.0083	1529	1178
P335A	0.0342	0.0070	1043	1594
P79V	0.0278	0.0072	2016	1396
P124V	0.0225	0.0080	2171	1510
WT	0.0307	0.0068	1600	1700

Conditions for the kinetic experiments were 25° C, pH 6 MES 5mM, NaCl 50mM.

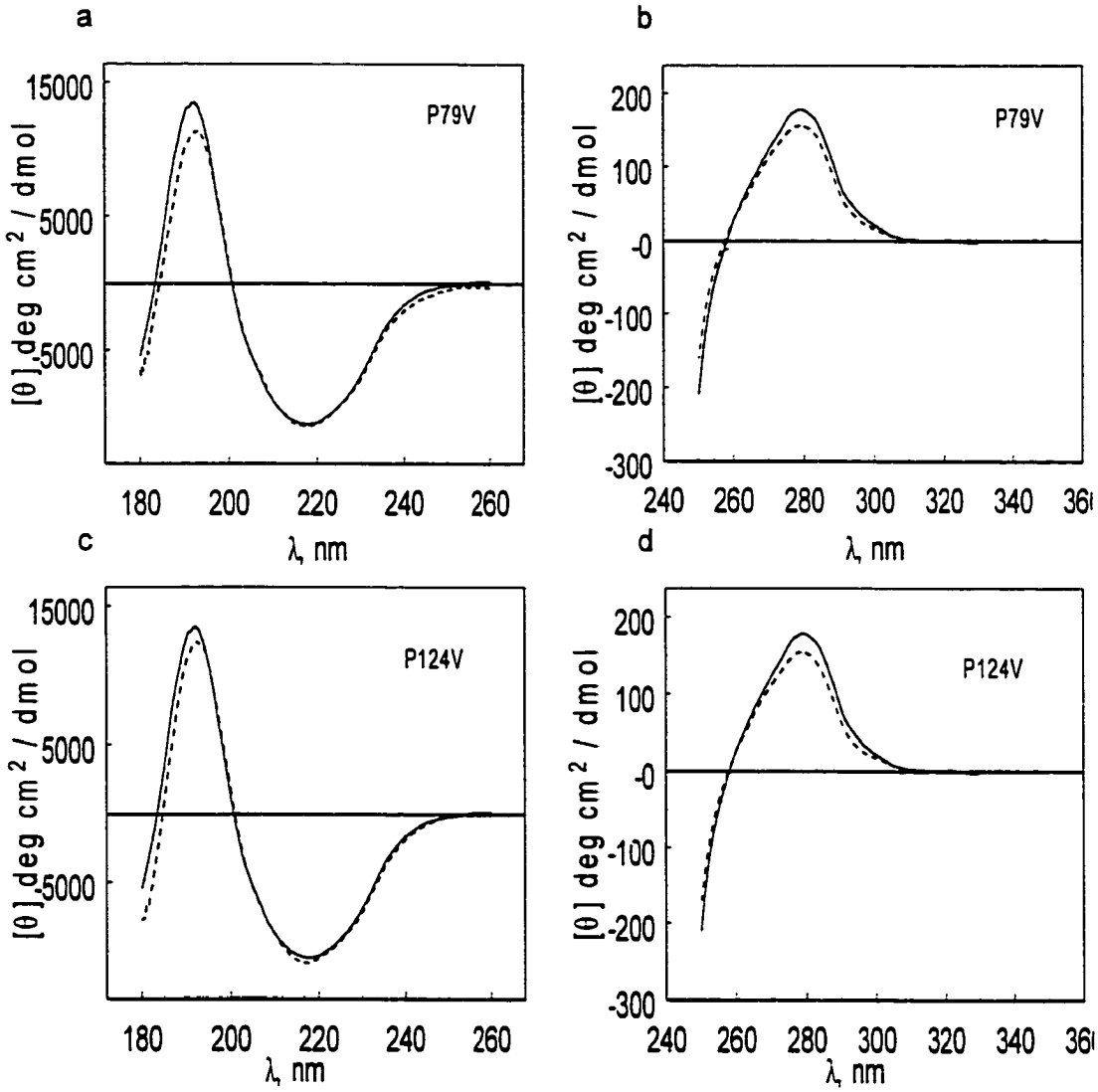


Figure 5.15: CD spectra of P79V (a, b) and P124V (c, d) compared to the spectra of wild-type pelC. The dashed lines represent the mutants and the solid lines represent the wild-type protein.

the structure of pelC to the point that kinetic experiments are likely to show a substantial change in the folding mechanism.

Folding kinetics of P79V and P124V. Refolding kinetic experiments on the Pro→Val mutants were carried out identically to those for the Pro→Ala mutants. Figure 5.16 shows the refolding kinetic experiments monitored by far-UV CD. In both cases, as with the previous Pro→Ala mutations, the data fit to the sum of two exponential terms with rate constants and amplitudes similar to those for the wild-type protein. Actually, the amplitude for the intermediate slow phase is larger for these two mutants. It is this phase that should be eliminated upon mutation of an essential proline. It appears that neither of these prolines is responsible for the intermediate slow phase.

5.4 Discussion

The slowest step in folding is due to the isomerization of Leu219-Pro220 peptide bond. The hypothesis that slow folding results from proline isomerization first was proposed in 1975 (Brandts et al. 1975). It was suggested that a general intermediate in folding was the result of incorrect *cis-trans* prolyl peptide bonds. In many small- to medium-sized proteins lacking *cis* peptide bonds, no kinetic phases slower than 1-5 seconds are observed. But in almost all proteins containing *cis* prolyl peptide bonds, slow phases are observed ranging from tens

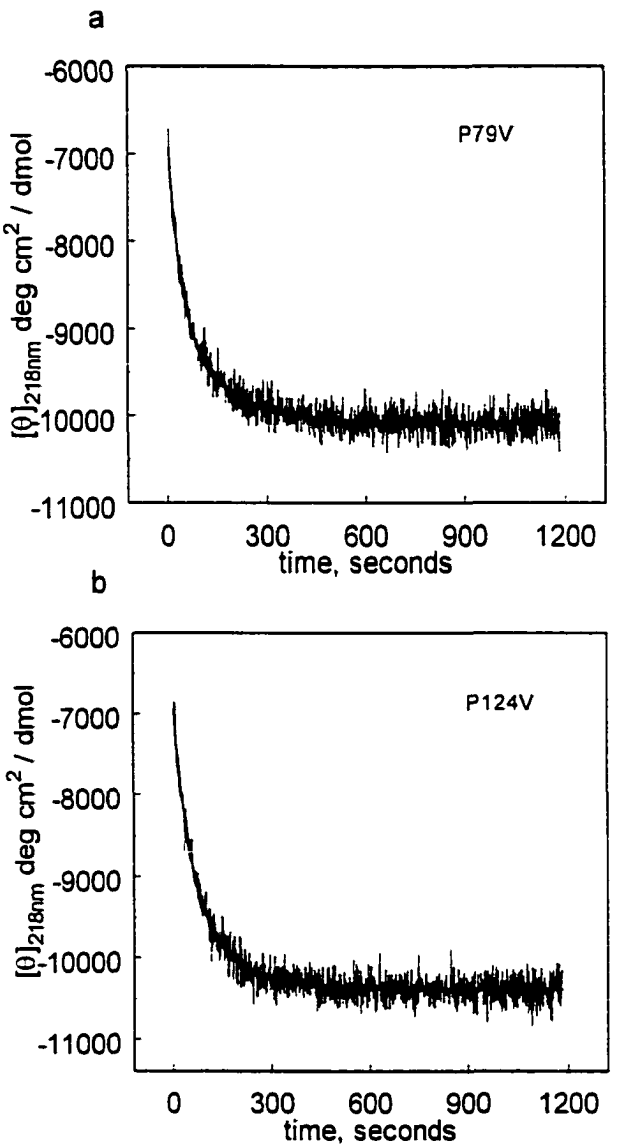


Figure 5.16: Folding kinetics of (a) P79V and (b) P124V. The experiments show the CD at 218 nm as a function of time. The solid line in both figures represents a double-exponential fit. Final conditions are 0.3 M gdn-HCl, 5 mM MES pH 6, 50 mM NaCl, and 25° C.

of seconds to hours. In several cases, proteins containing prolyl peptide bonds in the *trans* conformation also show slow kinetics (Eyles and Gierasch 2000; Jackson and Fersht 1991; Jullien and Baldwin 1981; Ogasahara and Yutani 1997; Veeraraghavan and Nall 1994; Wood et al. 1988). The mutagenesis results presented conclusively prove that the slowest folding phase for pelC is due to isomerization of the Leu219-Pro220 peptide bond. Mutation to alanine eliminates that slow phase. Aside from that, this mutation has little effect on the overall secondary and tertiary structure and almost no effect on the refolding mechanism.

The catalytic activity of P220A is reduced to levels uncharacteristic of the native enzyme. This is not surprising. Heffron *et al* carried out a structure-based sequence alignment of 14 pectate lyases from five organisms (Heffron et al. 1995). Pro220 is strictly conserved in all of these proteins. In fact it is the only proline so strictly conserved out of the 12 in *E. chrysanthemi* pelC. This amino acid is obviously functionally important. It maintains the *cis* form in order to align properly another invariant residue, Arg218, for important catalytic function (Heffron et al. 1995).

The free energy of unfolding, $\Delta G°_{H_2O}$, is also significantly reduced in P220A, $\Delta\Delta G° \approx -4$ kcal /mol. This most likely results from the loss of some intramolecular interactions in the local region of the protein around the mutation. This proline is the first residue of a β-strand. Forcing the trans conformation could significantly distort the backbone. The *m* value is only slightly reduced by 0.6 kcal / mol M. This would suggest that the overall denaturation mechanism

remains unchanged. The only differences lie in the region of the mutation, where some buried surface area would be exposed to the solvent. It is also noted that the denaturation midpoint changes by -0.28 M gdn-HCl to 0.71 M. This is a result of the lower free energy of unfolding at essentially constant *m*.

The slow step in P220A folding is coupled to an intermediate. In simpler proteins, such as RNAase A, that exhibit slow steps in folding, the rate and amplitude of the slow steps are independent of the final folding conditions, such as denaturant concentration (Schmid and Baldwin 1979). This is because the population distribution in the unfolded protein is solely dependent on the *cis:trans* ratio in the unfolded protein. The percentage of molecules with the incorrect isomeric configuration is established in the unfolded protein. For wild-type pelC and P220A the rate and amplitude for the 21-second phase is strongly dependent on final denaturant concentration. The only explanation for this is that the isomerization is coupled with an intermediate (Jackson and Fersht 1991; Schmid 1986). At low gdn-HCl concentrations (0.1 M) the folding amplitude of the 21-second phase in WT pelC is about 11% of the total. One might expect a prolyl peptide bond to be 10-30% *cis* if unfolded. However, the amplitude increases dramatically at 0.6 M gdn-HCl to over 50%. In P220A these percentages are 25% and 90% respectively. At low denaturant, the intermediate is stable and most of the molecules populate this form. As with WT pelC, this intermediate is in rapid equilibrium with unfolded protein. At higher denaturant concentrations, a fraction of the intermediate can fold to the native state.

However, for the major population, formation of the intermediate might be the rate-limiting step, rather than the proline isomerization.

The 21-second phase is due to multiple proline isomerizations. Mutation of Pro220 results in elimination of the slowest kinetic phase. All evidence points to the notion that the intermediate slow phase is also due to proline isomerization. This evidence includes the slowly interconverting population of unfolded molecules, the 16 kcal / mol activation enthalpy, the relaxation time of 21 seconds and the effect of PPI. It is interesting that no single proline mutant results in elimination of this phase. In some cases the kinetic parameters are slightly different. P335A has an amplitude for this phase that is lower than the wild-type. Three mutants have rates that are noticeably lower than the wild-type rate. These are P139A/P220A, P302A, and P124V. They are only reduced by 30 or 40% however. It is puzzling that none of these mutants show a significant effect on the kinetics. Several groups have reported the elimination of slow kinetics by mutation of prolines (Eyles and Gierasch 2000; Kelley and Richards 1987; Wood et al. 1988) and results from the P220A mutation agree with these. Several other groups have reported that mutation of prolines does not eliminate slow phases (Schindler et al. 1996) or that certain slow phases are not the result of proline isomerization (Herning et al. 1991). A very interesting paper has recently reported that slow-folding phases can result from isomerization of non-prolyl peptide bonds (Pappenberger et al. 2001). It was suggested that a very small population of peptide bonds could exist as *cis* when the protein is unfolded.

Therefore, as the size of the protein increases, slow kinetics resulting from non-prolyl peptide bond isomerization will also be more prevalent. Maybe this phenomenon occurs in pelC, but this is uncertain. Indeed many proteins that lack prolines show no slow phases at all. Another possible scenario, similar to the above, is that the intermediate phase results from isomerization of more than one proline with similar activation energies. If several prolines are undergoing isomerization with similar transition state energies, and hence similar rates, their individual processes might not be resolved with the techniques employed. Considering the evidence in favor of an isomerization in the unfolded state, these two scenarios seem equally likely. One piece of evidence that tips the scales in favor of the multiple proline isomerization argument is the fact that the intermediate slow phase is catalyzed by cyclophilin (see Chapter 4). Considering this evidence, it is most likely that the intermediate slow folding phase results from an ensemble of prolines isomerizing from the non-native *cis* isomer to the native *trans* isomer.

Chapter 6

Quantum Chemical Calculations of the CD Spectrum of PelC

6.1 Circular Dichroism, an introduction

Circular dichroism is a form of optical spectroscopy that depends upon the interaction of light with chiral molecules. More specifically it is the phenomenon of differential absorption of right and left circularly polarized light by a chiral molecule. This technique is very similar to absorption spectroscopy in that absorption of a photon of a specific energy by a molecule will result in promotion of the molecule to an excited state of the same energy. The main difference is that in CD circularly polarized light is used.

Plane-polarized light can be decomposed into two circular components, left- and right-handed. Circularly polarized light can be constructed by combining two plane-polarized beams oriented perpendicular to each other and 90° out of phase. The tip of the resultant vector of light traces out a helix in space, and therefore is chiral itself. This is the foundation for its differential interaction with chiral molecules.

CD is a powerful technique for studying biological molecules. With respect to proteins, its usefulness is in determining low-resolution structures (Bienkiewicz et al. 2000), secondary structure analysis (Johnson 1999; Provencher and Glockner 1981; Sreerama et al. 1999), and observing structural transitions in the events of folding and ligand binding (Woody and Dunker 1996). The most abundant chromophore in proteins is the amide group. The amide group gives rise to at least three electronic transitions in the far-UV (Manning and Woody 1991; Clark 1995): two $\pi\pi^*$ transitions (NV_1 at 190 nm and NV_2 at 140

nm) and one $n\pi^*$ transition (220 nm). In proteins the exact wavelength, intensity and sign of each transition can vary. The nature of this variability lies in the conformation of each amide group. In other words, the far-UV CD spectrum is dependent on the secondary structure of the protein. This observation is crucial to one of CD's main applications, secondary structure analysis. It turns out that the experimentally observed far-UV CD spectrum of a protein is a linear combination of contributions from component secondary structures (Greenfield and Fasman 1969). This has allowed several groups to develop methods to obtain secondary structure fractions from CD (Johnson 1999; Provencher and Glöckner 1981; Sreerama et al. 1999).

In addition to the amide group, the side chains of aromatic amino acids are chromophores in proteins. It is well established that the side-chains of tryptophan, tyrosine and phenylalanine have absorption and CD bands in the near-UV region (250 - 350 nm) (Woody and Dunker 1996). CD in this region is a good probe for protein folding and ligand binding. One of the simplest critical tests for a two-state denaturation transition is to measure the transition with far- and near-UV CD. Since each probe monitors a distinct part of the protein (2° and 3° structure, respectively), coincidence of the transitions indicates a two-state process. Unfortunately, coincident transitions only suggest the possibility of a two-state transition, which must be proven in other ways.

Not only is side-chain CD observed in the near UV, but significant contributions to far-UV CD are made by tryptophan side chains. The indole B_b transition occurs near 220 nm and the magnitude can exceed 10^5 deg cm^2 / dmol

(Woody 1994). Often, these largely cancel, but in proteins with low helix content this can be significant. When coupling of two tryptophan side-chains is considered, noticeable contributions are easily made to the far-UV CD (Grishina and Woody 1994).

Extrinsic chromophores are important in the CD analysis of proteins. Porphyrin derivatives, such as heme (Blauer et al. 1993; Hsu and Woody 1971) are found in proteins. Small organic dyes, such as ANS (Sato and Woody 1980), congo red and cibacron blue (Edwards and Woody 1977; Edwards and Woody 1979) mimic the binding of cofactors to enzymes. Molecules such as these show little or no CD when alone in solution. When bound to a protein, a significant induced CD is often observed. In addition to distortion of the ligand, coupling with aromatic side-chains and peptide groups are responsible for the induced optical activity (Hsu & Woody 1971; Kiefl et al submitted). CD can reveal information about the structure of the binding site and conformational changes in the protein and ligand upon binding.

6.2 Theoretical calculations of protein CD spectra

Although CD is widely used experimentally, many important aspects of CD need to be described theoretically. When CD spectra can be calculated and quantitatively agree with experiments in all aspects, then there will be a complete understanding of the phenomena. Because CD deals with transition of electrons, quantum mechanics must be used to describe it mathematically. The

fundamental theoretical parameter in CD is the rotational strength. This parameter, experimentally derived from the area under a CD band, is analogous to the oscillator strength in absorption spectroscopy. In classical physics, this has the physical significance of the number of electrons undergoing the transition. The theoretical definition of the rotational strength is:

$$\mathbf{R} = \mathrm{Im}\{\mu_{0i} \cdot \mathbf{m}_{i0}\}$$

The rotational strength is the imaginary (Im) part of the scalar product of the electric dipole transition moment (μ_{0i}) and the magnetic dipole transition moment ($\mathbf{m}_{i0}$). The electric dipole transition moment can be thought of as a linear displacement of charge occurring upon transition from the ground to excited state. The magnetic dipole transition moment is a circular motion of charge associated with a transition. In order to have CD, a transition must have both linear and circular movement of charge. The axis of the circular motion must have a component along the linear displacement direction. The resultant displacement of charge is actually helical, with the helical sense being determined by the relative signs of **m** and μ. This helical sense of a transition is fundamental in the differential absorption of left- and right- circularly polarized light.

In proteins, the peptide and aromatic side-chain chromophores are achiral by themselves. Because they have some center or plane of symmetry, the rotational strength will vanish for all transitions. But, because they are part of a chiral molecule, i.e. a folded protein, the chromophores are distorted in their electronic structures and are no longer symmetric. Mixing of excited states, or

coupling of oscillators, can also give rise to optical activity in symmetric chromophores. For example, the peptide $\pi\pi^*$ transition is magnetically forbidden. It is a purely linear displacement of charge. But with respect to axes outside of the chromophore, the magnetic dipole transition moment for a transition with a nonzero electric dipole transition moment will be non-vanishing. The result is a coupled transition with both electric and magnetic dipole transition moments. In this case, the rotational strength will be non-vanishing.

Protein CD calculations have been extensively studied and reviewed (Koslowski et al. 2000; Woody 1996). For the most part, CD spectra calculated for α-helix-rich proteins using semi-empirical methods (Woody and Sreerama 1999) show reasonable agreement with experiment. Agreement is generally poorer with β-sheet-rich proteins. In addition, contributions of aromatic side-chains can be calculated. Absolute spectra often show rather poor agreement for β-rich proteins, but difference spectra in which individual aromatic side-chains are replaced with weaker or non-chromophoric side chains are in much better agreement with experiment (Grishina and Woody 1994).

The goals of the work presented in this chapter are two-fold. First, the far-UV CD spectrum of pelC appears to have a rather broad long-wavelength component. It actually might have a slight shoulder at ~230 nm. Calculations were carried out to determine if this is due to aromatic side-chain CD. Second, the folding kinetics of pelC in the near-UV indicate a change of sign in the fastest observed folding phase. Also, at pH 2 and high salt concentrations, the near-UV CD spectrum is negative, opposite in sign to the native spectrum. Calculations

were performed to see if any individual or cluster of aromatic side chains has a significant negative contribution to the CD. This will test the hypothesis that the intermediates have native character.

6.3 Methods

Calculations were performed as described by Woody and Sreerama (1999). The matrix method (Bayley et al. 1969) in its origin-independent form (Goux and Hooker 1980) was used. The calculations consider three peptide transitions: $n\pi^*$, $\pi\pi^*$ (NV_1), and $\pi\pi^*$ (NV_2). Transition moment directions for the secondary amide of N-acetylglycine were used for the $\pi\pi^*$ transitions (Clark 1995). Transition monopole charges for the $n\pi^*$ transition and for transitions connecting the $\pi\pi^*$ and $n\pi^*$ excited states were calculated from INDO/S wavefunctions (Ridley and Zerner 1973) for N-methylacetamide. The ground-state monopole charges were from Woody (1968). Side-chain transitions of Trp and Tyr were included using the parameters described by Grishina and Woody (1994). The monopole charges were located at atomic centers for aromatic side chains and at the positions given by the semiempirical Z values for amide transitions (Woody 1968).

6.4 Results and discussion

CD spectrum of pelC. The CD spectrum of pelC was calculated in the wavelength region of 299 - 120 nm. The calculation was done on pelC with all chromophores included, with peptide groups only, and with aromatic groups only. Figure 6.1a shows the calculated far-UV spectrum of pelC from 260 – 180 nm with all chromophores compared to the experimental spectrum. Overall the agreement is good for a beta-rich protein. The position of the short-wavelength band in the experiment is at 192 nm with an intensity of 13,500 deg cm^2 / dmol. The calculation puts this band at 188 nm with almost twice the intensity at 25,900 deg cm^2 / dmol. In the long-wavelength region the intensity is in much better agreement but the position of the minimum is further off. The calculation yields –11,500 deg cm^2 / dmol at 210 nm whereas the experiment yields –10,500 deg cm^2 / dmol at 218 nm. If only peptide transitions are considered, the agreement between theory and experiment is actually improved (Figure 6.1b). The 188 nm band intensity drops to 21,900 deg cm^2 / dmol and the long-wavelength band shifts to 215 nm with an intensity of –8650 deg cm^2 / dmol. Qualitatively, the spectra look more similar in the long-wavelength region as well. The calculated band appears broader, similar to the experiment. The calculation for pelC agrees well with the experimental spectra, relative to the calculation for other β-sheet-rich proteins. The reason probably lies in the regularity of the parallel β-helix structure, and the fact that there is not a lot of twist to the β-sheets. This provides a homogeneous environment, which the calculation assumes.

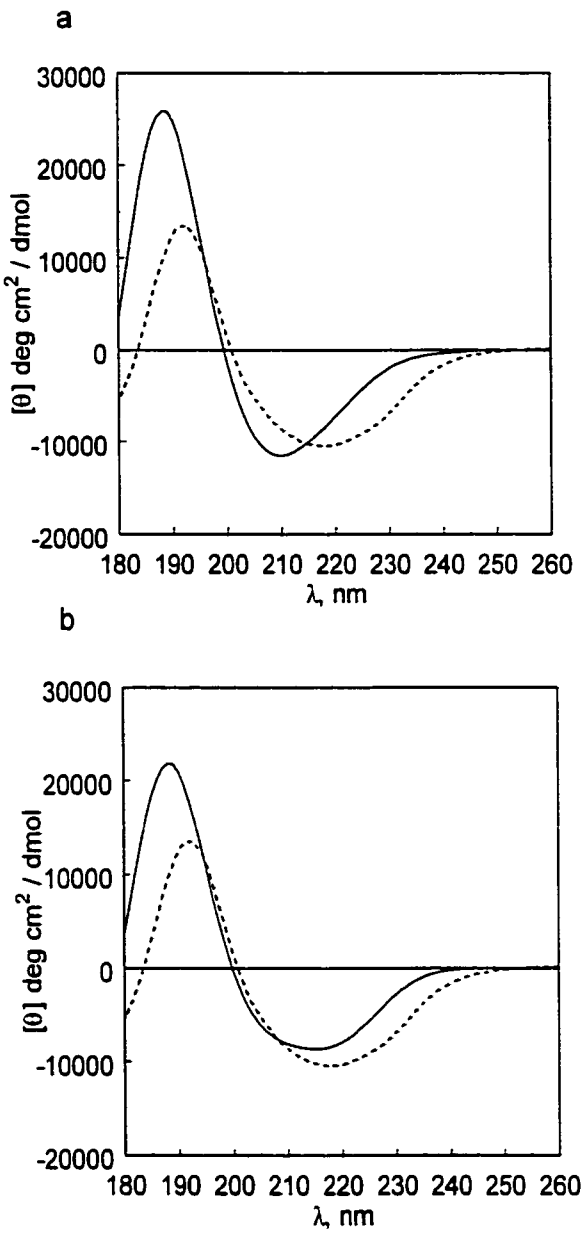


Figure 6.1: Calculated CD spectra of pelC including all chromophores (a), and excluding aromatic side-chains (b); calculated spectra (——) and experimental spectra (----).

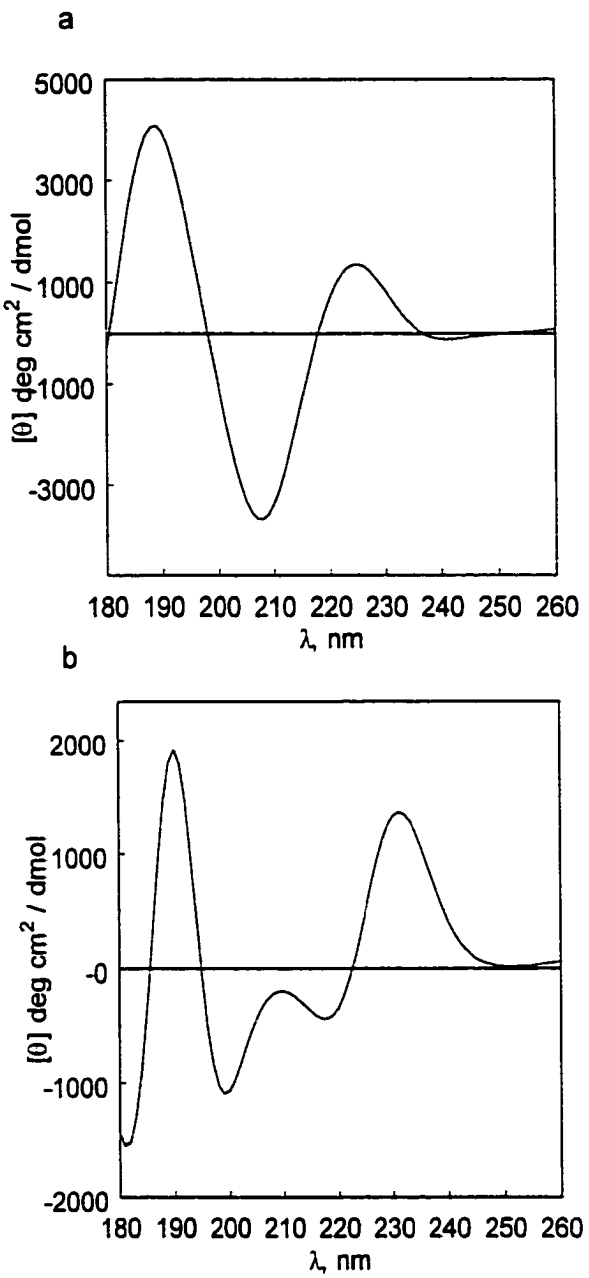


Figure 6.2: Aromatic side-chain contribution to the far-UV CD spectrum. Difference between all chromophores and peptides only (a), and calculation for aromatic side-chains only (b).

The contribution of aromatic side-chains to the far-UV CD was also investigated. Figure 6.2 shows the difference spectrum for all chromophores minus the peptides only, and the calculation for pelC excluding peptides. Qualitatively, the two spectra show some similarities. The short-wavelength band is at almost the same position for both spectra: 4080 deg cm^2 / dmol at 189 nm for the difference spectrum and 1920 deg cm^2 / dmol at 190 nm for the calculated aromatic side chain spectrum. The longest wavelength components agree well in intensity but differ in position. The difference spectrum approach is more accurate because it includes coupling of peptides with aromatics. Calculating the aromatics only is usually qualitatively correct only in the near UV and in the far UV at wavelengths above ~220 nm. Figures 6.3 and 6.4 show difference spectra indicating the contributions to the far-UV CD for individual tryptophan or tyrosine side-chains. In most cases for tryptophan, the major feature is a couplet above 220 nm. For tyrosine, the long-wavelength region of these difference spectra is more complex.

These results predict a positive contribution from aromatics at ~230 nm. Also, the peptides-only spectrum is in better agreement with experiment. These two factors suggest that the broadness of the band and the slight shoulder is probably not due to aromatic contributions. Actually, if one looks at the first derivative of the pelC spectrum, it appears that the maximum is in a similar position as the maximum in a model α-helix (Figure 6.5). In pelC there is a peak at 232.5 nm. In the α-helix, the peak is at 233 nm, and in a model β-sheet the peak is at 230 nm. It is therefore possible that there are noticeable contributions

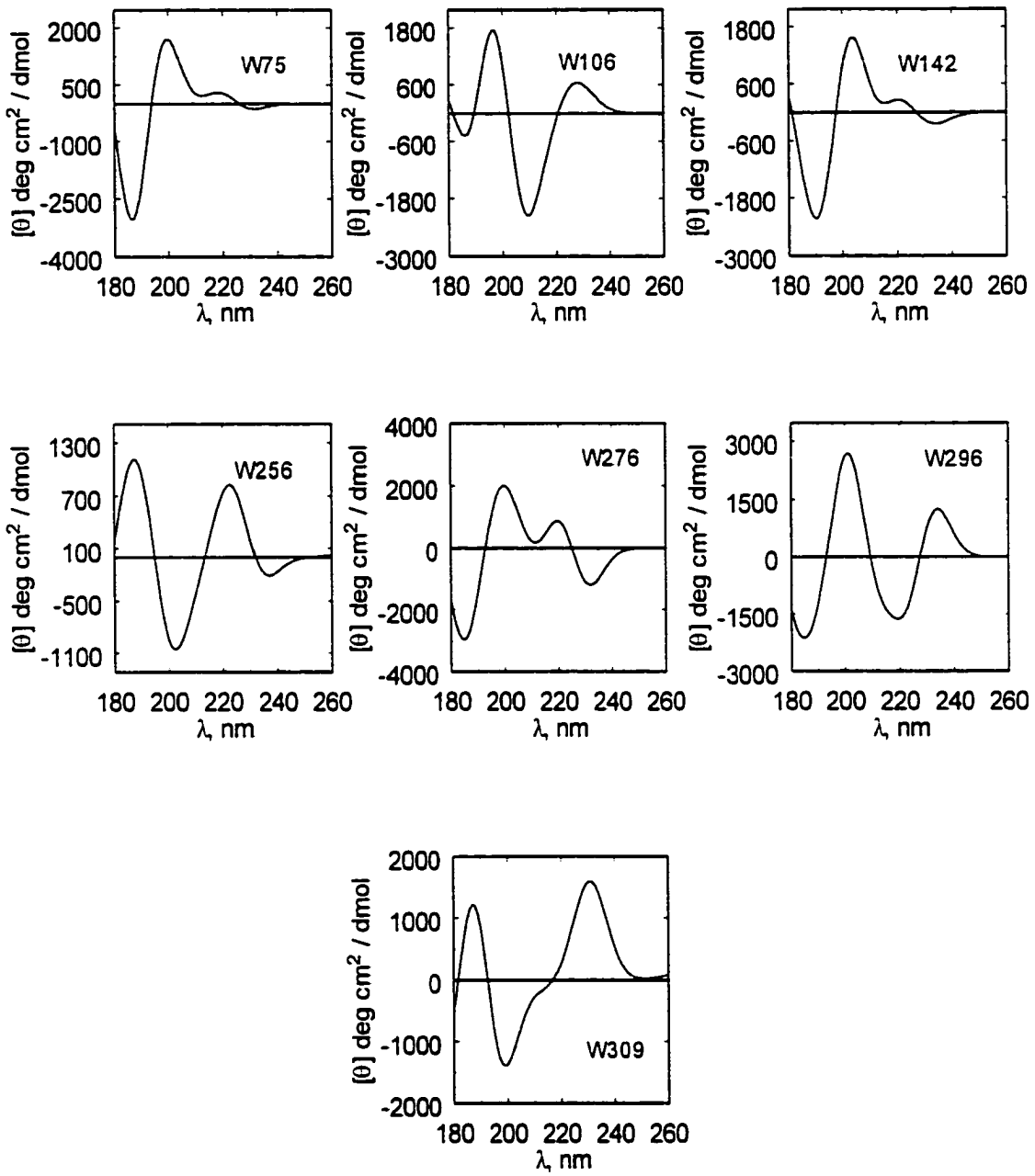


Figure 6.3: Contribution of individual tryptophan side-chains to the far-UV CD spectrum of pelC. Each spectrum is a difference spectrum between pelC with all chromophores and pelC with the specified side chain deleted.

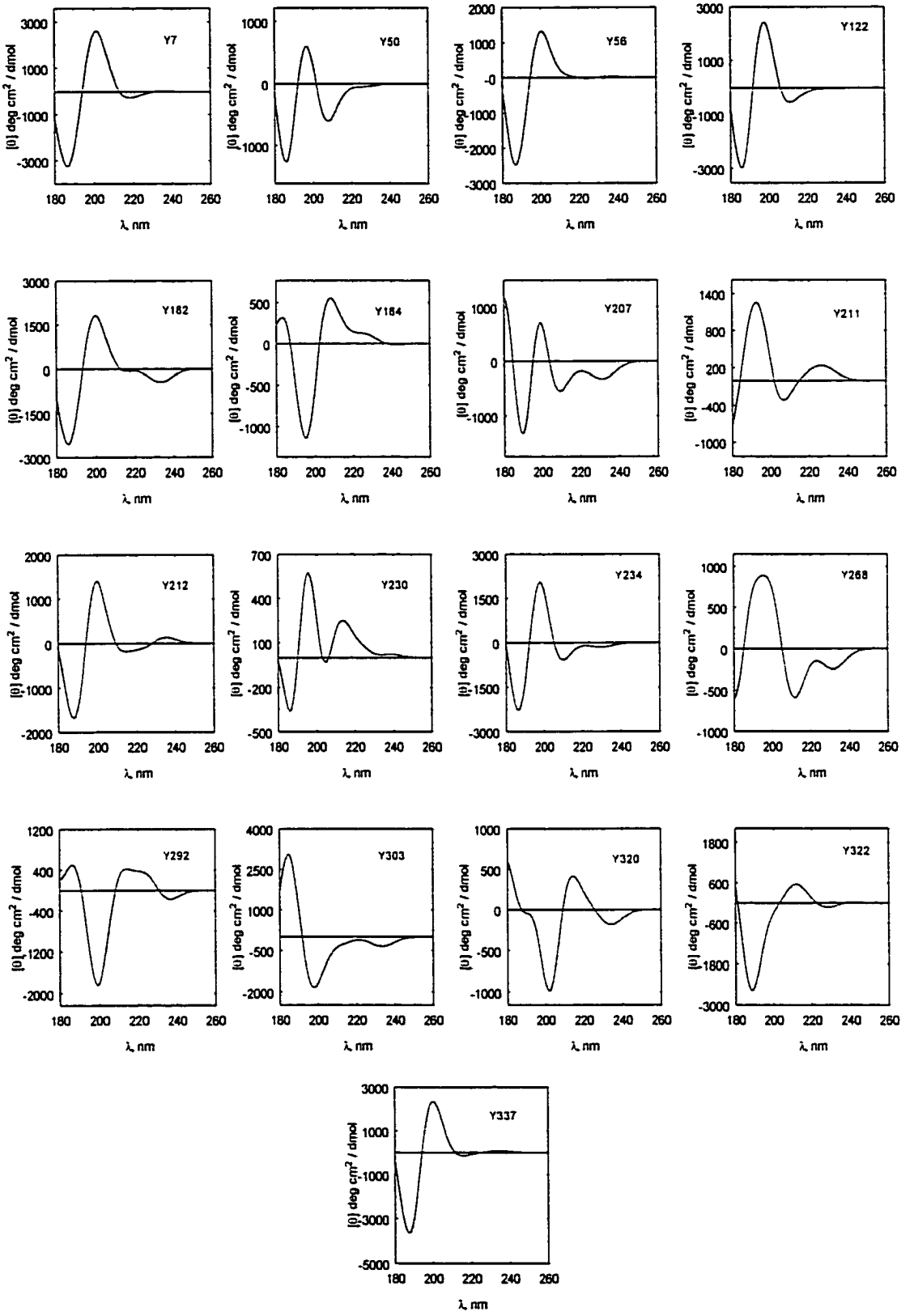


Figure 6.4: Contribution of individual tyrosine side-chains to the far-UV CD spectrum of pelC.

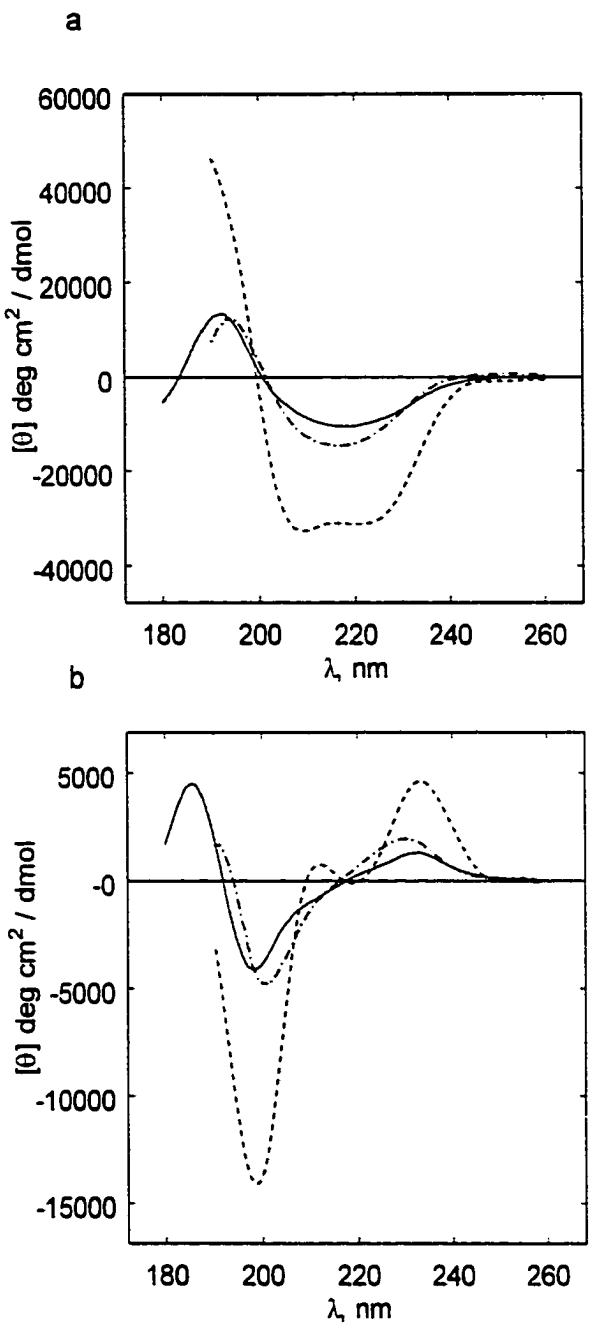


Figure 6.5: (a) CD spectra for pelC (——) and for poly(L-lysine) in α-helix conformation (----) and β-sheet conformation (—·—·—). (b) First derivative for the spectra in (a).

from α-helix, even though there is only ~10% helix content, compared to 30% β-sheet. It is also possible that the β-sheet transitions are spread out over a broader wavelength range, compared to the poly (L-lysine) model, resulting in a broad band. Aromatic side chains on the surface of the protein are more dynamic than buried ones. This may be the reason why the calculation with the peptides only agrees better with the experiment. The calculation uses a static crystal structure. Using a molecular dynamics simulation trajectory might improve the agreement.

The near-UV calculation is also in good agreement with the experiment (Figure 6.6). The experimental maximum is at 279 nm with an intensity of 174 deg cm^2 / dmol. The calculation predicts 231 deg cm^2 / dmol at 273 nm. Qualitatively, the features are also quite good. The shoulder at 290 nm is not predicted by the theory. However, this probably results from disulfides, which are not considered in the calculations.

Individual aromatic side-chain contributions to the near-UV CD. CD in the near-UV is much less predictable than the far-UV. In the far-UV the major chromophore is the peptide group. The predictability lies in the secondary structures that can be populated. In the near-UV the chromophores are primarily the side chains of tryptophan, tyrosine, and phenylalanine. These side chains can combine in all kinds of geometries and couple in an almost infinite variety of ways. Therefore, one can not relate the sign, shape or intensity of the CD

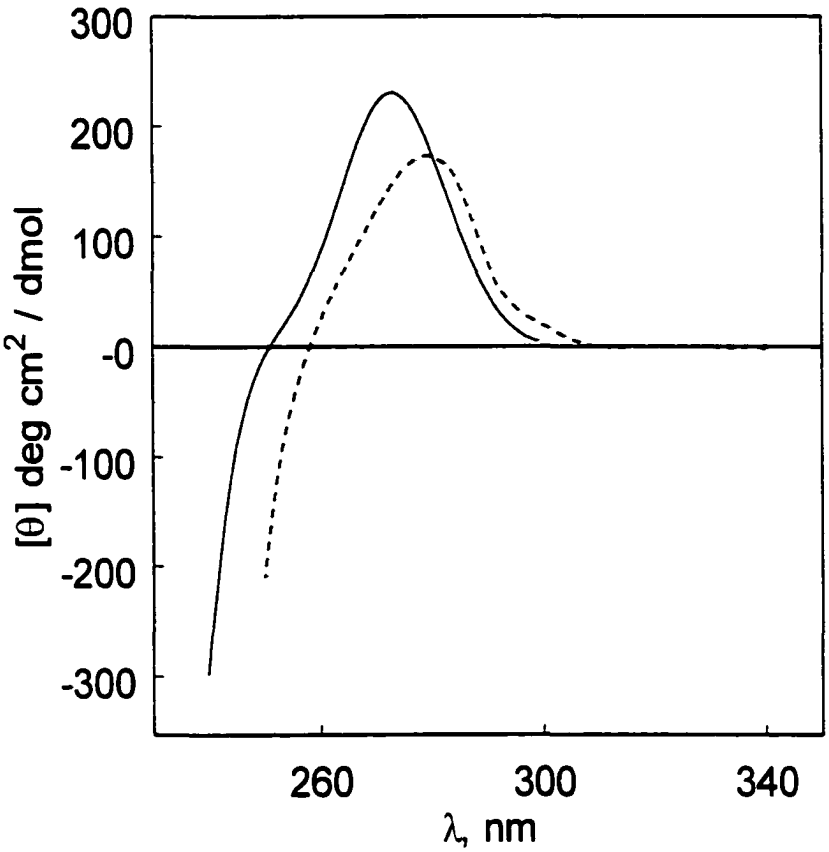


Figure 6.6: Calculated near-UV CD spectrum for pelC (——) and experimental spectrum for wild-type pelC (----).

spectrum to any particular structure. However, one can look at particular contributions, and changes to the near-UV CD spectrum with conditions.

It was noted in chapter 4 that when the folding kinetics are followed by near-UV CD there is a change of sign in the fastest observed folding phase. This phase occurs with a rate constant that is identical to that measured with far-UV CD. The experimental evidence so far can not distinguish between this species being native-like or not. It was proposed that a portion of the folding protein has native-like structure, and this portion makes a negative contribution to an overall positive CD. To possibly resolve this dilemma, calculations on pelC with each individual tyrosine and tryptophan changed to alanine were carried out. The difference spectra between the wild-type and mutant were then obtained.

Figures 6.7 and 6.8 show the difference spectra for all the mutations. The difference spectra show the individual contributions to the near-UV CD. Several side-chains make a negative contribution. If one looks at only the residues in the β-helix region one sees that the sum of the CD values at 277 nm is -8 deg cm^2 / dmol (Table 6.1). This is consistent with the magnitude of the fast phase in the refolding process. This supports, but does not prove, the hypothesis that the fastest phase in pelC folding results from the rapid formation of the β-helix domain. In a single step, the β-sheet backbone forms, and at the same rate the aromatic side-chains form a native-like structure. In a slower phase, the remainder of the side-chains relax to their minimum energy conformations. A corresponding phase is not observed in the far-UV because the majority of the rest of the molecule is unordered structure and there is no observable change in

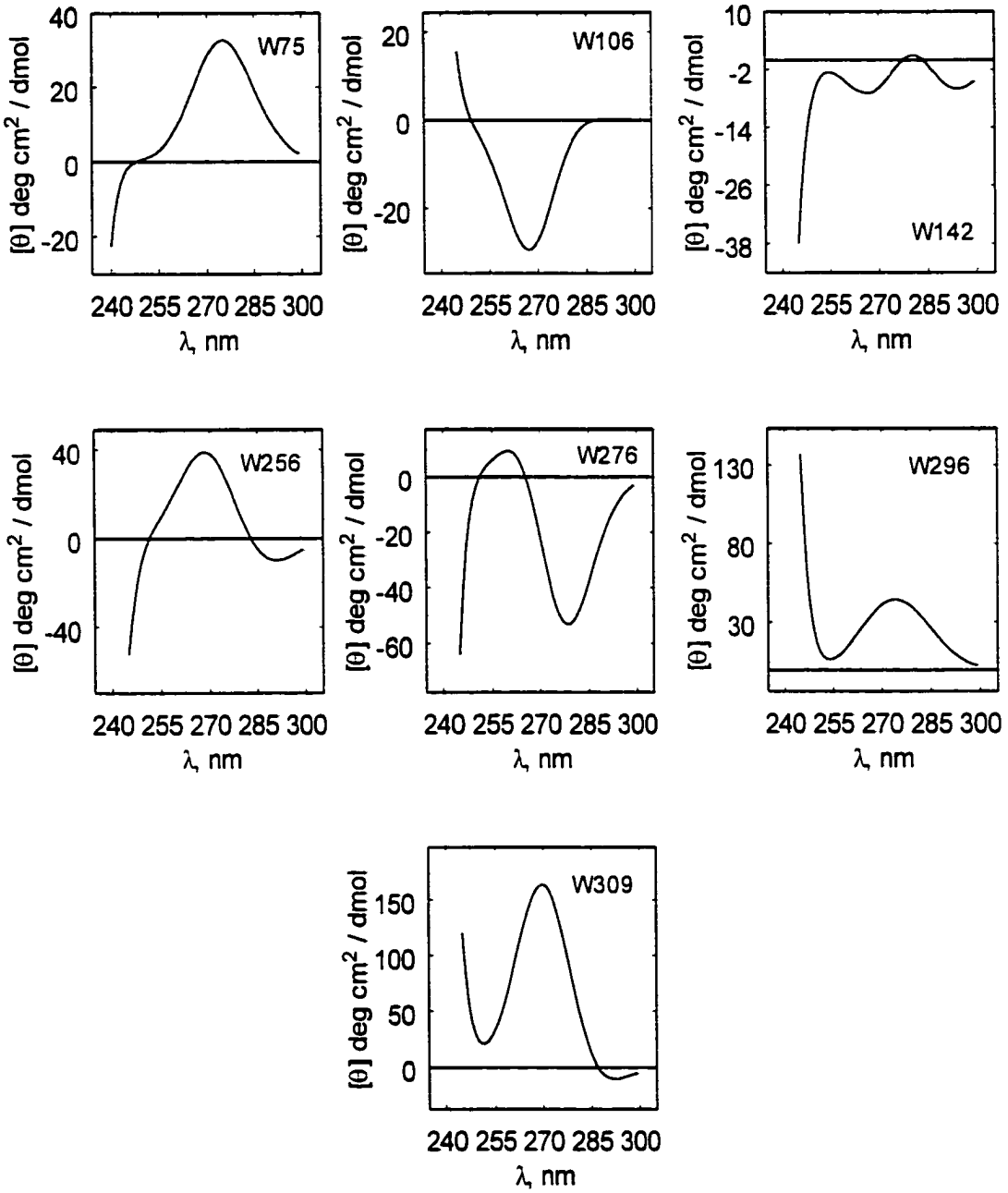


Figure 6.7: Contribution of individual tryptophan side-chains to the near-UV CD spectrum of pelC.

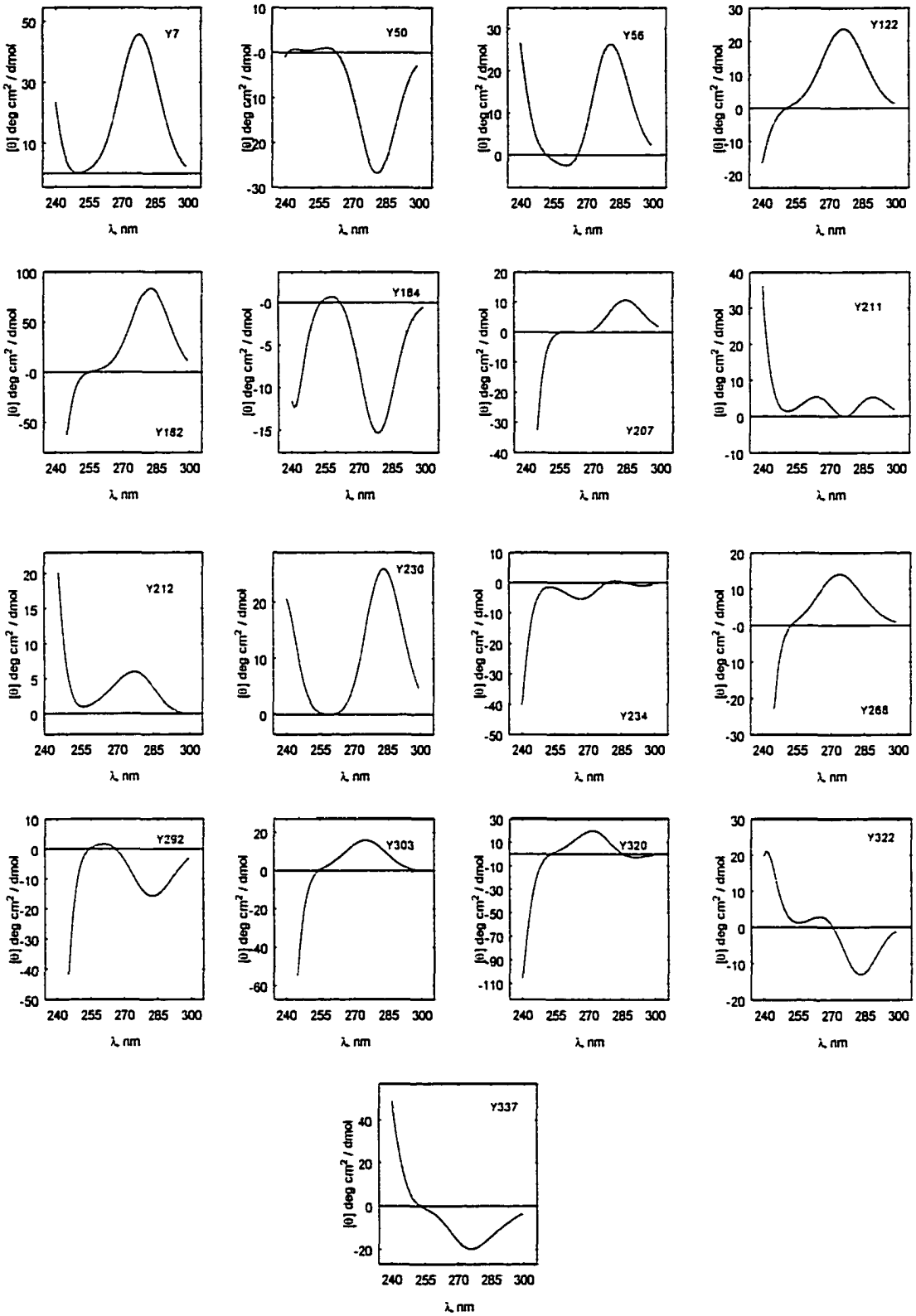


Figure 6.8: Contribution of individual tyrosine side-chains to the near-UV CD spectrum of pelC.

Table 6.1: Molar ellipticity values at 277nm calculated for aromatic amino acids found in the β-helix region.

Residue number	$[\theta]_{277}$ nm (deg cm^2 / dmol)
W106	-11
W142	-1
W256	19
W276	-52
Y56	23
Y184	-15
Y207	6
Y211	0
Y212	6
Y230	18
Y234	-1

going from complete random coil to the unordered loops. This is spectroscopically silent in the far-UV.

It has also been noted that at pH 2 and NaCl concentrations greater than 100 mM, an intermediate is stabilized with a far-UV CD spectrum very similar to native pelC but a near-UV CD opposite in sign, with an intensity of about -60 deg cm^2 / dmol at 280 nm. The side-chains in the β-helix domain alone can not account for this amplitude. There are some individual contributions that could account for the amplitude but no specific assignments are possible. It is likely that the kinetic and equilibrium intermediates are not as similar as originally thought. It is possible that binding of anions induces a structure with only partial native structure and that some side-chains might assume orientations different from those in native pelC.

Chapter 7

Summation and Future Directions

7.1 Summation

The work presented in this dissertation represents the initial studies on the folding of the parallel β-helix protein pelC. The structural properties, stability and folding of this class of proteins have received little attention. With the exception of some studies of the tailspike protein from bacteriophage P22, pelC is the only parallel β-helix protein to have been studied by biophysical methods. The thermodynamic experiments described in Chapter 2 have demonstrated that pelC has a thermodynamic stability of folding of approximately 12 kcal / mol. This value puts pelC among the most stable proteins observed. This stability is possibly a result, at least in part, of the cooperative interaction of two energetic domains in pelC. It is also probably a result of the arrangement of side chains in the hydrophobic core of pelC and the highly ordered arrangement of the β-sheets. This ordered arrangement might contribute to favorable side chain packing. The gdn-HCl denaturation transition at pH 7 appears to be approximated well by a two-state transition. The same experiment at pH 5 shows distinctly different properties. Subsequent calorimetry experiments indicated that a two-state denaturation transition is not observed. However, the only way to justify all the experimental observations is to model the denaturation transition as the sum of two individual two-state processes that show some degree of independence. This is supported by temperature-dependent CD experiments.

There is evidence for the presence of intermediate conformations that might in some way be related to some type of domain structure in pelC. The binding of anions to acid-denatured pelC stabilizes an intermediate. This intermediate has far-UV CD properties very similar to those of the native protein. The CD in the near UV differs significantly. The magnitude is about one third of the native signal and the sign is opposite. Calculated CD spectra suggest that there are several aromatic amino acids that have a negative near-UV CD signal. However, these data do not suggest that any obvious domains exist with a substantial negative CD. The binding of anions to pelC does suggest some interesting things. ANS is certainly not an innocuous probe of hydrophobic surfaces in proteins. This probe in fact induces structure in pelC.

How are equilibrium properties of proteins related to their kinetic folding mechanisms? In some cases this question can be answered. Some proteins will populate a molten globule intermediate at equilibrium. In several cases, kinetic intermediates have similar properties as these molten globules (Kuwajima et al. 1985). Molten globules have been proposed to be a general intermediate in protein folding (Christensen and Pain 1994). PelC does not populate a classic molten globule under conditions that commonly would stabilize them. Addition of anions to pelC stabilizes an intermediate that is similar but not identical to a molten globule. Molten globules lack tertiary structure. As demonstrated in chapter 3, the complex of pelC and anionic species leads to a protein with a significant near-UV CD signal. Semi-empirical CD calculations suggest that this intermediate results from at least some population of non-native structure.

The kinetic experiments imply that some collapsed intermediate forms in the early stages of folding. This intermediate has a very native-like secondary structure, but the tertiary structure is different from that of the native protein. At least that is what one might think at first. Calculations support the idea that the β-helix region is the domain that forms in the fast kinetic phase (see Chapters 4 and 6). This domain forms rapidly in the folding process, followed by the slower arrangement of aromatic side chains in the loop regions of pelC. It is certainly possible that the β-helix region is in one of the domains observed in the equilibrium experiments and the loop regions are in the other domain. The two domains are not observed in unfolding experiments because they are interconnected. They can be resolved in the refolding experiments because of the sequential nature of the pathway. The collapse of the β-helix domain must occur before the side chains in the loop regions are in close enough contact to interact and form the minimum energy conformation.

7.2 Future Directions

This project is far from complete. Many other projects can be spawned from the work done so far. Site-directed mutagenesis has become one of the most powerful and lucrative tools in protein chemistry and structural biology. There are many examples of amino acids in pelC that could potentially be mutated and the effects on the structure and folding investigated. Hydrophobic side chains involved in the extensive stacking networks could be mutated to less bulky side

chains. An example would be leucine to valine, or valine to alanine. Such mutations would disrupt the packing but not change the hydrophobic nature of the protein interior. The effects of the highly ordered packing on stability could be investigated. The interior polar stacks would be a challenge to mutate. Generally, mutation of a buried polar group to alanine would be so drastic that the protein would probably not fold. The serine stack might be a good candidate for study by mutating all three to alanine. In this way, polar interactions will be eliminated completely and the hydrophobic core would be even more homogeneous. The asparagine ladder would not be an easy set of amino acids to mutate due to the extensive interactions, but it might be worth considering Asn→Ser mutations. I have often speculated that the polar stacks are formed early in the folding of pelC and confer structural specificity.

Aromatic amino acids are good targets in pelC for mutations. It would be important to alter the stacks in order to test if these interactions make substantial contributions towards the stability of pelC. Mutation of the stacks would also test if they are involved in the low pH, salt-induced intermediate or the kinetic intermediate formed in the fast phase of folding. In addition to the stacks, mutating the aromatics found in the β-helix region would test the idea that this region is structured in the fast-phase intermediate. The best mutation in most cases would be tyrosine or tryptophan to phenylalanine. Phenylalanine is a much weaker chromophore, yet the mutation incurs only a modest change in the chemistry of the protein. Mutations such as these would also allow one to test the accuracy of the CD calculations on the aromatic contributions. The best way

to do this is to compare the theoretical and experimental difference spectra between the wild-type and mutant proteins.

One very good way to characterize the early kinetic stages is with hydrogen-deuterium exchange techniques. In this method, the protein is unfolded in a regular buffer. The refolding is initiated in deuterated buffer. Labeling occurs under alkaline conditions where H/D exchange is catalyzed. Any amide protons involved in hydrogen bonding (secondary structure) will be protected from exchange with deuterium. One can identify the exchanged protons because they are silent in NMR experiments. Using different labeling pulse sequences one can label the protein at any time period of the reaction. Thus, one can construct a model of secondary structure formation on the millisecond time scale. The only drawback to this technique is the NMR part. Assignment of the majority of the backbone resonances is absolutely necessary. In a protein the size of pelC (37,676 Da) this is very difficult. But, new methods are constantly being developed. TROSY (Pervushin et al. 1997) is the latest NMR method and has already been successful in making resonance assignments in proteins with M_r > 100kDa.

Other parallel β-helix proteins have been observed (see Table 1.1). It would be of great interest to study the folding of some other members of this family. In recent years, several small proteins or protein domains have been observed. The two antifreeze proteins are the smallest. These might be difficult to study considering the probable importance of disulfides in their stabilization. The most promising parallel β-helix protein is the β-helix domain of the human

insulin-like growth factor receptor. This protein is only about 16.5 kDa and a high percentage is β-sheet. The attraction to this protein is mainly its small size. This would make it very amenable to NMR and protein engineering techniques.

I hope that the work that I performed with pelC teaches people a little more about proteins. Hopefully the work will continue with pelC. I also would like to see more parallel β-helix proteins studied. These simple proteins have the potential to unveil quite a bit on the properties of proteins, especially in stability and folding.

Chapter 8

References

Adler, M., and Scheraga, H. A. (1990). "Nonnative isomers of proline-93 and -114 predominate in heat-unfolded ribonuclease A." *Biochemistry*, 29(36), 8211-6.

Ahmad, F., and Bigelow, C. C. (1982). "Estimation of the free energy of stabilization of ribonuclease A, lysozyme, alpha-lactalbumin, and myoglobin." *J Biol Chem*, 257(21), 12935-8.

Ahmad, F., and Bigelow, C. (1986). "Estimation of the stability of globular proteins." *Biopolymers*, 25, 1623-1633.

Akiyama, S., Takahashi, S., Ishimori, K., and Morishima, I. (2000). "Stepwise formation of alpha-helices during cytochrome c folding." *Nat Struct Biol*, 7(6), 514-20.

Ali, V., Prakash, K., Kulkarni, S., Ahmad, A., Madhusudan, K. P., and Bhakuni, V. (1999). "8-anilino-1-naphthalene sulfonic acid (ANS) induces folding of acid unfolded cytochrome c to molten globule state as a result of electrostatic interactions." *Biochemistry*, 38(41), 13635-42.

Anderson, S. R. (1969). "The circular dichroism spectra of the complexes of 1-anilino-8- naphthalenesulfonate with bovine serum albumin. Evidence for heterogeneity of binding." *Biochemistry*, 8(12), 4838-42.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." *Science*, 181(4096), 223-230.

Anfinsen, C. B., Haber, E., and Jr., F. H. W. (1961). "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain." *Proc Natl Acad Sci, USA*, 47(9), 1309-1314.

Anson, M. L. (1945). "Protein denaturation and the properties of protein groups." Advances in Protein Chemistry, Vol 2, 361-386.

Ayed, A., and Duckworth, H. W. (1999). "A stable intermediate in the equilibrium unfolding of Escherichia coli citrate synthase." *Protein Sci*, 8(5), 1116-26.

Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995). "Protein folding intermediates: native-state hydrogen exchange." *Science*, 269(5221), 192-7.

Baker, D. (2000). "A surprising simplicity to protein folding." *Nature*, 405(6782), 39-42.

Baldwin, R. L. (1996). "On-pathway versus off-pathway folding intermediates." *Fold Des*, 1(1), R1-8.

Ballew, R. M., Sabelko, J., and Gruebele, M. (1996a). "Direct observation of fast protein folding: the initial collapse of apomyoglobin." *Proc Natl Acad Sci U S A*, 93(12), 5759-64.

Ballew, R. M., Sabelko, J., and Gruebele, M. (1996b). "Observation of distinct nanosecond and microsecond protein folding events." *Nat Struct Biol*, 3(11), 923-6.

Barras, F., Gijsegem, F. v., and Chatterjee, A. K. (1994). "Extracellular enzymes and pathogenesis of soft-rot Erwinia." Ann Rev Phytopath, 32, 201-234.

Barrick, D., and Baldwin, R. L. (1993). "Three-state analysis of sperm whale apomyoglobin folding." *Biochemistry*, 32(14), 3790-6.

Bayley, P. M., Nielsen, E. B., and Schellman, J. A. (1969). "Rotatory properties of molecules containing two peptide groups-Theory." *J. Phys. Chem.*, 73, 228-243.

Beaman, T. W., Binder, D. A., Blanchard, J. S., and Roderick, S. L. (1997). "Three-dimensional structure of tetrahydrodipicolinate N-succinyltransferase." *Biochemistry*, 36(3), 489-94.

Beaman, T. W., Sugantino, M., and Roderick, S. L. (1998). "Structure of the hexapeptide xenobiotic acetyltransferase from Pseudomonas aeruginosa." *Biochemistry*, 37(19), 6689-96.

Betts, S. D., and King, J. (1998). "Cold rescue of the thermolabile tailspike intermediate at the junction between productive folding and off-pathway aggregation." *Protein Science*, 7, 1516-1523.

Betts, S., and King, J. (1999). "There's a right way and a wrong way: *in vivo* and *in vitro* folding, misfolding and subunit assembly of the P22 tailspike." *Structure*, 7, R131-R139.

Betts, S. D., Kreisberg, J. F., and King, J. (1999). "Importance of Side-Chain Stacks in beta-helix Folding and Stability." *Biophys. J.*, 76, A142.

Bienkiewicz, E. A., Moon Woody, A., and Woody, R. W. (2000). "Conformation of the RNA polymerase II C-terminal domain: Circular dichroism of long and short fragments." *J Mol Biol*, 297(1), 119-33.

Bilsel, O., Zitzewitz, J. A., Bowers, K. E., and Matthews, C. R. (1999). "Folding mechanism of the alpha-subunit of tryptophan synthase, an alpha/beta barrel protein: global analysis highlights the interconversion of multiple

native, intermediate, and unfolded forms through parallel channels." *Biochemistry*, 38(3), 1018-29.

Biltonen, R. L., and Freire, E. (1978). "Thermodynamic Characterization of Conformational States of Biological Macromolecules Using Differential Scanning Calorimetry." *CRC Crit. Rev. Biochem.*, 5, 85-124.

Blauer, G., Sreerama, N., and Woody, R. W. (1993). "Optical activity of hemoproteins in the Soret region. Circular dichroism of the heme undecapeptide of cytochrome c in aqueous solution." *Biochemistry*, 32(26), 6674-9.

Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975). "Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues." *Biochemistry*, 14(22), 4953-63.

Brockwell, D. J., Smith, D. A., and Radford, S. E. (2000). "Protein folding mechanisms: new methods and emerging ideas." *Curr Opin Struct Biol*, 10(1), 16-25.

Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C., and Bourne, Y. (1999). "Crystal structure of the bifunctional N-acetylglucosamine 1-phosphate uridyltransferase from Escherichia coli: a paradigm for the related pyrophosphorylase superfamily." *EMBO J*, 18(15), 4096-4107.

Bryngelson, J. D., and Wolynes, P. G. (1987). "Spin glasses and the statistical mechanics of protein folding." *Proc Natl Acad Sci U S A*, 84(21), 7524-8.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995). "Funnels, pathways, and the energy landscape of protein folding: a synthesis." *Proteins*, 21(3), 167-95.

Capaldi, A. P., and Radford, S. E. (1998). "Kinetic studies of beta-sheet protein folding." *Curr Opin Struct Biol*, 8(1), 86-92.

Chan, C. K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A., and Hofrichter, J. (1997). "Submillisecond protein folding kinetics studied by ultrarapid mixing." *Proc Natl Acad Sci U S A*, 94(5), 1779-84.

Chan, H. S., and Dill, K. A. (1998). "Protein folding in the landscape perspective: chevron plots and non- Arrhenius kinetics." *Proteins*, 30(1), 2-33.

Chen, E., Wood, M. J., Fink, A. L., and Kliger, D. S. (1998). "Time-resolved circular dichroism studies of protein folding intermediates of cytochrome c." *Biochemistry*, 37(16), 5589-98.

Chen, E., Wittung-Stafshede, P., and Kliger, D. S. (1999). "Far-UV Time-Resolved Circular Dichroism Detection of Electron-Transfer-Triggered Cytochrome c Folding." *J. Am. Chem. Soc.*, 121, 3811-3817.

Chen, Y., and Barkley, M. D. (1998). "Toward understanding tryptophan fluorescence in proteins." *Biochemistry*, 37(28), 9976-82.

Cheng, H. N., and Bovey, F. A. (1977). "Cis-trans equilibrium and kinetic studies of acetyl-L-proline and glycyl-L-proline." *Biopolymers*, 16(7), 1465-1472.

Chou, K. C., Pottle, M., Nemethy, G., Ueda, Y., and Scheraga, H. A. (1982). "Structure of beta-sheets. Origin of the right-handed twist and of the

increased stability of antiparallel over parallel sheets." *J Mol Biol*, 162(1), 89-112.

Christensen, H., and Pain, R. H. (1994). "The Contribution of the Molten Globule Model." Mechanisms of Protein Folding, R. H. Pain, ed., Oxford University Press, New York, pp 55-79.

Clark, L. B. (1995). "Polarization assignments in the vacuum UV spectra of the primary amide, carboxyl, and peptide groups." *J. Am. Chem. Soc.*, 117, 7974-7986.

Clarke, D. T., Doig, A. J., Stapley, B. J., and Jones, G. R. (1999). "The alpha-helix folds on the millisecond time scale." *Proc Natl Acad Sci U S A*, 96(13), 7232-7.

Clarke, J., and Itzhaki, L. S. (1998). "Hydrogen exchange and protein folding." *Curr Opin Struct Biol*, 8(1), 112-8.

Cook, K. H., Schmid, F. X., and Baldwin, R. L. (1979). "Role of proline isomerization in folding of ribonuclease A at low temperatures." *Proc Natl Acad Sci U S A*, 76(12), 6157-61.

Cota, E., and Clarke, J. (2000). "Folding of beta-sandwich proteins: three-state transition of a fibronectin type III module." *Protein Sci*, 9(1), 112-20.

Creighton, T. E. (1978). "Experimental studies of protein folding and unfolding." *Prog Biophys Mol Biol*, 33(3), 231-97.

Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*, W.H. Freeman and Company, New York.

Daniel, E., and Yang, J. T. (1973). "Analysis of the circular dichroism of the complexes of 8-anilino-1- naphthalenesulfonate with bovine serum albumin." *Biochemistry*, 12(3), 508-12.

Dill, K. A. (1999). "Polymer principles and protein folding." *Protein Sci*, 8(6), 1166-80.

Dill, K. A., and Chan, H. S. (1997). "From Levinthal to pathways to funnels." *Nat Struct Biol*, 4(1), 10-19.

Dinner, A. R., Sali, A., and Karplus, M. (1996). "The folding mechanism of larger model proteins: role of native structure." *Proc Natl Acad Sci U S A*, 93(16), 8356-61.

Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M. (2000). "Understanding protein folding via free-energy surfaces from theory and experiment." *Trends Biochem Sci*, 25(7), 331-9.

Dobson, C. M., and Karplus, M. (1999). "The fundamentals of protein folding: bringing together theory and experiment." *Curr Opin Struct Biol*, 9(1), 92-101.

Dobson, C. M., Sali, A., and Karplus, M. (1998). "Protein Folding: A Perspective from Theory and Experiment." *Angew Chem Int Ed*, 37, 868-893.

Durchschlag, H. (1986). Thermodynamic Data for Biochemistry and Biotechnology, Springer-Verlag, New York, 45.

Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R., and Hofrichter, J. (2000). "Fast kinetics and mechanisms in protein folding." *Annu Rev Biophys Biomol Struct*, 29, 327-59.

Edelhoch, H. (1967). "Spectroscopic determination of tryptophan and tyrosine in proteins." *Biochemistry*, 6(7), 1948-54.

Edwards, R. A., and Woody, R. W. (1977). "Induced circular dichroism as a probe of Cibacron Blue and Congo Red bound to dehydrogenases." *Biochem Biophys Res Commun*, 79(2), 470-6.

Edwards, R. A., and Woody, R. W. (1979). "Spectroscopic studies of Cibacron Blue and Congo Red bound to dehydrogenases and kinases. Evaluation of dyes as probes of the dinucleotide fold." *Biochemistry*, 18(23), 5197-204.

Elöve, G. A., Chaffotte, A. F., Roder, H., and Goldberg, M. E. (1992). "Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy." *Biochemistry*, 31(30), 6876-83.

Emsley, P., Charles, I. G., Fairweather, N. F., and Isaacs, N. W. (1996). "Structure of Bordetella pertussis virulence factor P.69 pertactin." *Nature*, 381(6577), 90-2.

Eyles, S. J., and Gierasch, L. M. (2000). "Multiple roles of prolyl residues in structure and folding." *J Mol Biol*, 301(3), 737-47.

Eyring, H. (1935). "The activated complex and the absolute rate of chemical reactions." *Chem Rev*, 17(1), 65-77.

Fersht, A. R. (1993). "The sixth Datta Lecture. Protein folding and stability: the pathway of folding of barnase." *FEBS Lett*, 325(1-2), 5-16.

Fersht, A. R. (1995a). "Characterizing transition states in protein folding: an essential step in the puzzle." *Curr Opin Struct Biol*, 5(1), 79-84.

Fersht, A. R. (1995b). "Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications." *Proc Natl Acad Sci U S A*, 92(24), 10869-73.

Fersht, A. R. (1997). "Nucleation mechanisms in protein folding." *Curr Opin Struct Biol*, 7(1), 3-9.

Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991). "The energy landscapes and motions of proteins." *Science*, 254(5038), 1598-603.

Freire, E., van Osdol, W. W., Mayorga, O. L., and Sanchez-Ruiz, J. M. (1990). "Calorimetrically determined dynamics of complex unfolding transitions in proteins." *Annu Rev Biophys Biophys Chem*, 19, 159-88.

From, N. B., and Bowler, B. E. (1998). "Urea denaturation of staphylococcal nuclease monitored by Fourier transform infrared spectroscopy." *Biochemistry*, 37(6), 1623-31.

Fuchs, A., Seiderer, C., and Seckler, R. (1991). "*In vitro* folding pathway of phage P22 tailspike protein." *Biochemistry*, 30, 6598-6604.

Garel, J. R., and Baldwin, R. L. (1973). "Both the fast and slow refolding reactions of ribonuclease A yield native enzyme." *Proc Natl Acad Sci U S A*, 70(12), 3347-51.

Garel, J. R., and Baldwin, R. L. (1975). "The heat-unfolded state of ribonuclease A is an equilibrium mixture of fast and slow refolding species." *J Mol Biol*, 94(4), 611-20.

Garel, J. R., Nall, B. T., and Baldwin, R. L. (1976). "Guanidine-unfolded state of ribonuclease A contains both fast- and slow- refolding species." *Proc Natl Acad Sci U S A*, 73(6), 1853-7.

Garrett, T. P., McKern, N. M., Lou, M., Frenkel, M. J., Bentley, J. D., Lovrecz, G. O., Elleman, T. C., Cosgrove, L. J., and Ward, C. W. (1998). "Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor." *Nature*, 394(6691), 395-9.

Gilmanshin, R., Williams, S., Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1997). "Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin." *Proc Natl Acad Sci U S A*, 94(8), 3709-13.

Goto, Y., and Fink, A. L. (1989). "Conformational states of beta-lactamase: molten-globule states at acidic and alkaline pH with high salt." *Biochemistry*, 28(3), 945-52.

Goto, Y., Calciano, L. J., and Fink, A. L. (1990a). "Acid-induced folding of proteins." *Proc Natl Acad Sci U S A*, 87(2), 573-7.

Goto, Y., Takahashi, N., and Fink, A. L. (1990b). "Mechanism of acid-induced folding of proteins." *Biochemistry*, 29(14), 3480-8.

Goux, W. J., and Hooker, T. M. (1980). "Chiroptical properties of proteins .1. Near-ultraviolet circular dichroism of ribonuclease-S." *J. Am. Chem. Soc.*, 102, 7080-7087.

Graether, S. P., Kuiper, M. J., Gagne, S. M., Walker, V. K., Jia, Z., Sykes, B. D., and Davies, P. L. (2000). "Beta-helix structure and ice-binding properties

of a hyperactive antifreeze protein from an insect." *Nature*, 406(6793), 325-8.

Grantcharova, V., Alm, E. J., Baker, D., and Horwich, A. L. (2001). "Mechanisms of protein folding." *Curr Opin Struct Biol*, 11(1), 70-82.

Greenfield, N., and Fasman, G. D. (1969). "Computed circular dichroism spectra for the evaluation of protein conformation." *Biochemistry*, 8(10), 4108-16.

Griko Yu, V., Venyaminov, S., and Privalov, P. L. (1989). "Heat and cold denaturation of phosphoglycerate kinase (interaction of domains)." *FEBS Lett*, 244(2), 276-8.

Grishina, I. B., and Woody, R. W. (1994). "Contributions of tryptophan side chains to the circular dichroism of globular proteins: Exciton couplets and coupled oscillators." *Faraday Discuss*, 99, 245-62.

Gruebele, M. (1999). "The Fast Protein Folding Problem." *Annu. Rev. Phys. Chem.*, 50, 485-516.

Gruebele, M., Sabelko, J., Ballew, R., and Ervin, J. (1998). "Laser Temperature Jump Induced Protein Refolding." *Acc. Chem. Res*, 31, 699-707.

Gutin, A. M., Abkevich, V. I., and Shakhnovich, E. I. (1995). "Is burst hydrophobic collapse necessary for protein folding?" *Biochemistry*, 34(9), 3066-76.

Haase-Pettingell, C., and King, J. (1997). "Prevalence of temperature sensitive folding mutations in the parallel beta coil domain of the phage P22 tailspike endorhamnosidase." *J Mol Biol*, 267, 88-102.

Haezebrouck, P., Noyelle, K., and Van Dael, H. (1998). "Equilibrium and kinetic folding of pigeon lysozyme." *Biochemistry*, 37(19), 6772-80.

Harding, M. W., Handschumacher, R. E., and Speicher, D. W. (1986). "Isolation and amino acid sequence of cyclophilin." *J Biol Chem*, 261(18), 8547-55.

Harrison, S. C., and Durbin, R. (1985). "Is there a single pathway for the folding of a polypeptide chain?" *Proc Natl Acad Sci U S A*, 82(12), 4028-30.

Heffron, S., Henrissat, B., Yoder, M. D., Lietzke, S., and Jurnak, F. (1995). "Structure-based multiple alignment of extracellular pectate lyase sequences." *Mol Plant Microbe Interact*, 8(2), 331-4.

Heffter, A. (1907). *Chem Ztg*, 11, 822.

Heidary, D. K., Gross, L. A., Roy, M., and Jennings, P. A. (1997). "Evidence for an obligatory intermediate in the folding of interleukin-1 beta." *Nat Struct Biol*, 4(9), 725-31.

Herning, T., Yutani, K., Taniyama, Y., and Kikuchi, M. (1991). "Effects of proline mutations on the unfolding and refolding of human lysozyme: the slow refolding kinetic phase does not result from proline cis-trans isomerization." *Biochemistry*, 30(41), 9882-91.

Houry, W. A., Rothwarf, D. M., and Scheraga, H. A. (1996). "Circular dichroism evidence for the presence of burst-phase intermediates on the conformational folding pathway of ribonuclease A." *Biochemistry*, 35(31), 10125-33.

Hsu, M. C., and Woody, R. W. (1971). "The origin of the heme Cotton effects in myoglobin and hemoglobin." *J Am Chem Soc*, 93(14), 3515-25.

Hu, C. Q., Sturtevant, J. M., Thomson, J. A., Erickson, R. E., and Pace, C. N. (1992). "Thermodynamics of ribonuclease T1 denaturation." *Biochemistry*, 31(20), 4876-82.

Huang, W., Matte, A., Li, Y., Kim, Y. S., Linhardt, R. J., Su, H., and Cygler, M. (1999). "Crystal structure of chondroitinase B from Flavobacterium heparinum and its complex with a disaccharide product at 1.7 A resolution." *J Mol Biol*, 294(5), 1257-69.

Hurle, M. R., and Matthews, C. R. (1987). "Proline isomerization and the slow folding reactions of the alpha subunit of tryptophan synthase from Escherichia coli." *Biochim Biophys Acta*, 913(2), 179-84.

Ivanov, M. V., and Nagradova, N. K. (1974). "Optical characteristics of the interaction of 1-anilinonaphthalene-8-sulfonate with yeast glyceraldehyde-3-phosphate dehydrogenase." *Int J Biochem*, 5, 569-578.

Jackson, S. E., and Fersht, A. R. (1991a). "Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition." *Biochemistry*, 30(43), 10428-35.

Jackson, S. E., and Fersht, A. R. (1991b). "Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding." *Biochemistry*, 30(43), 10436-43.

Jaenicke, R. (1987). "Folding and association of proteins." *Prog Biophys Mol Biol*, 49(2-3), 117-237.

Jenkins, J., Mayans, O., and Pickersgill, R. (1998). "Structure and evolution of parallel β-helix proteins." *J Struct Biol*, 122, 236-246.

Jenkins, J., Mayans, O., Smith, D., Worboys, K., and Pickersgill, R. W. (2001). "Three-dimensional Structure of Erwinia chrysanthemi Pectin Methylesterase Reveals a Novel Esterase Active Site." *J Mol Biol*, 305(4), 951-960.

Jennings, P. A., and Wright, P. E. (1993). "Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin." *Science*, 262(5135), 892-6.

Johnson, W. C. (1999). "Analyzing protein circular dichroism spectra for accurate secondary structures." *Proteins*, 35(3), 307-12.

Jones, C. M., Henry, E. R., Hu, Y., Chan, C. K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J., and Eaton, W. A. (1993). "Fast events in protein folding initiated by nanosecond laser photolysis." *Proc Natl Acad Sci U S A*, 90(24), 11860-4.

Jullien, M., and Baldwin, R. L. (1981). "The role of proline residues in the folding kinetics of the bovine pancreatic trypsin inhibitor derivative RCAM(14-38)." *J Mol Biol*, 145(1), 265-80.

Jurnak, F., Yoder, M. D., Pickersgill, R., and Jenkins, J. (1994). "Parallel beta-domains: a new fold in protein structures." *Curr Opin Struct Biol*, 4(6), 802-6.

Kabsch, W., and Sander, C. (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." *Biopolymers*, 22(12), 2577-637.

Kamen, D. E., Griko, Y., and Woody, R. W. (2000). "The stability, structural organization, and denaturation of pectate lyase C, a parallel beta-helix protein." *Biochemistry*, 39(51), 15932-43.

Kelley, R. F., and Richards, F. M. (1987). "Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding." *Biochemistry*, 26(21), 6765-74.

Kelly, J. W. (1998). "The alternative conformations of amyloidogenic proteins and their multi- step assembly pathways." *Curr Opin Struct Biol*, 8(1), 101-6.

Khurana, R., and Fink, A. L. (2000). "Do parallel β-helix proteins have a unique fourier transform infrared spectrum?" *Biophysical Journal*, 78, 994-1000.

Kiefhaber, T. (1995). "Kinetic traps in lysozyme folding." *Proc Natl Acad Sci U S A*, 92(20), 9029-33.

Kiefhaber, T., Quaas, R., Hahn, U., and Schmid, F. X. (1990). "Folding of ribonuclease T1. 1. Existence of multiple unfolded states created by proline isomerization." *Biochemistry*, 29(12), 3053-61.

Killick, T. R., Freund, S. M., and Fersht, A. R. (1999). "Real-time NMR studies on a transient folding intermediate of barstar." *Protein Sci*, 8(6), 1286-91.

Kim, P. S., and Baldwin, R. L. (1990). "Intermediates in the folding reactions of small proteins." *Annu Rev Biochem*, 59, 631-60.

King, J., Haase-Pettingell, C., Robinson, A. S., Speed, M., and Mitraki, A. (1996). "Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates." *FASEB J*, 10, 57-66.

Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. C. (1996). "A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon Methanosarcina thermophila." *EMBO J*, 15(10), 2323-30.

Kita, N., Boyd, C. M., Garrett, M. R., Jurnak, F., and Keen, N. T. (1996). "Differential effect of site-directed mutations in pelC on pectate lyase activity, plant tissue maceration, and elicitor activity." *J Biol Chem*, 271(43), 26529-35.

Kobe, B., and Kajava, A. V. (2000). "When protein folding is simplified to protein coiling: the continuum of solenoid protein structures." *Trends Biochem Sci*, 25(10), 509-15.

Koslowski, A., Sreerama, N., and Woody, R. W. (2000). "Theoretical Approach to Electronic Optical Activity." Circular Dichroism: Principles and Applications, N. Berova, K. Nakanishi, and R. W. Woody, eds., Wiley-VCH, NY, pp 55-96.

Kuhlman, B., Boice, J. A., Fairman, R., and Raleigh, D. P. (1998a). "Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding." *Biochemistry*, 37(4), 1025-32.

Kuhlman, B., Luisi, D. L., Evans, P. A., and Raleigh, D. P. (1998b). "Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9." *J Mol Biol*, 284(5), 1661-70.

Kuwajima, K. (1977). "A folding model of alpha-lactalbumin deduced from the three-state denaturation mechanism." *J Mol Biol*, 114(2), 241-58.

Kuwajima, K. (1996). "The molten globule state of alpha-lactalbumin." *FASEB J*, 10(1), 102-9.

Kuwajima, K., Nitta, K., Yoneyama, M., and Sugai, S. (1976). "Three-state denaturation of alpha-lactalbumin by guanidine hydrochloride." *J Mol Biol*, 106(2), 359-73.

Kuwajima, K., Hiraoka, Y., Ikeguchi, M., and Sugai, S. (1985). "Comparison of the transient folding intermediates in lysozyme and alpha- lactalbumin." *Biochemistry*, 24(4), 874-81.

Kuwajima, K., Garvey, E. P., Finn, B. E., Matthews, C. R., and Sugai, S. (1991). "Transient intermediates in the folding of dihydrofolate reductase as detected by far-ultraviolet circular dichroism spectroscopy." *Biochemistry*, 30(31), 7693-703.

Lang, K., Schmid, F. X., and Fischer, G. (1987). "Catalysis of protein folding by prolyl isomerase." *Nature*, 329(6136), 268-70.

Laurents, D. V., and Baldwin, R. L. (1998). "Protein folding: matching theory and experiment." *Biophys J*, 75(1), 428-34.

Lazo, N. D., and Downing, D. T. (1997). "β-helical fibrils from a model peptide." *Biochem biophys res commun*, 235, 675-679.

Lazo, N. D., and Downing, D. T. (1998). "Amyloid fibrils may be assembled from β-helical protofibrils." *Biochemistry*, 37(7), 1731-1735.

Levinthal, C. (1968). "Are there pathways for protein folding?" *J Chim Phys Phys-Chem Biolog*, 65, 44-45.

Levinthal, C. "Mossbauer Spectroscopy in Biological Systems." *Proceedings of a Conference Held at Allerton House*, Monticello, Illinois, 22-24.

Lietzke, S. E., Yoder, M. D., Keen, N. T., and Jurnak, F. (1994). "The three-dimensional structure of pectate lyase E, a plant virulence factor from *Erwinia chrysanthemi*." *Plant Physiology*, 106, 849-862.

Lin, L. N., and Brandts, J. F. (1983). "Isomerization of proline-93 during the unfolding and refolding of ribonuclease A." *Biochemistry*, 22(3), 559-63.

Lin, L. N., and Brandts, J. F. (1984). "Involvement of prolines-114 and -117 in the slow refolding phase of ribonuclease A as determined by isomer-specific proteolysis." *Biochemistry*, 23(24), 5713-23.

Liou, Y. C., Tocilj, A., Davies, P. L., and Jia, Z. (2000). "Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein." *Nature*, 406(6793), 322-4.

Lumry, R., and Eyring, H. (1954). "Conformational changes of proteins." *J. Phys. Chem.*, 58, 110-120.

Mackenzie, D. (2001). "Virtual molecules nail bacteria's weapon." *Science*, 291, 809.

Manning, M. C., and Woody, R. W. (1991). "Theoretical CD studies of polypeptide helices: examination of important electronic and geometric factors." *Biopolymers*, 31(5), 569-86.

Martinez, J. C., and Serrano, L. (1999). "The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved." *Nat Struct Biol*, 6(11), 1010-1016.

Matagne, A., Chung, E. W., Ball, L. J., Radford, S. E., Robinson, C. V., and Dobson, C. M. (1998). "The origin of the alpha-domain intermediate in the folding of hen lysozyme." *J Mol Biol*, 277(5), 997-1005.

Matagne, A., Jamin, M., Chung, E. W., Robinson, C. V., Radford, S. E., and Dobson, C. M. (2000). "Thermal unfolding of an intermediate is associated with non-Arrhenius kinetics in the folding of hen lysozyme." *J Mol Biol*, 297(1), 193-210.

Mateo, P. L., and Privalov, P. L. (1981). "Pepsinogen denaturation is not a two-state transition." *FEBS Lett*, 123(2), 189-92.

Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E., and Fersht, A. R. (1995). "Movement of the position of the transition state in protein folding." *Biochemistry*, 34(41), 13656-62.

Matulis, D., and Lovrien, R. (1998). "1-Anilino-8-naphthalene sulfonate anion-protein binding depends primarily on ion pair formation." *Biophys J*, 74(1), 422-9.

Matulis, D., Baumann, C. G., Bloomfield, V. A., and Lovrien, R. E. (1999). "1-anilino-8-naphthalene sulfonate as a protein conformational tightening agent." *Biopolymers*, 49(6), 451-8.

Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R., and Jenkins, J. (1997). "Two crystal structures of pectin

lyase A from Aspergillus reveal a pH driven conformational change and striking divergence in the substrate- binding clefts of pectin and pectate lyases." *Structure*, 5(5), 677-89.

Mayne, L., and Englander, S. W. (2000). "Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange." *Protein Sci*, 9(10), 1873-7.

McMahon, L. P., Colucci, W. J., McLaughlin, M. L., and Barkley, M. D. (1992). "Deuterium Isotope Effects in Constrained Tryptophan Derivatives: Implications for Tryptophan Photophysics." *J. Am. Chem. Soc.*, 114, 8442-8448.

Miller, S., Schuler, B., and Seckler, R. (1998). "A reversibly unfolding fragment of P22 tailspike protein with native structure: The isolated β-helix domain." *Biochemistry*, 37, 9160-9168.

Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997). "Folding dynamics and mechanism of beta-hairpin formation." *Nature*, 390(6656), 196-9.

Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." *Protein Sci*, 4(10), 2138-48.

Nall, B. T., Garel, J. R., and Baldwin, R. L. (1978). "Test of the extended two-state model for the kinetic intermediates observed in the folding transition of ribonuclease A." *J Mol Biol*, 118(3), 317-30.

Ng, K. K., and Weis, W. I. (1998). "Coupling of prolyl peptide bond isomerization and Ca2+ binding in a C- type mannose-binding protein." *Biochemistry*, 37(51), 17977-89.

Ng, K. K., Park-Snyder, S., and Weis, W. I. (1998). "Ca^{2+}-dependent structural changes in C-type mannose-binding proteins." *Biochemistry*, 37(51), 17965-76.

O'Connell, T. M., Wang, L., Tropsha, A., and Hermans, J. (1999). "The "random-coil" state of proteins: comparison of database statistics and molecular simulations." *Proteins*, 36(4), 407-18.

Ogasahara, K., and Yutani, K. (1997). "Equilibrium and kinetic analyses of unfolding and refolding for the conserved proline mutants of tryptophan synthase alpha subunit." *Biochemistry*, 36(4), 932-40.

Oliveberg, M. (1998). "Alternative explanations for "Multistate" kinetics in protein folding: Transient aggregation and changing transition state ensembles." *Acc Chem Res*, 31, 765-772.

Oliveberg, M. (2001). "Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis." *Curr Opin Struct Biol*, 11(1), 94-100.

Otzen, D. E., Itzhaki, L. S., elMasry, N. F., Jackson, S. E., and Fersht, A. R. (1994). "Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding." *Proc Natl Acad Sci U S A*, 91(22), 10422-5.

Otzen, D. E., Kristensen, O., Proctor, M., and Oliveberg, M. (1999). "Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots." *Biochemistry*, 38(20), 6499-511.

Pace, C. N. (1988). "Measuring the Conformational Stability of Proteins." Protein Structure: A Practical Approach 1st Ed, IRC Press, New York, pp 311-330.

Panchenko, A. R., Luthey-Schulten, Z., and Wolynes, P. G. (1996). "Foldons, protein structural modules, and exons." *Proc Natl Acad Sci U S A*, 93(5), 2008-13.

Pappenberger, G., Aygun, H., Engels, J. W., Reimer, U., Fischer, G., and Kiefhaber, T. (2001). "Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics." *Nat Struct Biol*, 8(5), 452-8.

Pascher, T., Chesick, J. P., Winkler, J. R., and Gray, H. B. (1996). "Protein folding triggered by electron transfer." *Science*, 271(5255), 1558-60.

Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997). "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." *Proc Nat Acad Sci USA*, 94, 12366-12371.

Petersen, T. N., Kauppinen, S., and Larsen, S. (1997). "The crystal structure of rhamnogalacturonase A from Aspergillus aculeatus: a right-handed parallel beta helix." *Structure*, 5(4), 533-44.

Pflumm, M., Luchins, J., and Beychok, S. (1986). "Stopped-flow circular dichroism." Meth Enzym, 130, 519-534.

Pickersgill, R., Jenkins, J., Harris, G., Nasser, W., and Robert-Baudouy, J. (1994). "The structure of Bacillus subtilis pectate lyase in complex with calcium." *Nat Struct Biol*, 1(10), 717-23.

Pickersgill, R., Smith, D., Worboys, K., and Jenkins, J. (1998). "Crystal structure of polygalacturonase from Erwinia carotovora ssp. carotovora." *J Biol Chem*, 273(38), 24660-4.

Plaxco, K. W., Simons, K. T., and Baker, D. (1998). "Contact order, transition state placement and the refolding rates of single domain proteins." *J Mol Biol*, 277(4), 985-94.

Privalov, P. L. (1979). "Stability of proteins. Small globular proteins." Adv. Protein Chem, 33, 167-241.

Privalov, P. L., and Makhatadze, G. I. (1990). "Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects." *J Mol Biol*, 213(2), 385-91.

Privalov, P. L., and Potekhin, S. A. (1986). "Scanning calorimetry in studying temperature-induced changes in proteins." Meth Enzym, Academic Press, 131, 4-51.

Provencher, S. W., and Glöckner, J. (1981). "Estimation of globular protein secondary structure from circular dichroism." *Biochemistry*, 20(1), 33-7.

Prusiner, S. B. (1997). "Prion diseases and the BSE crisis." *Science*, 278(5336), 245-51.

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I. (1990). "Evidence for a molten globule state as a general intermediate in protein folding." *FEBS Lett*, 262(1), 20-4.

Radford, S. E. (2000). "Protein folding: progress made and promises ahead." *Trends Biochem Sci*, 25(12), 611-618.

Radford, S. E., Dobson, C. M., and Evans, P. A. (1992). "The folding of hen lysozyme involves partially structured intermediates and multiple pathways." *Nature*, 358(6384), 302-7.

Raetz, C. R., and Roderick, S. L. (1995). "A left-handed parallel beta helix in the structure of UDP-N- acetylglucosamine acyltransferase." *Science*, 270(5238), 997-1000.

Rao, M. N., Kembhavi, A. A., and Pant, A. (1996). "Role of lysine, tryptophan and calcium in the β-elimination activity of low-molecular-mass pectate lyase from *Fusarium moniliformae*." *Biochemical Journal*, 319, 159-164.

Raschke, T. M., and Marqusee, S. (1997). "The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions." *Nat Struct Biol*, 4(4), 298-304.

Reid, K. L., Rodriguez, H. M., Hillier, B. J., and Gregoret, L. M. (1998). "Stability and folding properties of a model beta-sheet protein, Escherichia coli CspA." *Protein Sci*, 7(2), 470-9.

Reinstadler, D., Fabian, H., Backmann, J., and Naumann, D. (1996). "Refolding of thermally and urea-denatured ribonuclease A monitored by time-resolved FTIR spectroscopy." *Biochemistry*, 35(49), 15822-30.

Rexova-Benkova, L., and Markovic, O. (1976). "Pectic enzymes." Advan Carbohyd Chem Biochem, 33, 323-385.

Richardson, J. S. (1977). "Beta-sheet topology and the relatedness of proteins." *Nature*, 268(5620), 495-500.

Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999). "Experiment and theory highlight role of native state topology in SH3 folding." *Nat Struct Biol*, 6(11), 1016-1024.

Ridley, J., and Zerner, M. C. (1973). "Intermediate neglect of differential overlap technique for spectroscopy - pyrrole and azines." *Theor. Chim. Acta*, 32, 111-134.

Robinson, A. S., and King, J. (1997). "Disulphide-bonded intermediate on the folding and assembly pathway of a non-disulphide bonded protein." *Nat Struct Biol*, 4(6), 450-455.

Sackett, D. L., and Wolff, J. (1987). "Nile red as a polarity-sensitive fluorescent probe of hydrophobic protein surfaces." *Anal Biochem*, 167(2), 228-34.

Sanchez-Ruiz, J. M. (1992). "Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry." *Biophys. J.*, 61, 921-935.

Sato, Y., and Woody, R. W. (1980). "Circular dichroism of N-phenylnaphthylamine derivatives complexed with the beta-form of poly(L-lysine)." *Biopolymers*, 19, 2021-2031.

Scheraga, H. A. (1996). "Recent developments in the theory of protein folding: searching for the global energy minimum." *Biophys Chem*, 59(3), 329-39.

Schiene, C., and Fischer, G. (2000). "Enzymes that catalyse the restructuring of proteins." *Curr Opin Struct Biol*, 10(1), 40-5.

Schindler, T., Mayr, L. M., Landt, O., Hahn, U., and Schmid, F. X. (1996). "The role of a trans-proline in the folding mechanism of ribonuclease T1." *Eur J Biochem*, 241(2), 516-24.

Schmid, F. X. (1986). "Proline isomerization during refolding of ribonuclease A is accelerated by the presence of folding intermediates." *FEBS Lett*, 198(2), 217-20.

Schmid, F. X., and Baldwin, R. L. (1978). "Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization." *Proc Natl Acad Sci U S A*, 75(10), 4764-8.

Schmid, F. X., and Baldwin, R. L. (1979a). "Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons against exchange." *J Mol Biol*, 135(1), 199-215.

Schmid, F. X., and Baldwin, R. L. (1979b). "The rate of interconversion between the two unfolded forms of ribonuclease A does not depend on guanidinium chloride concentration." *J Mol Biol*, 133(2), 285-7.

Schonbrunner, N., Pappenberger, G., Scharf, M., Engels, J., and Kiefhaber, T. (1997). "Effect of preformed correct tertiary interactions on rapid two-state tendamistat folding: evidence for hairpins as initiation sites for beta- sheet formation." *Biochemistry*, 36(29), 9057-65.

Schuler, B., and Seckler, R. (1998). "P22 tailspike folding mutants revisited: Effects on the thermodynamic stability of the isolated β-helix domain." *Journal of Molecular Biology*, 281, 227-234.

Schultz, D. A., Schmid, F. X., and Baldwin, R. L. (1992). "Cis proline mutants of ribonuclease A. II. Elimination of the slow- folding forms by mutation." *Protein Sci*, 1(7), 917-24.

Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J., and Ptitsyn, O. B. (1987). "Sequential mechanism of refolding of carbonic anhydrase B." *FEBS Lett*, 224(1), 9-13.

Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Gilmanshin, R. I. (1991). "Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe." *Biopolymers*, 31(1), 119-28.

Shakhnovich, E. I. (1997). "Theoretical studies of protein-folding thermodynamics and kinetics." *Curr Opin Struct Biol*, 7(1), 29-40.

Sheridan, R. P., Lee, R. H., Peters, N., and Allen, L. C. (1979). "Hydrogen-bond cooperativity in protein secondary structure." *Biopolymers*, 18, 2541-2548.

Sheshadri, S., Lingaraju, G. M., and Varadarajan, R. (1999). "Denaturant mediated unfolding of both native and molten globule states of maltose binding protein are accompanied by large ΔCp's." *Protein Sci*, 8(8), 1689-95.

Sieber, V., Jurnak, F., and Moe, G. R. (1995). "Circular dichroism of the parallel β-helical proteins pectate lyase C and E." *Proteins: Structure, Function, and Genetics*, 23, 32-37.

Silow, M., and Oliveberg, M. (1997). "High-energy channeling in protein folding." *Biochemistry*, 36(25), 7633-7.

Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994). "The barriers in protein folding." *Nat Struct Biol*, 1(3), 149-56.

Sosnick, T. R., Shtilerman, M. D., Mayne, L., and Englander, S. W. (1997). "Ultrafast signals in protein folding and the polypeptide contracted state." *Proc Natl Acad Sci U S A*, 94(16), 8545-50.

Sreerama, N., and Woody, R. W. (1993). "A self-consistent method for the analysis of protein secondary structure from circular dichroism." *Anal Biochem*, 209(1), 32-44.

Sreerama, N., and Woody, R. W. (1994). "Poly(pro)II helices in globular proteins: identification and circular dichroic analysis." *Biochemistry*, 33(33), 10022-5.

Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (1999). "Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy." *Protein Sci*, 8(2), 370-80.

Stafford, W. F., 3rd. (1992). "Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile." *Anal Biochem*, 203(2), 295-301.

Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994). "Crystal structure of P22 tailspike protein: Interdigitated subunits in a thermostable trimer." *Science*, 265, 383-386.

Straub, J. E., and Thirumalai, D. (1993). "Exploring the energy landscape in proteins." *Proc Natl Acad Sci U S A*, 90(3), 809-13.

Stryer, L. (1965). "The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites." *J Mol Biol*, 13(2), 482-95.

Sulzenbacher, G., Gal, L., Peneff, C., Fassy, F., and Bourne, Y. (2001). "Crystal structure of *Streptococcus pneumoniae N*-acetyl-glucosamine-1-phosphate uridyltrasferase bound to acetylcoenzyme A reveals a novel active site architecture." *J Biol Chem*, 276, 11844-51.

Sutherland, I. W. (1995). "Polysaccharide lyases." FEMS Microbiol Rev, 16(4), 323-347.

Takahashi, S., Yeh, S. R., Das, T. K., Chan, C. K., Gottfried, D. S., and Rousseau, D. L. (1997). "Folding of cytochrome c initiated by submillisecond mixing." *Nat Struct Biol*, 4(1), 44-50.

Tamaki, S. J., Gold, S., Robeson, M., Manulis, S., and Keen, N. T. (1988). "Structure and organization of the *pel* genes from *Erwinia chrysanthemi*." *Journal of Bacteriology*, 170(8), 3468-3478.

Ternstrom, T., Mayor, U., Akke, M., and Oliveberg, M. (1999). "From snapshot to movie: phi analysis of protein folding transition states taken one step further." *Proc Natl Acad Sci U S A*, 96(26), 14854-9.

Thompson, P. A., Eaton, W. A., and Hofrichter, J. (1997). "Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model." *Biochemistry*, 36(30), 9200-10.

Tsai, C. J., Kumar, S., Ma, B., and Nussinov, R. (1999). "Folding funnels, binding funnels, and protein function." *Protein Sci*, 8(6), 1181-90.

Turner, D. C., and Brand, L. (1968). "Quantitative estimation of protein binding site polarity. Fluorescence of N-arylaminonaphthalenesulfonates." *Biochemistry*, 7(10), 3381-90.

Udgaonkar, J. B., and Baldwin, R. L. (1988). "NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A." *Nature*, 335(6192), 694-9.

van Santen, Y., Benen, J. A., Schroter, K. H., Kalk, K. H., Armand, S., Visser, J., and Dijkstra, B. W. (1999). "1.68-A crystal structure of endopolygalacturonase II from Aspergillus niger and identification of active site residues by site-directed mutagenesis." .

Vanek, C. F. (1975). "Circular dichroism of ANS and TNS complexes with horse liver and yeast alcohol dehydrogenases," M.S. Thesis, Arizona State University.

Veeraraghavan, S., and Nall, B. T. (1994). "Characterization of folding intermediates using prolyl isomerase." *Biochemistry*, 33(3), 687-92.

Vitali, J., Schick, B., Kester, H. C., Visser, J., and Jurnak, F. (1998). "The three-dimensional structure of *Aspergillus niger* pectin lyase B at 1.7-A resolution." *Plant Physiol*, 116(1), 69-80.

Weber, G., and Young, L. B. (1964). "Fragmentation of bovine serum albumin by pepsin." *J. Biol. Chem.*, 239, 1415-1423.

Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1996). "Fast events in protein folding: helix melting and formation in a small peptide." *Biochemistry*, 35(3), 691-7.

Witholt, B., Boekhout, M., Brock, M., Kingma, J., Heerikhuizen, H. V., and Leij, L. D. (1976). "An efficient and reproducible procedure for the formation of spheroplasts from variously grown Escherichia coli." *Anal Biochem*, 74(1), 160-70.

Wolynes, P. G. (1996). "Symmetry and the energy landscapes of biomolecules." *Proc Natl Acad Sci U S A*, 93(25), 14249-55.

Wood, L. C., White, T. B., Ramdas, L., and Nall, B. T. (1988). "Replacement of a conserved proline eliminates the absorbance-detected slow folding phase of iso-2-cytochrome c." *Biochemistry*, 27(23), 8562-8.

Woody, R. W. (1968). "Improved calculation of the $n\pi^*$ rotational strength in polypeptides." *J. Chem. Phys.*, 49, 4797-4806.

Woody, R. W. (1994). "Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins." *Eur Biophys J*, 23(4), 253-62.

Woody, R. W. (1996). "Theory of circular dichroism of proteins." Circular Dichroism and the Conformational Analysis of Biomolecules, G. D. Fasman, ed., Plenum Press, NY, pp 25-67.

Woody, R. W., and Dunker, A. K. (1996). "Aromatic and cystine side-chain circular dichroism in proteins." Circular Dichroism and the Conformational

Analysis of Biomolecules, G. D. Fasman, ed., Plenum Press, NY, pp 109-157.

Woody, R. W., and Sreerama, N. (1999). "Comment on "Improving protein circular dichroism calculations in the far-ultraviolet through reparametrizing the amide chromophore"[J. Chem. Phys. 109, 782 (1998)]." *J. Chem. Phys.*, 111(6), 2844-2845.

Wu, H. (1931). "Studies on denaturation of proteins XIII. A theory of denaturation." *Chinese Journal of Physiology*, 5(4), 321-344.

Yeh, S. R., Takahashi, S., Fan, B., and Rousseau, D. L. (1997). "Ligand exchange during cytochrome c folding." *Nat Struct Biol*, 4(1), 51-6.

Yoder, M. D., and Jurnak, F. (1995). "Protein motifs. 3. The parallel beta helix and other coiled folds." *FASEB J*, 9(5), 335-42.

Yoder, M. D., Keen, N. T., and Jurnak, F. (1993a). "New domain motif: The structure of pectate lyase C, a secreted plant virulence factor." *Science*, 260, 1503-1507.

Yoder, M. D., Lietzke, S. E., and Jurnak, F. (1993b). "Unusual structural features in the parallel β-helix in pectate lyases." *Structure*, 1, 241-251.

Zitzewitz, J. A., and Matthews, C. R. (1999). "Molecular dissection of the folding mechanism of the alpha subunit of tryptophan synthase: an amino-terminal autonomous folding unit controls several rate-limiting steps in the folding of a single domain protein." *Biochemistry*, 38(31), 10205-14.