Structural studies of the alpha-lactalbumin molten globule: implications for protein folding

by
Brenda Arlene Schulman

Submitted to the Department of Biology
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in Biology
at the

Massachusetts Institute of Technology
December 1995

Signature of Author

Department of Biology
December 5, 1995

Certified by

Peter S. Kim
Professor, Department of Biology
Thesis Supervisor

Accepted by

Frank Solomon
Chairman, Biology Graduate Committee

Dedication

To Mom and Dad,

Thank you for your endless love and support.

Acknowledgments

I was inspired by three excellent teachers early in my career, who encouraged my interest in structural biology. I thank Prof. Mark R. Haussler for the tremendous opportunity to work in his lab as a high school student. Mark gave me my first taste of biology research and got me hooked. Through his daily activities, Mark showed me the importance of reading. In addition, a research associate Mark's his lab, Dr. Barry Komm, instructed me in technique, and his lessons are at the heart of my daily lab practices. I thank Prof. Robert J. Maier for being interested in both my results and my person, in college and now. I am grateful to Rob also for making it possible for me to meet with Peter Kim when he gave the departmental seminar my senior year at Johns Hopkins. Prof. Jeremy Berg's Protein Chemistry class stirred the interest in proteins that I've had ever since. In addition to being a superb educator, I found Jeremy Berg to be a wonderful mentor, and I appreciate his spending time outside of class to talk to me about science and other things.

Many good friends have made my years at M.I.T. fun. Peggy Goodell has been a close friend, role model, and cheerleader. I hope that some of Peggy's wit and wonderful bold personality has rubbed off on me over the years! I thank Peggy also for many fun times, including one of the best vacations of my life. I thank Debbie Fass for being so generous to me--for giving and giving and giving--for sharing struggles with me and enthusiastically helping me through mine. (I thank Debbie also for reading a number of drafts of this thesis.) Jonathan Loeb, my longest and most sage friend, was my companion in climbing hundreds of miles of stairs. My best apartment-mates ever, Juli Klemm and Neal Silverman, have made the day-to-day a lot better, even after we stopped living together. Kevin Lumb, in addition to getting me acquainted with the highlights of the nightlife of Boston, was like family to me. I thank Sarah Ades for encouraging me and accompanying me in running hundreds of miles, and for the soothing conversations along the way. I thank Greg Marcus for being a very caring friend, ready with the cocoa and a bedtime story when I was sick. I thank Samantha Calqoun, David Goodman, Rob Brink, Cynthia Lean, and Stephanie and Stan Watowich for many excellent (especially holiday) dinners. Lawrence McIntosh, in addition to teaching me NMR, taught me a lot about life. Logan Donaldson was an excellent companion over the network. I thank Jamie McKnight for Kim-lab bluegrass jams, for teaching me NMR, and for being such a nice person. I thank Pehr Harbury for being my late night companion in the lab and for wonderful conversations about protein folding and people that took place by the HPLC in the wee hours of the morning. Pehr's challenges, bets, explanations, ideas and unceasing excitement about his work (and my work) make science fun. I thank Dan Minor for being a loyal friend, and for sharing his tremendous strength with me. I thank Martha Oakley for being there when I wanted to talk things over, for rides home on late nights and for some of my favorite CDs. I thank

Andrea Cochran for helping me hash out all of my late-breaking, late-night ideas, for her love of good food, and for inspiring creativity. Steve Blacklow has been a terrific baymate, always willing to listen to my ideas and offering good advice. I thank Ton Schumaker for sharing with me his excellent sense of humor and reality, and Min Lu for much enthusiastic support. I thank Angelika Amon and Julie Segre for fun times and good advice, and for taking me away from it all, at least in spirit if not in reality, when it was good to get away. I thank my penultimate baymate, Jon Staley, for being a caring, lasting friend, and for many helpful discussions about BPTI. I thank Chave Carr for being positive in all circumstances, and for much friendship when I was a young graduate student. I thank Erin O'Shea for fun and illuminating discussions, and for being an excellent role model. I owe special acknowledgment to Jonathan Weissman, for sharing so many good things, especially his outstanding sense of humor, with me during my early years in the lab--Jonathan has been a special friend. In addition, I thank Jonathan for many exciting scientific discussions, which led to some of the ideas presented in this thesis.

For expert technical assistance in synthesizing the peptides described in Appendix III, and for and keeping things going smoothly in the lab, I thank Rheba Rutkowski, Mike Burgess, and Mike Milhollen. I appreciate the behind-the-scenes work of Pam Baud and Susan Doka that makes life easier.

Zheng-yu Peng and Lawren Wu (the rest of the "α-LA subgroup") have consistently contributed to the way I think about scientific problems. I am especially grateful to Peng for being extremely generous toward me, by not only allowing, but encouraging me to work on the rich and interesting model system that he developed. I am an α-LA junky--I love thinking and talking about what all of our (and others') work on α-LA really means! By constantly discussing results (both expected and unexpected), experiments and ideas Peng and Lawren have made my thesis project challenging, exciting and fun. I also thank my thesis committee, Prof. Robert Sauer and Prof. Carl Pabo, for taking the time to think about my progress and giving me good advice.

I thank Chris Dobson for the attention he has given me and my work. Chris' enthusiasm for understanding protein folding is contagious, and I'm one of the many people in the field who has benefited from it. Chris has been a great mentor, and our brief meetings were among the most intense and inspiring discussions in my graduate career. I appreciate the opportunity to work with Christina Redfield in Chris' lab in Oxford, which not only was very fun, but also taught me state-of-the-art NMR techniques.

I owe the greatest acknowledgments to the two Peters in my life:

I am grateful for the privilege of working with my adviser, Peter S. Kim, in his exceptional lab. Peter will be a lifelong role model for me, with many rare talents that I suspect are unequaled. I hope that in my 5 1/2 years in the lab I have absorbed some of Peter's extraordinary ability to identify the really meaningful information in data, even when experiments have been done for an unrelated purpose. I am a beneficiary of Peter's high standards, which make his lab an exciting and challenging environment. Most of all, I thank Peter for being honest, for teaching me to think critically, for showing me how much I love biology, and for providing me with a very rich graduate experience.

I thank Peter J. Murray for being the love of my life and my best friend. I have reached greater heights than I thought possible, both intellectually and creatively, because of Peter's interest and support. I am grateful to Peter for help in many aspects of this thesis, for teaching me how to be more relaxed, for encouraging me to read more, and for making me have fun. Peter is my inspiration, as well as my companion.

Structural studies of the alpha-lactalbumin molten globule: implications for protein folding

by

Brenda Arlene Schulman

Submitted to the Department of Biology on December 5, 1995
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Biology

Abstract

The molten globule formed by the helical domain of α-lactalbumin (α-LA) has a native-like overall fold, even though it lacks fixed tertiary interactions and its structure is highly fluctuating. To identify the minimal requirements for forming a native-like overall topology, the helices that are folded in the molten globule of α-LA were identified by a combination of amide hydrogen exchange studies (Chapter 2) and a new mutagenesis approach, proline scanning mutagenesis (Chapter 3). The four helices that are folded in the molten globule form a subdomain in the structure of native α-LA. Unlike structure in native proteins, however, the helices in the molten globule assemble non-cooperatively, and no single helix is required for forming the native-like overall fold. Mutant proteins with prolines inserted into multiple helices still adopt the native-like topology with a significant, albeit diminished, preference when compared to a random coil. These results suggest that the key determinants of a protein's overall fold may not be of the all-or-none type.

The hydrogen exchange studies (Chapter 2) also permit investigation of the differences between the molten globule and the native state. These studies, together with disulfide exchange and NMR studies (Appendix I) indicate that the calcium-dependent folding of the β- and interdomain regions is important for forming specific tertiary interactions in α-LA.

Thesis Supervisor: Dr. Peter S. Kim, Professor of Biology

Table of Contents

CHAPTER 1

PARTIALLY FOLDED STRUCTURES OF PROTEIN FOLDING INTERMEDIATES

Intermediates in protein folding

The Levinthal paradox implies that it is impossible for proteins to fold via a random search: most proteins fold on the order of milliseconds to hours, but it would take an astronomical amount of time for even a small protein to randomly sample all accessible conformations (1). This observation has guided efforts in protein folding, with the view that identification of the pathway of protein folding and the intermediates along this path would elucidate the mechanism of folding. Indeed, over the past 15 years many proteins have been shown to fold via intermediates, and much progress has been made in assessing their structural and physical properties (2).

Residual structure in protein folding intermediates--identification of cores

In order to identify the important forces in protein folding, residual structure has been studied in partially folded forms of a number of proteins. In general, these studies are aimed at locating the structured regions and correlating them with high resolution structural information about the native state. Three main experimental approaches have been used to delineate structure in folding intermediates. The first approach is to measure protection from amide hydrogen exchange. In native proteins, amide protons sequestered from solvent, such as those involved in hydrogen bonds, are protected from hydrogen exchange. It is thought that protection from hydrogen exchange in intermediates reflects the presence of secondary structures (3).

The second way to identify structured regions in folding intermediates is mutagenesis. Mutation of side chains that are involved in specific tertiary interactions is highly destabilizing (4). Mutagenesis has been used to identify tertiary interactions present in both equilibrium and kinetic intermediates (5, 6).

Finally, protein dissection has been used to localize both residual structure and regions of the protein that are unfolded or extraneous in partially folded forms of proteins (7, 8).

A number of small proteins have been developed as model systems for protein folding. Equilibrium and kinetic intermediates observed during the folding/unfolding of these proteins have been well characterized by the approaches described above, and fall roughly into two classes. The first class is exemplified by the intermediates observed during the oxidative refolding of bovine pancreatic trypsin inhibitor (BPTI). They are cooperatively folded, with specific tertiary interactions stabilizing at least a subdomain of the protein. The second class is exemplified by the molten globule forms of apo-myoglobin. Molten globules are more loosely structured, partially folded forms of proteins that are populated at equilibrium under mildly denaturing conditions (9, 10). Apo-myoglobin forms two different molten globules, each with a high degree of secondary structure, but no fixed tertiary interactions. To explain what is and what is not folded in these two classes of intermediates (and the often weak distinctions between them), the residual structure found in the intermediates of four proteins (BPTI, cytochrome c, barnase and apo-myoglobin) is discussed in more detail below.

BPTI

BPTI is a 58-residue protein stabilized by three disulfide bonds. The oxidative refolding of BPTI has been studied extensively. Of the 15 possible single disulfide bonded isomers, only two predominate early in refolding: [30-51] and [5-55] (11, 12). The cysteines involved in these disulfide bonds are in the middle of secondary structure elements: Cys 5 is in the N-terminal 3_{10}-helix, Cys 30 is in the β-sheet, and Cys 51 and Cys 55 are in the C-terminal α-helix. Peptide models were made to test the idea that native-like structure stabilizes these disufide bonds (7, 8).

To identify the residual structure in [30-51], peptides were studied that correspond to the helix (Pα) and the sheet (Pβ), containing the cysteines involved in the 30-51 disulfide bond (7). The peptide sequences correspond to residues around the 30-51 disulfide in native BPTI that make extensive contacts with one another. Pα and Pβ fail to fold individually. However, they form a cooperatively folded subdomain in the presence of the 30-51

disulfide bond. NMR experiments indicate that a large number of specific tertiary interactions from both hydrophobic cores stabilize this subdomain.

To identify the structure that is important for forming [5-55], peptides with sequences corresponding to the two helices were studied. Unlike PαPβ, this peptide model fails to fold. The peptide model of [5-55] does fold with substantial stability, however, if the central β-sheet region, containing many of the aromatic side-chains in the two hydrophobic cores, is included (8).

Hydrogen exchange experiments on full-length BPTI with only the two cysteines corresponding to either the 30-51 or the 5-55 disulfide bonds (and the other 4 cysteines in each variant mutated to alanine) confirmed that in [30-51], the central sheet and C-terminal helix are folded, while in [5-55] the N-terminal 3_{10}-helix is folded as well (13, 14). The major hydrophobic cores are folded in both [30-51] and [5-55]. [5-55] is highly sensitive to destabilizing mutations in this core (15).

The studies on early intermediates in the folding of BPTI show that secondary structure elements come together to form a subdomain early in folding. These subdomains are cooperatively folded and are stabilized by specific interactions in the hydrophobic cores.

Cyt c

Equine cytochrome c (cyt c) is a 104 residue, single domain, heme binding protein. Pulsed hydrogen exchange experiments showed that an early kinetic intermediate contains two helices (the N- and C-terminal helices) that fold coincidentally (16). In the structure of native cyt c, both of these helices interact with the heme. Equilibrium denaturation of cyt c has been monitored for each residue by hydrogen exchange, and these same two helices are most resistant to denaturant-induced unfolding, and unfold cooperatively (17). Peptide models corresponding to these helices associate roughly stoichiometrically in the presence of heme, suggesting that tertiary interactions stabilize this intermediate (18).

These same two helices, plus another, are protected from hydrogen exchange in the relatively stable equilibrium molten globule formed by cyt c in acid and salt (19, 20). While this form of the protein displays poor chemical shift dispersion reflecting conformational heterogeneity (19), it also displays a sigmoidal thermal transition, which is evidence for cooperatively folded structure (21). The acid-destabilized form of yeast cyt c is cooperatively folded and sensitive to mutations that disrupt interactions between the N- and C-terminal helices (6). It is not yet clear how the structure in the "molten globule" and the kinetic intermediate are related, even though they both contain the same secondary structure elements.

The studies of the cyt c intermediate, like those of BPTI, demonstrate that secondary structures formed by disparate parts of the protein come together early in folding. The helices that are present in this intermediate surround and interact with heme, and fold cooperatively with one another. While the molten globule form of cyt c appears to be stabilized by specific tertiary interactions, it also appears to be somewhat fluctuating.

Barnase

Barnase is a small bacterial RNAse, whose folding has been studied extensively by both NMR and mutagenesis (22, 23). Fersht's group has studied both an early intermediate and the transition state for folding in barnase. Both amide proton exchange experiments and mutagenesis indicate that the kinetic intermediate contains substantial secondary structure. However, the intermediate is insensitive to mutations that destabilize the hydrophobic core, so it may resemble the molten globule intermediates observed at equilibrium for a number of proteins (23). On the other hand, the transition state is somewhat sensitive to these mutations, indicating that the rate-limiting step in the refolding of barnase is the formation of fixed structure within this core.

Again, extensive secondary structure is folded in the kinetic intermediate observed during the refolding of barnase. However, fixed tertiary interactions do not appear to be required for its folding. Instead, fixed tertiary interactions form in the rate-limiting step.

Apo-myoglobin

Apo-myoglobin can adopt 2 different molten globule forms. In the best studied form, three helices (A, G and H) are protected from hydrogen exchange (24). Another molten globule form of the protein, in which the B-helix is also folded, exists in low concentrations of trichloroacetate and is observed during kinetic folding experiments (25, 26). In the structure of native myoglobin, the A, G, and H helices pack together and form a subdomain. However, mutating the hydrophobic residues between the helices has little effect on the stability of the molten globule (27, 28). Thus, the apo-myoglobin molten globule contains extensive secondary structure in the absence of fixed tertiary interactions.

The studies on BPTI, cyt c, barnase, and apo-myoglobin, as well as similar studies on many other proteins, lead to several general conclusions regarding the early events in protein folding. First, the bulk of the structure observed in folding intermediates is native-like. Thus, studying protein folding intermediates can identify both the interactions that determine the overall architecture of proteins and the interactions lacking in partially folded forms of proteins. In addition, secondary structure elements are formed in intermediates. Because the secondary structures that are folded come from disparate parts of the polypeptide, and surround hydrophobic cores in the structures of the corresponding native proteins, it is thought that formation of some sort of core, even without fixed tertiary interactions, is important for folding.

The kinetic role of protein folding intermediates

Originally, the intermediates observed along protein folding pathways were thought to promote folding. A number of intermediates are kinetically competent for folding to the native state (10, 29). For example, equine cyt c folds rapidly to the native state from the molten globule. In addition, mutants that may destabilize the intermediate in T4 lysozyme fold more slowly (30).

Recently, however, the kinetic role of observed protein folding intermediates has been called into question. Several lines of evidence suggest that the intermediates that are detected by readily available techniques may hinder, rather than help folding. First, theoretical studies indicate that intermediates that accumulate during protein folding are trapped in local energy minima, rather than being on the most direct path to the native state (31-33). Second, rapid protein folding appears to correlate with a lack of detectable intermediates (34, 35). Finally, an increasing number of folding intermediates appear to be trapped in metastable conformations that are partially folded, but require some rearrangement in order to form the native structure. These intermediates slow down folding (for review, see 36). For example, intermediates observed during the refolding of cyt c and BPTI are kinetic traps. Forms of cyt c in which the heme is ligated to the wrong histidines can be stabilized and trapped by the formation of the subdomain containing the N- and C-terminal helices (29). The re-ligation of heme required for native folding is slowed by the presence of cooperatively folded native-like structure. In BPTI, which contains three disulfide bonds, formation of essentially the native structure can occur at the two-disulfide stage, with high stability. When the remaining free thiols are buried, the protein must unfold before it can form the third native disulfide bond (12).

The "classical view" versus the "new view" of protein folding--perhaps a convergence of the views is correct

The "classical" view of protein folding is that intermediates reduce the conformational search of a polypeptide and thus promote folding. However, the discovery that many of the classically studied intermediates may be kinetic traps has led to a "new" view of intermediates in protein folding (for review, see 37). In both views, it is still thought that the protein must search for the conformation with the lowest energy. However, in the "new view", the true obstacle in folding is not the search for the right conformation *per se*, but rather avoiding other local energy minima throughout the folding process. The protein folding intermediates that hinder folding are at such energy minima. In these intermediates, some native-like structure is cooperatively folded, but part of the protein is not properly folded. Native-like structure stabilizes these partially folded forms of proteins. However, in

order to attain the truly native structure, at least part of the structure that is already folded in such an "intermediate" must first unfold.

It is likely that the real role of protein folding intermediates is somewhere between the old view and the new view. While the observed pathway for folding may not be the fastest route to the native state, the intermediates along this path still reveal information about the mechanism of folding. Protein folding intermediates display strong preferences for forming native-like structure. Even the intermediates that slow down folding are very native-like. These intermediates are cooperatively folded and have fixed tertiary interactions, and may even be simpler model systems for identifying the requirements for forming a unique structure.

Two major steps in protein folding: (1) formation of native-like structure in the absence of specific tertiary interactions; and (2) formation of a unique structure

While there is a fine line between the molten globule-like and the native-like behavior in some protein folding intermediates, such as those formed by cyt c and barnase, there are obvious differences between the folding intermediates formed by apo-myoglobin and BPTI. The molten globule formed by apo-myoglobin lacks fixed tertiary interactions (27), and yet it still contains native-like secondary structures. In addition, the apo-myoglobin molten globule, like other classic molten globules, lacks a unique structure. On the other hand, the intermediates formed by BPTI are much more native-like, with fixed tertiary interactions in the hydrophobic cores. These two classes of intermediates provide simplified model systems for studying protein folding, and the differences in their behavior imply that protein folding can be simplified into studying two problems: 1) how can native-like structure fold in the absence of fixed side-chain interations? And 2) how is fixed tertiary structure formed?

The molten globule formed by α-lactalbumin

α-Lactalbumin (α-LA) is a two-domain, calcium binding protein that is stabilized by four disulfide bonds. α-LA forms one of the best studied molten globules under a wide variety of conditions, including at low pH, in the absence of calcium and other salts, and upon reduction of a number of its four disulfide bonds (38-41). Residual structure in the molten globule of α-LA is localized within one domain (the helical domain) (42). On the other hand, a number of studies have indicated that calcium binding by the other domain of the protein, which includes the β-sheet and inter-domain disulfide bonds, is important for the formation of fixed tertiary interactions (See Appendix I)(43, 44). Thus, in α-LA, it appears that the factors important for forming a molten globule are localized to one domain, and that the other domain contains the information important for directing formation of a unique structure.

The backbone topology of the helical domain of α-LA is native-like in the molten globule, even though NMR spectra (with extreme line-broadening and lack of chemical shift dispersion) indicate that it lacks fixed tertiary interactions (45-47). As with the other folding intermediates that have been well characterized structurally, the secondary structures folded in the molten globule of α-LA probably associate in some manner. Nonetheless, the important factors for bringing together secondary structures are poorly understood, as technical limitations have restricted high resolution structural studies of the molten globule. As a first step toward identifying the factors that are important for forming a native-like topology in the absence of specific tertiary interactions, the experiments described in this thesis use alternative approaches, including a new mutagenesis strategy presented in Chapter 3, to determine what parts of α-LA are folded in the molten globule. In addition, the results presented in Chapter 2 and Appendix I hint at ways to approach the next stage in folding: adopting a unique structure with fixed tertiary interactions.

References

1. C. Levinthal, *J. Chim. Phys.* **85**, 44-45 (1968).

2. P. S. Kim, R. L. Baldwin, *Annu Rev Biochem* **59**, 631-60 (1990).

3. P. S. Kim, *Methods Enzymol* **131**, 136-56 (1986).

4. C. R. Matthews, *Methods in Enzymology* **154**, 498-511 (1987).

5. A. R. Fersht, *Philos Trans R Soc Lond B Biol Sci* **348**, 11-5 (1995).

6. J. L. Marmorino, G. J. Pielak, *Biochemistry* **34**, 3140-3143 (1995).

7. T. G. Oas, P. S. Kim, *Nature* **336**, 42-8 (1988).

8. J. P. Staley, P. S. Kim, *Nature* **344**, 685-688 (1990).

9. O. B. Ptitsyn, in *Protein Folding* T. E. Creighton, Eds. (W. H. Freeman and Co., New York, 1992) pp. 243-300.

10. K. Kuwajima, M. Mitani, S. Sugai, *J. Mol. Biol.* **206**, 547-561 (1989).

11. T. E. Creighton, *J. Mol. Biol.* **87**, 603-624 (1974).

12. J. S. Weissman, P. S. Kim, *Science* **253**, 1386-1393 (1991).

13. J. P. Staley, Ph.D., Massachusetts Institute of Technology, Cambridge, Massachusetts (1993).

14. B. A. Schulman, P. S. Kim, *Protein Science* **3**, 2226-2232 (1994).

15. M.-h. Yu, J. S. Weissman, P. S. Kim, *J. Mol. Biol.* **249**, 388-397 (1995).

16. H. Roder, G. A. Elöve, S. W. Englander, *Nature* **335**, 700-704 (1988).

17. Y. Bai, T. R. Sosnick, L. Mayne, S. W. Englander, *Science* **269**, 192-197 (1995).

18. L. C. Wu, P. B. Laub, G. A. Elöve, J. Carey, H. Roder, *Biochemistry* **32**, 10271-10276 (1993).

19. M.-f. Jeng, S. W. Englander, G. A. Elöve, J. Wand, H. Roder, *Biochemistry* **29**, 10434-10437 (1990).

20. Y. Kuroda, S. Endo, K. Nagakayam, A. Wada, *J. Mol. Biol.* **247**, 682-688 (1995).

21. Y. Kuroda, S.-i. Kidokoro, A. Wada, *Journal of Molecular Biology* **223**, 1139-1153 (1992).

22. M. Bycroft, A. Matouschek, J. T. Kellis Jr., L. Serrano, A. R. Fersht, *Nature* **346**, 488-490 (1990).

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

24. F. M. Hughson, P. E. Wright, R. L. Baldwin, *Science* **249**, 1544-1548 (1990).

25. S. N. Loh, M. S. Kay, R. L. Baldwin, *Proc. Natl. Acad. Sci., USA* **92**, 5446-5450 (1995).

26. P. A. Jennings, P. E. Wright, *Science* **262**, 892-6 (1993).

27. F. M. Hughson, D. Barrick, R. L. Baldwin, *Biochemistry* **30**, 4113-8 (1991).

28. L. Lin, R. J. Pinker, K. Forde, G. D. Rose, N. R. Kallenbach, *Nature structural biology* **1**, 447-451 (1994).

29. T. R. Sosnick, L. Mayne, R. Hiller, S. W. Englander, *Nature Structural Biology* **1**, 149-156 (1994).

30. J. Lu, F. W. Dahlquist, *Biochemistry* **31**, 4749-4756 (1992).

31. J. D. Bryngelson, P. G. Wolynes, *Proc Natl Acad Sci U S A* **84**, 7524-8 (1987).

32. A. Sali, E. Shakhnovich, M. Karplus, *Nature* **369**, 248-51 (1994).

33. M. Karplus, A. Sali, *Curr Opin Struct Biol* **5**, 58-73 (1995).

34. T. Schindler, M. Herrler, M. A. Marahiel, F. X. Schmid, *Nature Structural Biology* **2**, 663-673 (1995).

35. G. S. Huang, T. G. Oas, *Proc. Natl. Acad. Sci., USA* **92**, 6878-6882 (1995).

36. J. S. Weissman, *Chemistry & Biology* **2**, 255-260 (1995).

37. R. L. Baldwin, *Journal of Biomolecular NMR* **5**, 103-109 (1995).

38. K. Kuwajima, *Proteins* **6**, 87-103 (1989).

39. K. Yutani, K. Ogasahara, K. Kuwajima, *J Mol Biol* **228**, 347-50 (1992).

40. J. J. Ewbank, T. E. Creighton, *Nature* **350**, 518-20 (1991).

41. Z.-y. Peng, L. C. Wu, P. S. Kim, *Biochemistry* **34**, 3248-3252 (1995).

42. C. M. Dobson, *Current Biology* **4**, 636-640 (1994).

43. J. J. Ewbank, T. E. Creighton, *Biochemistry* **32**, 3694-707 (1993).

44. T. M. Hendrix, Y. V. Griko, P. L. Privalov, *Protein Science* **4 (Supplement 2)**, 105 (1995).

45. J. Baum, C. M. Dobson, P. A. Evans, C. Hanley, *Biochemistry* **28**, 7-13 (1989).

46. Z.-y. Peng, P. S. Kim, *Biochemistry* **33**, 2136-41 (1994).

47. L. C. Wu, Z.-y. Peng, P. S. Kim, *Nature structural biology* **2**, 281-286 (1995).

CHAPTER 2

DIFFERENT SUBDOMAINS ARE MOST PROTECTED FROM HYDROGEN EXCHANGE IN THE MOLTEN GLOBULE AND NATIVE STATES OF α-LACTALBUMIN

α-lactalbumin (α-LA) is a two-domain, calcium-binding protein that forms one of the best studied molten globules. We present here amide hydrogen exchange studies of the molten globule formed by human α-LA at pH 2 and compare these results with a similar study of the native state at pH 6.3. The most persistent structure in the molten globule is localized in the helical domain, consistent with previous results. However, the helices most protected from hydrogen exchange in the molten globule are, in the native state, less protected from exchange than other regions of the protein. The molten globule appears to contain major elements of the native fold, but formation of the fully native state requires stabilization of structure around the calcium-binding site and domain interface.

Molten globules are partially folded forms of proteins postulated to be general intermediates in protein folding. Classic molten globules have few, if any, fixed tertiary interactions, and are observed at equilibrium for many proteins in mildly denaturing conditions (Kuwajima, 1989; Ptitsyn, 1992; Dobson, 1994). In addition, with rare exception (O'Shea et al., 1993), *de novo* designed proteins have properties more reminiscent of molten globules than native proteins (for review, see Betz et al., 1993). Thus, understanding what distinguishes partially folded states such as molten globules from fully folded proteins is critical for understanding protein folding and protein design.

The best characterized molten globule is that formed by α-lactalbumin (α-LA), a 123-residue, two-domain calcium-binding protein (Fig. 1). α-LA adopts a molten globule conformation readily under a wide variety of conditions including low pH, reduction of disulfide bonds, and the absence of calcium and other salts (Kuwajima, 1989; Ewbank & Creighton, 1991; Yutani et al., 1992). Numerous studies have led to the view that the α-LA molten globule has a bipartite structure, with (i) a disordered β-sheet domain and (ii) an α-helical domain containing substantial secondary structure and a native tertiary fold, even though it lacks extensive fixed tertiary interactions (Baum et al., 1989; Kuwajima, 1989; Alexandrescu et al., 1993; Peng & Kim, 1994; Xie & Freire, 1994; Wu et al., 1995).

It is important to understand the similarities and differences between the molten globule and a native protein. Amide-proton hydrogen exchange probes the structure and dynamics of proteins (Linderstrøm-Lang, 1955; Woodward et al., 1982; Englander & Kallenbach, 1983; Wagner & Wüthrich, 1986). We report here hydrogen exchange studies of both the molten globule and native states of human α-LA. For a number of proteins, a correlation has been found between the location of the amide protons most protected from solvent exchange in the native protein and the protein folding core (Woodward, 1993). For α-LA, however, we show here that the region of the native structure that is most protected from exchange in the molten globule is not the region most resistant to solvent exchange in the native protein. Our results can be attributed to the additional interactions necessary to stabilize the native state in this two-domain protein.

Exchange rates in the molten globule were measured indirectly because broad lines and poor dispersion complicate interpretation of the NMR spectra of the α-LA molten globule (Baum et al., 1989; Alexandrescu et al., 1993). Hydrogen exchange was carried out for a variable time period at pH* 2, 5°C, conditions at which α-LA is a molten globule. (pH* refers to meter readings in D_2O solutions using a glass pH electrode, without correction for isotope effects.) Solvent exchange was quenched by freezing and lyophilization, and the sample was analyzed subsequently under conditions where human α-LA is native (Nozaka et al., 1978) and gives rise to high quality ^{15}N-^{1}H HSQC spectra for which assignments are complete (Redfield, Peng, Schulman, Kim and Dobson, unpublished). In these experiments, only amide protons that are highly protected from hydrogen exchange in the *native* state can be probed. Nonetheless, 34 amide protons can be assayed in the molten globule: 24 from α-helices and 10 from the β-sheet domain (Fig. 2). None of the β-sheet amide protons that are assayed are significantly protected from solvent exchange in the molten globule. Protected protons are detected within the α-helical domain, with local differences in protection factors (Table 1, Fig. 3A). The B-helix contains the amide hydrogens most protected from exchange, with protection factors for four residues exceeding 100. The A-helix also persists in the molten globule, with three of its five probes protected more than 10-fold, and one is protected by a factor of 50. Only 4 of the 12 probes from the C-helix are protected more than 10-fold, and none is protected by more than a factor of 20.

Our data agree with previous work showing that specific helices are folded in the molten globule of α-LAs from guinea pigs, cows and humans (Baum et al., 1989; Alexandrescu et al., 1993; Chyan et al., 1993). An additional finding in the present study is that amides in the A-helix, which were not probed in previous ^{1}H NMR studies of α-LA, are significantly protected from exchange. The protection that we observe for a number of helices in the molten globule strongly supports a view of the molten globule with extensive native-like secondary structure. In previous studies, analysis of the propensities of cysteine residues to form disulfide bonds provided clear evidence that the helical domain of human α-LA retains an overall native-like fold even as a molten globule (Peng & Kim, 1994; Wu et al., 1995). The protection from amide proton exchange that we observe for the B-helix, taken

together with the strength of the 28-111 disulfide bond (Peng et al., 1995), suggest that there is a correlation between secondary and tertiary structure in α-LA, even in the absence of extensive sidechain packing interactions.

For native α-LA, amide proton exchange rates were measured directly (Table 1). Relatively similar numbers of amide protons are protected from exchange in the α-helical and the β-sheet domain regions. The most highly protected amide protons are located primarily around two disulfide bonds: 61-77, within the β-sheet domain; and 73-91, connecting the β-sheet domain to the long C-helix (Table 1, Fig. 3B). In the helical domain, the C-helix is most protected, followed by the B- and then the A-helix. The D- and 3_{10}-helices exchange too quickly to be measured. It is striking that the C-helix, which is only marginally protected from exchange in the molten globule, is the most highly protected from exchange in native α-LA (Fig. 4). Examination of the α-LA structure (Fig. 1), however, provides an explanation for this observation. The calcium binding site is located in a loop between the N-terminus of the C-helix and the β-sheet and interdomain disulfide bonds. In the molten globule state formed at low pH, calcium does not bind strongly. Loss of calcium, therefore, disrupts the domain interface and results both in loss of structure within the β-domain and in relative destabilization of the C-helix to which it is linked by many side-chain contacts and the 73-91 disulfide bond (cf. Loh et al., 1993).

Based on the large number of proteins which are molten globules under mildly denaturing conditions (Dobson, 1994), and the relative ease with which *de novo* design efforts yield partially folded proteins (Betz et al., 1993; Davidson & Sauer, 1994) it seems that the protein folding problem can be sub-divided into understanding: (i) how a protein can fold into a molten globule with a native-like tertiary fold in the absence of extensive packing interactions (Peng & Kim, 1994; Morozova et al., 1995; Wu et al., 1995) and (ii) how fixed tertiary structure is formed (Handel et al., 1993; O'Shea et al., 1993). The present results shed light on both of these issues. For α-LA, and perhaps for other proteins, the helices that can fold and still maintain dynamic flexibility form early in protein folding, without fixed tertiary packing. Importantly, studies of both α-LAs and the homologous lysozymes indicate that early kinetic folding intermediates resemble equilibrium molten

globules, with structure concentrated in the helical domain (Kuwajima et al., 1985; Ikeguchi et al., 1986; Radford et al., 1992; Hooke et al., 1994). In order to form the fully folded conformations, however, additional interactions are necessary to lock in the unique tertiary contacts of the native structure, which involve specific interdomain contacts for both lysozymes and α-LAs (Van Dael et al., 1993; Ewbank & Creighton, 1993a; Dobson et al., 1994; Pardon et al., 1995; Wu et al., 1995). In the case of α-LAs, these contacts are stabilized by the binding of calcium, which promotes the formation of the native state (Fig. 4). Interestingly, such additional specific interactions are crucial for forming native-like features in *de novo* designed proteins (Handel et al., 1993; O'Shea et al., 1993; Lumb & Kim, 1995).

Table 1: Amide proton exchange rates and protection factors for α-lactalbumin in the molten globule and native states.

	molten globule (pH* 2, 5°C)		native (pH* 6.3, 3mM Ca2+, 15°C)	
Residue	**k_{ex} (hr^{-1})**	**k_{int}/k_{ex}**	**k_{ex} (hr^{-1})**	**k_{int}/k_{ex} (10^{-3})**
Leu 8	.43	3.5	.080	6.8
Ser 9	.25	12	.56	7.6
Gln 10			.29	23
Leu 11	.078	18	.31	4.0
Leu 12	.037	50	.026	18
Lys 13	.35	6.0	.66	2.5
Glu 25	.14	13	.059	8.8
Leu 26	.0054	280	.018	31
Ile 27	.0039	220	.0039	86
Thr 29	.076	15	.18	41
Met 30	.010	120	.058	78
Thr 33	.0045	520	.071	230
Val 42	.24	1.9	.20	1.8
Tyr 50			.13	8.7
Phe 53			.047	22
Gln 54			.086	45
Ile 55	.34	2.0	.052	17
Ser 56			.52	7.8
Trp 60	.35	7.8	.040	17
Cys 61	3.0	.50	.035	230
Lys 62			.079	97
Ser 63			.097	93
Cys 73	1.0	1.1	.038	160
Ile 75	1.1	.97	.0099	37
Cys 77	2.6	1.1	.20	100
Lys 79	.33	9.2	.039	46
Phe 80	.59	2.8	.023	100
Leu 81			.028	32
Asp 82	1.1	1.7	.040	23
Asp 84			.084	12
Thr 86	1.7	.32		
Asp 87	1.0	2.9	.59	3.9
Asp 88	1.4	5.2	.025	39
Ile 89	.063	17	.011	33
Met 90	.099	7.2	.028	61
Cys 91	.17	14	.035	390
Ala 92	.44	7.6	.035	240
Lys 93	.20	14	.014	200
Lys 94	.089	19	.023	160
Leu 96	.94	.72	.0093	49
Asp 97	1.2	1.5	.037	25
Ile 98	.22	4.8	.023	16
Lys 99			.067	23
Asp 102	.20	6.3	.42	2.0
Trp 104(sc)			.06	.25

Table Legend

Amide proton exchange was performed on uniformly (>95%) ^{15}N-labeled recombinant α-LA. α-LA was expressed from BL21 (DE3) pLys S (Novagen) harboring the plasmid encoding human α-LA under T7 control (Studier et al., 1990; Peng et al., 1995; Wu et al., 1995), grown in minimal M9 media containing one gram/liter ^{15}N-ammonium sulfate (99.5% ^{15}N; Isotec; Miamisburg, Ohio), and supplemented with 0.5 mg/liter thiamine (McIntosh & Dahlquist, 1990). Inclusion bodies containing α-LA were solubilized, partially purified, oxidatively refolded and purified as described previously for a related protein (Peng & Kim, 1994). The identity of α-LA was confirmed by laser desorption mass spectrometry (Finnigan LASERMAT). All protein concentrations were determined by tryptophan, tyrosine and cystine absorbance (Edelhoch, 1967).

α-LA is monomeric under the conditions of hydrogen-deuterium exchange and NMR, as shown by equilibrium sedimentation, using Beckman 6-sector, 12 mm pathlength epoxy cells, and an An-60 Ti rotor in a Beckman XL-A 90 analytical ultracentrifuge. Samples were dialyzed exhaustively against the reference buffer [either 100 mM imidazole, pH 6.3, 3 mM $CaCl_2$ (native state) or 10 mM HCl (molten globule)]. For the native state, data were collected at 15°C and 25.5k rpm for three initial concentrations (20, 75 and 150 μM). The measured molecular weight is 15.0 kDa ± 5% (95% confidence; calculated, 14.1 kDa) with random residuals. The density of the buffer was measured manually and the partial specific volume was calculated to be 0.736 using constants from Laue et al. (1992). For the molten globule, nonideality was observed as a *decrease* in the apparent molecular weight as a function of increasing protein concentration, likely resulting from the large ratio of net charge to molecular weight at the low ionic strength of the buffer, thus requiring the use of the second virial coefficient as a fitting parameter (Williams et al., 1958). Therefore, data at 5°C from three wavelengths (251, 260 and 296 nm), rotor speeds (22, 26 and 30k rpm), and initial concentrations (20, 40 and 100 μM) were fit simultaneously using HID4000 (Johnson et al., 1981). Fifteen data sets were fit to a single molecular weight, fifteen intercepts, fourteen offsets and a single second virial coefficient, with no systematic variation in the residuals. The measured molecular weight is 13.9×10^3 Da (± 5%, 95% confidence) as compared to the calculated molecular weight of 14,054 Da.

Hydrogen-deuterium exchange was initiated by dissolving lyophilized samples either in D_2O, pH* 2 (molten globule) or native exchange buffer (100 mM d_4-imidazole [Cambridge Isotope Laboratories, >98% D], 3 mM $CaCl_2$ with a final pH* of 6.3). The protein concentration was 80 μM for exchange in the

molten globule and 800 μM for exchange in the native state. For the molten globule, exchange was quenched by freezing in liquid nitrogen and subsequent lyophilization at regular intervals from 10 seconds to 9 days. Dry samples were stored at -80°C and dissolved in native exchange buffer for analysis. Rates of H-D exchange were measured from ^{15}N-^{1}H HSQC spectra collected at 15°C, consisting of 48 increments of 16 transients defined by 1024 complex data points, and an ω2 spectral width of 6024 Hz, an ω1 spectral width of 1130 Hz, and a recycle delay of 1 second (Bax et al., 1990; Norwood et al., 1990). ^{15}N decoupling during acquisition was achieved with WALTZ-16 (Shaka et al., 1983). The residual HOD peak was suppressed with a low power purge pulse (Messerle et al., 1989). For the native state, rates of H-D exchange were measured similarly to those in the molten globule, except that one spectrum was collected for each time point with 64 transients per increment and 128 increments of t1.

Amide proton decays were followed by measuring peak volumes in ^{15}N-^{1}H HSQC spectra. The volumes were normalized to peak areas of non-exchangeable resonances (corresponding to the γ2 and δ protons of Ile 95, and likely to the α proton of Cys 61) in one-dimensional ^{1}H spectra collected immediately prior to each set of HSQC spectra for the molten globule, or before each spectrum of the native state. The amide proton exchange rates, k_{ex}, were determined (Kaleidograph, Abelbeck software) by fitting the data to the three parameter curve, $I(t) = I(\infty) + I(0) \exp(-k_{ex} * t)$, where I(t) is the intensity at time (t) after addition of deuterated solvent to the protein. Values for $I(\infty)$ were, on average, 10-fold lower than I(0) for the molten globule, and 10- to 30-fold lower for the native state. Predicted intrinsic rates were calculated according to Bai et al. (1993) for main chain amide protons and to Wüthrich (1986) for the Trp 104 indole.

Figure Legends

Figure 1: Schematic representation (Priestle, 1988) of human α-LA (Acharya et al., 1991). Helices are labeled A, B, C, D, and 3_{10}. Disulfide bonds are shown in grey and are labeled by the cysteines that they connect.

Figure 2: **(A)** ^{15}N-^{1}H HSQC spectra of native α-LA after exchange in the molten globule for roughly 5 minutes, one hour, 10 hours, and 8 days. (Resonance assignments, by Redfield, et al., are forthcoming.) **(B)** Time course of signal decay resulting from hydrogen exchange in the molten globule for the following representative amide protons: Leu 8, Leu 12, Met 30, Cys 73, Asp 88, and Cys 91.

Figure 3: Histograms showing the distribution of protection factors from amide hydrogen exchange for **(A)** the molten globule and **(B)** the native state of α-LA. The maximum protection factors shown are 100 for the molten globule and 10^5 for the native state to highlight differences.

Figure 4: The most protected amide protons in the molten globule and native states are highlighted on the structure of native human α-LA (Acharya et al., 1991). Residues with amide protons protected by factors greater than 50 in the molten globule are shown in red, and those protected by factors greater than 10^5 in the native state are shown in yellow. Thr 33, which is protected by these criteria under both conditions, is colored orange. Residues with protected amide protons are shown on a backbone drawing in **(A)** and in CPK models in **(B)**.

Acknowledgments

We thank Kevin Lumb and Rheba Rutkowski for useful comments regarding the manuscript, and Lawren Wu and our colleagues in the Dobson and Kim labs for helpful discussions. B.A.S. was supported by a National Science Foundation predoctoral fellowship and a National Institutes of Health training grant (T32AI07348). This work was supported by the Howard Hughes Medical Institute. The Oxford Centre for Molecular Sciences is supported by the U.K. Biotechnology and Biological Sciences Research Council and the Medical Research Council. The research of C.M.D. is supported in part by an International Research Scholar award from the Howard Hughes Medical Institute.

References

Acharya, K. R., Ren, J. S., Stuart, D. I., Phillips, D. C. & Fenna, R. E. (1991). Crystal structure of human alpha-lactalbumin at 1.7 A resolution. *J. Mol. Biol.* **221**, 571-81.

Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. & Dobson, C. M. (1993). Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: a two-dimensional NMR study. *Biochemistry* **32**, 1707-18.

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75-86.

Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry* **28**, 7-13.

Bax, A., Ikura, M., Kay, L. E., Torchia, D. A. & Tschudin, R. (1990). Comparison of Different Modes of Two-Dimensional Reverse-Correlation NMR for the Study of Proteins. *J. Mag. Reson.* **86**, 304-318.

Betz, S. F., Raleigh, D. P. & DeGrado, W. F. (1993). De novo protein design: from molten globules to native-like states. *Curr. Opin. Struct. Biol.* **3**, 601-610.

Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A. & Baum, J. (1993). Structure and stability of the molten globule state of guinea-pig alpha-lactalbumin: a hydrogen exchange study. *Biochemistry* **32**, 5681-91.

Davidson, A. R. & Sauer, R. T. (1994). Folded proteins occur frequently in libraries of random amino acid sequences. *Proc. Natl. Acad. Sci., U S A*, **91**, 2146-50.

Dobson, C. M. (1994). Solid evidence for molten globules. *Curr. Biol.* **4**, 636-640.

Dobson, C. M., Evans, P. A. & Radford, S. E. (1994). Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.* **19**, 31-7.

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

Englander, S. W. & Kallenbach, N. R. (1983). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **16**, 521-655.

Ewbank, J. J. & Creighton, T. E. (1991). The molten globule protein conformation probed by disulphide bonds. *Nature* **350**, 518-20.

Ewbank, J. J. & Creighton, T. E. (1993a). Pathway of disulfide-coupled unfolding and refolding of bovine alpha-lactalbumin. *Biochemistry* **32**, 3677-93.

Handel, T. M., Williams, S. A. & De Grado, W. F. (1993). Metal ion-dependent modulation of the dynamics of a designed protein. *Science* **261**, 879-85.

Hooke, S. D., Radford, S. E. & Dobson, C. M. (1994). The refolding of human lysozyme: a comparison with the structurally homologous hen lysozyme. *Biochemistry* **33**, 5867-76.

Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. (1986). Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: a comparative study of the folding reactions of alpha-lactalbumin and lysozyme. *Biochemistry* **25**, 6965-72.

Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* **36**, 575-88.

Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* **6**, 87-103.

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

Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugaton in Biochemistry and Polymer Science* pp. 90-125, The Royal Society of London, Cambridge.

Linderstrøm-Lang, K. U. (1955). Deuterium exchange between peptides and water. In: "Symposium on Peptide Chemistry.". *Chem. Soc. Spec. Publ.* **2**, 1-20.

Loh, S. N., Prehoda, K. E., Wang, J. & Markley, J. L. (1993). Hydrogen Exchange in Unligated and Ligated Staphylococcal Nuclease. *Biochemistry* **32**, 11022-11028.

Lumb, K. J. & Kim, P. S. (1995). A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**, 8642-8648.

McIntosh, L. P. & Dahlquist, F. W. (1990). Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q. Rev. Biophys.* **23**, 1-38.

Messerle, B. A., Wider, G., Otting, G., Weber, C. & Wüthrich, K. (1989). Solvent Suppression Using a Spin Lock in 2D and 3D NMR Spectroscopy with H2O Solutions. *J. Mag. Reson.* **85**, 608-613.

Morozova, L. A., Haynie, D. T., Arico-Muendel, C., Van Dael, H. & Dobson, C. M. (1995). Structural Basis of the Stability of a Lysozyme Molten Globule. *submitted.*

Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N. & Campbel, I. D. (1990). Comparison of techniques of 1H-detected heteronuclear 1H-15N spectrosopy. *J Magn Reson* **87**, 488-501.

Nozaka, M., Kuwajima, K., Nitta, K. & Sugai, S. (1978). Detection and characterization of the intermediate on the folding pathway of human alpha-lactalbumin. *Biochemistry* **17**, 3753-8.

O'Shea, E. K., Lumb, K. J. & Kim, P. S. (1993). Peptide "Velcro": design of a heterodimeric coiled coil. *Curr. Biol.* **3**, 658-667.

Pardon, E., Haezebrouck, P., De Baetseleier, A., Hooke, S. D., Fancourt, K. T., Desmet, J., Dobson, C. M., Van Dael, H. & Joniau, M. (1995). A Ca2+-binding chimeral of human lysozyme and bovine alpha-lactalbumin that can form a molten globule. *J.Biol.Chem.* **270**, 10514-10524.

Peng, Z.-y., Wu, L. C. & Kim, P. S. (1995). Local Structural Preferences in the α-Lactalbumin Molten Globule. *Biochemistry* **34**, 3248-3252.

Peng, Z. Y. & Kim, P. S. (1994). A protein dissection study of a molten globule. *Biochemistry* **33**, 2136-41.

Priestle, J. P. (1988). RIBBON: A stereo cartoon drawing program for proteins. *J Appl. Crystallogr.* **21**, 572-576.

Ptitsyn, O. B. (1992). The molten globule state. In *Protein Folding* (T. E. Creighton, eds.), pp. 243-300, W. H. Freeman and Co., New York.

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

Shaka, A. J., Keeler, J. & Freeman, R. (1983). Evaluation of a new broadband decoupling sequence: WALTZ-16. *J. Magn. Reson.* 313-340.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89.

Van Dael, H., Haezebrouck, P., Morozova, L., Arico-Muendel, C. & Dobson, C. M. (1993). Partially folded states of equine lysozyme. Structural characterization and significance for protein folding. *Biochemistry* **32**, 11886-94.

Wagner, G. & Wüthrich, K. (1986). Observation of internal motility of proteins by nuclear magnetic resonance in solution. *Methods Enzymol.* **131**, 307-26.

Williams, J. W., van Holde, K. E., Baldwin, R. L. & Fujita, H. (1958). The theory of sedimentation analysis. *Chem. Rev.* **58**, 715-806.

Woodward, C. (1993). Is the slow exchange core the protein folding core? *Trends Biochem. Sci.* **18**, 359-60.

Woodward, C., Simon, I. & Tuchsen, E. (1982). Hydrogen exchange and the dynamic structure of proteins. *Mol. Cell. Biochem.* **48**, 135-60.

Wu, L. C., Peng, Z.-y. & Kim, P. S. (1995). Bipartite structure of the α-lactalbumin molten globule. *Nature Struct. Biol.* **2**, 281-286.

Wu, L. C., Peng, Z.-y., Schulman, B. A. & Kim, P. S. (1995). Disulfide Determinants of Calcium-Induced Packing in alpha-Lactalbumin. *submitted.*

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids.* John Wiley & Sons, New York.

Xie, D. & Freire, E. (1994). Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. *Proteins* **19**, 291-301.

Yutani, K., Ogasahara, K. & Kuwajima, K. (1992). Absence of the thermal transition in apo-alpha-lactalbumin in the molten globule state. A study by differential scanning microcalorimetry. *J. Mol. Biol.* **228**, 347-50.

Fig. 1

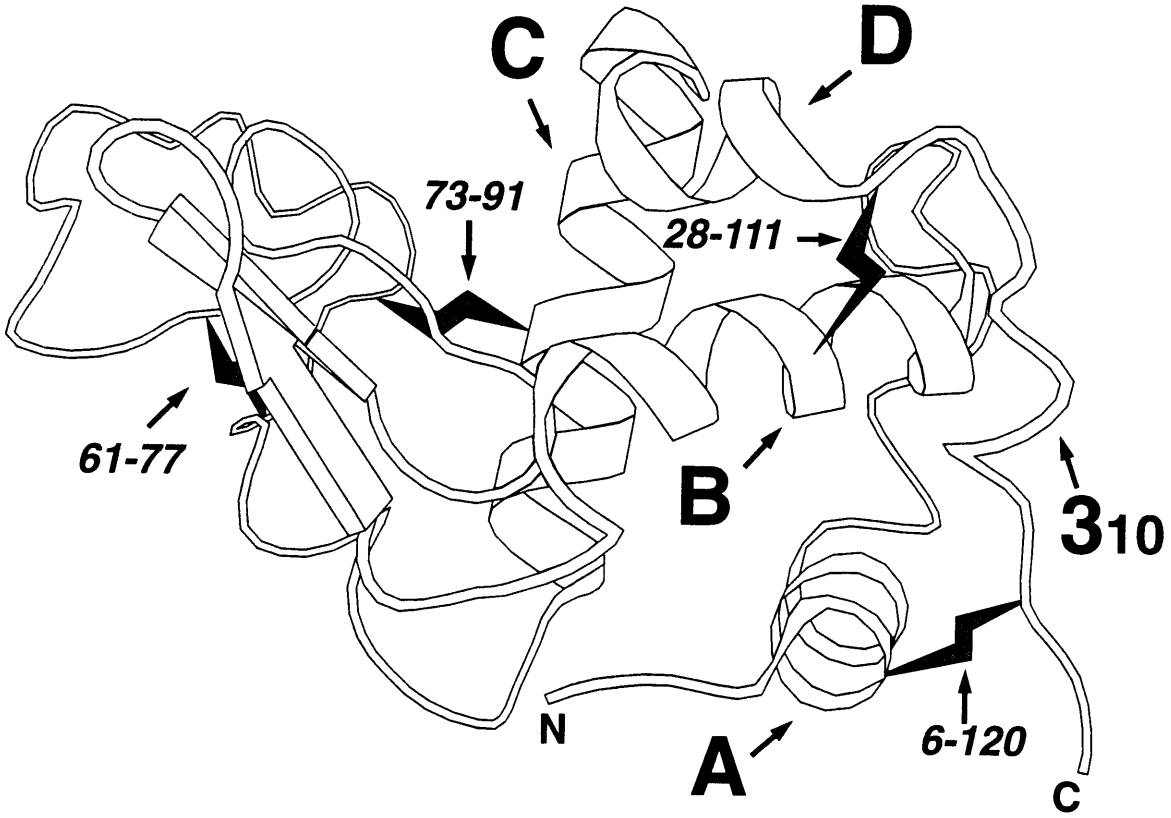


Fig. 2A

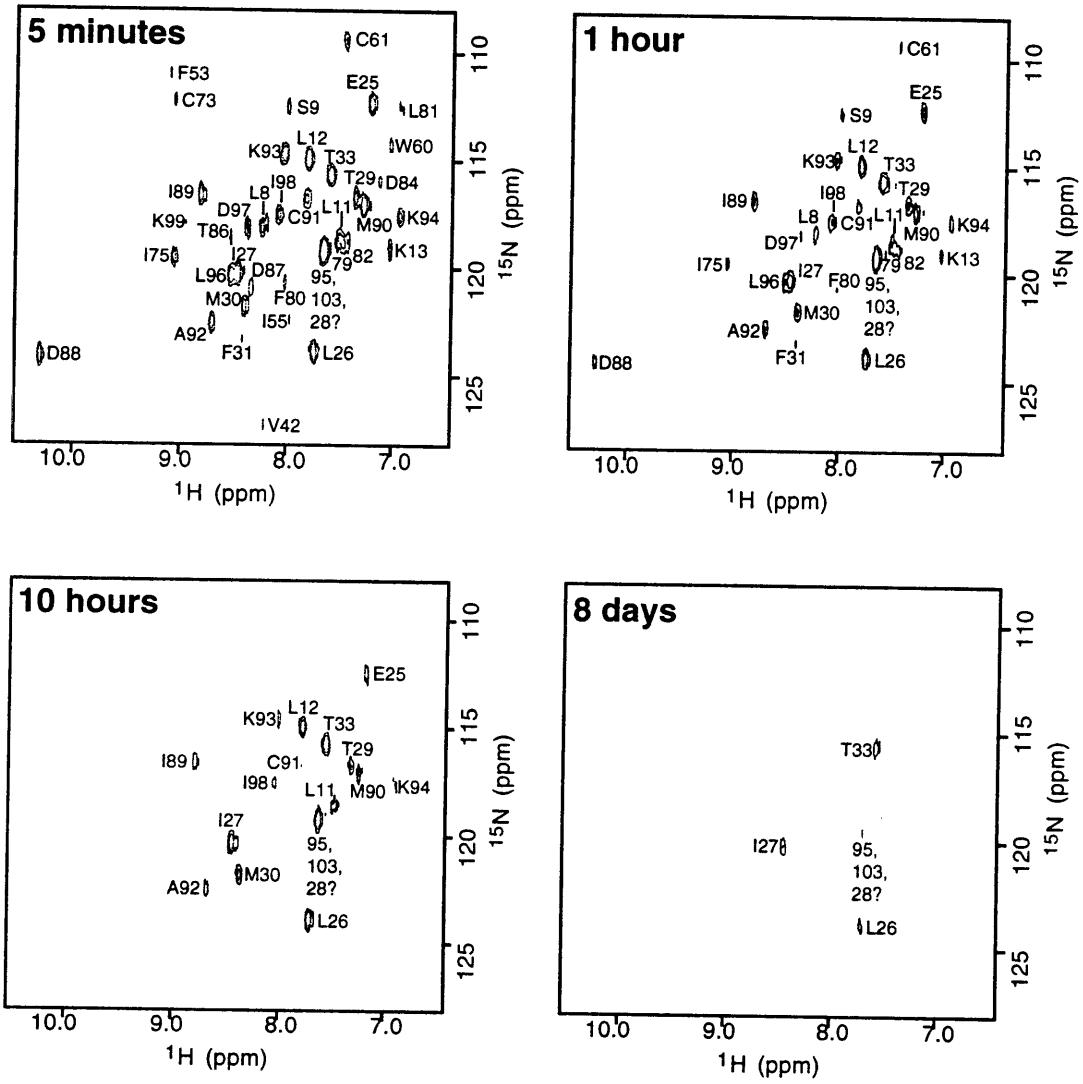
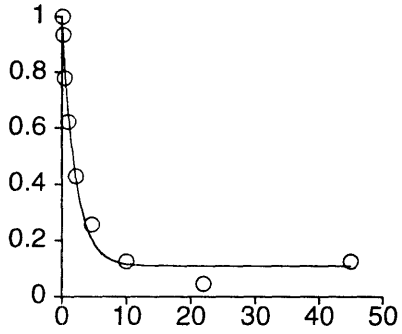
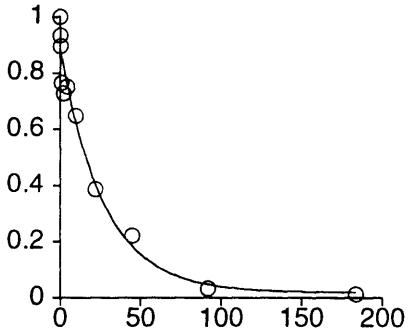

5 minutes
C61
F53
E25
C73
S9
L81
L12
K93
T33
W60
T29
D84
I89
I98
L8
L11
D97
C91
M90
K94
K99
T86
I75
I27
K13
L96
D87
95,
M30
F80
103,
I55
28?
A92
D88
F31
L26
V42
1 hour
C61
E25
S9
L12
K93
T33
T29
I89
I98
L8
C91
L11
M90
K94
D97
I75
I27
K13
L96
F80
95,
M30
103,
A92
28?
D88
F31
L26
10 hours
E25
K93
L12
T33
T29
I89
C91
I98
L11
M90
K94
I27
95,
M30
103,
A92
28?
L26
8 days
T33
I27
95,
103,
28?
L26
10.0
9.0
8.0
7.0
^{1}H (ppm)
110
115
120
125
^{15}N (ppm)

Fig. 2B

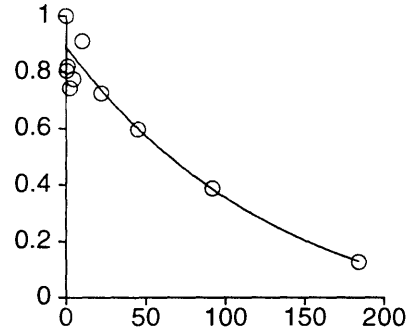
Leu 8



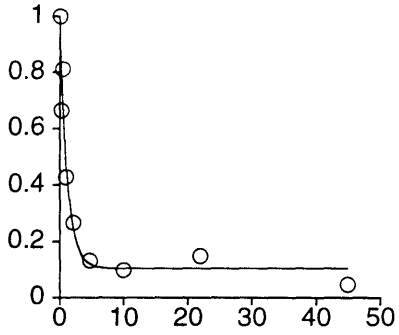
Leu 12



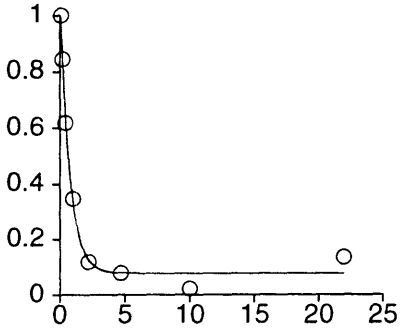
Met 30



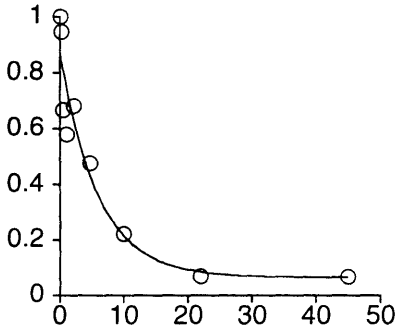
Cys 73



Asp 88



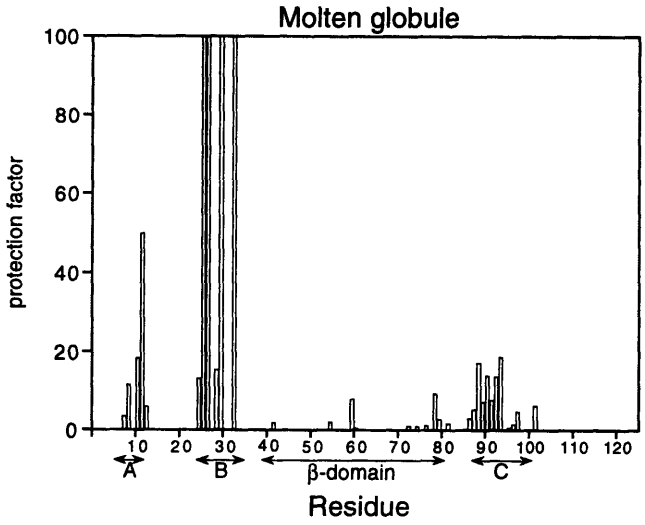
Cys 91



Time (hours)

Fig. 3

A)


Molten globule
protection factor
100
80
60
40
20
0
10 20 30 40 50 60 70 80 90 100 110 120
A
B
β-domain
C
Residue

B)

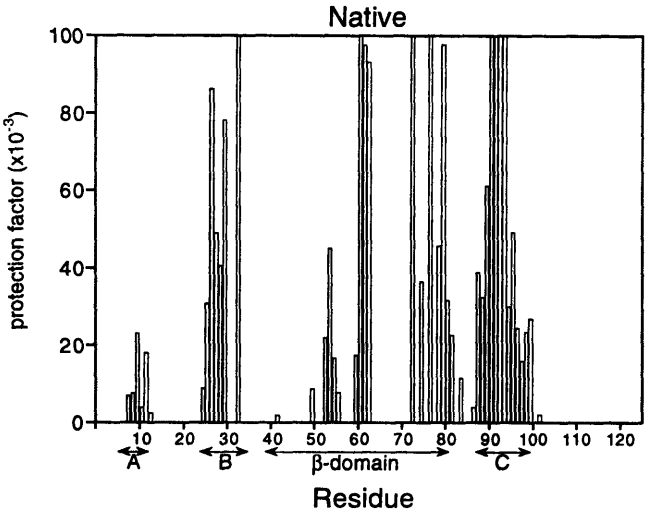
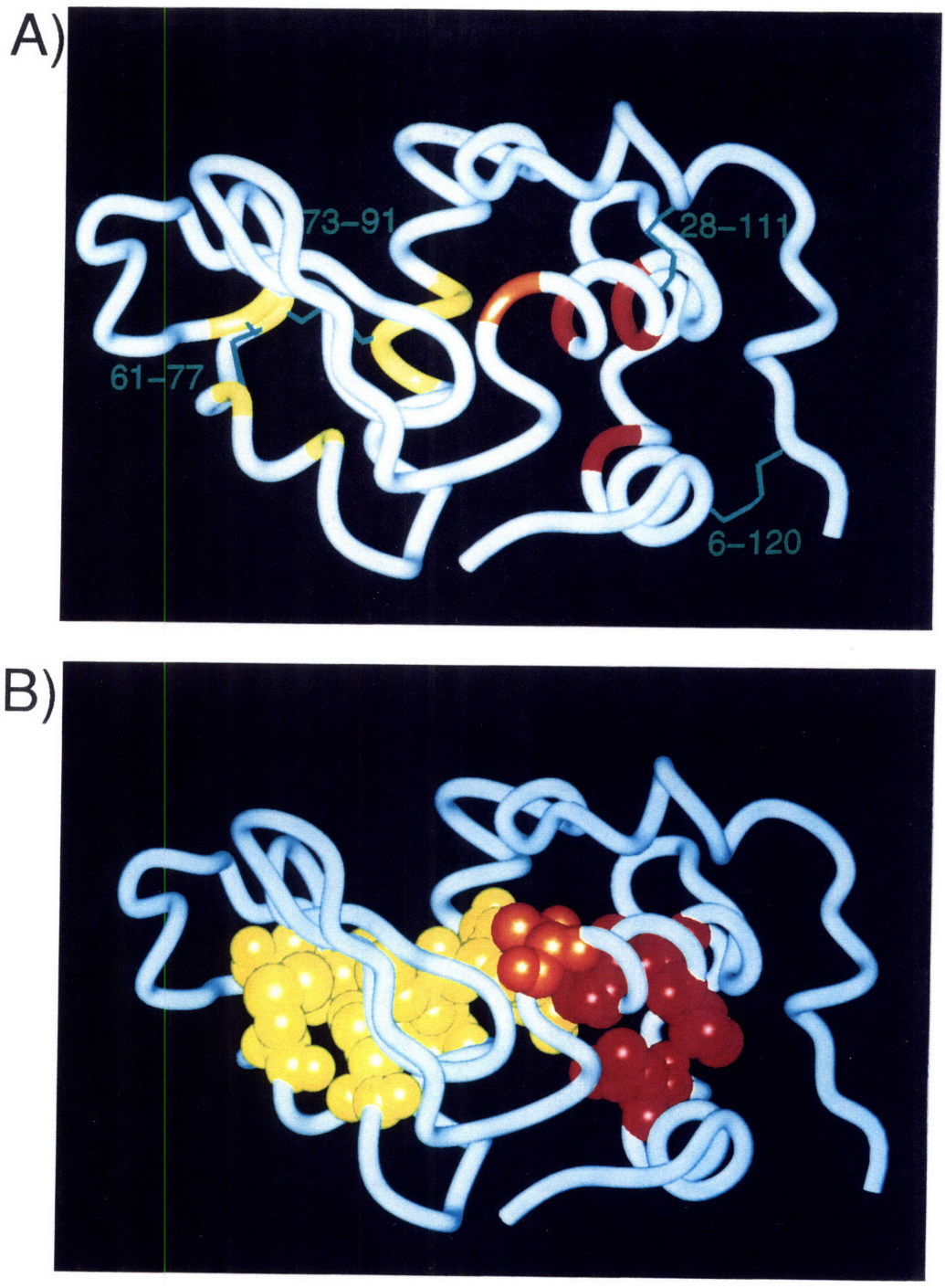

Native
protection factor (x10-3)
100
80
60
40
20
0
10 20 30 40 50 60 70 80 90 100 110 120
A
B
β-domain
C
Residue

Fig. 4



CHAPTER 3

NON-COOPERATIVE PROTEIN FOLDING

Small proteins generally fold cooperatively: disruption of significant parts of the folded structure leads to unfolding of the rest. We report here that a native-like overall fold is achieved by *non*-cooperative assembly of a subdomain within the α-lactalbumin (α-LA) molten globule. Evidence for non-cooperative folding was obtained from proline scanning mutagenesis. In contrast to the drastic destabilizing effects of proline substitutions in cooperatively folded proteins, proline mutations in the molten globule cause only individual helices to unfold, without substantially influencing the remainder of the subdomain. Thus, the key determinants of a protein's overall fold may not be of the all-or-none type.

One of the major problems in protein folding is to identify the critical features of a protein's primary sequence that code for its native overall fold. Molten globules provide simplified model systems for addressing this question. Although the term "molten globule" has been used to describe a wide variety of protein forms, classic molten globules, of which the A-state of α-lactalbumin (α-LA) is the paradigm (1), have the following characteristics: i) a near-native level of secondary structure; ii) compactness, with dimensions slightly larger than the native protein; and iii) a lack of extensive tertiary interactions, as judged by a number of criteria including a broad thermal transition and poor NMR chemical shift dispersion (1, 2).

Several pieces of evidence (3-7) indicate that the molten globule of α-LA has a native-like overall fold, even though it lacks the extensive fixed tertiary interactions that are characteristic of native proteins (3, 8). Thus, molten globules are likely to represent the simplest form of a protein in which there is substantial information transfer from the primary sequence to the three-dimensional fold. Detailed knowledge of the structures of molten globules is essential for understanding why they have a native-like fold. However, it is difficult to obtain detailed three-dimensional structural information about molten globules because they lack fixed packing interactions. Indeed, no high-resolution structure has been solved for any classic molten globule.

Here we use proline-scanning mutagenesis to probe the structure and folding of the α-LA molten globule. In this approach, individual residues are changed, one at a time, to proline (9). Proline substitutions are chosen because they unfold or greatly destabilize native proteins and structured peptides when inserted in the middle of secondary structures (10, 11). Because molten globules have a high degree of secondary structure, we reasoned that proline mutations would be useful probes for structure and folding (12).

Proline-sensitive regions were identified in the molten globule formed by a variant of α-LA, termed [28-111], by characterizing the circular dichroism (CD) spectra of proline mutants (13). [28-111] has been shown previously to retain the properties of the α-LA molten globule even near neutral pH (14). Numerous studies have shown that the native-like structure in the α-LA

molten globule is localized predominantly to the helical domain (3, 4, 8, 15, 16) and that the helices formed in the molten globule span the same, or nearly the same residues that are helical in native α-LA (17). Accordingly, the proline substitutions were targeted to residues that are helical in the native protein (Fig. 1A).

A number of sites are sensitive to proline mutations (Fig. 1B, C) (18). The loss of overall helix-content observed in the mutants appears to be specific for proline, as alanine substitutions at proline-sensitive sites have no significant effect (Fig. 1C). Remarkably, in no case do the proline mutations cause complete unfolding of the molten globule. This observation is surprising because the destabilizing effect of proline substitutions in native proteins (2.5-5.7 kcal/mol) (11) is comparable to the total free energy estimated to stabilize the molten globule of human α-LA (16, 19). Thus, if the molten globule were folded in an all-or-none cooperative manner, a proline mutation that disrupted a single helix would be expected to result in complete unfolding.

The small decrease in the helical CD signal observed in the proline mutants is roughly of the magnitude expected if only the helix containing the mutation, and not the entire protein, unfolds upon proline substitution (Fig. 2A). Indeed, the decrease in helix-content is similar for proline mutations at a number of different sites within each individual native helix (Fig. 1C). Moreover, the helix-content is almost unchanged when a second proline substitution is made within a single helix (Fig. 1C, hatched bars). Particularly since native-like helices are known to be formed in the α-LA molten globule (17), we conclude that, to a good approximation, proline mutations cause only individual helices within this molten globule to unfold.

Our results indicate that the A-, B-, D- and 3_{10}-helices are folded in the molten globule formed by human α-LA, but that the C-helix is not (Fig. 1C, 2A) (20). The structure of native α-LA suggests that the helices that are folded in the molten globule form a subdomain (Fig. 2B) (21). However, our observation that proline mutations cause only individual helices to unfold demonstrates that this subdomain is not formed in an all-or-none manner.

The ability to use proline mutations to knock out the helices in the molten globule, one at a time, allows us to evaluate the contribution of each helix toward forming the native-like overall fold of the molten globule. The preference of the polypeptide to form a native-like topology is reflected by the equilibrium distribution of disulfide pairings in the two-disulfide variant of α-LA, termed α-LA(α), with only the disulfides in the helical domain (4). There are three possible ways to form two-disulfide pairings in α-LA(α), and the equilibrium distribution of the three different disulfide isomers reflects the extent to which the chain prefers a particular backbone topology (Fig. 3A). Wild-type α-LA(α), which has the characteristics of the acid-induced molten globule of α-LA, has over a 25-fold greater preference for forming native disulfide bonds under native conditions than under denaturing conditions (4, 22).

The striking result is that none of the individual helices within the subdomain identified by proline-scanning mutagenesis is absolutely required by the remainder of the protein for forming a native-like overall fold: introduction of prolines into the A-, B-, D- or 3_{10}-helices does not abolish the preference for forming native disulfide bonds, although in some cases this preference is reduced slightly (Fig. 3B) (22, 23). Mutating additional helices reduces further the preference for forming native disulfide bonds, and the effects of multiple proline mutations are additive (Fig. 3B). The simultaneous introduction of prolines into the A-, B- and D-helices reduces the fraction of molecules that form native disulfide pairings from the wild-type value of 85% to 16%. Interestingly, this latter value is still significantly larger than the fraction predicted with a random-walk model (2%) or observed for α-LA(α) in denaturing conditions (~3%) (4).

Taken together with earlier studies, our results suggest a view of how the α-LA molten globule forms a native-like topology that differs dramatically from a cooperative all-or-none transition (24). It appears that the α-LA polypeptide has a significant preference for a native-like overall fold, even in the absence of substantial secondary structure formation. Hydrophobic collapse of the polypeptide chain is likely to be an important global feature for folding (25), and our results suggest that a loss of collapse

accompanies a decrease in the preference for forming a native-like topology (26). The inherent preference for a native-like overall fold in the collapsed α-LA polypeptide is enhanced by the formation of individual helices. These helices form independently of one another: individual helices unfold upon proline mutagenesis and no single helix is required to form a native-like overall fold (Fig. 4).

It is remarkable that a protein's overall three-dimensional fold, attained early and in the absence of extensive side-chain packing, is formed in a non-cooperative manner. It will be important to understand how a protein's fold is determined non-cooperatively: properties such as hydrophobic/hydrophilic patterning, side-chain volumes and secondary structure propensities are good candidates for investigation (3, 4, 27). The relationship between the non-cooperative folding observed here and the cooperative folding events described traditionally is likely to involve later consolidation of side-chain packing (28).

Figure Legends

Fig. 1: Proline scanning mutagenesis of the α-LA molten globule. **(A)** Sites of proline substitutions are shown on a schematic representation (29) of the structure of native human α-LA (30). The helices are labeled "A", "B", "C", "D" and "3_{10}"; the disulfide bonds are identified by the cysteine residues they bridge; and the region of α-LA shown previously to contain the essential elements of the molten globule is shaded (3). **(B)** Far-UV CD spectra of [28-111] (filled diamonds) (14) and some proline mutants (18). Some sites are insensitive to proline mutations (L96P--squares), whereas others are sensitive to prolines (M30P--circles). Substantial reduction in helix-content requires multiple mutations (L12P, M30P, L119P--pluses). **(C)** Effects of proline mutations on overall helix-content, as measured by CD at 222 nm. The bars are grouped according to the sites of mutation: "wt" refers to the value of $[\theta]_{222}$ in wild-type [28-111]; "loops/beta/Ala's" refers to mutations in loops, the β-sheet domain or alanine mutations in proline-sensitive sites; and the other labels refer to the helices containing the mutations. Hatched bars indicate that the effects of two proline mutations within a region corresponding to a single helix in the native structure are similar to the effects of a single mutation.

Fig. 2: Proline substitutions cause individual helices within a subdomain to unfold. **(A)** Comparison of the observed changes in helix-content for proline mutations (black bars) and the changes expected (31) for disrupting only individual helices (hatched bars). The helices containing mutations are indicated above the bars. For the A-, B-, C- and D-helices, the observed change in $[\theta]_{222}$ is averaged for all of the mutations within each helix. For the 3_{10}-helix, L119 was mutated to proline in a background with W118 mutated to leucine (32). The magnitude of the effects of proline mutations indicates that only individual helices (A, B, D and 3_{10}) unfold with proline substitutions. **(B)** Space-filling model of the helices and disulfide bonds in the helical domain of native α-LA (30). The helices folded in the molten globule (blue) comprise a subdomain in the native structure . The C-helix (yellow), which is insensitive to proline mutations in the molten globule, is not part of this subdomain (21).

Fig. 3: The effects of knocking out individual helices, with proline mutations, on forming the native-like overall fold (assayed by native disulfide bond formation) (3, 4, 22, 23). **(A)** HPLC analyses of equilibrium disulfide bond formation in proline mutants of α-LA(α) under native conditions (upper chromatograms) and in 6M GuHCl (lower chromatograms). [6-120; 28-111] is the native species, "NN1" refers to the non-native species [6-28; 111-120] and "NN2" refers to the non-native species [6-111; 28-120]. The numbers in brackets refer to cysteine residues that are disulfide bonded. The mutation

L96P has no significant impact on native disulfide bond formation, whereas the double mutant L12P, M30P is compromised in its ability to form the native disulfide bonds. Under denaturing conditions, for all of the proline mutants, the ratios of the disulfide species are similar to those predicted from a random-walk model of the polypeptide (3). **(B)** The effects of proline mutations in individual helices ("1P"), 2 helices ("2P's") and 3 helices ("3P's") on native disulfide bond formation. The observed frequencies of forming native disulfide bonds (black bars) are compared to the frequencies expected if the energetic consequences of mutating multiple helices were additive (hatched bars). The helices containing mutations are indicated: "A" refers to mutation of L12, "B" of M30, "C" of L96, "D" of A109, and "3_{10}" of L119. An alanine substitution in the B-helix (M30A) has no significant effect on rearrangement (data not shown). The predicted effects of multiple mutations were estimated by multiplying (% native bonds) observed for the constituent individual mutations.

Fig. 4: Non-cooperative formation of an overall native fold. Hypothetical structures are shown for the unfolded, collapsed (3 proline mutant), collapsed with some secondary structures (single proline mutants), and classic molten globule stages of folding.

Notes

1. K. Kuwajima, *Proteins* **6**, 87 (1989).

2. O. B. Ptitsyn, in *Protein Folding* T. E. Creighton, Ed. (W. H. Freeman and Co., New York, 1992) pp. 243-300; C. M. Dobson, *Current Biology* **4**, 636 (1994); D. Eliezer, P. A. Jennings, P. E. Wright, S. Doniach, K. O. Hodgson, H. Tsuruta, *Science* **270**, 487 (1995).

3. Z.-y. Peng, P. S. Kim, *Biochemistry* **33**, 2136 (1994).

4. L. C. Wu, Z.-y. Peng, P. S. Kim, *Nature structural biology* **2**, 281 (1995).

5. Z.-y. Peng, L. C. Wu, P. S. Kim, *Biochemistry* **34**, 3248 (1995).

6. P. P. de Laureto, V. D. De Filippis, M. Di Bello, M. Zambonin, A. Fontana, *Biochemistry* **34**, 12596 (1995).

7. G. Wilson, S. J. Ford, A. Cooper, L. Hecht, Z. Q. Wen, L. D. Barron, *J Mol Biol* **254**, 747 (1995).

8. A. T. Alexandrescu, P. A. Evans, M. Pitkeathly, J. Baum, C. M. Dobson, *Biochemistry* **32**, 1707 (1993).

9. Analogous to alanine-scanning mutagenesis, which identifies side-chains that contribute to protein-protein interactions [B. C. Cunningham and J. A. Wells, *Science* **244**, 1081 (1989)], proline-scanning mutagenesis is aimed at identifying regions of the backbone that specify a protein's fold.

10. J. S. Richardson, D. C. Richardson, *Science* **240**, 1648 (1988); K. G. Strehlow, A. D. Robertson, R. L. Baldwin, *Biochemistry* **30**, 5810 (1991); D. L. Minor Jr., P. S. Kim, *Nature* **367**, 660 (1994); C. K. Smith, J. M. Withka, L. Regan, *Biochemistry* **33**, 5510 (1994); D. L. Minor Jr., P. S. Kim, *Nature* **371**, 264 (1994); S. J. Wood, R. Wetzel, J. D. Martin, M. R. Hurle, *Biochemistry* **34**, 724 (1995).

11. K. T. O'Neil, W. F. De Grado, *Science* **250**, 646 (1990); A. Horovitz, J. M. Matthews, A. R. Fersht, *J Mol Biol* **227**, 560 (1992); U. H. Sauer, D. P. San, B. W. Matthews, *J Biol Chem* **267**, 2393 (1992); M. Blaber, X-j. Xhang, J. D. Lindstrom, S. D. Pepiot, W. A. Baase, B. W. Matthews, *J Mol Biol* **235**, 600 (1994).

12. Proline scanning mutagenesis will undoubtedly be useful for studying other protein folding intermediates, and could be used to improve our understanding of other proteins for which conformational heterogeneity hinders high-resolution structure determination.

13. Mutations were made as described previously (4), and were verified by restriction digests and sequencing the entire α-LA gene. Proline mutants were expressed, purified from inclusion bodies by ion exchange chromatography, refolded, and purified by reversed-phase HPLC as described previously (4, 5) except that NaCl was excluded from the ion exchange wash buffer and mutants of α-LA(α) were HPLC-purified in reduced form.

14. [28-111] contains a single disulfide bond between cysteines 28 and 111, with all other cysteines mutated to alanines. [28-111] is monomeric and retains the characteristic features of the acid-induced α-LA molten globule even near-neutral pH (5). The use of a single disulfide variant ensures formation of a particular disulfide bond, so formation of non-native disulfide isomers cannot obscure the effects of mutations on CD spectra.

15. J. Baum, C. M. Dobson, P. A. Evans, C. Hanley, *Biochemistry* **28**, 7 (1989); C. L. Chyan, C. Wormald, C. M. Dobson, P. A. Evans, J. Baum, *Biochemistry* **32**, 5681 (1993).

16. B. A. Schulman, Z.-y. Peng, C. Redfield, C. M. Dobson, P. S. Kim, *J Mol Biol* **253**, 651 (1995).

17. Numerous studies indicate that the helices folded in the α-LA molten globule are native-like: (i) A near-native level of helical structure is detected by CD, ORD and IR studies [reviewed in (1)]. (ii) The helical secondary structure is localized to the helical domain of the native protein (3). (iii) Raman optical activity measurements indicate that the helices in the molten globule are similar to those in native α-LA (7). And (iv), protection from amide hydrogen exchange is high for residues that are helical in the native protein (8, 15, 16). Hydrogen exchange is measured using a pH-jump method [8, 15; see also F. M. Hughson, P. E. Wright, R. L. Baldwin, *Science* **249**, 1544 (1990); M-F. Jeng, S. W. Englander, *J Mol Biol* **221**, 1045 (1991)], and thus is a stringent test for structure, because in order to observe protection, hydrogen exchange must be retarded both within the molten globule and also throughout the transition to the native state.

18. CD spectra were recorded in a 1-mm pathlength cell at protein concentrations of ~25 μM at 4°C. As a standard, $[\theta]_{222}$ was also measured for fresh samples of [28-111] at the same time as each variant, and the variation was <5%. All protein concentrations were determined by tryptophan, tyrosine and cystine absorbance [H. Edelhoch, *Biochemistry* **6**, 1948-54 (1967)]. An Aviv Model 60DS CD spectrophotometer equipped with a thermoelectric temperature controller was used. For [28-111] and 5 variants (L12P, M30P, L96P, A109P, and L119P), sedimentation equilibrium was performed with a Beckman XL-A analytical ultracentrifuge using an An-60 Ti rotor and Beckman epoxy 6-sector cells. Protein solutions were dialyzed overnight

versus 10 mM Tris, 0.5 mM EDTA, pH 8.5. Data were collected at 24 and 28 krpm at 4°C for three initial concentrations (15, 40 and 100 μM). The density of the buffer and the partial specific volumes were calculated with constants from T. M. Laue, B. D. Shah, T. M. Ridgeway, S. L. Pelletier, in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* , S. E. Harding, A. J. Rowe and J. C. Horton, Eds. (The Royal Society of Cambridge, Cambridge, 1992), pp. 90-125. Data sets were processed individually using the curve-fitting routine in Kaleidograph (Abelbeck software). Molecular weights were all within 10% of those calculated for ideal monomers, and the fits displayed random residuals.

19. M. Nozaka, K. Kuwajima, K. Nitta, S. Sugai, *Biochemistry* **17**, 3753 (1978).

20. These results agree with and extend an earlier hydrogen exchange study of the human α-LA molten globule, which indicated that the A- and B-helices, but not the C-helix, were folded (16).

21. Only four (A, B, D and 3_{10}) of the five helices within the helical domain of native α-LA are folded in the molten globule, and they are localized to a subdomain of the native protein. There is an extensive network of van der Waals contacts between the A-, B-, D- and 3_{10}-helices in the structure of native α-LA. On the other hand, these helices make only a small number of contacts to the C-helix and β-sheet domain (Fig. 2B). The extensive interactions localized within this region of the native protein (i.e., without the C-helix) was reported recently in a new algorithm for identifying protein domains [A. S. Siddiqui, G. J. Barton, *Protein Science* **4**, 872-884 (1995)].

22. Equilibrium disulfide exchange experiments were performed at room temperature in an anaerobic chamber (Coy Laboratory Products). Native buffer is 10 mM Tris, 0.5 mM EDTA, pH 8.5; denaturing buffer is 6 M guanidine-HCl, 10 mM Tris, 0.5 mM EDTA, pH 8.5. Exchange was initiated by adding reduced, HPLC-purified variants of α-LA(α) to exchange buffers containing 100 μM reduced glutathione and 20 μM oxidized glutathione. The final protein concentration was 5 μM. Exchange was quenched with 10% formic acid (v/v) after 24, 36 and 48 hours of equilibration and analyzed by reversed-phase HPLC as described previously (4). Time-courses indicate that disulfide exchange reached equilibrium for every variant by 36 hours, after which no further changes were observed. Peak areas are reproducible to within ± 2%. For wild-type α-LA(α) and 5 mutants (L96P, L12P M30P, L12P A109P, M30P A109P and L12P M30P A109P), exchange was also initiated from at least one purified oxidized species and the peak areas are reproducible to the same limits. All three major species (corresponding to the native plus the two non-native isomers) were completely oxidized, as assayed by the lack of

reaction with Ellman's reagent [5, 5'-dithio-bis(2-nitrobenzoic acid)] in 6M guanidine-HCl [G. L. Ellman, *Arch Biochem Biophys 82*, 70 (1959)].

23. Three different disulfide isomers are possible for α-LA(α) and three different species are observed. The non-native isomer with disulfides between cysteines 6-28 and 111-120 (denoted "NN1") was identified readily because it is the predominant species (>96%) at equilibrium in strongly denaturing conditions, as predicted by a random walk model (3). The other non-native species, with disulfides between cysteines 6-111 and 28-120 (denoted "NN2"), and the native species (6-120; 28-111), were assigned by proteolysis with endoproteinase-Glu-C (Endo-Glu), which cleaves after Asp and Glu residues, and mass spectrometry following purification by reversed-phase HPLC. Briefly, HPLC-purified native α-LA(α) or NN2 was dissolved in 50 mM sodium phosphate (pH 7.5) at approximately 1 mg/ml. Endo-Glu (Sigma) was dissolved in water at approximately 1 mg/ml. 1/10th volume of endo-Glu was added to the different isomers of variants of α-LA(α), and digestion proceeded for 2-3 hours at room temperature. Under these conditions, the fragments useful for disulfide assignment reflect preferential cleavage after Glu 7, Glu 25, Glu 43, Asp 97, Glu 113, Glu 116 and Glu 121, except in variants containing the L119P mutation, in which there is not substantial cleavage after Glu 121. The digests were desalted by reversed-phase HPLC with a C18 column (Vydac) and a fast gradient (4.5-54% acetonitrile in water over 6 minutes, with 0.1% TFA). Oxidized and reduced digestion products were analyzed directly after HPLC by laser-desorption mass spectrometry (Voyager Elite, Perseptive Biosystems). Programs that list all possible cleavage products and their corresponding masses, kindly provided by Dr. Zheng-yu Peng, allowed identification of the fragments with masses that are unambiguous for a particular disulfide isomer of each mutant. In all cases, the observed mass was within 2 Da of the mass expected for the fragment. For the native isomer, the 6-120 disulfide bond was distinguished by the mass of a fragment spanning from the N-terminal methionine (residue 0) through residue 7, disulfide-bonded to a fragment corresponding to residues 117 through 121 [denoted (0-7)-SS-(117-121)]. For mutants with L119P, the mass distinguishing 6-120 corresponds to (0-7)-SS-(117-123). The 28-111 disulfide bond was distinguished by two masses, corresponding to (26-43)-SS-(98-113) and (26-43)-SS-(98-116). In the reduced samples of the native isomers, the masses of these fragments disappear and new masses appear, corresponding to the individual fragments (26-43), (98-113), and (98-116). For the NN2 isomer, the 6-111 disulfide bond is distinguished by two masses, corresponding to disulfide-bonded fragments (0-7)-SS-(98-113) and (0-7)-SS-(98-116). The 28-120 disulfide bond is distinguished by the mass of the disulfide-bonded fragment (26-43)-SS-(117-121).

24. The heat and denaturant-induced unfolding transitions of the α-LA molten globule are also inconsistent with a two-state transition [D. Xie,

V. Bhakuni, E. Freire, *Biochemistry* **30**, 10673 (1991); K. Yutani, K. Ogasahara, K. Kuwajima, *J Mol Biol* **228**, 347 (1992); A. Shimizu, M. Ikeguchi, S. Sugai, *Biochemistry* **32**, 13198 (1993)]. In addition, two different molten globule forms of apo-myoglobin have been identified, with one form containing an additional helix that is not folded in the other [S. N. Loh, M. S. Kay, R. L. Baldwin, *Proc Natl Acad Sci, U.S.A.* **92**, 5446 (1995); T. Kiefhaber, R. L. Baldwin, *J Mol Biol* **252**, 122 (1995)].

25. Reviewed in K. A. Dill, *Biochemistry* **29**, 7133 (1990).

26. The following considerations suggest that the proline mutations of α-LA(α) diminish collapse of the polypeptide. In the equilibrium disulfide-exchange experiments with the mutants, decreases in the population of the native disulfide bonded isomer are accompanied by corresponding much greater increases in the NN1 ([6-28; 111-120]) isomer than the NN2 ([6-111; 28-120]) isomer (see Fig. 3A, for example). Comparison of the effective concentrations of the individual disulfide bonds in the native, NN1 and NN2 species, in native buffer and in guanidine-HCl, suggests that formation of the NN2 disulfide bonds is enhanced by collapse, whereas formation of the NN1 disulfide bonds is not (5). Briefly, in the absence of any proline mutations, the NN2 disulfide bonds have a 10- to 100-fold greater effective concentration in native buffer than in guanidine. This enhancement in effective concentration probably arises from the collapse of the polypeptide chain in aqueous buffer, which brings together the cysteines (6-111 and 28-120) that are far apart in the protein sequence. (The effective concentrations of the native disulfide bonds are enhanced ~100- to 1000-fold.) The effective concentrations of the NN1 and NN2 disulfide bonds are similar to each other in native buffer. However, in contrast to the NN2 disulfide bonds, the effective concentrations of the NN1 disulfides are the same in native buffer and guanidine, presumably because they are between cysteines that are nearest neighbors in the sequence (6-28 and 111-120), and thus are not enhanced by collapse. Therefore, in the proline mutants, the preferential accumulation of the NN1 isomer, with little change in NN2, likely reflects a loss of collapse as well as specificity.

27. K. F. Lau, K. A. Dill, *Proc Natl Acad Sci, U. S. A.* **87**, 638 (1990); M. J. Behe, E. E. Lattman, G. D. Rose, *Proc Natl Acad Sci, U. S. A.* **88**, 4195 (1991); J. U. Bowie, R. Luthy, D. Eisenberg, *Science* **253**, 164 (1991).

28. O. B. Ptitsyn, *FEBS Lett* **285**, 176 (1991).

29. J. P. Priestle, *J Appl Crystallogr* **21**, 572 (1988).

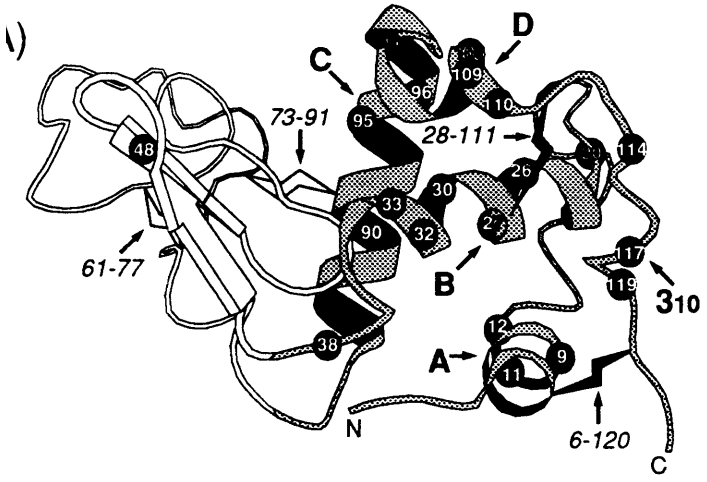
30. K. R. Acharya, J. S. Ren, D. I. Stuart, D. C. Phillips, R. E. Fenna, *J Mol Biol* **221**, 571 (1991).

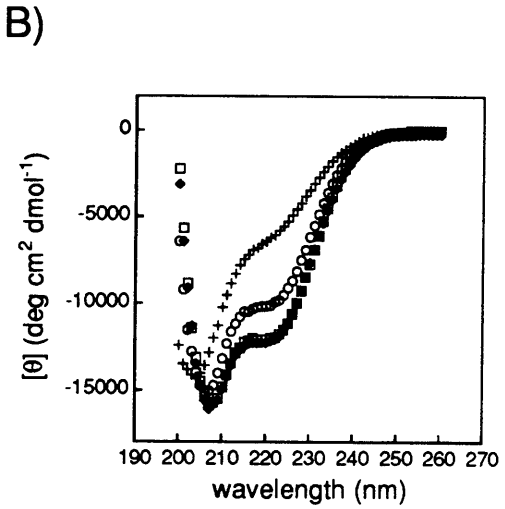
31. In order to estimate the decrease in $-[\theta]_{222}$ expected for knocking out a single native-like helix, the helices in the native structure are defined as follows (30): "A" (the A α-helix plus a contiguous turn of 3_{10}-helix) spans residues 5 through 16; "B" spans residues 23 through 34; "C" spans residues 86 through 99; "D" spans residues 105 through 109; and "3_{10}" spans residues 115 through 119. The predicted decreases in $-[\theta]_{222}$ were calculated by assuming a value of -29,400 deg cm^2 $dmol^{-1}$ for residues in short α-helices [Y. H. Chen, J. T. Yang, K. H. Chau, *Biochemistry* **13**, 3350 (1974)] and taking account of 123 total residues in α-LA. As an example, the B-helix contains 12 residues of helix, so the value of $[\theta]_{222}$ contributed by the B-helix is estimated as: (12 / 124) * -29,400 = -2845 deg cm^2 $dmol^{-1}$.

32. The loss of $-[\theta]_{222}$ with proline mutations in the 3_{10}-helix is slightly greater than predicted. W118, in the middle of this region, may make an aromatic contribution to $[\theta]_{222}$ [R. W. Woody, in *The Peptides: Analysis, Synthesis, Structure*, V. J Hruby, Ed. (Academic Press, Orlando, 1985), vol. 7, pp. 15-114; M. C. Manning and R. W. Woody, *Biochemistry* **28**, 8609 (1989); A. Chakrabartty, T. Kortemme, S. Padmanabhan, R. L. Baldwin, *Biochemistry* **32**, 5560 (1993)]. To test this, we made a mutant of [28-111] with W118 mutated to leucine. The effects of the L119P mutation are different in the wild-type and W118L backgrounds (compare Fig. 1C with Fig. 2A). On the other hand, the effects of a different proline mutation (M30P) are roughly the same in the two backgrounds (data not shown). Therefore, for comparison in Fig. 2, we use the results for the 3_{10}-helix obtained in the W118L background.

33. We thank P. Harbury, D. Minor, Z-y. Peng, J. Weissman and L. Wu for exciting discussions which led to some of the ideas presented in this manuscript, and P. Murray and members of the Kim lab for other helpful comments. B. A. S. was supported by a National Institutes of Health predoctoral training grant (T32AI07348). This work was supported by the Howard Hughes Medical Institute.

Fig. 1


A)
C
D
73-91
28-111
61-77
B
310
A
N
6-120
C


B)
[θ] (deg cm2 dmol-1)
0
-5000
-10000
-15000
190 200 210 220 230 240 250 260 270
wavelength (nm)

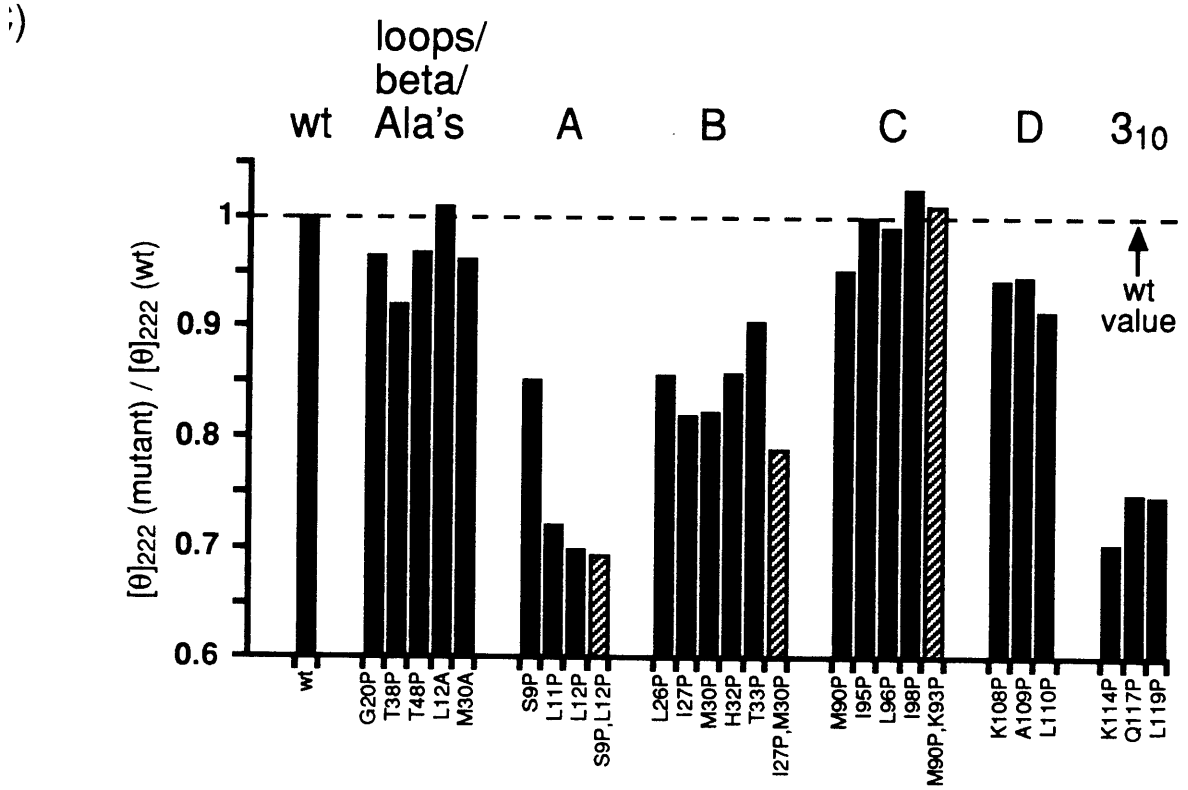
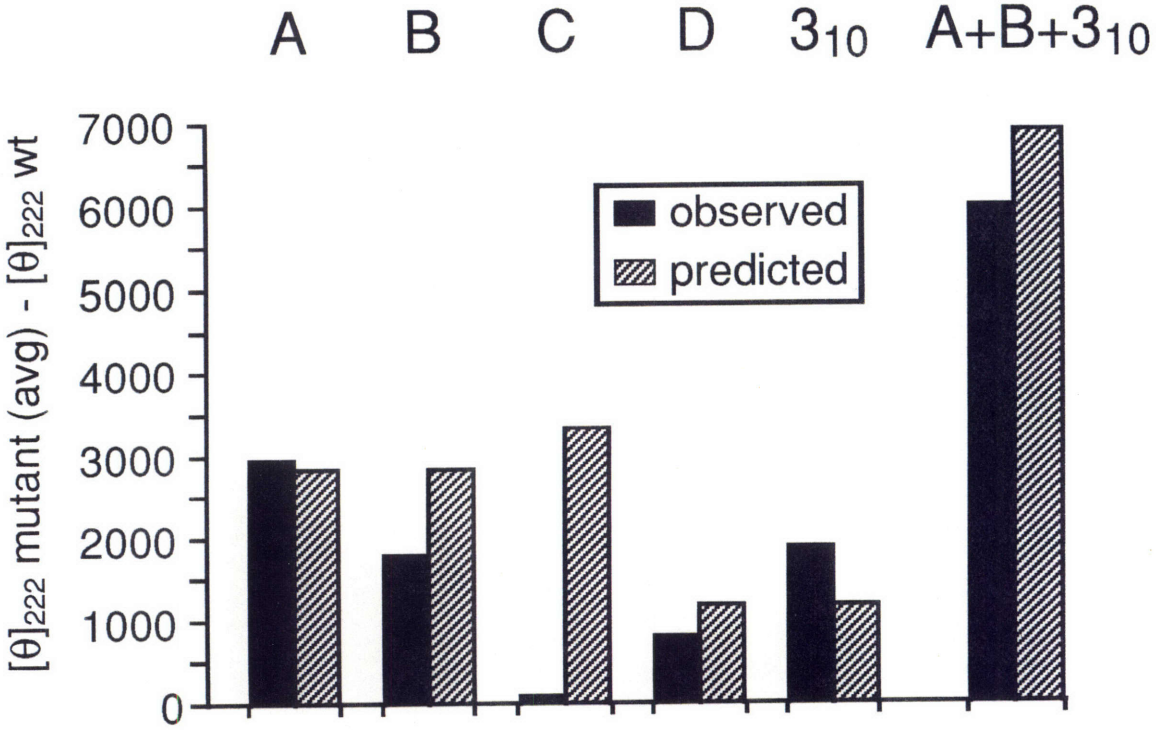

C)
wt
loops/
beta/
Ala's
A
B
C
D
310
[θ]222 (mutant) / [θ]222 (wt)
1
0.9
0.8
0.7
0.6
wt
value
wt
G20P
T38P
T48P
L12A
M30A
S9P
L11P
L12P
S9P,L12P
L26P
I27P
M30P
H32P
T33P
I27P,M30P
M90P
I95P
L96P
I98P
M90P,K93P
K108P
A109P
L110P
K114P
Q117P
L119P

Fig. 2

A)


A
B
C
D
3_{10}
A+B+3_{10}
$[\theta]_{222}$ mutant (avg) - $[\theta]_{222}$ wt
7000
6000
5000
4000
3000
2000
1000
0
observed
predicted

B)

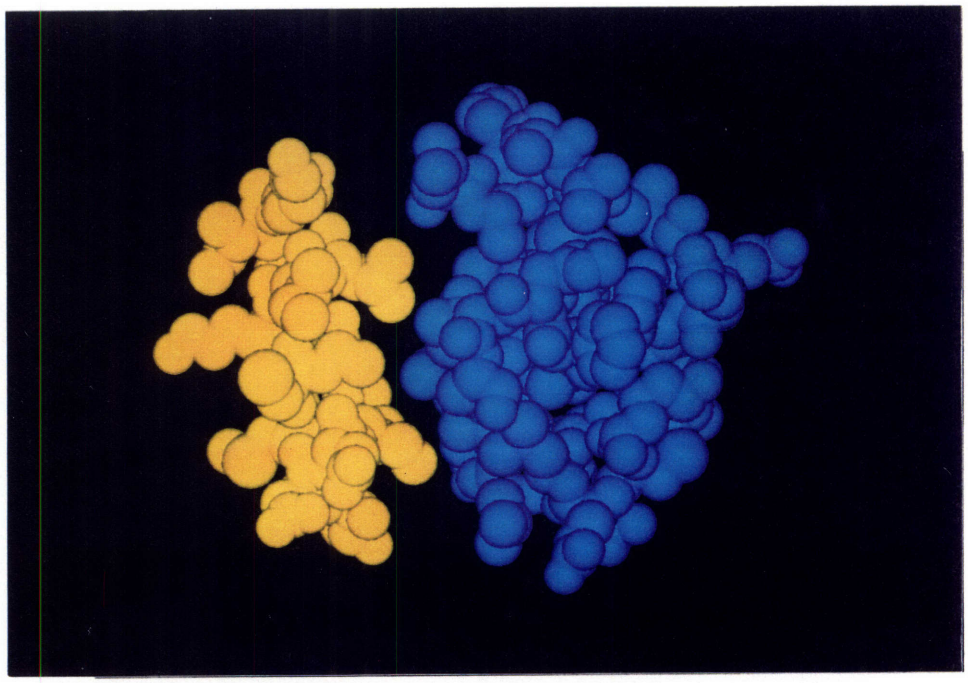
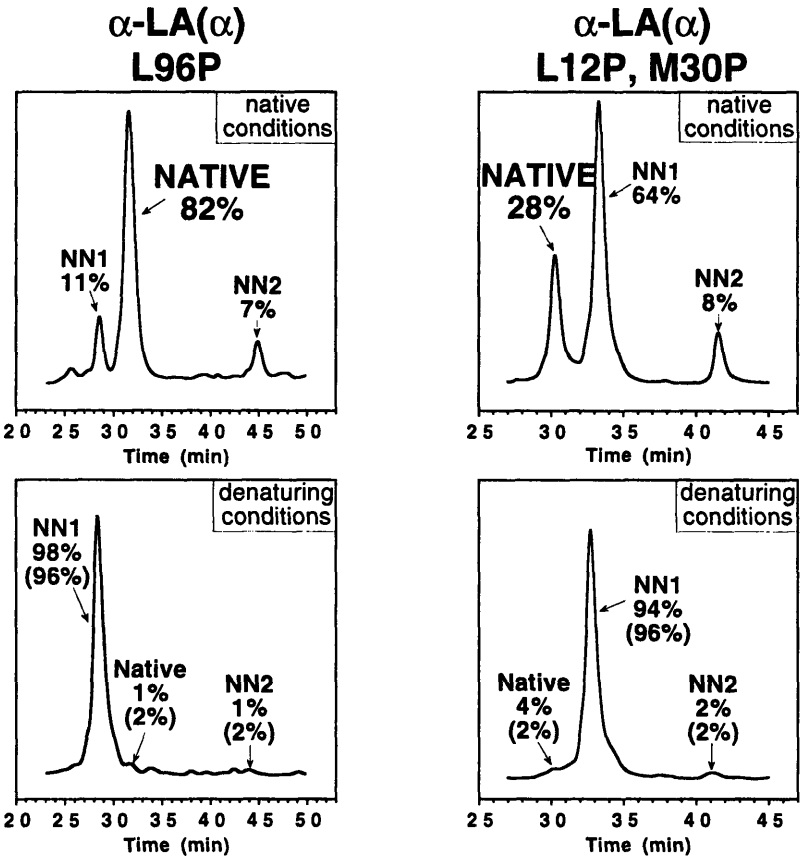


Fig. 3

A)


α-LA(α)
L96P
native conditions
NATIVE
82%
NN1
11%
NN2
7%
20 25 30 35 40 45 50
Time (min)
denaturing conditions
NN1
98%
(96%)
Native
1%
(2%)
NN2
1%
(2%)
20 25 30 35 40 45 50
Time (min)
α-LA(α)
L12P, M30P
native conditions
NATIVE
28%
NN1
64%
NN2
8%
25 30 35 40 45
Time (min)
denaturing conditions
NN1
94%
(96%)
Native
4%
(2%)
NN2
2%
(2%)
25 30 35 40 45
Time (min)

B)

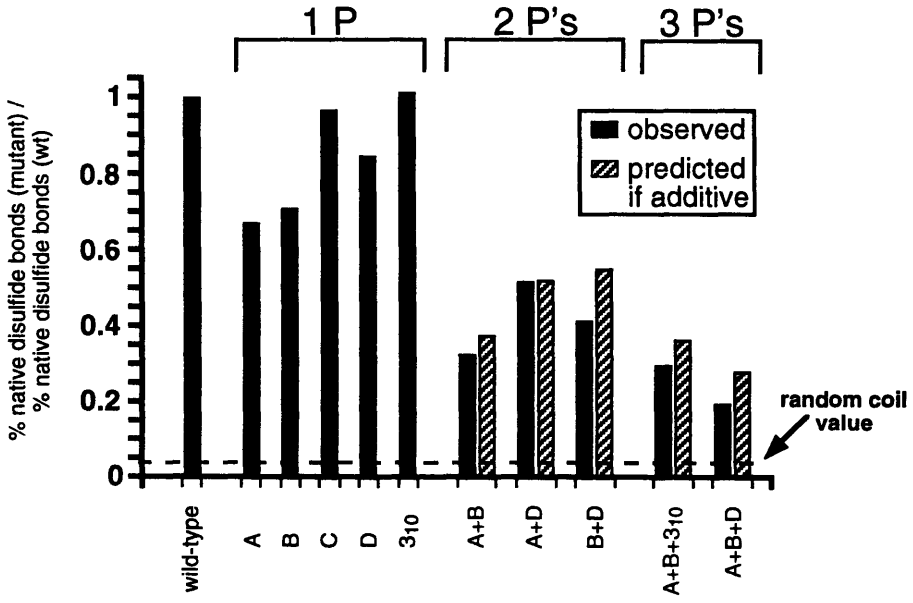
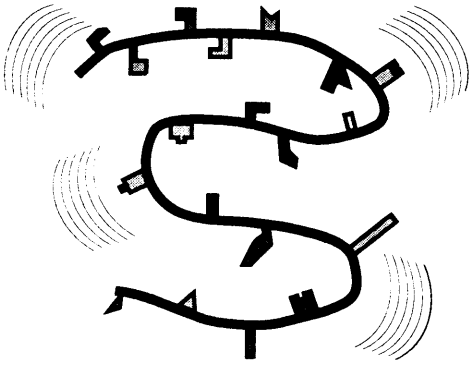

1 P
2 P's
3 P's
observed
predicted if additive
% native disulfide bonds (mutant) / % native disulfide bonds (wt)
0
0.2
0.4
0.6
0.8
1
random coil value
wild-type
A
B
C
D
3₁₀
A+B
A+D
B+D
A+B+3₁₀
A+B+D

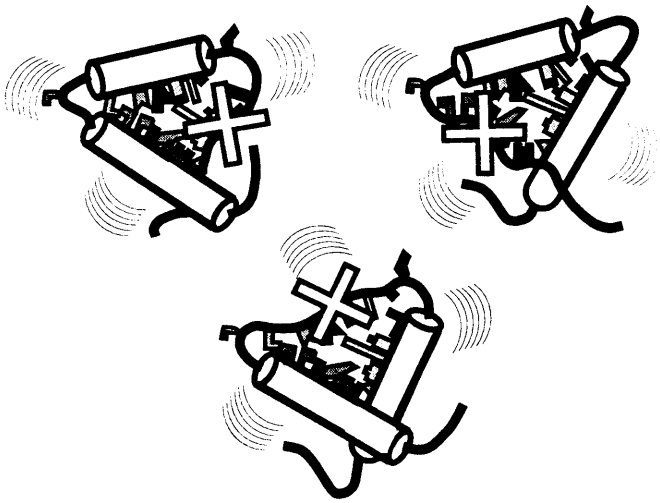
Fig. 4

Unfolded

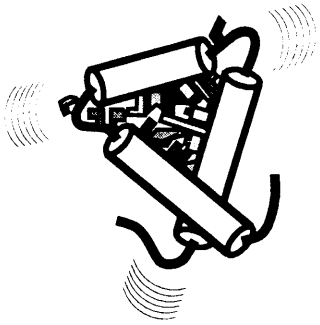


Hydrophobic Collapse

Single Proline Mutants



Classic Molten Globule



CHAPTER 4

NON-COOPERATIVE FORMATION OF A NATIVE-LIKE TOPOLOGY AND FORMATION OF FIXED TERTIARY INTERACTIONS IN α-LACTALBUMIN

Are molten globules folded non-cooperatively?

Whether or not molten globules fold in an all-or-none manner is highly debated. On one hand, Ptitsyn and his colleagues present two pieces of evidence that the molten globule is a discrete thermodynamic state of proteins. First, they observe a correlation between the slope of denaturant-induced unfolding and the molecular weights of a large number of proteins. Second, they monitor the change in elution volume during denaturation of native proteins by size-exclusion chromatography, and observe a bi-modal change.

Ptitsyn and his colleagues argue that molten globules will unfold in a manner that is proportional to their molecular weights if and only if the entire polypeptide chain is cooperatively folded in an all-or-none manner (1). The slopes of the denaturant-induced unfolding transitions for molten globules are found to be proportional to the molecular weights of the proteins, and these results are interpreted as indicating that molten globules unfold in a two-state manner. The underlying assumption of this argument is that guanidine binds an unfolded protein roughly proportionally to its molecular weight.

For the size exclusion chromatography experiment, Ptitsyn and Uversky observe a bi-modal change in elution volume upon addition of guanidine, reflecting the existence of two distinct states with different degrees of compactness (2). For β-lactamase, a protein known to unfold via a molten globule intermediate, the elution volume changes at guanidine concentrations *above* that required for unfolding (3). Thus, the two-state transition is thought to coincide with the transition from one denatured state to another. The intermediate is thought to be as compact as the native state, and it appears to become less compact in an all-or-none manner.

While Ptitsyn's arguments seem reasonable, there are several drawbacks with both of these studies. Three issues are unresolved regarding the correlation between denaturant-induced unfolding transitions and molecular weight. First, the mechanism of guanidine binding is poorly understood, and may not be random for any given protein sequence (4-6).

Second, the authors state that the correlation they find would only be possible if the entire polypeptide chain unfolds in an all-or-none manner. In several molten globules, however, residual structure appears to be localized to one subdomain of the protein (7, 8). Third, the "control" native proteins, thought to unfold in a two-state cooperative manner, include some proteins such as cyt c, that unfold via intermediates, termed partially unfolded forms (PUFS) (9). Cyt c unfolds with guanidine in a progressive, rather than all-or-none manner, and yet it still aggrees with Ptitsyn's predictions. Therefore, it is not clear what the correlation between guanidine-induced unfolding and molecular weight really means.

In addition, limiting the number of states observed in the size exclusion chromatography experiments may not be possible for two reasons. First, each of the two elution peaks apparently corresponds to at least two states (the "compact" peak contains both the native and molten globule states and the "non-compact" peak contains the pre-molten globule and unfolded states). Thus, it is plausible that numerous states have the same level of compactness. In addition, molten globules have a strong tendency to aggregate in salt, and the observed bi-modal transition occurs in a range of guanidine concentrations where the α-LA molten globule aggregates significantly (1-2 M guanidine, Schulman, Wu and Kim, unpublished results).

On the other hand, data are amassing that are consistent with the notion that molten globules are *not* folded cooperatively. For example, the thermal denaturation of the molten globule formed by apo-α-lactalbumin occurs with no measurable change in heat capacity (10). Second, the urea-induced unfolding of bovine α-LA was examined using 1D ^{1}H-NMR (11). The transitions of different aromatic residues were monitored by changes in chemical shift with increasing concentrations of urea. The transitions for different resonances are non-coincidental. Because the chemical shifts of unfolded peptides generally do not vary with urea concentration, these results suggest that unfolding of the molten globule is not a two-state process. However, interpretation of these data is complicated for two reasons. First, none of the transitions has clean baselines, so it is difficult to compare them.

Second, even the transitions of adjacent protons on a single tryptophan side chain do not appear to unfold simultaneously.

Further evidence suggesting that molten globules are non-cooperatively folded comes from the different ways to stabilize equilibrium molten globules formed by apo-myoglobin. High resolution experiments indicate that 3 native-like helices (A, G, H) are folded in one form, and a fourth helix (B) is added on in another (7, 12, 13). The fact that the B-helix is not always part of the molten globule shows that it does not fold cooperatively with the A, G, and H-helices.

The proline mutagenesis data presented in Chapter 3 demonstrate that one of the best studied molten globules, that formed by human α-LA, contains native-like secondary structures that assemble with a native-like topology in a non-cooperative manner. In the molten globule, proline substitutions unfold only individual helices, and no single helix is required for forming a more-or-less native-like overall fold in the helical domain. In addition, combinations of proline mutations have additive effects on native disulfide bond formation.

On the other hand, in native proteins, proline substitutions destabilize the entire protein. When a proline substitution does not completely unfold a protein, its structure remains more-or-less the same but its stability is dramatically decreased (14, 15). This is because native proteins are cooperatively folded. The all-or-none formation of structure in native proteins has been compared with the structure in a house of cards. If one wall is removed, the rest of the structure will fall down. On the other hand, the assembly of structure in the molten globule somehow is more flexible, and it is not known what causes the helices to assemble with a native-like overall fold.

How can a native-like tertiary fold be achieved non-cooperatively?

Two different extreme models might explain the key steps for establishing a protein's fold in a non-cooperative manner. The first model emphasizes the role of local interactions: early formation of secondary

structures might restrict the number of conformations available to the polypeptide, and thus limit the number of topologies accessible to the chain (16, 17). In contrast, the second model suggests that the pattern of hydrophobic and hydrophilic residues dictates the fold (18-20). The polypeptide chain would be restricted to forming a native-like structure by the limited number of ways that hydophobic side chains could come together, with polar side chains exposed .

Hydrophobic collapse probably drives formation of the native fold

Because the most primitive folding intermediates that have been studied (i.e. molten globules) appear to contain subdomains with folded secondary structures surrounding a core, it is difficult to deconvolute the roles of secondary structure formation from hydrophobic collapse. Nonetheless, native-like local structures do appear to be formed in the unfolded states of proteins. For example, an "unfolded state" of lysozyme, at low pH and in TFE, has substantial secondary structure (21, 22). NMR and CD experiments indicate that both the full length protein and individual peptides adopt similar structures under these conditions, with most native helices at least partially folded. Thus, secondary structures could restrict the conformations available to the chain. Indeed, the α-helix propensity of the sequence of the B-helix in the apo-myoglobin molten globule is a factor determining whether or not it is folded (13).

The most convincing evidence that local structure plays a role in dictating the three-dimensional fold of proteins comes from the use of structure prediction algorithms based on local interactions like secondary structure propensities, such as LINUS (Local Independently Nucleated Units of Structure) (23). The moderate success of this program suggests that local interactions are important for determining a protein's fold.

However, numerous studies, both experimental and theoretical, indicate that hydrophobic collapse is very important (18). Theoretical studies, championed by Dill, suggest that hydrophobic/polar (h/p) patterning in proteins directs formation the native-like topology, and also gives rise to two-state cooperativity in folding (24). Dill and his co-workers have modeled the

polypeptide backbone onto a lattice, and have discovered that some sequences fold into unique native conformations while others do not prefer any particular conformation. The h/p patterning dictates both the nature of the structure and whether a sequence has a preference for a single native conformation.

Two elegant experiments provide compelling evidence that hydrophobic collapse or tertiary interactions direct a protein's fold. In the first experiment, Xiong et al. have synthesized four short peptides to test whether secondary structure propensitites or h/p patterning are most important for dictating which secondary structures are formed in oligomeric complexes (19). One set of the peptides was biased for forming either helix or sheet based on secondary structure propensities, and the other set of peptides was biased on the basis of h/p patterning. Regardless of the propensities of the peptide sequences to form a given secondary structure, the peptides formed the structure predicted based on h/p patterning.

The work of Minor and Kim demonstrates that in at least some cases, secondary structure formation does not depend on local determinants (Minor and Kim, unpublished results). They have made two different mutants of the 57 residue immunoglobulinG binding protein domain, GB1, that contain an 11-residue sequence they call a "chameleon". In one part of the protein, this sequence forms an α-helix, but when placed in another part of the protein, "chameleon" forms a β-strand and a turn. Thus, the rest of the protein determines the secondary structure of this 11-residue sequence.

Secondary structure formation is important for forming the native topology in a non-cooperative manner

The molten globule formed by the helical domain of α-LA has a >20-fold greater preference for forming native disulfide bonds in native buffer, compared to disulfide bond formation in denaturant or predicted by a random-walk model (8, 25). Thus, monitoring disulfide bond formation in the helical domain of α-LA is a good way to examine the most important factors for forming a native-like tertiary fold. The results presented in Chapters 2 and 3 of this thesis, as well as protein dissection experiments by

Peng and Kim (unpublished data) indicate that only the A-, B-, D- and 3_{10}-helices are folded in the molten globule of α-LA. Originally, we set out to identify folded helices because we thought they would reveal the features of the protein that direct folding. In the structure of native α-LA, the helices that are folded in the molten globule form a subdomain, with an extensive network of van der Waals contacts between the helices (as opposed to the C-helix and the β-domain) (see Chapter 3). On the other hand, other features of the sequence and structure of α-LA, such as hydrophobicity or intrinsic helix propensity, do not correlate with the part of the protein that is folded in the molten globule.

In the structures of native α-LA and the homologous lysozymes, there is a major hydrophobic core between the A-, B- and 3_{10}-helices, containing aromatic side chains from the B-helix (F31, H32) and the 3_{10}-helix (W118) as well as Y36 from a loop just after the B-helix. This core fills much of the space between the 6-120 and 28-111 disulfide bonds of the helical domain. The finding that the helices folded in the molten globule surround and contribute to this core, and that more or less the same helices are folded in folding intermediates formed by human and equine lysozymes (26, 27), suggested that this core might be important for establishing the native-like overall fold of the molten globule.

It is now clear, however, that the specific interactions observed in the native state are not important for folding in the molten globule. First, mutation of Trp 118 to alanine or aspartic acid has no effect on the amount of native disulfide bond formation (unpublished results). Thus, if there is a hydophobic core in this region, it is likely that no specific residue is required. Far more compelling evidence comes from unpublished results of Wu and Kim, who find that replacing many of the hydrophobic residues in the helical domain of α-LA with leucine does not affect formation of native disulfide bonds. These mutations result in a loss of side-chain specificity and a large decrease in side-chain volume. Thus, if formation of a core is important, it does not involve specific side-chain contacts or a particular volume to be filled by side-chains. However, leucine is a good helix-former, so in this mutant, helix formation cannot be ruled out as a factor that is important native-like folding.

The only way that we have been able to somewhat destroy formation of native disulfide bonds in the molten globule of α-LA is by proline substitution. It is clear that the location of the proline substitution, and not simply the nature of the mutation, is important, because in some cases L→P mutations are deleterious and in others they are not. In addition, we tested that the effect was due to the proline side chain, and not just the decrease in hydrophobic volume: mutation of Met 30 to alanine has no significant effect. The effects of mutations on CD signal are roughly similar for all of the residues within a single native-like helix (Chapter 3). These data suggest strongly that it is the destruction of helices that is important. While the effects of these mutations are small, they add up and are significantly destabilizing when multiple helices are destroyed. Thus, secondary structure formation is one factor involved in forming the native-like tertiary fold.

However, these experiments do not test whether or not secondary structure sets up the native-like architecture, because it is possible, for example, that in order for the correct hydrophobic interactions to form, the correct secondary structure elements must be able to fold. Two results suggest that there is a relationship between collapse and secondary structure formation. First, as described in Chapter 3, there is probably decreased collapse in the proline mutants. Second, while reduced α-LA is helical (and this helical structure is sensitive to proline mutations--data not shown), the CD signal at 222 nm increases significantly when native disulfide bonds are formed (25, 28). On the other hand, the CD signal decreases when non-native disulfide bonds are formed. Thus, formation of the native-like overall fold (i.e. in the presence of the correct disulfide bonds) promotes secondary structure formation (8, 25).

While native-like secondary structure formation is at least one of the factors important for forming the native fold, collapse also appears to be critical. It is likely that neither of the extreme models accurately accounts for all of the features necessary for forming a native-like topology. Hydrophobic collapse is probably the driving force for folding, but local interactions, such as helix formation, also play a role.

What causes proteins to adopt fixed structure? (Calcium-binding in α-LA)

While the foundation for the native structure appears to be built by the molten globule stage, it is still not known what is required for adopting a unique, fixed structure. This is a major problem in protein design, as well as an unanswered question in protein folding. In designed proteins with features more reminiscent of molten globules than native proteins, charge-metal interactions have been incorporated to make the protein more native (29). Buried polar interaction have been shown to be critical for forming a unique structure in designed coiled-coils, as well (30). For α-LA, folding of the calcium binding domain has been implicated as one factor that can induce fixed packing and the transition from the molten globule to the native state (10, 31-33). α-LA is a molten globule in the absence of calcium and other salts. In contast to formation of the helical domain disulfide bonds, formation of the β- and inter-domain disulfide bonds is sensitive to calcium-induced structural changes (Appendix I). Disulfide bond reduction, 1D ^{1}H NMR, far and near UV CD, and thermal denaturation experiments have suggested that, in the presence of calcium, variants of α-LA containing only the β- and inter-domain disulfide bonds, termed α-LA(β) (see Appendix I) have a cooperatively folded calcium-binding domain with fixed tertiary interactions. In addition, NMR experiments suggest that the calcium binding region, including the C-helix, forms a discrete domain in the native protein (Chapter 2, Redifield, Schulman, Kim and Dobson, unpublished, and (33)). The same delineation of the domains is ascertained by an algorithm that predicts protein domains based on packing (34).

If the calcium-binding domain is responsible for forming fixed tertiary interactions, what is the role of the helical domain in folding? Results from kinetic refolding experiments on both α-LA and the homologous protein, lysozyme, lead us to question the importance of forming the helical intermediate. In both the presence and absence of calcium, α-LA folds via a helical intermediate. Although bovine α-LA can adopt the native structure in the absence of calcium, folding is accelerated up to 1000-fold in the presence of calcium (35), suggesting that formation of the calcium binding domain is rate-limiting.

In addition, the role of the helical intermediate that accumulates during the kinetic refolding of lysozyme, which resembles the intermediate formed by α-LA, is unclear. In the original and in recent kinetic experiments, a significant population of the molecules has been found to fold fast, while the remaining ~85% accumulates as the helical intermediate (36, 37). Furthermore, a mutant of hen lysozyme, with the cysteines involved in the 6-127 disulfide bond (corresponding to the 6-120 disulfide bond in α-LA) reduced and carboxymethylated, folds into essentially the native structure without accumulation of the helical intermediate (38).

Calcium-dependent folding of a-LA is not localized to the calcium binding domain and requires the helical domain

We have begun to dissect the interplay between the calcium-binding domain and the helical domain involved in adopting fixed structure (i.e. calcium-dependent folding) in α-LA. The studies of α-LA(β) described above, and a number of other studies of similar variants of α-LA, suggest that the important part of α-LA for calcium-dependent folding is the calcium-binding domain. To test this notion, we have studied the folding of a mutant of α-LA containing only the residues thought to be important both chemically and structurally for calcium-dependent folding. The boundaries reported by Siddiqui et al. for the two domains in a-LA were used to design this truncated version of α-LA, termed "β-domain", comprising residues 38-104 (34). This sequence contains all of the residues known to be folded in α-LA(β): the calcium binding ligands are the carbonyls of K79 and D84, and the carboxylates of D82, D87 and D88; the β- and inter-domain disulfide bonds are between cysteines 61-77, and 73-91; Tyr 103 and Trp 104 are involved in an aromatic cluster with Phe 53 and Trp 60 in the β-domain. In addition, in the 1D ^1H nmr spectrum of α-LA(β), there are three methyl peaks shifted upfield of TMSP. These resonances probably correspond to the γ methyl protons of isoleucines 55 and 95, which have similar chemical shifts in native α-LA (Redfield, Schulman, Peng, Kim and Dobson, unpublished). In the native structure, Ile 55 and Ile 95 pack in between the β- and inter-domain disulfide bonds and the aromatic cluster.

Unlike α-LA(β), however, β-domain fails to form native disulfide bonds in the presence of calcium. This leads to two possible conclusions. Either 1) formation of the molten globule, which has been removed in β-domain, is required for forming a cooperatively-folded calcium-binding domain, or 2) the calcium-binding domain extends into part of the helical domain. Experiments are in progress to test these possibilities. If the first model is correct, then formation of the molten globule would establish both the native-like architecture in the helical domain, and also the fixed tertiary interactions in the calcium-binding domain. On the other hand, if a region of the helical domain is part of the cooperatively folded calcium binding domain, this would explain why binding of calcium leads to fixed tertiary interactions in both the β- and the α-domains.

Summary

In molten globules, a native-like overall fold is achieved in a non-cooperative manner. The folding of individual helices within the molten globule plays a role in dictating the native-like topology, while the formation of specific tertiary interactions probably does not. Specific tertiary interactions are imparted to α-LA by the folding of the calcium-binding region. Formation of the β- and inter-domain disulfide bonds is calcium-sensitive, and thus provides a sensitive assay for studying the formation of specific tertiary interactions. Preliminary results indicate that the helical domain plays a role in calcium-dependent structure formation.

References

1. O. B. Ptitsyn, V. N. Uversky, *FEBS Letters* **341**, 15-18 (1994).

2. O. B. Ptitsyn, V. E. Bychkova, V. N. Uversky, *Philos Trans R Soc Lond B Biol Sci* **348**, 35-41 (1995).

3. V. N. Uversky, O. B. Ptitsyn, *Biochemistry* **33**, 2782-91 (1994).

4. T. E. Creighton, *Proteins: structures and molecular principles* (W. H. Freeman and Co., New York, 1983).

5. G. I. Makhatadze, P. L. Privalov, *J. Mol. Biol.* **226**, 491-505 (1992).

6. J. A. Schellman, *Ann. Rev. Biophys. Biophys. Chem.* **16**, 115-37 (1987).

7. F. M. Hughson, P. E. Wright, R. L. Baldwin, *Science* **249**, 1544-1548 (1990).

8. L. C. Wu, Z.-y. Peng, P. S. Kim, *Nature structural biology* **2**, 281-286 (1995).

9. Y. Bai, T. R. Sosnick, L. Mayne, S. W. Englander, *Science* **269**, 192-197 (1995).

10. K. Yutani, K. Ogasahara, K. Kuwajima, *J Mol Biol* **228**, 347-50 (1992).

11. A. Shimizu, M. Ikeguchi, S. Sugai, *Biochemistry* **32**, 13198-203 (1993).

12. S. N. Loh, M. S. Kay, R. L. Baldwin, *Proc. Natl. Acad. Sci., USA* **92**, 5446-5450 (1995).

13. T. Kiefhaber, R. L. Baldwin, *J. Mol. Biol.* **252**, 122-132 (1995).

14. U. H. Sauer, D. P. San, B. W. Matthews, *J Biol Chem* **267**, 2393-9 (1992).

15. M. Blaber, et al., *J Mol Biol* **235**, 600-24 (1994).

16. P. S. Kim, R. L. Baldwin, *Annu Rev Biochem* **51**, 459-89 (1982).

17. G. D. Rose, *J. Mol. Biol.* **134**, 447-470 (1979).

18. K. A. Dill, *Biochemistry* **29**, 7133-55 (1990).

19. H. Xiong, B. L. Buckwalter, H. M. Shieh, M. H. Hecht, *Proc Natl Acad Sci U S A* **92**, 6349-53 (1995).

20. M. W. West, M. H. Hecht, *Protein Science* **4**, 2032-2039 (1995).

21. J. J. Yang, et al., *J. Mol. Biol.* **252**, 483-491 (1995).

22. M. Buck, H. Schwalbe, C. M. Dobson, *Biochemistry* **34**, 13219-13232 (1995).

23. R. Srinivasan, G. D. Rose, *Proteins: Structure, Function and Genetics* **22**, 81-99 (1995).

24. H. S. Chan, S. Bromberg, K. A. Dill, *Phil. Trans. R. Soc. Lond. B* **348**, 61-70 (1995).

25. Z. Y. Peng, P. S. Kim, *Biochemistry* **33**, 2136-41 (1994).

26. S. D. Hooke, S. E. Radford, C. M. Dobson, *Biochemistry* **33**, 5867-76 (1994).

27. L. A. Morozova, D. T. Haynie, C. Arico-Muendel, H. van Dael, D. M. Dobson, *Nature Structural Biology* **2**, 871-875 (1995).

28. Z.-y. Peng, L. C. Wu, P. S. Kim, *Biochemistry* **34**, 3248-3252 (1995).

29. S. F. Betz, D. P. Raleigh, W. F. DeGrado, *Curr. Op. Struct. Biol.* **3**, 601-610 (1993).

30. K. J. Lumb, P. S. Kim, *Biochemistry* **34**, 8642-8648 (1995).

31. J. J. Ewbank, T. E. Creighton, *Biochemistry* **32**, 3677-93 (1993).

32. J. J. Ewbank, T. E. Creighton, *Biochemistry* **32**, 3694-707 (1993).

33. B. A. Schulman, Z.-y. Peng, C. Redfield, C. M. Dobson, P. S. Kim, *J. Mol. Biol.* **253**, 651-657 (1995).

34. A. S. Siddiqui, G. J. Barton, *Protein Science* **4**, 872-884 (1995).

35. K. Kuwajima, M. Mitani, S. Sugai, *J. Mol. Biol.* **206**, 547-561 (1989).

36. S. E. Radford, C. M. Dobson, P. A. Evans, *Nature* **358**, 302-7 (1992).

37. T. Kiefhaber, *Proc. Natl. Acad. Sci., USA* **92**, 9029-9033 (1995).

38. S. J. Eyles, S. E. Radford, C. V. Robinson, C. M. Dobson, *Biochemistry* **33**, 13038-48 (1994).

APPENDIX I

DISULFIDE DETERMINANTS OF CALCIUM-INDUCED PACKING IN α-LACTALBUMIN

α-Lactalbumin (α-LA) is a two-domain calcium-binding protein that folds through a molten globule intermediate. Calcium binding to the wild-type α-LA molten globule induces a transition to the native state. Here we assess the calcium-binding properties of the α-LA molten globule by studying two disulfide variants. α-LA(α) contains only the two disulfide bonds in the α-helical domain of α-LA, while α-LA(β) contains only the β-sheet domain and interdomain disulfide bonds. We find that only α-LA(β) binds calcium, leading to the cooperative formation of substantial tertiary interactions. In addition, the β-sheet domain acquires a native-like backbone topology. Thus, specific interactions within α-LA imposed by the β-sheet domain and interdomain disulfide bonds, as opposed to the two α-helical domain disulfides, are necessary for the calcium-induced progression from the molten globule towards more native-like structure. Our results suggest that organization of the β-sheet domain, coupled with calcium binding, comprises the locking step in the folding of α-LA from the molten globule to the native state.

Introduction

Native proteins are distinguished from partially folded forms by rigid side chain packing and extensive tertiary interactions. Many proteins fold through the molten globule, a compact but highly dynamic species with a native-like backbone topology (Kuwajima, 1989; Dobson, 1992; Ptitsyn, 1992; Ptitsyn, 1995). Native packing and tertiary interactions are subsequently acquired from the molten globule (Matthews, 1993), but our understanding of the structural and mechanistic bases of these events is poor.

α-Lactalbumin (α-LA)[1] is a widely studied calcium-binding protein composed of two structural domains. Folding of α-LA proceeds in two major steps (Ikeguchi et al., 1986; Kuwajima et al., 1990; Ptitsyn, 1992; Dobson et al., 1994; Peng & Kim, 1994; Wu et al., 1995). Collapse of the polypeptide chain yields a molten globule intermediate, which then acquires rigid side chain packing and tertiary interactions to form the native protein. We previously assessed the structure of the α-LA molten globule by studying the properties of two disulfide variants, α-LA(α) and α-LA(β) (Wu et al., 1995). Our studies indicated that the molten globule of α-LA is a bipartite structure in which the α-helical domain adopts a native-like backbone topology, while the β-sheet domain is largely unstructured (Peng & Kim, 1994; Wu et al., 1995).

Formation of the α-LA molten globule is independent of calcium (Ikeguchi et al., 1986; Kuwajima, 1989). In addition, the equilibrium molten globule of α-LA is studied in the absence of calcium (for reviews, see Kuwajima, 1989; Ptitsyn, 1992). Calcium binding to α-LA appears to induce native structure from the molten globule and affects the rate of folding from the molten globule to the native state (Dolgikh et al., 1981; Hiraoka & Sugai, 1984; Ikeguchi et al., 1986; Kuwajima et al., 1990; Ewbank & Creighton, 1993a; Ewbank & Creighton, 1993b). Furthermore, calorimetric and structural analyses of equine lysozyme, a calcium-binding lysozyme structurally homologous to α-LA, indicates that the protein unfolds in two stages, the first of which is calcium-dependent and results in the loss of specific tertiary interactions and side chain packing (Van Dael et al., 1993; Griko et al., 1995). Here we study the calcium-binding properties of α-LA(α) and α-LA(β), in order to localize the disulfide determinants of native packing and tertiary interactions in α-LA.

Materials and methods

Production of α-LA Variants

α-Lactalbumin variants were produced and purified as described previously from expression plasmids pALA-A2 and pALA-B2 for α-LA(α) and α-LA(β), respectively (Wu et al., 1995).

Sedimentation Equilibrium

Sedimentation equilibrium experiments were performed on a Beckman XL-A analytical ultracentrifuge as described previously (Wu et al., 1995). Our data indicate that slight aggregation of α-LA(α) and α-LA(β) occurs at high concentrations of $CaCl_2$ (~15% aggregation at 1 mM $CaCl_2$). Based on this observation and the results of the calcium titration (see below), a concentration of 200 μM $CaCl_2$ was used for all further studies of α-LA(α) and α-LA(β) in the presence of calcium. Protein solutions were dialyzed overnight against 10 mM Tris, pH 8.5, with either 0.5 mM EDTA or 200 μM $CaCl_2$. Initial protein concentrations of 100, 40 and 15 μM were analyzed at 23 and 27 krpm. The data for α-LA(α) and α-LA(β) in both the presence and absence of calcium fit well (within 8%) to a model for an ideal monomer, with no systematic deviation of the residuals. There is no concentration dependence of the observed molecular weight for either variant.

Circular Dichroism (CD) Spectroscopy

CD spectroscopy was performed with an Aviv 62 DS spectrometer as described previously (Wu et al., 1995). Samples were dissolved in 10 mM Tris, pH 8.5, with either 0.5 mM EDTA or 200 μM $CaCl_2$. The spectra of native α-LA were taken in 10 mM Tris, 1 mM $CaCl_2$, pH 8.5. Thermal denaturation data were collected at 208 nm at 2° intervals, allowing 1.5 minutes equilibration time and 60 seconds data averaging, using samples dissolved to a concentration of 2.5 μM in 10 mM Tris, pH 8.5, with either 0.5 mM EDTA or 200 μM $CaCl_2$. Denaturation curves were smoothed by least squares fitting to a third-order polynomial, using a window of ten data points. All thermal melts are >90% reversible with no hysteresis. Protein concentrations were

determined by absorbance in 6 M GuHCl, 20 mM sodium phosphate, pH 6.5, using an extinction coefficient at 280 nm of 22,430 (Edelhoch, 1967).

Calcium Titration

Changes in the far-UV CD signal of α-LA(α) and α-LA(β) upon addition of $CaCl_2$ were monitored at 208 nm. Samples were dissolved to a concentration of 4 μM in 10 mM Tris, pH 8.5, pre-treated with a chelating agent (Chelex, Bio-Rad). $CaCl_2$ was added in small aliquots from 300 μM, 3 mM, 30 mM, 300 mM, and 3 M stocks, and CD signal was normalized for volume changes. The titration data for α-LA(β) were fit to a model for a single binding site with a non-linear least squares fitting program (Kaleidagraph, Abelbeck Software) to yield the dissociation constant.

Disulfide Exchange

Disulfide exchange studies were performed at 4°C as described previously (Wu et al., 1995). Native buffer consisted of 10 mM Tris, pH 8.5, with either 0.5 mM EDTA or 200 μM $CaCl_2$. Denaturing buffer consisted of 6 M GuHCl, 10 mM Tris, 200 μM $CaCl_2$, pH 8.5.

Nuclear Magnetic Resonance (NMR) Spectroscopy

Protein samples were dissolved in D_2O to a concentration of ~100 μM at pH 8.5 (uncorrected for isotope effects), and either 0.5 mM deuterated EDTA or 200 μM $CaCl_2$. Both α-LA(α) and α-LA(β) are monomers under these conditions, as determined by sedimentation equilibrium. ^1H NMR spectroscopy was preformed at 500.1 MHz on a Bruker AMX spectrometer. 1D spectra were acquired at 4°C using a spectral width of 7812.5 Hz, 4096 complex points, 32,768 transients, and a recycle delay of 1.5 seconds. The residual water peak was suppressed by mild presaturation. Chemical shifts were referenced to 0 ppm with internal (trimethylsilyl)-propionate (TMSP).

Results

α-LA(α) corresponds to human α-LA, with the β-sheet domain and interdomain cysteines replaced by alanines, leaving the native 6-120 and 28-111 disulfide bonds intact (Fig. 1a). Conversely, α-LA(β) corresponds to human α-LA, with the α-helical domain cysteines replaced by alanines, leaving the native 61-77 and 73-91 disulfide bonds intact. In the absence of calcium, α-LA(α) and α-LA(β) are both molten globules with structural properties very similar to the widely studied pH 2 molten globule (A-state) of α-LA (Wu et al., 1995).

Calcium does not affect the structural properties of α-LA(α), as judged by far-UV CD titrations of up to 1 mM calcium (Fig. 1b). Indeed, the far-UV and near-UV CD spectra (Fig. 2a) of α-LA(α) in the presence of calcium are superimposable on those of the calcium-free molten globule of α-LA(α). Moreover, the ^{1}H NMR spectra of α-LA(α) are identical in the presence and absence of calcium, and resemble closely that of the A-state molten globule of α-LA (Fig. 3). Finally, disulfide exchange studies (Peng & Kim, 1994; Wu et al., 1995) under native conditions in both the presence and absence of calcium give identical results (data not shown).

On the other hand, α-LA(β) binds calcium with a K_d of 6.6 ± 0.3 μM (Fig. 1b). For comparison, the K_d of wild-type α-LA is 2-10 nM, depending on solution conditions. (We determined the K_d of calcium binding to α-LA(β), a variant of human α-LA, in 10 mM Tris, pH 8.5, 0 °C. Reported K_d's for wild-type bovine α-LA are 1.7×10^{-9} M, measured in 5 mM Tris, pH 8.04, 20 °C (Permyakov et al., 1981), and 1×10^{-8} M, measured in 60 mM Tris, pH 8.4, 25 °C (Hamano et al., 1986). The K_d of human α-LA is ~20% greater than that of bovine α-LA under identical conditions (Segawa & Sugai, 1983). (Permyakov et al., 1981; Segawa & Sugai, 1983; Hamano et al., 1986; Mitani et al., 1986). The far-UV CD spectrum of calcium-bound α-LA(β) resembles closely that of native α-LA (Fig. 2b). The near-UV CD of calcium-bound α-LA(β) is more intense than that of the α-LA(β) molten globule, indicative of tighter side chain packing, but it is substantially less intense than that of native α-LA (Fig. 2b). These results suggest that, upon addition of calcium, α-LA(β) acquires more native-like structure, but does not become fully native. Thermal

denaturation studies indicate that the formation of this structure is cooperative (Fig. 4).

In the molten globule of α-LA, the β-sheet domain is largely unstructured, lacking a marked preference for native disulfide pairings in disulfide exchange assays, although the α-helical domain has a native-like tertiary fold (Peng & Kim, 1994; Wu et al., 1995). Under denaturing conditions (6 M GuHCl) in the presence of calcium, only 18% of the α-LA(β) molecules assume the native disulfide pairings, in agreement with the behavior predicted for an unfolded polypeptide, using a random-walk model (Kauzman, 1959; Snyder, 1987). Strikingly, addition of calcium to α-LA(β) under native conditions permits 90% of the α-LA(β) molecules to assume the native disulfide pairings (Fig. 5a), in sharp contrast to the molten globule (Fig. 5b) and unfolded (Fig. 5c) forms of α-LA(β).

Although the CD and disulfide exchange studies indicate that calcium binding induces more native structure in the β-sheet domain, the lack of a significant increase in near-UV CD intensity in calcium-bound α-LA(β) suggests that the molecule is still flexible. The NMR spectrum of α-LA(β) in the absence of calcium is broad, lacks chemical shift dispersion, and resembles closely the ^{1}H NMR spectrum of the A-state molten globule of α-LA (Fig. 3). However, the NMR spectrum of calcium-bound α-LA(β) (Fig. 3) contains extensive chemical shift dispersion, including resonances shifted upfield of TMSP, indicative of substantial tertiary interactions. These features suggest that a significant amount of folded structure exists in calcium-bound α-LA(β), albeit less than in native α-LA.

Discussion

Previous studies indicate that in the molten globule of α-LA, the α-helical domain is a dynamic, native-like structure, while the β-sheet domain is largely unstructured (Wu et al., 1995). Calcium induces the transition between the molten globule and the native state of α-LA. We find that calcium binding to α-LA(β) introduces specific structure, while α-LA(α) remains a molten globule in the presence of calcium. Thus, the β-sheet domain (61-77) and interdomain (73-91) disulfide bonds, as opposed to the

α-helical domain disulfides, are crucial for the calcium-induced progression from the α-LA molten globule towards native structure.

The folding of lysozyme, a protein homologous to α-LA, proceeds via a kinetic molten globule intermediate in which the α-helical domain is native-like, while the β-sheet domain remains largely disordered (Radford et al., 1992; Miranker et al., 1993; Dobson et al., 1994, Balbach et al., 1995). Acquisition of near-UV CD signal and enzymatic activity are single kinetic events late in the folding pathway, with rates identical to the rate of folding of the β-sheet domain determined by hydrogen-exchange NMR (Radford et al., 1992; Dobson et al., 1994; Itzhaki et al., 1994).

Thus, in both α-LA and lysozyme, although the α-helical domain folds first, interactions outside of the α-helical domain are important for the acquisition of native packing and tertiary interactions. Our studies, taken together with previous work (Radford et al., 1992; Miranker et al., 1993; Dobson et al., 1994; Peng & Kim, 1994; Wu et al., 1995), suggest the following pathway for the folding of the structurally homologous α-lactalbumins and lysozymes. Initial formation of the molten globule yields a species in which the α-helical domain is native-like, while the β-sheet domain is predominantly unfolded. Subsequently, a locking step requiring organization of the β-sheet domain, and calcium binding in α-LA, yields the unique native structure.

It is interesting that the NMR spectrum of calcium-bound α-LA(β) suggests significant tertiary contacts, while the near-UV (aromatic) CD spectrum lacks much of the intensity of native α-LA. Our studies indicate that the β-sheet domain has a native-like fold in calcium-bound α-LA(β). However, the detailed structure of calcium-bound α-LA(β) remains unclear. Many possibilities are apparent, in which the individual domains of α-LA have varying degrees of native structure.

At one extreme is the possibility that calcium binding to α-LA(β) yields slightly more structure throughout the entire molecule, converting α-LA(β) from a molten globule to a "highly ordered molten globule", a flexible, partially folded species in which the native secondary structures are largely

formed, but loop regions are disordered (Feng et al., 1994; Redfield et al., 1994). At the other extreme is the possibility that calcium binding to α-LA(β) causes the β-sheet domain to fold entirely, while the α-helical domain remains dynamic; the relatively small near-UV CD signal may result from the fact that only one of three tryptophans and one of four tyrosines in α-LA are present in the β-sheet domain. This second possibility is consistent with the previous identification of a stable two-disulfide species of α-LA involving the β-sheet domain and interdomain disulfide bonds, which is transiently populated during reduction of α-LA in the presence of calcium (Ewbank & Creighton, 1993a; Ewbank & Creighton, 1993b). High-resolution structural characterization or protein dissection of α-LA(β) in the presence of calcium should resolve this issue.

Figure legends

Figure 1: Calcium induces structural changes in α-LA(β), but not α-LA(α). (a) Schematic representation of human α-LA (Acharya et al., 1989; Acharya et al., 1991) produced with the program RIBBON (Priestle, 1988). The α-helical domain (white) contains the four α-helices. The β-sheet domain (shaded) contains two short β-strands and several loop structures. A single calcium ion (black ball) binds to the calcium-binding loop, comprised of residues 78-89. α-LA(α) contains the two disulfide bonds in the α-helical domain (6-120 and 28-111; white), with the β-sheet domain and interdomain cysteines replaced by alanines. α-LA(β) contains the disulfide bond in the β-sheet domain (61-77; black) and the interdomain disulfide bond (73-91; black), with the cysteines in the α-helical domain replaced by alanines. (b) Calcium titration of α-LA(α) (open circles) and α-LA(β) (filled circles) at 4°C, pH 8.5, monitored by the CD signal at 208 nm.

Figure 2: Far-UV and near-UV CD spectra of α-LA(α) and α-LA(β) in the presence (filled circles) and absence (open circles) of calcium at 4°C, pH 8.5. Spectra of native α-LA (pluses) are also shown for comparison. (a) α-LA(α). (b) α-LA(β). The spectra of native α-LA and α-LA(α) and α-LA(β) in the absence of calcium have been published previously (Wu et al., 1995).

Figure 3: ^{1}H NMR spectra of α-LA(α) and α-LA(β) in the presence and absence of calcium at 4°C, pH 8.5. For comparison, spectra of the pH 2 A-state and native α-LA (Peng & Kim, 1994) are also shown.

Figure 4: Thermal denaturation studies of α-LA(β) in the presence of calcium indicate that formation of native-like structure is cooperative. CD signal of α-LA(β) in the presence (filled circles) and absence (open circles) of calcium are shown.

Figure 5: Disulfide exchange studies of α-LA(β) at 4°C, pH 8.5. (a) Native conditions with calcium. (b) Native conditions without calcium. (c) Denaturing conditions. The expected equilibrium populations (calculated with a random-walk model) for α-LA(β) under denaturing conditions are given in parentheses in (c). The numbers in brackets denote the disulfide-bonded residues in each species.

Acknowledgments
We thank members of the Kim lab for helpful comments regarding the manuscript.

References

Acharya, K. R., Ren, J. S., Stuart, D. I., Phillips, D. C. & Fenna, R. E. (1991) *J. Mol. Biol. 221*, 571-581.

Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M. & Phillips, D. C. (1989) *J. Mol. Biol. 208*, 99-127.

Balbach, J., Forge, V., van Nuland, N. A. J., Winder, S. L., Hore, P. J. & Dobson, C. M. (1995) *Nature Struct. Biol. 2*, 865-869.

Dobson, C. M. (1992) *Curr. Opin. Struct. Biol. 2*, 6-12.

Dobson, C. M., Evans, P. A. & Radford, S. E. (1994) *Trends Biochem. Sci. 19*, 31-37.

Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y. & Ptitsyn, O. B. (1981) *FEBS Lett. 136*, 311-315.

Edelhoch, H. (1967) *Biochemistry 6*, 1948-1954.

Ewbank, J. J. & Creighton, T. E. (1993a) *Biochemistry 32*, 3677-3693.

Ewbank, J. J. & Creighton, T. E. (1993b) *Biochemistry 32*, 3694-3707.

Feng, Y., Sligar, S. G. & Wand, A. J. (1994) *Nature Struct. Biol. 1*, 30-35.

Griko, Y. V., Freire, E., Privalov, G., Van Dael, H., and Privalov, P. L. (1995) *submitted.*

Hamano, M., Nitta, K., Kuwajima, K. & Sugai, S. (1986) *J. Biochem. (Tokyo) 100*, 1617-1622.

Hiraoka, Y. & Sugai, S. (1984) *Int. J. Peptide Protein Res. 23*, 535-542.

Ikeguchi, M., Kuwajima, K. & Sugai, S. (1986) *J. Biochem. (Tokyo) 99*, 1191-1201.

Itzhaki, L. S., Evans, P. A., Dobson, C. M. & Radford, S. E. (1994) *Biochemistry 33*, 5212-5220.

Kauzman, W. (1959) in *Sulphur in Proteins* (Benesch, R., Benesch, R.E., Boyer, P.D., Klotz, I.M., Middlebrook, W.R., Szent-Gyorgyi, A.G. & Schwarz, D.R., Eds.) pp 93-108, Academic Press, New York.

Kuwajima, K. (1989) *Proteins 6*, 87-103.

Kuwajima, K., Ikeguchi, M., Sugawara, T., Hiraoka, Y. & Sugai, S. (1990) *Biochemistry 29*, 8240-8249.

Matthews, C. R. (1993) *Annu. Rev. Biochem. 62*, 653-683.

Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T. & Dobson, C. M. (1993) *Science 262*, 896-900.

Mitani, M., Harushima, Y., Kuwajima, K., Ikeguchi, M. & Sugai, S. (1986) *J. Biol. Chem. 261*, 8824-8829.

Peng, Z. Y. & Kim, P. S. (1994) *Biochemistry 33*, 2136-2141.

Permyakov, E. A., Yarmolenko, V. V., Kalinichenko, L. P., Morozova, L. A. & Burstein, E. A. (1981) *Biochem. Biophys. Res. Commun. 100*, 191-197.

Priestle, J. P. (1988) *J. Appl. Crystallogr. 21*, 572-576.

Ptitsyn, O. B. (1992) in *Protein Folding* (Creighton, T.E., Eds.) pp 243-300, W. H. Freeman and Co., New York.

Ptitsyn, O. B. (1995) *Curr. Opin. Struct. Biol. 5*, 74-78.

Radford, S. E., Dobson, C. M. & Evans, P. A. (1992) *Nature 358*, 302-307.

Redfield, C., Smith, R. A. G. & Dobson, C. M. (1994) *Nature Struct. Biol. 1*, 23-29.
Segawa, T. & Sugai, S. (1983) *J. Biochem. (Tokyo) 93*, 1321-1328.
Snyder, G. H. (1987) *Biochemistry 26*, 688-694.
Van Dael, H., Haezebrouck, P., Morozova, L., Arico-Muendel, C., and Dobson, C. M. (1993) *Biochemistry 32*, 11886-11894.
Wu, L. C., Peng, Z.-y. & Kim, P. S. (1995) *Nature Struct. Biol.* 2, 281-286.

Fig. 1A

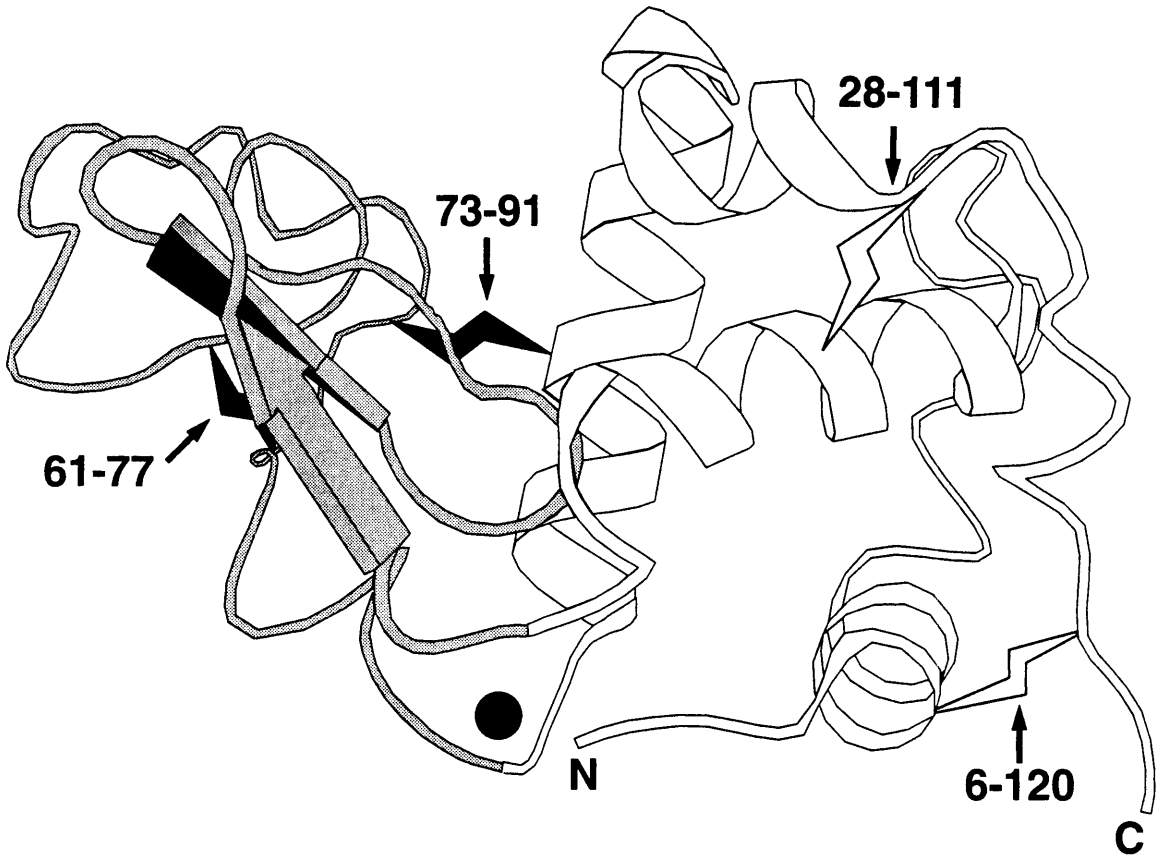


Fig. 1B

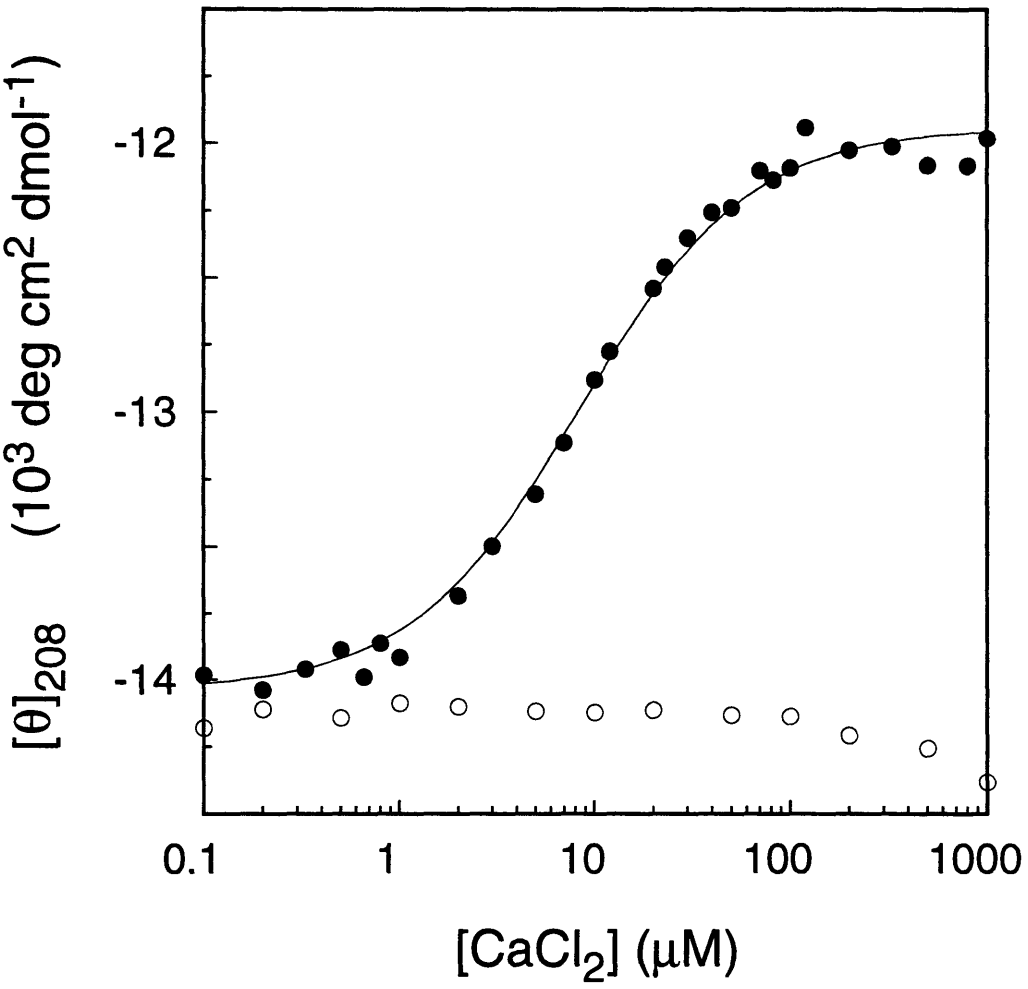
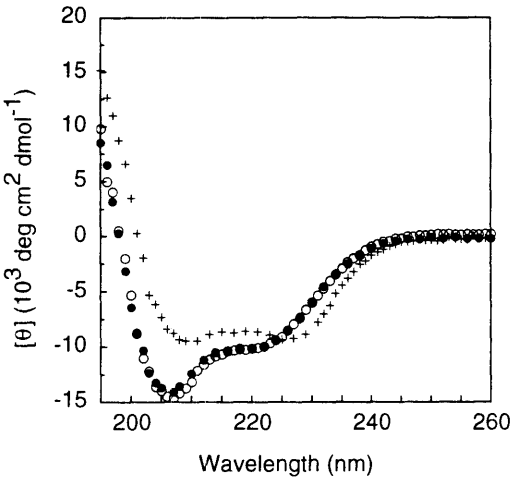
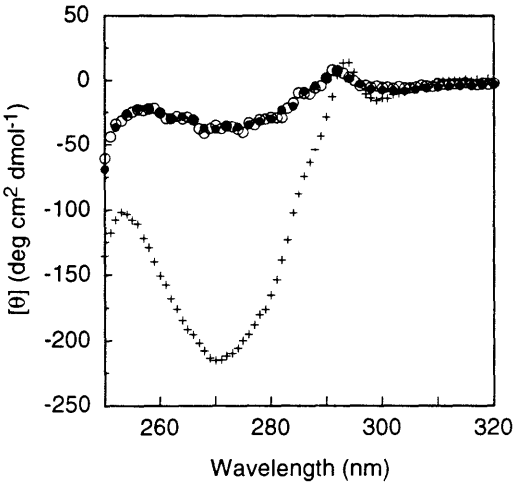


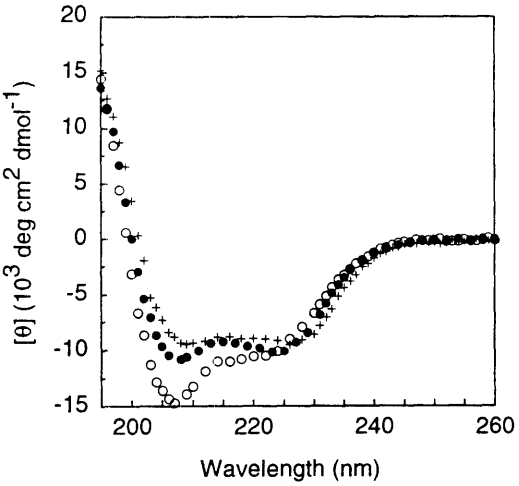
Fig. 2

A) α-LA(α)





B) α-LA(β)



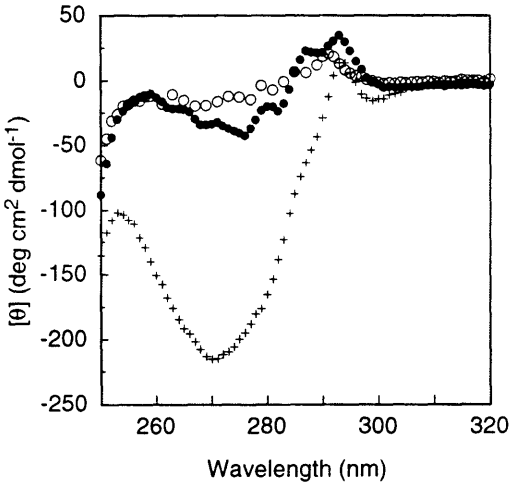
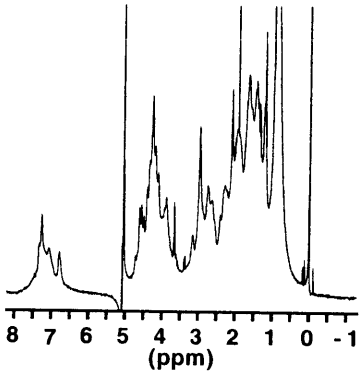
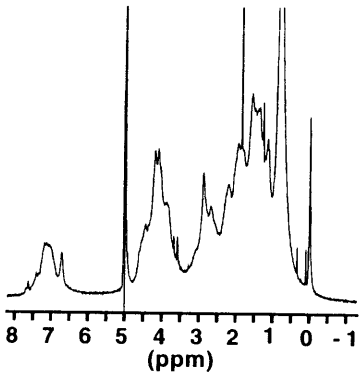


Fig. 3

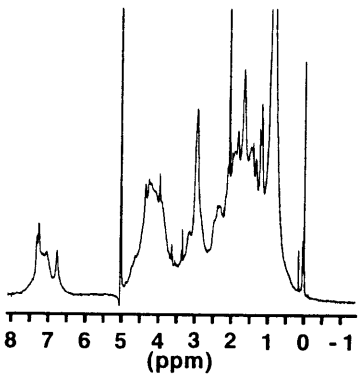
α-LA(α)



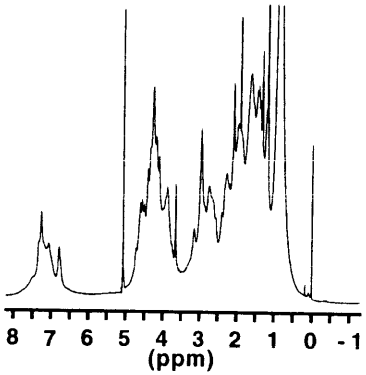
α-LA(β)



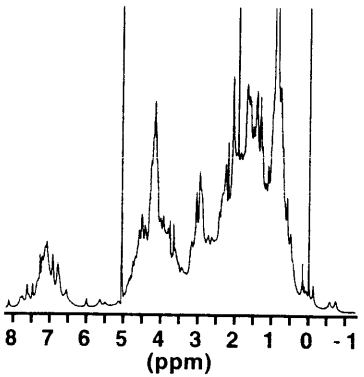
A-state
(molten globule)



α-LA(α)
+ calcium



α-LA(β)
+ calcium



Native

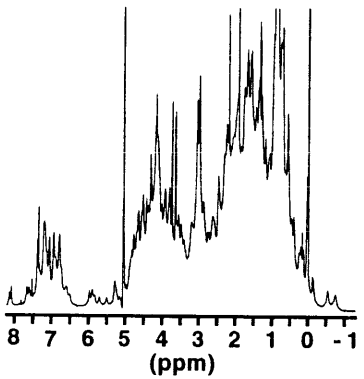


Fig. 4

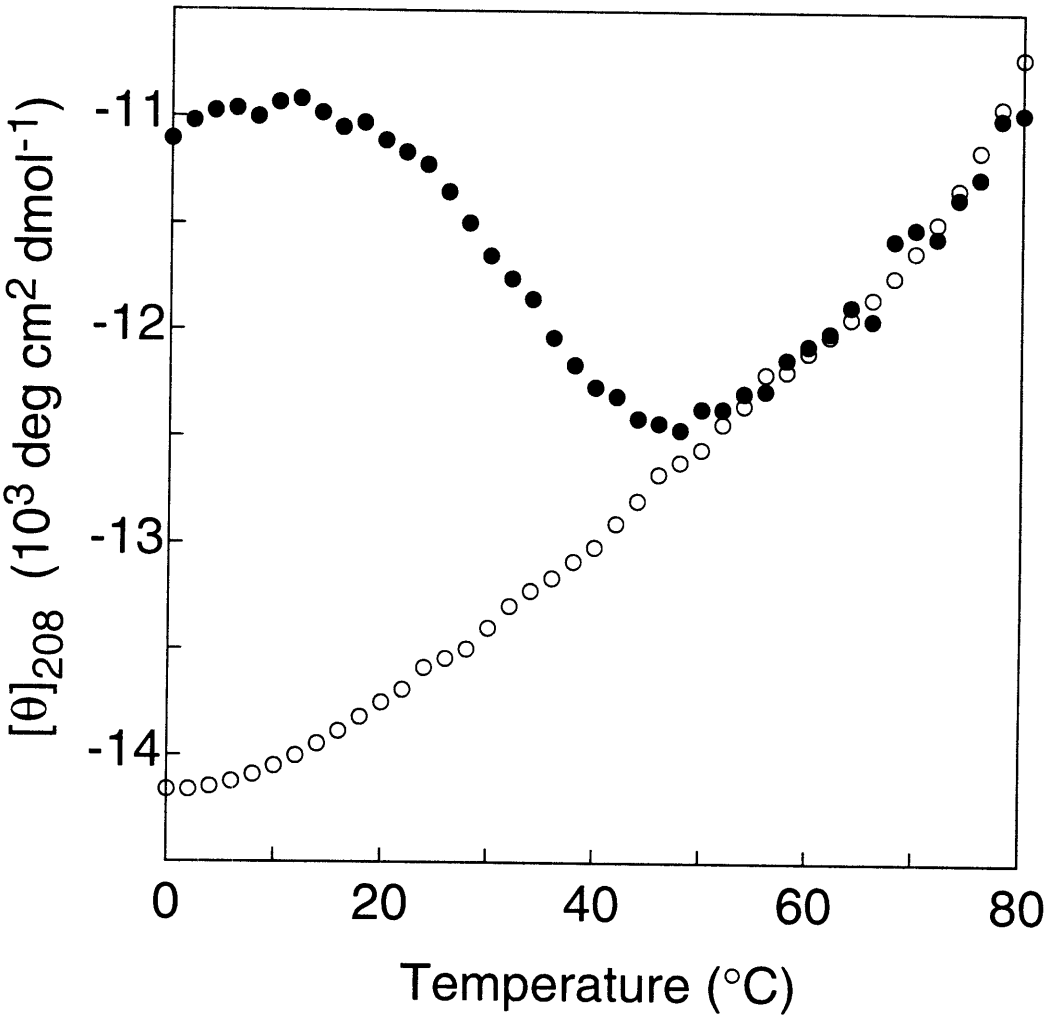
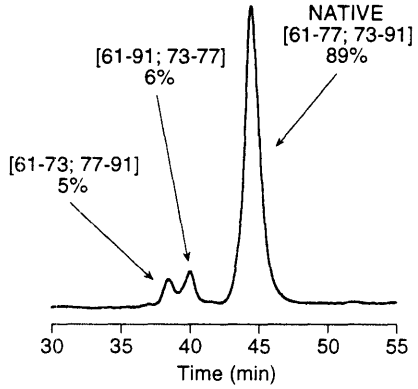

-11
-12
-13
-14
$[\theta]_{208}$ (10^3 deg cm^2 dmol^{-1})
0
20
40
60
80
Temperature (°C)

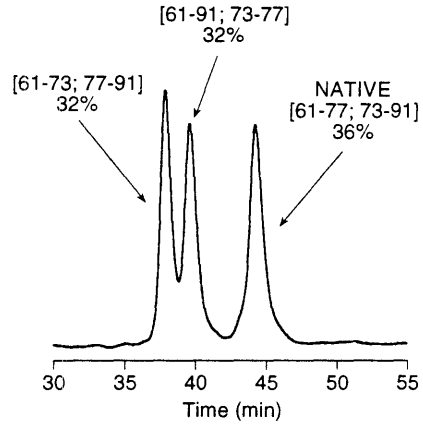
Fig. 5

A)



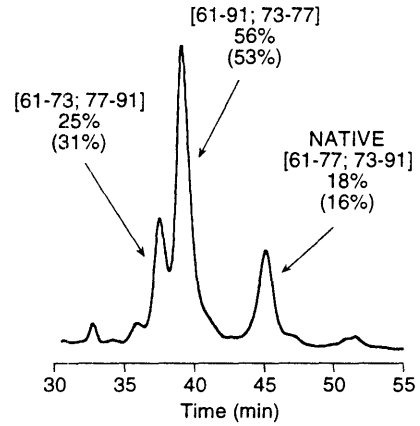
α-LA(β)
+
Calcium

B)



α-LA(β)

C)



Unfolded
α-LA(β)

APPENDIX II

HYDROGEN EXCHANGE IN BPTI VARIANTS THAT DO NOT SHARE A COMMON DISULFIDE BOND

Bovine pancreatic trypsin inhibitor (BPTI) is stabilized by three disulfide bonds, between cysteines 30-51, 5-55, and 14-38. To better understand the influence of disulfide bonds on local protein structure and dynamics, we have measured amide proton exchange rates in two folded variants of BPTI, $[5\text{-}55]_{Ala}$ and $[30\text{-}51; 14\text{-}38]_{V5A55}$, which share no common disulfide bonds. Essentially the same amide hydrogens are protected from exchange in both of the BPTI variants studied here as in native BPTI, demonstrating that the variants adopt fully folded, native-like structures in solution. However, the most highly protected amide protons in each variant differ, and are contained within the sequences of previously studied peptide models of related BPTI folding intermediates containing either the 5-55 or the 30-51 disulfide bond.

Disulfide bonds stabilize substantially the native state of many proteins. This stabilizing effect has been attributed both to the restriction of conformational entropy in the unfolded polypeptide and to enthalpic stabilization of the folded state (for a review, see Betz, 1993). A quantitative understanding of the effects of disulfide bonds on protein stability will therefore require knowledge of the effects of cross-links on the folded state as well as the unfolded state.

BPTI is a very stable 58 residue protein stabilized by three disulfide bonds, between residues 30-51, 5-55, and 14-38 (Fig. 1A). BPTI unfolds spontaneously upon reduction of these disulfide bonds, even in the absence of denaturants. Thus, the folding of BPTI is thermodynamically coupled to disulfide bond formation (Creighton, 1977; Creighton & Goldenberg, 1984; Weissman & Kim, 1991, 1992b).

The folding pathway of BPTI (Fig. 1B) has been studied extensively in terms of the disulfide-bonded intermediates that accumulate during folding (Creighton, 1977; Creighton & Goldenberg, 1984; Weissman & Kim, 1991, 1992b). Two interesting intermediates which accumulate during folding are [5-55] and [30-51; 14-38]. Both of these intermediates have been shown to fold into native-like conformations (Altman, et al., 1991; Weissman & Kim, 1991; van Mierlo, et al., 1991a, 1991b; Staley & Kim, 1992). Thus, no particular disulfide bond is essential for adopting the BPTI fold.

We focus here on the amide proton exchange properties of two completely folded recombinant variants of BPTI: $[5\text{-}55]_{Ala}$, retaining only the 5-55 disulfide bond, with Cys 14, 30, 38 and 51 replaced by Ala (Staley & Kim, 1992); and $[30\text{-}51; 14\text{-}38]_{V5A55}$, containing both the 30-51 and 14-38 disulfide bonds, with Cys 5 replaced by Val and Cys 55 by Ala (Altman, et al., 1991). Our amide proton exchange results confirm that both $[5\text{-}55]_{Ala}$ and $[30\text{-}51; 14\text{-}38]_{V5A55}$ adopt native-like structures in solution and indicate that the arrangement of disulfide bonds affects local structural fluctuations in the folded state of BPTI.

Results

NMR Assignments of [30-51; 14-38]$_{V5A55}$

The high-resolution crystal structure of [30-51; 14-38]$_{V5A55}$ is very similar to that of native BPTI (Kossiakoff, personal communication). The similarity in the chemical shifts of the alpha and amide protons in [30-51; 14-38]$_{V5A55}$ and BPTI (Fig. 2A, B) and a number of unambiguous, readily identifiable, nonsequential NOEs (Fig. 2C) indicate that [30-51; 14-38]$_{V5A55}$ also folds in solution into a conformation very similar to native BPTI.

Aside from resonances from the mutated residues, only five alpha or amide proton chemical shifts differ by greater than 0.2 ppm between [30-51; 14-38]$_{V5A55}$ and BPTI, and all of these protons are close to sites of mutation. These five protons arise from residues that include Tyr 23, Thr 54 and Gly 56, which are within 6Å of the 5-55 disulfide bond in native BPTI (Wlodawer, et al., 1984). The other two residues with large changes in chemical shift are Gly 57 and Ala 58; these differences may result from the addition of a methionine residue to the N-terminus of [30-51; 14-38]$_{V5A55}$, potentially interfering with a hydrogen-bonded salt bridge between the α-carboxylate of Ala 58 and the free amino terminus of Arg 1 in native BPTI (Brown, et al., 1978).

Amide Proton Exchange

The rate constants of exchange at pH 4.6 for amide protons in [5-55]$_{Ala}$ and [30-51; 14-38]$_{V5A55}$ were measured at 5° and 20°C, and additionally at 30°C for [30-51; 14-38]$_{V5A55}$ (Fig. 3; Table 1). Both [5-55]$_{Ala}$ and [30-51; 14-38]$_{V5A55}$ are monomeric in exchange buffer at 5°C as determined by equilibrium sedimentation (see Materials and Methods). The patterns of protection from amide proton exchange (Fig. 4) in [5-55]$_{Ala}$ and [30-51; 14-38]$_{V5A55}$ are very similar to each other and to BPTI (Wagner & Wüthrich, 1982a; Wagner, et al., 1984). All of the amide hydrogens involved in regular secondary structure in BPTI are significantly protected in both disulfide variants.

Although almost all of the amide protons protected from exchange are the same in $[5\text{-}55]_{Ala}$ and $[30\text{-}51; 14\text{-}38]_{V5A55}$, some protons are protected to different extents in the two molecules. The most striking overall difference is that most of the protected amide hydrogens in $[5\text{-}55]_{Ala}$ are protected to a similar degree, whereas there is much greater variation in the protection factors for amide hydrogens in native BPTI (Wagner & Wüthrich, 1982a; Wagner, et al., 1984) and $[30\text{-}51; 14\text{-}38]_{V5A55}$ (Fig. 4).

In both $[5\text{-}55]_{Ala}$ and $[30\text{-}51; 14\text{-}38]_{V5A55}$, all of the amide protons are more protected at 5°C than at 20°C. This difference is greater in $[5\text{-}55]_{Ala}$, which is slightly less stable than $[30\text{-}51; 14\text{-}38]_{V5A55}$; the Tm values are 42°C and 60°C, respectively (Staley & Kim, 1992 and data not shown). The extent of this increase in protection, however, is clearly different for different residues. In $[5\text{-}55]_{Ala}$, there is no obvious correlation between the temperature dependence of protection from amide proton exchange and structural features of the protein. In $[30\text{-}51; 14\text{-}38]_{V5A55}$, on the other hand, the amide hydrogens from the N-terminal 3_{10}-helix and from residues 54 and 55 in the C-terminal α-helix display a smaller change in protection factor with temperature than other residues involved in secondary structure. This difference is even more pronounced between 20° and 30°C (Fig. 4).

Stable formation of native-like structures

In both $[5\text{-}55]_{Ala}$ and $[30\text{-}51; 14\text{-}38]_{V5A55}$, 25 of the 29 backbone amide hydrogens involved in H-bonds in the crystal structures of native BPTI (Deisenhofer & Steigemann, 1975; Wlodawer, et al., 1984, 1987) are protected from exchange by greater than 100-fold at 5°C (Table 1). Of the remaining four H-bonded residues, three (Phe 4, Lys 26 and Asp 50) are also not protected significantly in native BPTI (Wagner, et al., 1984). The additional H-bonded amide hydrogens that are not protected substantially from exchange are Val 5 in the case of $[30\text{-}51; 14\text{-}38]_{V5A55}$ and Ala 38 in $[5\text{-}55]_{Ala}$. In both cases these residues are sites of mutation and the amide protons are protected slightly (over 30-fold at 5°C).

Ten additional amide protons are highly protected in both $[5\text{-}55]_{Ala}$ and $[30\text{-}51; 14\text{-}38]_{V5A55}$. All of these protons are either considerably buried in the

crystal structure of native BPTI (Deisenhofer & Steigemann, 1975; Wlodawer, et al., 1984, 1987) or are protected in native BPTI despite exposure to solvent in the crystal structures (Wagner & Wüthrich, 1982a; Wagner, et al., 1984; Tüchsen & Woodward, 1985). The amide proton of Ser 47 is significantly protected from exchange in both variants described here, and is among the most protected amide protons in a partially-folded variant of BPTI retaining only the 30-51 disulfide bond (Staley, 1993). Although Ser 47 has not been shown previously to be protected from exchange in native BPTI, it is buried in the crystal structure (Chothia & Janin, 1975; Wagner & Wüthrich, 1982a). Additionally, in $[5\text{-}55]_{Ala}$, Ala 30 is also protected more than 100-fold at 5°C.

X-ray crystallographic studies have shown that $[30\text{-}51; 14\text{-}38]_{V5A55}$ has the same global fold as native BPTI (Kossiakoff, personal communication). Similarly, a previous ^{1}H NMR study showed that $[5\text{-}55]_{Ala}$ retains native structure (Staley & Kim, 1992). Our results show that stable, native-like folding occurs in both variants in solution.

Discussion

Comparison of Amide Proton Exchange Properties

One remarkable feature of amide proton exchange in $[5\text{-}55]_{Ala}$, as compared to native BPTI and $[30\text{-}51; 14\text{-}38]_{V5A55}$, is that all of the amide hydrogens involved in H-bonds are protected to a similar extent, suggesting that exchange occurs predominantly via a route that involves global unfolding. It is particularly striking that at 5°C, the N-terminal 3_{10}-helix and residues 54-56 of the C-terminal α-helix are more protected from exchange in $[5\text{-}55]_{Ala}$ than in $[30\text{-}51; 14\text{-}38]_{V5A55}$, even though $[5\text{-}55]_{Ala}$ is less stable than $[30\text{-}51; 14\text{-}38]_{V5A55}$. The differences in the extent of protection from exchange can be explained by differences in the subdomains that are likely to be populated in $[30\text{-}51; 14\text{-}38]_{V5A55}$ and $[5\text{-}55]_{Ala}$ (see also Clarke, et al., 1993; Kim, et al., 1993; Woodward, 1993).

BPTI contains a subdomain comprising the anti-parallel β-sheet and C-terminal α-helix (Oas & Kim, 1988). Peptides corresponding to this subdomain, termed PαPβ, are folded in the presence of the 30-51 disulfide

bond (Oas & Kim, 1988), and this subdomain is present in partially folded variants of BPTI containing only the 30-51 disulfide bond (Staley, 1993; van Mierlo, et al., 1993; Staley & Kim, 1994). The N-terminal 3_{10}-helix is not part of the PαPβ subdomain. In $[30\text{-}51; 14\text{-}38]_{V5A55}$, the high degree of protection in part of the α-helix and the anti-parallel β-sheet, taken together with the more local unfolding behavior of the 3_{10}-helix (Fig. 4B), suggests that the PαPβ region is a stable subdomain in $[30\text{-}51; 14\text{-}38]_{V5A55}$.

The 5-55 disulfide bond connects the N-terminal 3_{10}-helix to the C-terminal α-helix. Peptide models containing these secondary structures, together with the 5-55 disulfide bond, fail to fold without inclusion of a substantial portion of the central β-sheet of BPTI (Staley & Kim, 1990). Thus, while a peptide model, termed PαPγ, of a subdomain of BPTI containing the 5-55 disulfide can be made, at least part of each unit of BPTI secondary structure is required for folding (Staley & Kim, 1990). This requirement might explain the increased contribution of global unfolding to the exchange properties of $[5\text{-}55]_{Ala}$.

Another difference between the patterns of protection in $[5\text{-}55]_{Ala}$ and $[30\text{-}51; 14\text{-}38]_{V5A55}$ is the rate of exchange of residue 30 in the β-sheet. In native BPTI and $[30\text{-}51; 14\text{-}38]_{V5A55}$, Cys 30 is disulfide-bonded to Cys 51. In $[5\text{-}55]_{Ala}$, Cys 30 is mutated to alanine. The amide proton of Cys 30 exchanges rapidly in native BPTI (Wagner, et al., 1984), $[30\text{-}51; 14\text{-}38]_{V5A55}$ (Fig. 3B, 4B), and a modified form of BPTI in which cysteines 14 and 38 have been blocked by iodoactetamide (Wagner, et al., 1984). In $[5\text{-}55]_{Ala}$, however, the amide proton of Ala 30 is highly protected from exchange (Fig. 3A, 4A). This difference might arise from a change in dynamics near residue 30 in $[5\text{-}55]_{Ala}$, or from a conformational change upon removal of the 30-51 disulfide bond. Both solution and x-ray structures have been solved for a mutant BPTI in which both Cys 30 and Cys 51 have been mutated to alanines (Eigenbrot, et al., 1990, 1992; Hurle, et al., 1992). Although the mutation results in slight changes in conformation at the interface between the α-helix and β-sheet (Eigenbrot, et al., 1990), there does not appear to be an obvious structural explanation for the dramatic increase in protection from amide proton exchange for Ala 30 in $[5\text{-}55]_{Ala}$.

Implications for the folding pathway of BPTI

A key feature of the oxidative folding pathway of BPTI (Creighton, 1977; Creighton & Goldenberg, 1984; Weissman & Kim, 1991, 1992b) is that an initially formed two-disulfide intermediate, [30-51; 14-38], cannot oxidize directly to native BPTI, but instead rearranges to another two-disulfide intermediate before forming a third disulfide bond (Creighton, 1977; Creighton & Goldenberg, 1984; Weissman & Kim, 1991). It has been proposed (Weissman & Kim, 1991, 1992a) that premature formation of native structure during the folding of BPTI buries the free thiols of Cys 5 and Cys 55, rendering them inaccessible to external oxidizing agents such as glutathione, thereby inhibiting formation of a third native disulfide bond in [30-51; 14-38] (see also Goto & Hamaguchi, 1981). The amide proton exchange properties of $[30\text{-}51; 14\text{-}38]_{V5A55}$ suggest that native structure, including that around residues 5 and 55, inhibits amide proton exchange much like it might inhibit disulfide chemistry.

Similarly, the one-disulfide folding intermediate, [5-55], does not readily form the 30-51 disulfide bond. Again, this behavior can be explained (Staley & Kim, 1990, 1992; van Mierlo, et al., 1991b) by native structure burying Cys 30 and Cys 51, inhibiting access of external oxidizing agents to these thiols. A previous NMR study of $[5\text{-}55]_{Ala}$ (Staley & Kim, 1992), and a related study of $[5\text{-}55]_{Ser}$ in which the cysteines other than 5 and 55 were replaced with serines (van Mierlo, et al., 1991b), demonstrated that [5-55] folds into essentially the same conformation as native BPTI. The pattern of protection from amide proton exchange in $[5\text{-}55]_{Ala}$ (Fig. 4A) strengthens this conclusion.

In summary, our results demonstrate that variants of BPTI containing different disulfide bonds show local differences in the extent of protection from amide proton exchange, indicating that there are substantial local differences in the dynamics of the folded structure. Nonetheless, the high degree of protection of the same residues in $[5\text{-}55]_{Ala}$, $[30\text{-}51; 14\text{-}38]_{V5A55}$ and native BPTI underscores the conclusion that none of the disulfide bonds is absolutely required for BPTI to adopt its native fold in solution.

Figure Legends

Fig. 1: **(a)** Schematic representation (adapted from Creighton, 1975) of the crystal structure of BPTI (Deisenhofer & Steigemann, 1975; Wlodawer, et al., 1984; Wlodawer, et al., 1987), with labels for the cysteine residues involved in disulfide bonds and the first and last residues. The disulfide bond 14-38 is accessible to solvent, exposing 50% of its total surface area, whereas the disulfide bonds 30-51 and 5-55 are inaccessible, exposing 0% of their total surface area (Lee & Richards, 1971). **(b)** Schematic diagram of the folding pathway of BPTI at 25°C, pH 7.3 (Weissman & Kim, 1991; Weissman & Kim, 1992b). Intermediates are designated by the disulfide bonds that they contain. R refers to reduced BPTI; N, to native BPTI; N_{SH}^{SH}, to the precursor to native BPTI; and N*, to a kinetically trapped intermediate. The folding of N* can be accelerated ~6000-fold by protein disulfide isomerase (Weissman & Kim, 1993). The dashed arrows indicate that the major one-disulfide intermediates do not form directly from reduced protein but rather from rearrangement of other one-disulfide intermediates (Creighton, 1977). The relative rates of intramolecular transitions at pH 7.3 are indicated; "very fast" rates are on the order of milliseconds, while "very slow" rates are on the order of months (Weissman & Kim, 1992c). N_{SH}^{SH}, N*, [30-51; 14-38] and [5-55] fold into essentially the same conformation as native BPTI (Stassinopoulou, et al., 1984; States, et al., 1984; States, et al., 1987; Eigenbrot, et al., 1990; Altman, et al., 1991; van Mierlo, et al., 1991a; van Mierlo, et al., 1991b; Eigenbrot, et al., 1992; Hurle, et al., 1992; Staley & Kim, 1992; Kossiakoff, personal communication).

Fig. 2: Differences in chemical shift ($\Delta\delta$) between native BPTI (Staley, 1993) and $[30\text{-}51; 14\text{-}38]_{V5A55}$ at pH 4.6, 20°C for the **(a)** amide and **(b)** alpha protons, plotted versus amino acid residue number. A positive $\Delta\delta$ value represents an upfield shifted proton in $[30\text{-}51; 14\text{-}38]_{V5A55}$ as compared to native BPTI. **(c)** Unambiguous nonsequential NOEs arise from pairs of protons distributed throughout $[30\text{-}51; 14\text{-}38]_{V5A55}$ and are consistent with the formation of native secondary and tertiary structure. Some of these proton pairs are displayed on a schematic representation (Richardson, 1985; Staley & Kim, 1992) for native BPTI.

Fig. 3: The decay in intensity of the ^{15}N-^{1}H HSQC correlation for the amide protons of residues 17 (squares), 29 (circles) and 30 (triangles) in **(a)** $[5\text{-}55]_{Ala}$ at 5°C, pH* 4.6 and **(b)** $[30\text{-}51; 14\text{-}38]_{V5A55}$ at 20°C, pH* 4.6.

Fig. 4: Relative amide proton exchange rates plotted as log (k_{ex}/k_{int}) versus residue number at pH* 4.6 in **(a)** $[5\text{-}55]_{Ala}$ at 5°C (triangles) and 20°C (circles) and **(b)** $[30\text{-}51; 14\text{-}38]_{V5A55}$ at 5°C (triangles), 20°C (circles) and 30°C (diamonds), and at pH* 4.5 in **(c)** BPTI at 36°C (circles) and 56°C (triangles). The values for

BPTI are taken from Wagner, et al., 1984. Exchange rates that are too fast to measure are shown, in filled symbols, at the minimum rate of exchange for these residues.

Materials and Methods

Protein Expression and Purification

The gene encoding [30-51; 14-38]$_{V5A55}$ was produced by oligonucleotide-directed mutagenesis (Kunkel, et al., 1987) of a gene encoding BPTI (D. Nguyen, J. P. Staley and P. S. Kim, unpublished), which was synthesized with convenient restriction sites and optimal codon usage for *E. coli* and ligated into the NdeI/BamHI site of pAED4 , a pUC-based T7 expression plasmid with an F1 origin (Studier, et al., 1990; Doering, 1992). The resulting plasmid is called pV5A55BPTI. The plasmid encoding [5-55]$_{Ala}$, denoted p5-55, is described elsewhere (M. -H. Yu, J. S. Weissman and P. S. Kim, submitted). Unlabeled [5-55]$_{Ala}$ and [30-51; 14-38]$_{V5A55}$ were expressed essentially as described previously for [5-55]$_{Ala}$ (Staley & Kim, 1992) except that rifampicin was omitted. ^{15}N-labeled proteins were expressed in minimal M9 media, containing one gram/liter ^{15}N-ammonium sulfate (99.7% ^{15}N; Isotec, Ohio) and 0.5 mg/liter thiamine (McIntosh & Dahlquist, 1990). Cells were induced at an OD_{600} of ~0.8 and harvested three hours after induction. Both variants of BPTI form inclusion bodies upon expression in *E. coli*. Inclusion bodies containing [30-51; 14-38]$_{V5A55}$ were recovered and solubilized as for [5-55]$_{Ala}$ (Staley & Kim, 1992). Reduced [30-51; 14-38]$_{V5A55}$ was then purified by reversed-phase HPLC (Weissman & Kim, 1991) and the pure, reduced protein was refolded at a concentration of 10 μM in degassed folding buffer (100 mM Tris, pH 8.7, 1 mM EDTA, 5 mM reduced glutathione, 0.5 mM oxidized glutathione) containing 15% glycerol (Cleland, 1991) at room temperature and was subsequently repurified by reversed-phase HPLC. [5-55]$_{Ala}$ was purified and refolded as described previously (Staley & Kim, 1992). The identity of each variant was confirmed by laser desorption mass spectrometry on a Finnegan Lasermat and found to be within 2 daltons of the calculated mass.

NMR Spectroscopy and Resonance Assignments

Spectra were collected at protein concentrations of ~4mM (pH 4.6) in the absence of buffer or salt. ^{1}H chemical shifts were referenced to trimethylsilylpropionic acid (DeMarco, 1977), and the ^{15}N shifts referenced indirectly to NH_4^+ (Levy & Lichter, 1979). Data were collected on a Bruker AMX 500-MHz spectrometer. Water was presaturated for 1 second and 1024 data points were collected in the t2 dimension. 150-256 increments were used in the t1 dimension. Data from 2D heteronuclear nuclear Overhauser spectroscopy (HSMQC-NOESY) experiments were collected with a mixing

time of 150 ms; data from heteronuclear 2D total correlation spectroscopy (HSQC-TOCSY) experiments were collected with mixing times of 80 and 110 msec (Gronenborn, et al., 1989; McIntosh, et al., 1990).

Strong similarity with spectra of BPTI (Wagner & Wüthrich, 1982b; Tüchsen & Woodward, 1987; Wagner, et al., 1987; van Mierlo, et al., 1991a) facilitated the assignment of resonances in spectra of $[30\text{-}51; 14\text{-}38]_{V5A55}$. At least one unambiguous sequential NOE (Wüthrich, 1986) was observed for each residue.

Equilibrium Sedimentation

Measurements were made in Beckman 6-sector, 12 mm pathlength epoxy cells in a Beckman XL-A 90 analytical ultracentrifuge. Samples were dialyzed against 20 mM acetic acid, pH 4.6. Data were collected at 5°C at three wavelengths (249, 261 and 272 nm), rotor speeds (37, 41 and 45 krpm), and initial concentrations (50, 100 and 200 μM) as determined by tyrosine and cystine absorbance (Edelhoch, 1967). The final concentrations varied continuously between 25μM and 400μM. The program HID4000 (Johnson, et al., 1981) was used to calculate molecular weights by simultaneously fitting fifteen data sets to a single molecular weight, fifteen intercepts, fourteen offsets and a single second virial coefficient. The residuals showed no systematic variation. Nonideality was observed as a decrease in the apparent molecular weight as a function of increasing protein concentration, requiring the use of the second virial coefficient as a fitting parameter. This nonideality most likely results from the large ratio of net charge to molecular weight (Williams, et al., 1958) at the low ionic strength of the buffer used for NMR. The density of solutions and partial specific volumes of protein species were calculated using values from (Laue, et al., 1992). An analysis of $[5\text{-}55]_{Ala}$ yields a molecular weight of 6246 Da ± 5% (95% confidence; calculated, 6519 Da) and $[30\text{-}51; 14\text{-}38]_{V5A55}$ yields a molecular weight of 6571 Da ± 4% (95% confidence; calculated 6610 Da). A similar analysis of native BPTI yielded comparable results (Staley & Kim, 1994).

Amide Proton Exchange

Fully protonated samples of $[30\text{-}51; 14\text{-}38]_{V5A55}$ or $[5\text{-}55]_{Ala}$ were adjusted to pH 4.6 and lyophilized prior to initiating H-D exchange by dissolving the samples in exchange buffer (20 mM $NaCD_3COO$, pH* 4.6). (pH* refers here to meter readings in D_2O solutions using a glass pH electrode, without correction for isotope effects.) The temperature was maintained in circulating water baths between later time points. The protein concentration was ~3 mM.

Rates of H-D exchange were measured by recording ^{15}N-^{1}H HSQC spectra with 2-4 transients per increment, 2048 real data points, an ω2 spectral width of 6250 Hz, a total recycle delay of 1 second, and 128 increments of t1 (Norwood, et al., 1990). ^{15}N decoupling during the acquisition was achieved with WALTZ-16 (Shaka, et al., 1983). The residual HOD peak was suppressed with low power presaturation during the recycle delay.

Data Analysis

Amide proton decays were followed by measuring peak volumes in ^{15}N-^{1}H HSQC spectra. The volumes were normalized to peak areas of the non-exchangeable resonances of the Y21 alpha or Y23 epsilon protons in 1D ^{1}H spectra collected immediately prior to each HSQC spectrum except the first six. One 1D spectrum was collected immediately following the first six experiments. The amide proton exchange rates, k_{ex}, were determined using the curve-fitting routine in Kaleidograph (Abelbeck software) by fitting the data to the three parameter fit, $I(t) = I(\infty) + I(0) \exp(-k_{ex} * t)$, where I(t) is the intensity at time (t) after addition of deuterated solvent to the protein. Values for $I(\infty)$ were 60 to 200-fold lower than I(0).

Predicted constants for the intrinsic rate of base-catalyzed exchange are based on data for H-D exchange in poly-D,L-alanine (Englander, et al., 1979) with corrections for the effects of nearest neighbor side chains on intrinsic exchange rates (Molday, et al., 1972; Robertson & Baldwin, 1991). The second-order rate constant for base-catalyzed exchange was modified to account for the isotope effect on the ionization of D_2O (Covington, et al., 1966; Roder, et al., 1985; Robertson & Baldwin, 1991) . For calculation of intrinsic rates of exchange at temperatures other than 20°C, the activation enthalpy was taken to be 2.8 kcal/mol. The pK_{D_2O} was calculated at different temperatures using the relationship (Covington, et al., 1966) $pK_{D_2O} = 4913(T)^{-1} - 7.60 + 0.02009(T)$.

Acknowledgments

We thank A.A. Kossiakoff for sharing with us the X-ray crystal structure coordinates of $[30\text{-}51; 14\text{-}38]_{V5A55}$ prior to publication, L.P. McIntosh, C.J. McKnight and Z.-Y. Peng for help with NMR spectroscopy and P.B. Harbury, K.J. Lumb, D.L. Minor, J.P. Staley and J.S. Weissman for helpful discussions. B.A.S. was supported by the National Science Foundation predoctoral fellowship program and by a National Institutes of Health training grant (T32AI07348). This research was supported by a grant from the National Institutes of Health (GM41307).

References

Altman JD, Henner D, Nilsson B, Anderson S, Kuntz ID. 1991. Intracellular expression of BPTI fusion proteins and single column cleavage/affinity purification by chymotrypsin. Protein Eng 4: 593-600.

Betz SF. 1993. Disulfide bonds and the stability of globular proteins. Protein Sci 2: 1551-8.

Brown LR, De Marco A, Richarz R, Wagner G, Wüthrich K. 1978. The influence of a single salt bridge on static and dynamic features of the globular solution conformation of the basic pancreatic trypsin inhibitor. 1H and 13C nuclear-magnetic-resonance studies of the native and the transaminated inhibitor. Eur J Biochem 88: 87-95.

Chothia C, Janin J. 1975. Principles of protein-protein recognition. Nature 256: 705-8.

Clarke J, Hounslow AM, Bycroft M, Fersht AR. 1993. Local breathing and global unfolding in hydrogen exchange of barnase and its relationship to protein folding pathways. Proc Natl Acad Sci U S A 90: 9837-41.

Cleland JL. 1991. Mechanisms of Protein Aggregation and Refolding. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.

Covington AK, Robinson RA, Bates RG. 1966. The Ionization Constant of Deuterium Oxide from 5 to 50°. J. Phys. Chem. 12: 3820-3824.

Creighton TE. 1975. The two-disulphide intermediates and the folding pathway of reduced pancreatic trypsin inhibitor. J Mol Biol 95: 167-199.

Creighton TE. 1977. Conformational restrictions on the pathway of folding and unfolding of the pancreatic trypsin inhibitor. J Mol Biol 113: 275-93.

Creighton TE, Goldenberg DP. 1984. Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. J Mol Biol 179: 497-526.

Deisenhofer J, Steigemann W. 1975. Crystallographic Refinement of the Structure of Bovine Pancreatic Trypsin Inhibitor at 1.5 Å Resolution. Acta Cryst B31: 238-250.

DeMarco A. 1977. pH Dependence of Internal References. J Magn Reson 26: 527-528.

Doering D. 1992. Functional and structural studies of a small f-actin binding domain. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.

Edelhoch H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 6: 1948-54.

Eigenbrot C, Randal M, Kossiakoff AA. 1990. Structural effects induced by removal of a disulfide-bridge: the X-ray structure of the C30A/C51A mutant of basic pancreatic trypsin inhibitor at 1.6 A. Protein Eng 3: 591-8.

Eigenbrot C, Randal M, Kossiakoff AA. 1992. Structural effects induced by mutagenesis affected by crystal packing factors: the structure of a 30-51 disulfide mutant of basic pancreatic trypsin inhibitor. Proteins 14: 75-87.

Englander JJ, Calhoun DB, Englander SW. 1979. Measurement and Calibration of Peptide Group Hydrogen-Deuterium Exchange by Ultraviolet Spectroscopy. Analytical Biochemistry 92: 517-524.

Goto Y, Hamaguchi K. 1981. Formation of the intrachain disulfide bond in the constant fragment of the immunoglobulin light chain. J Mol Biol 146: 321-40.

Gronenborn AM, Bax A, Wingfield PT, Clore GM. 1989. A powerful method of sequential proton resonance assignment in proteins using relayed 15N-1H multiple quantum coherence spectroscopy. Febs Lett 243: 93-8.

Hurle MR, Eads CD, Pearlman DA, Seibel GL, Thomason J, Kosen PA, Kollman P, Anderson S, Kuntz ID. 1992. Comparison of solution structures of mutant bovine pancreatic trypsin inhibitor proteins using two-dimensional nuclear magnetic resonance. Protein Sci 1: 91-106.

Johnson ML, Correia JJ, Yphantis DA, Halvorson HR. 1981. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. Biophys J 36: 575-88.

Kim KS, Fuchs JA, Woodward CK. 1993. Hydrogen exchange identifies native-state motional domains important in protein folding. Biochemistry 32: 9600-8.

Kunkel TA, Roberts JD, Zakour RA. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol 154: 367-82.

Laue TM, Shah BD, Ridgeway TM, Pelletier SL. 1992. Computer aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton JC, ed. Analytical Ultracentrifugation in biochemistry and polymer science. Cambridge, UK: Royal Society of Chemistry, pp. 90-125.

Lee B, Richards FM. 1971. The interpretation of protein structures: estimation of static accessibilty. J Mol Biol 55: 379-400.

Levy GC, Lichter RL. 1979. Nitrogen-15 Nuclear Magnetic Resonance Spectroscopy. New York: John Wiley & Sons.

McIntosh LP, Dahlquist FW. 1990. Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. Q Rev Biophys 23: 1-38.

McIntosh LP, Wand AJ, Lowry DF, Redfield AG, Dahlquist FW. 1990. Assignment of the backbone 1H and 15N NMR resonances of bacteriophage T4 lysozyme. Biochemistry 29: 6341-62.

Molday RS, Englander SW, Kallen RG. 1972. Primary Structure Effects on Peptide Group Hydrogen Exchange. Biochemistry 11: 150-158.

Norwood TJ, Boyd J, Heritage JE, Soffe N, Campbell ID. 1990. Comparison of techniques for 1H-detected heteronuclear 1H-15N spectroscopy. J. Magn. Reson. 87: 488-501.

Oas TG, Kim PS. 1988. A peptide model of a protein folding intermediate. Nature 336: 42-8.

Richardson JS. 1985. Schematic drawings of protein structures. Methods Enzymol 115: 359-80.

Robertson AD, Baldwin RL. 1991. Hydrogen exchange in thermally denatured ribonuclease A. Biochemistry 30: 9907-14.

Roder H, Wagner G, Wüthrich K. 1985. Individual amide proton exchange rates in thermally unfolded basic pancreatic trypsin inhibitor. Biochemistry 24: 7407- 11.

Shaka AJ, Keeler J, Freeman R. 1983. Evaluation of a New Broadband Decoupling Sequence: WALTZ-16. J Magn Reson 53: 313-340.

Staley JP. 1993. Structural studies of early intermediates in the folding pathway of bovine pancreatic trypsin inhibitor. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.

Staley JP, Kim PS. 1990. Role of a subdomain in the folding of bovine pancreatic trypsin inhibitor. Nature 344: 685-8.

Staley JP, Kim PS. 1992. Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond. Proc Natl Acad Sci U S A 89: 1519-23.

Staley JP, Kim PS. 1994. Formation of a Native-Like Subdomain in a Partially-Folded Intermediate of Bovine Pancreatic Trypsin Inhibitor. Protein Science 3: 1822-32.

Stassinopoulou CI, Wagner G, Wüthrich K. 1984. Two-dimensional 1H NMR of two chemically modified analogs of the basic pancreatic trypsin inhibitor. Sequence-specific resonance assignments and sequence location of conformation changes relative to the native protein. Eur J Biochem 145: 423-30.

States DJ, Creighton TE, Dobson CM, Karplus M. 1987. Conformations of intermediates in the folding of the pancreatic trypsin inhibitor. J Mol Biol 195: 731-9.

States DJ, Dobson CM, Karplus M, Creighton TE. 1984. A new two-disulphide intermediate in the refolding of reduced bovine pancreatic trypsin inhibitor. J Mol Biol 174: 411-8.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol 185: 60-89.

Tüchsen E, Woodward C. 1985. Mechanism of surface peptide proton exchange in bovine pancreatic trypsin inhibitor. Salt effects and O-protonation. J Mol Biol 185: 421-30.

Tüchsen E, Woodward C. 1987. Assignment of asparagine-44 side-chain primary amide 1H NMR resonances and the peptide amide N1H resonance of glycine-37 in basic pancreatic trypsin inhibitor. Biochemistry 26: 1918-25.

van Mierlo CP, Darby NJ, Keeler J, Neuhaus D, Creighton TE. 1993. Partially folded conformation of the (30-51) intermediate in the disulphide folding pathway of bovine pancreatic trypsin inhibitor. 1H and 15N resonance assignments and determination of backbone dynamics from 15N relaxation measurements. J Mol Biol 229: 1125-46.

van Mierlo CP, Darby NJ, Neuhaus D, Creighton TE. 1991a. (14-38, 30-51) double-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor: a two-dimensional 1H nuclear magnetic resonance study. J Mol Biol 222: 353-71.

van Mierlo CP, Darby NJ, Neuhaus D, Creighton TE. 1991b. Two-dimensional 1H nuclear magnetic resonance study of the (5-55) single-disulphide folding intermediate of bovine pancreatic trypsin inhibitor. J Mol Biol 222: 373-90.

Wagner G, Bruhwiler D, Wüthrich K. 1987. Reinvestigation of the aromatic side-chains in the basic pancreatic trypsin inhibitor by heteronuclear two-dimensional nuclear magnetic resonance. J Mol Biol 196: 227-31.

Wagner G, Stassinopoulou CI, Wüthrich K. 1984. Amide-proton exchange studies by two-dimensional correlated 1H NMR in two chemically modified analogs of the basic pancreatic trypsin inhibitor. Eur J Biochem 145: 431-6.

Wagner G, Wüthrich K. 1982a. Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance. J Mol Biol 160: 343-61.

Wagner G, Wüthrich K. 1982b. Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. J Mol Biol 155: 347-66.

Weissman JS, Kim PS. 1991. Reexamination of the folding of BPTI: predominance of native intermediates. Science 253: 1386-93.

Weissman JS, Kim PS. 1992a. The Disulfide Folding Pathway of BPTI: Response. Science 256: 112-114.

Weissman JS, Kim PS. 1992b. Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A 89: 9900-4.

Weissman JS, Kim PS. 1992c. The pro region of BPTI facilitates folding. Cell 71: 841-51.

Weissman JS, Kim PS. 1993. Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. Nature 365: 185-8.

Williams JW, van Holde KE, Baldwin RL, Fujita H. 1958. The theory of sedimentation analysis. Chem Rev 58: 715-806.

Wlodawer A, Nachman J, Gilliland GL, Gallagher W, Woodward C. 1987. Structure of form III crystals of bovine pancreatic trypsin inhibitor. J Mol Biol 198: 469-80.

Wlodawer A, Walter J, Huber R, Sjölin L. 1984. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. J Mol Biol 180: 301-29.

Woodward C. 1993. Is the slow exchange core the protein folding core? Trends Biochem Sci 18: 359-60.

Wüthrich K. 1986. NMR of Proteins and Nucleic Acids. New York: John Wiley & Sons.

Table 1: Amide Proton Exchange Rates ($x10^{-4}min^{-1}$) of BPTI Variants at pH* 4.6

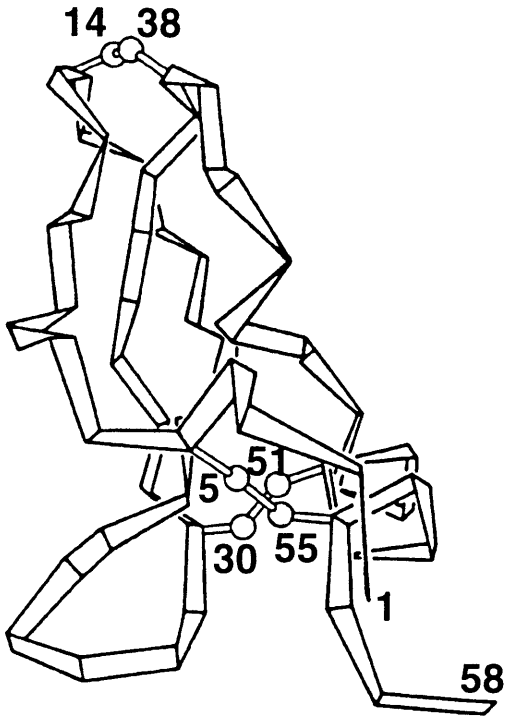
			[30-51; 14-38]$_{V5A55}$			[5-55]$_{Ala}$	
Residue‡		**H-bonded†**	**5°C**	**20°C**	**30°C**	**5°C**	**20°C**
Cys-5	(Val)	+	12	96	390	1.1	72
Leu-6		+	14	110	430	0.70	40
Glu-7		+	10	84	290	0.36	23
Tyr-10			2.6	24	140	4.3	160
Gly-12			26	220	930	13	390
Cys-14	(Ala)		10	1300	1100	350	fast
Ala-16		+	1.6	22	200	18	440
Arg-17			140	900	5900	190	fast
Ile-18		+	0.16	2.6	25	1.1	60
Ile-19			7.6	52	240	4.5	89
Arg-20		+	0.094	1.9	25	1.1	99
Tyr-21		+	0.076	1.2	17	0.96	88
Phe-22		+	0.22	1.5	20	1.1	95
Tyr-23		+	0.14	3.2	35	1.4	110
Asn-24		+	0.16	4.8	56	3.0	250
Ala-25			74	570	2800	13	480
Lys-26		+	400	fast	fast	230	fast
Ala-27		+	13	83	330	8.6	190
Gly-28		+	1.8	13	84	2.5	190
Leu-29			0.096	2.2	24	1.1	91
Cys-30	(Ala)		35	3100	1300	2.1	120
Gln-31		+	0.20	2.8	34	1.5	150
Thr-32			3.1	25	160	3.8	250
Phe-33		+	0.068	1.5	18	1.5	130
Val-34			3.4	25	110	2.7	90
Tyr-35		+	0.29	4.2	37	1.1	92
Gly-36		+	3.8	43	370	6.2	460
Gly-37			1.9	28	290	7.9	560
Cys-38	(Ala)	+	15	180	1400	97	fast
Ala-40			340	fast	fast	190	fast
Lys-41			8.9	110	870	6.5	520
Arg-42			230	fast	fast	120	fast
Asn-43		+	55	620	6500	18	700
Asn-44		+	2.3	45	500	6.4	660
Phe-45		+	0.60	14	170	2.9	260
Ser-47			17	110	500	4.3	260
Ala-48			110	790	630	20	630
Glu-49			500	fast	fast	130	2100
Asp-50		+	210	1600	5400	19	370
Cys-51	(Ala)	+	0.32	5.2	63	0.78	63
Met-52		+	0.28	3.4	36	0.58	43
Arg-53		+	0.42	5.9	61	0.86	71
Thr-54		+	5.6	39	200	1.9	130
Cys-55	(Ala)	+	6.9	46	250	2.3	210
Gly-56		+	0.19	150	680	2.9	210
Gly-57			430	fast	fast	290	fast
Ala-58			fast	fast	fast	300	fast

‡Residues which are mutated in either [5-55]$_{Ala}$ or [30-51; 14-38]$_{V5A55}$ are denoted in parentheses.

†Observed to participate in an H-bond in at least one of the crystal structures of native BPTI (Deisenhofer & Steigemann, 1975; Wlodawer, et al., 1984; Wlodawer, et al., 1987; Oas & Kim, 1988).

Fig. 1

A)


14
38
51
5
30
55
1
58

B)

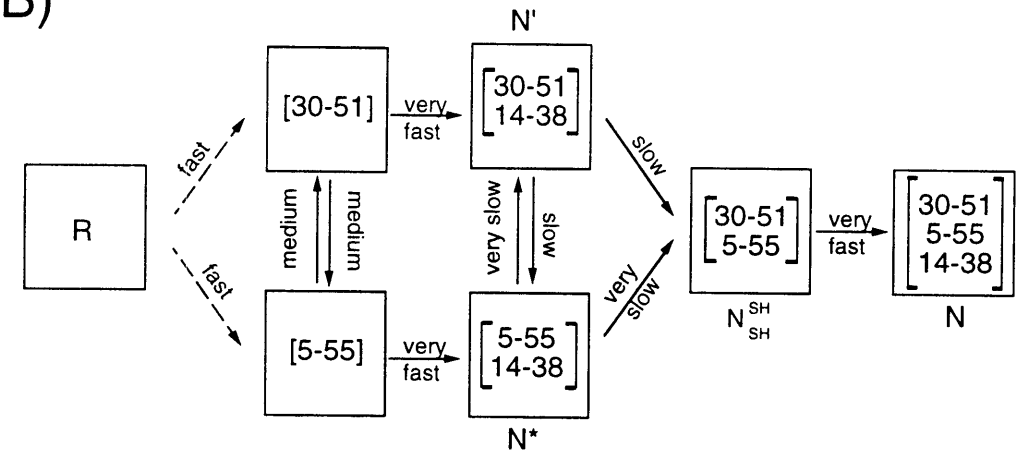
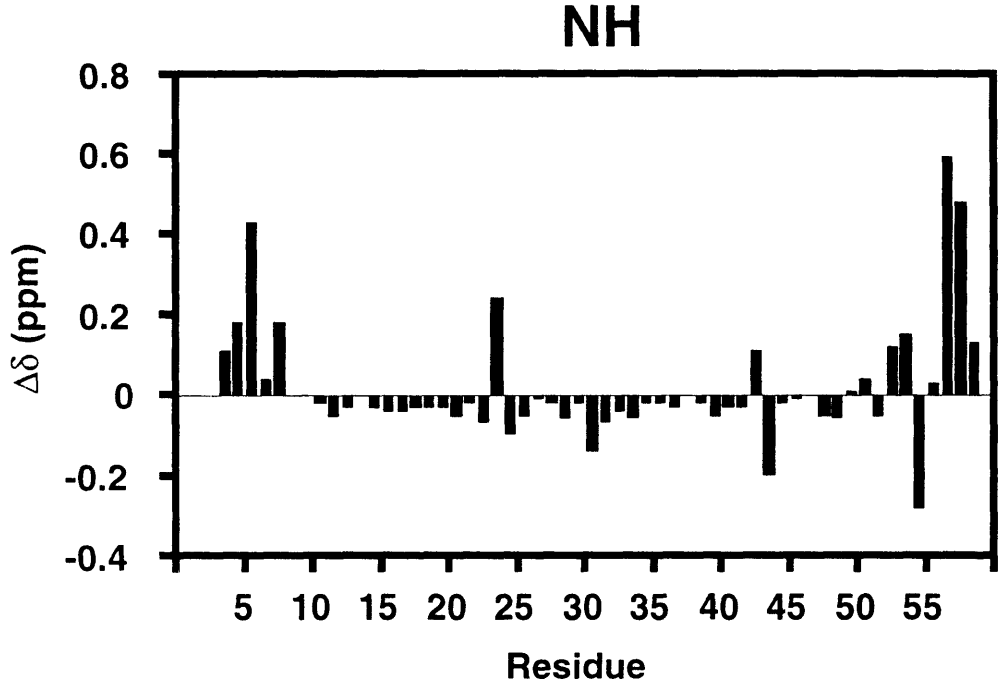

R
fast
fast
[30-51]
very fast
N'
[30-51 14-38]
medium
medium
very slow
slow
[5-55]
very fast
[5-55 14-38]
N*
slow
very slow
[30-51 5-55]
N SH SH
very fast
[30-51 5-55 14-38]
N

Fig. 2

A)


NH
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Δδ (ppm)
5 10 15 20 25 30 35 40 45 50 55
Residue

B)

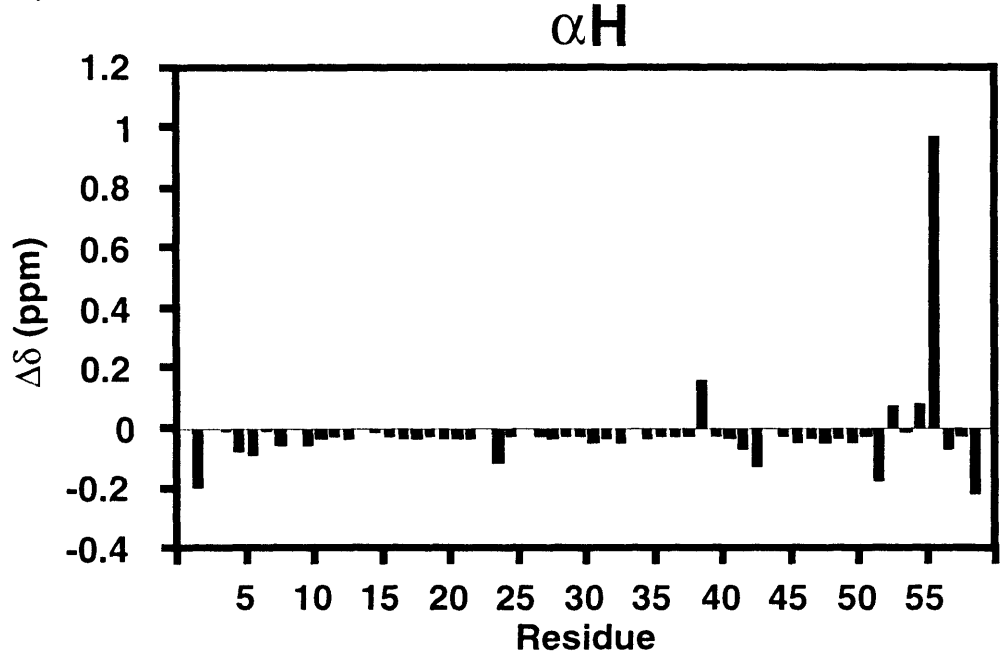

αH
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Δδ (ppm)
5 10 15 20 25 30 35 40 45 50 55
Residue

Fig. 2C

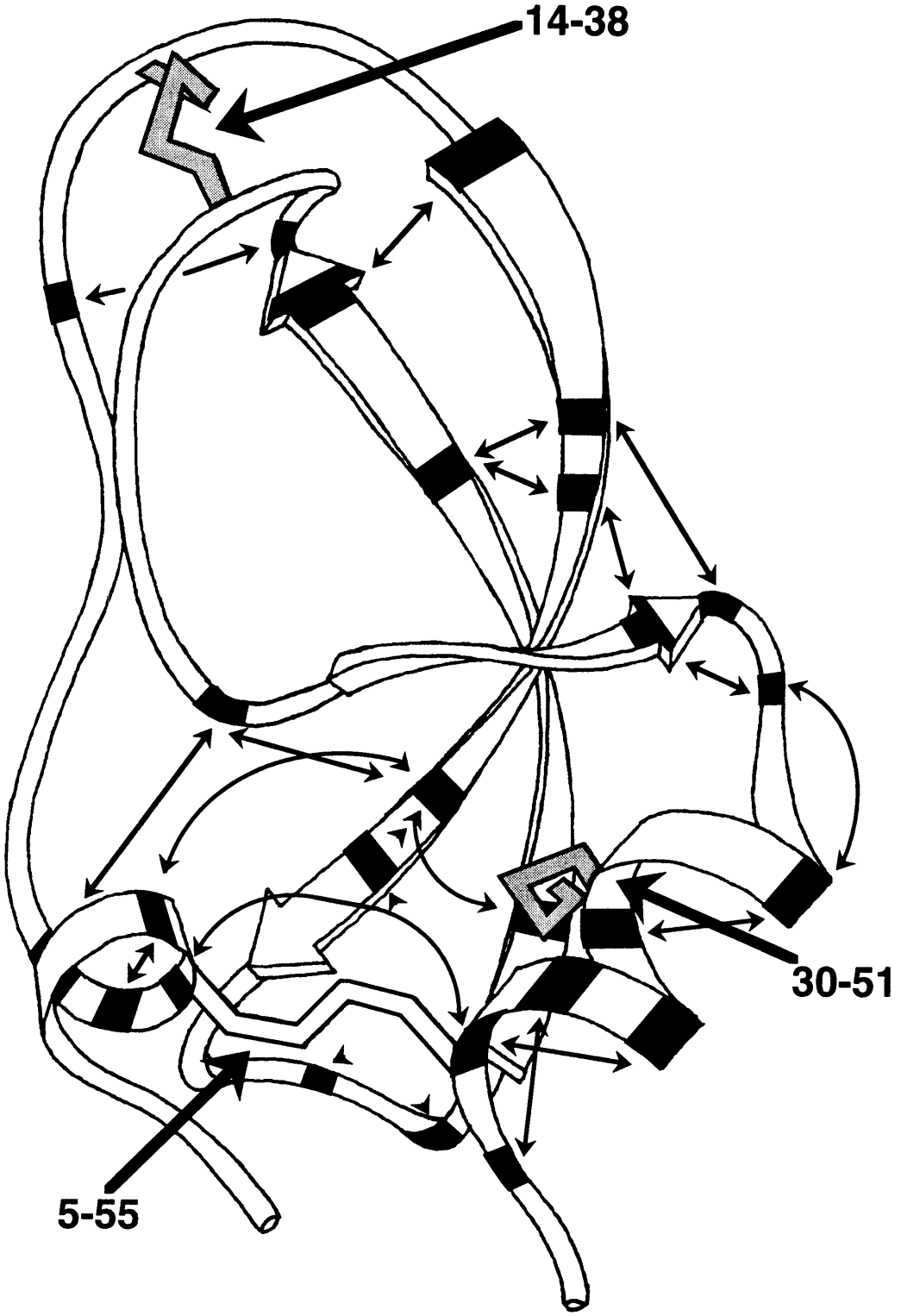


Fig. 3

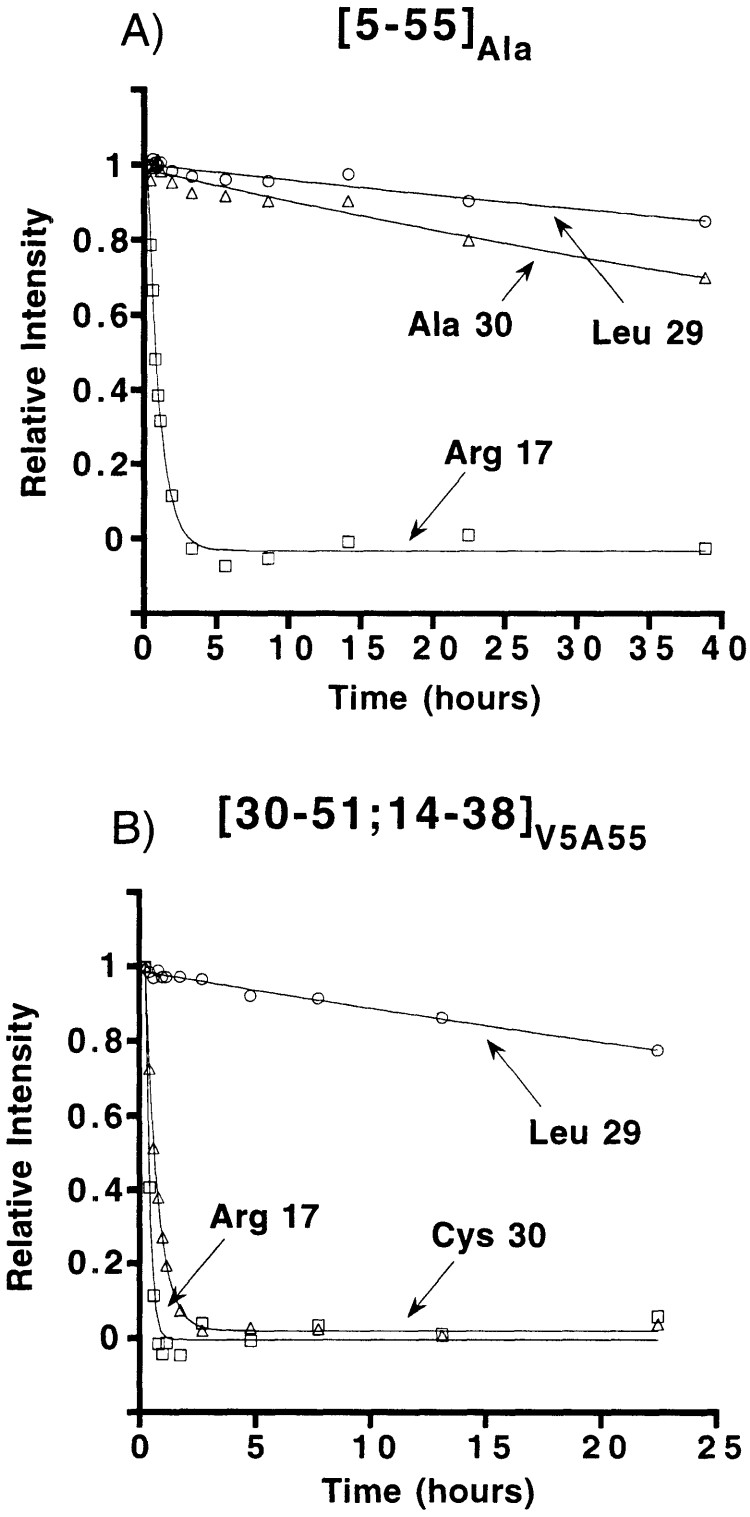
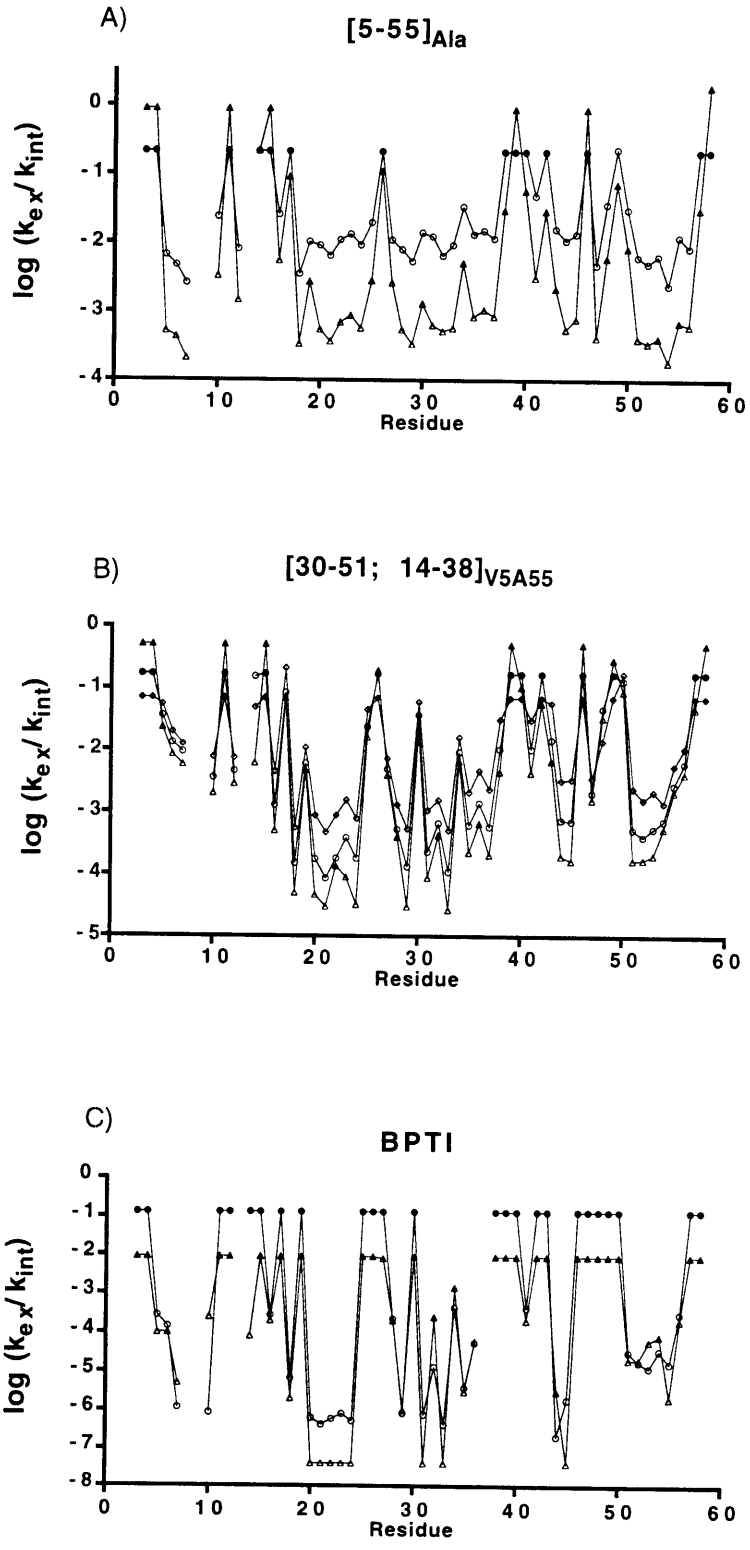

A)
[5-55]Ala
Relative Intensity
1
0.8
0.6
0.4
0.2
0
Ala 30
Leu 29
Arg 17
0 5 10 15 20 25 30 35 40
Time (hours)
B)
[30-51;14-38]V5A55
Relative Intensity
1
0.8
0.6
0.4
0.2
0
Leu 29
Arg 17
Cys 30
0 5 10 15 20 25
Time (hours)

Fig. 4


A)
[5-55]Ala
log (kex/kint)
0
-1
-2
-3
-4
0
10
20
30
40
50
60
Residue
B)
[30-51; 14-38]V5A55
log (kex/kint)
0
-1
-2
-3
-4
-5
0
10
20
30
40
50
60
Residue
C)
BPTI
log (kex/kint)
0
-1
-2
-3
-4
-5
-6
-7
-8
0
10
20
30
40
50
60
Residue

APPENDIX III

POSITION-DEPENDENCE OF HELICAL STRUCTURE IN ALANINE-BASED PEPTIDES

Introduction

The protein folding problem is difficult to approach because the structures of proteins are diverse and they are stabilized by a combination of many local and global interactions within the protein, and by interactions with solvent. Nevertheless, proteins are composed of a hierarchy of structures: proteins contain compact subdomains, which, in turn, consist of units of secondary structure packing together (1). In principle, understanding the folding of each of these simpler units will provide the first step towards explaining how a protein folds. Toward this end, we are trying to understand the folding transition of one of the simplest units of structure, the α-helix.

Several features of the α-helix make it an important and tractable model system. First, the α-helix is the most abundant form of secondary structure in proteins (2). Second, it is a uniform, repeating structure. Despite differences in hydrogen bonding between the middle and the ends of a helix, in the model α-helix, the phi and psi angles of the backbone are -57° and -47°, respectively (2). Third, isolated helices are stabilized primarily by local interactions. And, perhaps most importantly, short monomeric peptides, both corresponding to protein helices and of *de novo* design, can form stable helices in solution (3-7).

Three pieces of evidence suggest that the structure and stability of α-helices in proteins are at least partially determined by local sequence. First, some factors that influence the stability of isolated helices, such as the helix-propensities of the constituent amino acids and charge-helix dipole interactions, influence the stability of proteins, as well (8-22). Second, helix propensities of different amino acids are similar in helices formed by both short peptides and proteins (11, 13, 23-26). In addition, the moderate success of secondary structure prediction methods suggests that local structure is at least one determinant of a protein's overall fold (27). Thus, as a model system for understanding local structure in proteins, isolated α-helices provide information that may be useful for protein and peptide design.

In addition to providing information about the forces that stabilize helices, isolated helical peptides are a model system for studying the

mechanism of a polypeptide folding transition. The helix-coil transition of model peptides cannot be described as a simple two-state equilibrium. Helical peptides are not only either fully folded or fully unfolded, but rather can adopt a range of partially folded structures (28, 29). Thus, the average helix content of a peptide is not directly related to a single thermodynamic state function such as ΔG. An understanding of the mechanism of the helix-coil transition is crucial for quantifying the thermodynamic contributions of the factors stabilizing helices.

A number of statistical mechanical models, such as the "zipper", Zimm-Bragg, and Lifson-Roig models, describe the helix-coil transition (30-32). While these models differ in detail, they are based on similar assumptions. They assume that each residue can exist in either of two states, helix or unfolded, and assign statistical weights for the conformation of each residue depending on the conformations of neighboring residues. The Zimm-Bragg model, for example, uses two parameters: σ for nucleating a helix and s for propagating a helix (31). The value of these models is that they use only a small number of statistical parameters to describe the helix-coil transition. If these models are proven valid, the statistical parameters can be measured for different peptide helices and compared in order to evaluate the relative contributions of the forces that stabilize helices.

The statistical mechanical models have been successful in predicting qualitatively the length- and temperature-dependent behavior of isolated helices. The theories also predict details of helical structure, such as a position dependence of structure: residues in the middle of a peptide are expected to be found in the helical conformation more frequently than residues at the ends.

There are limited ways to measure the fraction of peptides in which a given residue is helical because assays for site-specific structure, such as x-ray crystallography or NMR, traditionally detect structure in fully folded molecules. Other methods such as CD measure average properties of helices, so details of the specific nature of the helix-coil transition only can be inferred from how well experiments agree with theoretical predictions. One microscopic property that can be measured at each position in a peptide is

chemical shift. Proton chemical shifts are sensitive to the electromagnetic environment, including the electronic properties of the atom itself and external factors that influence the local magnetic environment. If the electronic properties of the probes are the same, then changes in chemical shift reflect changes in structure.

Alanine methyl protons are good probes of helical structure for several reasons. First, alanine has been shown to have very high helix propensity (11, 13, 24, 33). Second, in crystal structures of helical polymers, the β-methylene or methyl group of a residue is near the amide of the following residue (34, 35). This geometry is unique to helices, and is not observed in other forms of secondary structure (36, 37). Finally, because the methyl protons do not participate in hydrogen bonds, are not polar and are not in exchange with solvent, they will be relatively insensitive to solvent conditions. Thus, in the case of alanine-based helical peptides, alanine methyl proton chemical shifts should be strongly influenced by the structure of a peptide. We have measured the changes in alanine side-chain methyl proton chemical shifts from 5-80°C for a number of individual residues in alanine-based helical peptides. The helical structure in the peptides unfolds over this temperature range. The change in alanine methyl proton chemical shift varies with the position of the residue in the peptide sequence. In agreement with helix-coil transition theory, the data suggest that the residues at the ends of helical peptides are substantially folded, but are less likely to be folded than residues at the center.

Results

Resonance Assignments for alanine-based helical peptides

Alanine has a high propensity for adopting a helical conformation, and alanine-based peptides have been shown to be highly helical (33). The 17-, 22- and 27-residue peptides, with the sequences Ac-WA$(RA_4)_n$-NH_2 and Ac-AA$(RA_4)_n$-NH_2 (n=3-5) used for this study were based on model peptides used by Baldwin and colleagues to study the helix-coil transition. (The Trp was included at the N-terminus for CD studies, in order to facilitate concentration determination.) Fig. 1 shows the CD spectrum of a

representative peptide, the 22-mer Ac-WA$(RA_4)_5$-NH_2. All of the peptides used in this study are highly helical at low temperature (data not shown).

Part of the appeal of studying alanine-based peptides is their homogeneity and anticipated lack of sequence-specific effects. However, in a typical 1D ^{1}H-NMR spectrum of alanine-rich polypeptides, assignment of resonances is difficult because of chemical shift overlap. To circumvent this problem, we used peptides with side-chain deuterated alanine incorporated in most positions. Deuterium is invisible in ^{1}H-NMR spectra, so the methyl deuterons do not appear in the spectrum. Up to three carefully selected positions are resolved per peptide, and these are distinguished by using different ratios of ^{1}H-alanine:^{2}H-alanine in each position (Fig. 2). Substitution of deuterons for protons in the methyl groups is not expected to influence the stability of these peptides because there is little difference amongst the helix propensities of many straight-side-chain, nonpolar amino acids, which have even bulkier substituents attached to the β-carbon (14, 38).

Position dependent changes in chemical shift over the temperature range of 5-80°C

Changes in alanine methyl proton chemical shift over the temperature range of 5 to 80°C were monitored for several positions in the 17-, 22- and 27-residue peptides. All if the data are compared to the chemical shift at 80°C, a temperature at which CD measurements indicate that the peptides are unfolded (data not shown). The change in alanine methyl chemical shift ($\Delta\delta$) (that is, the difference in δ at a given temperature from δ at 80°C) is an observable that depends both on peptide length and position within the peptide.

The relative values of $\Delta\delta$ have been compared, qualitatively, for many positions along the sequence of each the 17-, 22- and 27-residue peptides, Ac-AA$(RA_4)_{3\text{-}5}$-NH_2. There is a smooth curve for the position dependence of $\Delta\delta$ at low temperature: $\Delta\delta$ is largest for central alanines and decreases progressively from the middle to the ends of the peptide (Fig. 3). $\Delta\delta$ clearly depends on formation of the helical structure, because ~8M urea, which

unfolds the peptides, essentially eliminates the differences in $\Delta\delta$ for different residues (Fig. 4).

Changes in $\Delta\delta$ reflect changes in helix-content

To test whether the temperature-dependent change in chemical shift in the helical peptides is related to helical structure, we compared the results obtained by NMR with the results obtained for a method known to measure helix-content: CD at 222 nm. CD measures the average helix-content. The overall features of the thermal transitions observed by CD are similar to the thermal transitions observed by NMR (Fig. 5), suggesting that the changes in chemical shift reflect changes in helix content.

Length-dependent differences in the thermal transitions observed for different positions

Many studies have suggested that helices are frayed at the ends. Indeed, for residues at the ends, there is a smaller change in $\Delta\delta$ upon unfolding than there is for the middle. In addition to the difference in the magnitude of the transition for the middle and the ends, there is also a length-dependent difference in the nature of the thermal transitions for the middle and the ends. In both the 22- and 27-residue peptides, which are nearly completely helical at low temperature as judged by CD, the thermal transition for the middle alanine is more S-shaped, approaching baselines at low and high temperatures On the other hand, the terminal alanine undergoes a more broad transition and never approaches a fully folded baseline at low temperature (Figure 6). Thus, the nature of the transition for the middle residue is different from the ends, and depends on peptide length.

Discussion

Are the ends of helices frayed?

Theoretical models for the helix-coil transition predict a strong position-dependence of helical structure (30). 2D NMR studies of isolated peptides show that the interproton distances and angles are consistent with α-

helical structure. These studies also indicate, qualitatively, that the ends of helices are frayed (39-41).

Baldwin's group has examined the loss of overall helix-content arising from amino acid substitutions at various positions within a helical sequence. Studies on the C-peptide of RNAse A and on a streamlined alanine-based peptide show that there is a position-dependent loss of helix-content upon substituting an amino acid with low helix-forming propensity (glycine or proline) for one with high helix-forming propensity (alanine) (42-44). There is little effect of substituting residues at the termini, but there is a large effect for mutating residues near the middle of the sequence. These data have been interpreted to mean that the ends of the helices are substantially frayed.

Recently, the protection from hydrogen exchange has been measured for different residues in ^{15}N-labeled alanine-based peptides (45). Rohl and Baldwin observe greater protection from hydrogen exchange for residues in the middle of the peptide than for residues at the end, consistent with the predictions of helix-coil transition theory.

To compare our method with other techniques, we also examined the changes in $\Delta\delta$ in a 21-residue helical peptide that is substantially folded at the N-terminus. Diagnostic NOEs indicate that the N-terminal four residues of Suc-$A_5(A_3RA)_3A$-NH_2 are helical (46, 47). Nonetheless, there is a difference in $\Delta\delta$ for the middle and ends in this peptide, and the difference is comparable to that described above for the 22-residue peptide (data not shown--$\Delta\delta$ for the N-terminal alanine is ~60% of that for the central alanine).

However, it is unclear what fraction of the population must be helical in order to give rise to helical NOEs. Sequential NOEs have been observed for the first four residues in two salt-bridge-stabilized helical peptides, termed "EAK" and "ELK", which are not as streamlined as the alanine-based peptides (41). However, unlike studies of the 21-mer Suc-$A_5(A_3RA)_3A$-NH_2, which were limited by extreme spectral overlap, in the EXK peptides NOE intensities were measured for the middle residues as well as the ends. The intensities of NOEs in the middle of these peptides are 3-4 times greater than NOEs between end residues. The differences in NOE intensity and position-

dependent differences in J-coupling constants indicate that the ends of the EXK peptides are frayed (41). Thus, residues at the termini of frayed helices can give rise to NOEs that are diagnostic for helical structure.

We show here that for helical alanine-based peptides, $\Delta\delta$ for each alanine increases progressively from the end to the middle. In addition, the ends undergo a less cooperative thermal transition than the residues in the middle of the peptide. These data not only agree with previous experiments suggesting that the ends of helices are frayed, but also demonstrate that the ends and middle undergo different unfolding transitions.

Helix-coil transition theory predicts the behavior of helical peptides

The simplest statistical mechanical model of the helix-coil transition is the "zipper model" (30), which assumes that only one stretch of helix can be found within a given peptide. Only the statistical parameters of the Zimm-Bragg theory ("σ" for nucleating a helix and "s" for adding a new helical unit to the end of an existing helix) (31) are required for describing helix formation in this model. Thus, the partition function for a chain of n units is:

$$q = 1 + \sum_{k=1}^{n} \Omega_k \sigma s^k ,$$

where Ω_k is the number of ways of putting k helical units together in one sequence in a chain of n residues (30). The fraction-folded for each residue in a helical peptide can be predicted from the zipper model as described below.

RATIONALE

I. The penalty for initiating a helix is σ.
III. The equilibrium constant for adding a residue of helix to a pre-existing helix is s.

IV. The number of different ways of putting i helical units together in sequence within a chain of n residues is (n+1-i). The statistical weights for forming such helices is F[i_].

V. The partition function is Q. $(Q = \sum_{k=1}^{n} 1+F[i\_])$

VI. The number of ways to put a helix of length i in a sequence of length n with its jth residue in the helical conformation is given by:

(1) If $i \leq j$ and $i \leq (n + 1 - j)$, then there are i helices (▭) containing the jth residue: (e.g. i =4, j = 5)

(2) If $i \geq j$ and $i \leq (n + 1 - j)$, then there are j helices containing the jth residue: (e.g. i = 7, j = 5)

(3) If $i \leq j$ and $i \geq (n + 1 - j)$, then there are (n - j + 1) helices containing the jth residue: (e.g. i = 7, j = 7)

(4) If $i \geq j$ and $i \geq (n + 1 - j)$, then there are (n - i + 1) helices containing the jth residue: (e.g. i = 9, j = 7)

VII. The propensity of a given residue (l) to be in the helical conformation is the ratio of helices with the lth residue in the helical conformation compared to the total number of helices possible for a peptide of length n.

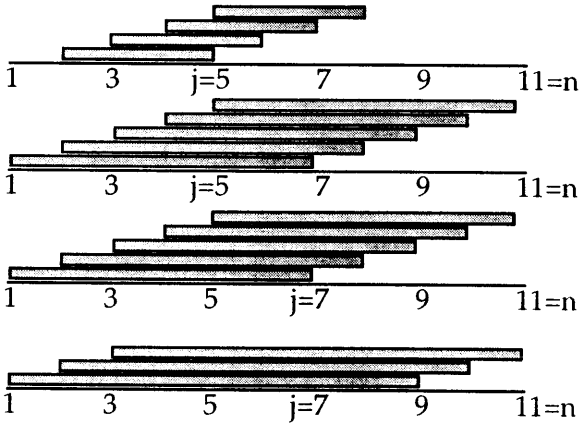
MATHEMATICA* PROGRAM

I. σ=.0029
II. n=length of peptide
III. s=1.56

IV. F[i_]:=σ^i*s*(n+1-i)

V. Q:=(Sum[F[i],{i,1,n}]+1)

VI. nmb[i_,j_]:=i+Min[j-i,0]+Min[n+1-j-i,0]



VII. pos [l_]:=Sum [nmb[i,l]S*s^i, {i,1,n}]/Q

(*Wolfram Research)

Given our observation that the changes in $\Delta\delta$ upon unfolding depend on the position of a residue within the peptide, we compared the properties of $\Delta\delta$ with the residue-specific properties of helical peptides predicted with the zipper model. Fig. 7 shows that the position-dependence of $\Delta\delta$ at 5°C is similar to the position-dependence of helix content predicted by the zipper model.

In addition, although temperature is not an explicit parameter in this model, "s", the equilibrium constant for adding a residue to a pre-existing helix, in principle depends on temperature (30). Thus, a temperature dependent transition can be represented by a transition as a function of s. On a qualitative level, the observed temperature dependent behavior of the middle and the end agrees strikingly well with theoretical predictions (Fig. 8). Although likely to be an oversimplification, the zipper model predicts relatively accurately the position- and temperature-dependent behavior of $\Delta\delta$ for short alanine-based peptides.

Conclusions

The temperature-dependent changes in alanine methyl proton chemical shift appear to reflect changes in helical structure in isolated alanine-based peptides. Both the magnitude and the nature of the thermal transition differ for the central and end residues. These differences are predicted by a simple helix-coil transition theory.

Figure Legends

Fig. 1: CD spectrum at 0°C for the 22-residue peptide Ac-WA$(RA_4)_4$-NH_2.

Fig. 2: NMR spectra at 5°C of two 17-residue alanine based peptides shown to scale. (A) The peptide Ac-WARAAAARAAAARBAAB-NH_2 gives rise to 13 methyl resonances between 1.54 and 1.36 ppm. All of these methyl resonances except that of Ala 2, which is ring-current shifted upfield by Trp 1, are overlapping. (B) The peptide Ac-AA$(RA_4)_3$-NH_2 contains 14 alanine methyl groups. However, in this peptide, every alanine has a deuterated side chain except for Ala 7, Ala 14, and Ala17, which are easily identified because they are labeled to an extent of 50, 100 and 25 percent, respectively, with normal, protonated L-alanine.

Fig. 3: Position dependence of $\Delta\delta$ (from 278K to 353K) for the 17-residue peptide Ac-AA$(RA_4)_3$-NH_2 and the 22-residue peptide Ac-AA$(RA_4)_4$-NH_2. For the 17-mer, Ala 9 is the central alanine. For the 22-mer, Ala 12 is considered to be 0.5 residue from the middle of the peptide.

Fig. 4: The temperature dependence of $\Delta\delta$ (δ at a given temperature - δ at 353K) for three residues in the 22-residue peptide Ac-AA$(RA_4)_4$-NH_2. Filled symbols show the changes in chemical shifts in aqueous buffer (150 mM NaCl, 10 mM sodium phosphate, pH* 7.0). Open symbols show the changes in chemical shifts in the presence of 7.8 M deuterated urea.

Fig. 5: The thermal melts by CD of the peptides Ac-WA$(RA_4)_{3\text{ or }4}$-NH_2 are compared with thermal melts by NMR for the peptides Ac-AA$(RA_4)_{3\text{ or }4}$-NH_2. The overall fraction helix is estimated for the NMR melts as follows. The $\Delta\delta$ value at ~5° C for the central residue in the 22- and 27-residue peptides is nearly identical (~.156 ppm). Thus, it is assumed that (.156*the number of residues, 17 or 22) is equal to 100% helix. The fraction of helical structure at a given temperature was determined by: 1) fitting the position-dependence data for each temperature to a fourth order polynomial (which recapitulates the overall shape of the curve), 2) integrating under the curve for the length of the peptide, and 3) comparing the area of $\Delta\delta$ with the value for 100% helix.

Fig. 6: Comparison of the temperature dependence of $\Delta\delta$ (δ at a given temperature - δ at 353K) for the middle and the terminal alanines of the 17-, 22- and 27-residue peptides, Ac-AA$(RA_4)_{3\text{-}5}$-NH_2. The values of $\Delta\delta$ for the end residues (squares) are normalized to the scale observed for the central residues (circles). The difference in the transition for the middle and the end residues of a peptide increases with peptide length.

Fig. 7: A) The fraction folded for different positions in the 17-residue peptide Ac-AA$(RA_4)_3$-NH_2 is estimated by comparing $\Delta\delta$ with the value of $\Delta\delta$ observed for the central residue in the 22- and 27-residue helices (0.156 ppm is considered as the value for 100% helix.) B) The fraction folded for different positions in a 17-residue helical peptide as predicted by the zipper model for the helix-coil transition (30). Values of σ (0.0029) and s (1.56) obtained for alanine were taken from (44).

Fig. 8: The s-dependence of the site-specific equilibrium constant for the center and end positions of a 17-, 21-, and a 27-residue peptide as approximated by the "zipper model" for the helix-coil transition (30). The value for σ is 0.0029, determined for alanine-based peptides by (44). All residues are assumed to contribute the same values of σ and s when in the helical conformation. The value for s of 1.56 is used for low temperature (44).

Materials and Methods

The peptides were synthesized using solid-phase t-Boc methods (for review, see (48)) on an Applied Biosystems Model 430A peptide synthesizer and were cleaved from the resin using TFMSA (49, 50). The peptides were desalted on a Sephadex G-10 column in 5% acetic acid and purified by reversed-phase HPLC on a Vydac C18 preparative column using a linear water/acetonitrile gradient containing 0.1% TFA. N-t-Boc L-Alanine [3,3,3,-D3] was obtained from Tracer Technologies (Somerville, MA). Peptide identity was confirmed by mass spectrometry (M-scan, Westchester, PA).

CD experiments were performed on an AVIV Model 60DS CD spectometer using a 1 cm path-length cell. The peptide concentrations for CD were determined by absorption of tryptophan under denaturing conditions (51), and were ~30 μM, in 150mM NaCl and 10 mM sodium phosphate, pH 7.0.

NMR spectroscopy was performed on a Bruker AMX spectrometer operating at 500.1 MHz for ^1H. Free induction decays were averaged over 2048 scans with 8 K complex data points. Data were acquired using a spectral width of 5556 Hz. Samples were ~3 mM peptide in D_2O, with 150mM NaCl and 10 mM sodium phosphate, adjusted to pH* 7.0. (pH* refers to pH-meter readings uncorrected for isotope effects.) Chemical shifts were determined relative to an internal (trimethylsilyl) propionate (TMSP) standard (52).

Acknowledgments

We are grateful to R. Rutkowski and M. Burgess for peptide synthesis, and J. Weissman for many helpful discussions and critical comments on earlier versions of this chapter. In addition, we thank J. Weissman, Z-y. Peng, and L. Nekludova for help with the program to test the zipper model for the helix-coil transition.

References

1. P. S. Kim, R. L. Baldwin, *Annu Rev Biochem* **51**, 459-89 (1982).

2. T. E. Creighton, *Proteins* (W.H. Freeman and Co., New York, 1984).

3. J. E. Brown, W. A. Klee, *Biochemistry* **10**, 470-476 (1971).

4. A. Bierzynski, P. S. Kim, R. L. Baldwin, *Proc Natl Acad Sci U S A* **79**, 2470-4 (1982).

5. K. R. Shoemaker, et al., *Cold Spring Harb Symp Quant Biol* **52**, 391-8 (1987).

6. S. Marqusee, R. L. Baldwin, *Proc Natl Acad Sci U S A* **84**, 8898-902 (1987).

7. E. M. Goodman, P. S. Kim, *Biochem.* **28**, 4343-4347 (1989).

8. K. R. Shoemaker, et al., *Proc Natl Acad Sci U S A* **82**, 2349-53 (1985).

9. K. R. Shoemaker, P. S. Kim, E. J. York, J. M. Stewart, R. L. Baldwin, *Nature* **326**, 563-7 (1987).

10. R. Fairman, K. R. Shoemaker, E. J. York, J. M. Stewart, R. L. Baldwin, *Proteins* **5**, 1-7 (1989).

11. S. Padmanabhan, S. Marqusee, T. Ridgeway, T. M. Laue, R. L. Baldwin, *Nature* **344**, 268-70 (1990).

12. G. Merutka, W. Lipton, W. Shalongo, S.-h. Park, E. Stellwagen, *Biochem.* **29**, 7511-7515 (1990).

13. P. C. Lyu, M. I. Liff, M. A. Marky, N. R. Kallenbach, *Science* **250**, 669-673 (1990).

14. P. C. Lyu, J. C. Sherman, A. Chen, N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA* **88**, 5317-5320 (1991).

15. M. H. Hecht, J. M. Sturtevant, R. T. Sauer, *Proteins* **1**, 43-6 (1986).

16. C. Mitchinson, R. L. Baldwin, *Proteins* **1**, 23-33 (1986).

17. J. S. Richardson, D. C. Richardson, *Science* **240**, 1648-1652 (1988).

18. D. Sali, M. Bycroft, A. R. Fersht, *Nature* **335**, 740-743 (1988).

19. H. Nicholson, W. J. Becktel, B. W. Matthews, *Nature* **336**, 651-6 (1988).

20. H. Nicholson, D. E. Anderson, S. Dao-pin, B. W. Matthews, *Biochemistry* **30**, 9816-28 (1991).

21. S. Dao-Pin, W. A. Baase, B. W. Matthews, *Proteins* **7**, 198-204 (1990).

22. X. J. Zhang, W. A. Baase, B. W. Matthews, *Biochemistry* **30**, 2012-7 (1991).

23. G. D. Fasman, *Prediction of Protein Structure and the Principles of Protein Conformation* (Plenum Press, New York, 1989).

24. K. T. O'Neil, W. F. De Grado, *Science* **250**, 646-51 (1990).

25. A. Horovitz, J. M. Matthews, A. R. Fersht, *J. Mol. Biol.* **227**, 560-568 (1992).

26. M. Blaber, X.-j. Zhang, B. W. Matthews, *Science* **260**, 1637-1640 (1993).

27. K. A. Dill, *Biochemistry* **29**, 7133-55 (1990).

28. J. M. Scholtz, et al., *Proc Natl Acad Sci U S A* **88**, 2854-8 (1991).

29. J. M. Scholtz, R. L. Baldwin, *Annu Rev Biophys Biomol Struct* **21**, 95-118 (1992).

30. C. R. Cantor, P. R. Schimmel, in *Biophysical Chemistry*. (W.H. Freeman and Co., New York, 1980), vol. 3, pp. 1041-1073.

31. B. H. Zimm, J. K. Bragg, *J. Chem. Phys.* **31**, 526-535 (1959).

32. S. Lifson, A. Roig, *J. Chem. Phys.* **34**, 1963-1974 (1961).

33. S. Marqusee, V. H. Robbins, R. L. Baldwin, *Proc Natl Acad Sci U S A* **86**, 5286-90 (1989).

34. L. Pauling, R. B. Corey, H. R. Branson, *Proc. Natl. Acad. Sci. USA* **37**, 205-210 (1951).

35. S. Arnott, S. D. Dover, *J Mol Biol* **30**, 209-12 (1967).

36. K. Wüthrich, M. Billeter, W. Braun, *J. Mol. Biol.* **180**, 715-740 (1984).

37. K. Wüthrich, *NMR of Proteins and Nucleic Acids* (John Wiley and Sons, New York, 1986).

38. S. Padmanabhan, R. L. Baldwin, *J Mol Biol* **219**, 135-7 (1991).

39. J. Osterhout Jr., et al., *Biochemistry* **28**, 7059-64 (1989).

40. E. K. Bradley, J. F. Thomason, F. E. Cohen, P. S. Kosen, I. D. Kuntz, *J. Mol. Biol.* **215**, 607-622 (1990).

41. M. I. Liff, P. C. Lyu, N. R. Kallenbach, *J. Am. Chem. Soc.* **113**, 1014-1019 (1991).

42. K. G. Strehlow, R. L. Baldwin, *Biochemistry* **28**, 2130-3 (1989).

43. K. G. Strehlow, A. D. Robertson, R. L. Baldwin, *Biochemistry* **30**, 5810-4 (1991).

44. A. Chakrabartty, J. A. Schellman, R. L. Baldwin, *Nature* **351**, 586-8 (1991).

45. C. A. Rohl, R. L. Baldwin, *Biochemistry* **33**, 7760-7767 (1994).

46. D. J. Lockhart, P. S. Kim, *Science* **257**, 947-951 (1992).

47. D. J. Lockhart, P. S. Kim, *Science* **260**, 198-202 (1993).

48. S. B. H. Kent, *Annu. Rev. Biochem.* **57**, 959-989 (1988).

49. H. Yajima, N. Fujii, in *The Peptides* E. Gross, J. Meienhofer, Eds. (Academic Press, New York, 1983), vol. 5, pp. 65-109.

50. J. P. Tam, W. F. Heath, R. B. Merrifield, *J. Am. Chem. Soc.* **108**, 5242-5252 (1986).

51. H. Edelhoch, *Biochemistry* **6**, 1948-1953 (1967).

52. A. DeMarco, *J. Magn. Reson.* **26**, 527-528 (1977).

Fig. 1

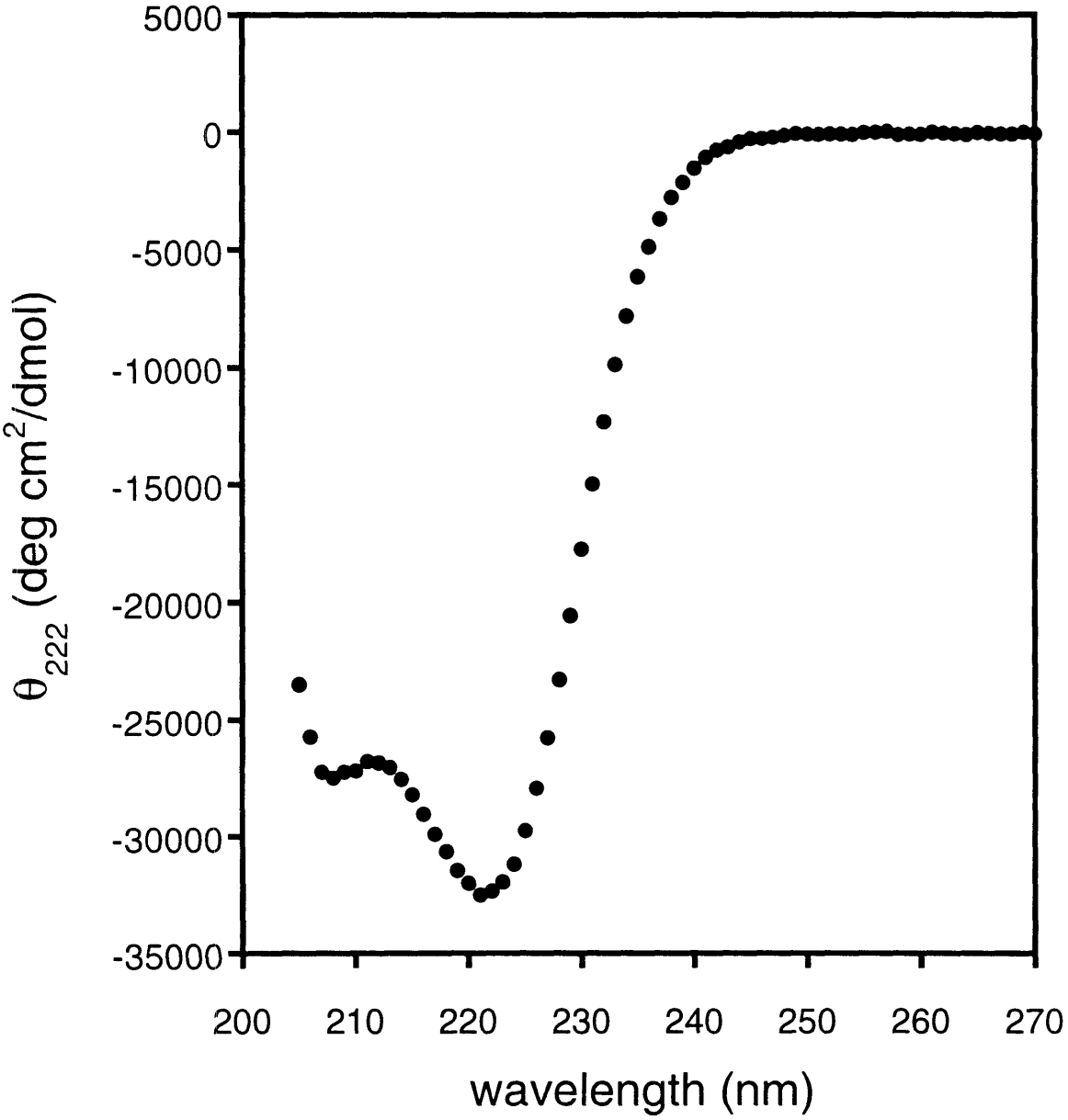

5000
0
-5000
-10000
-15000
-20000
-25000
-30000
-35000
θ222 (deg cm2/dmol)
200
210
220
230
240
250
260
270
wavelength (nm)

Fig. 2

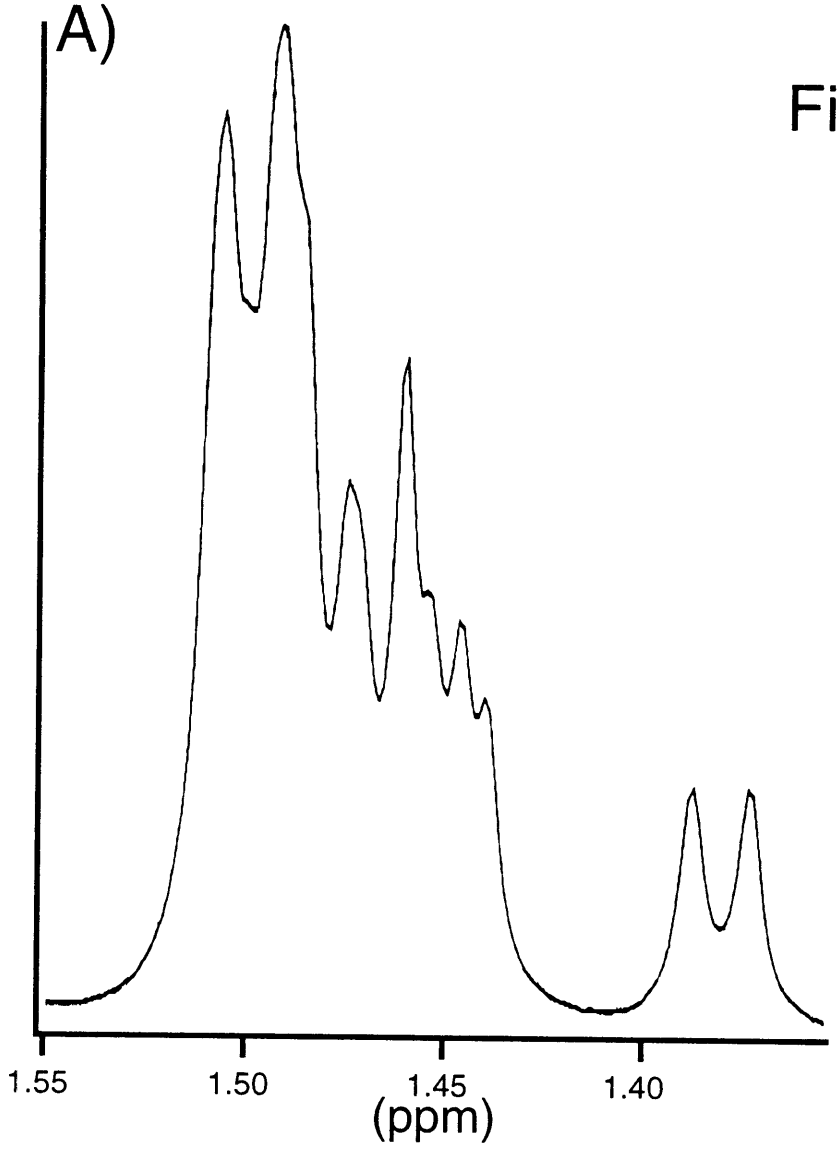


Fig. 3

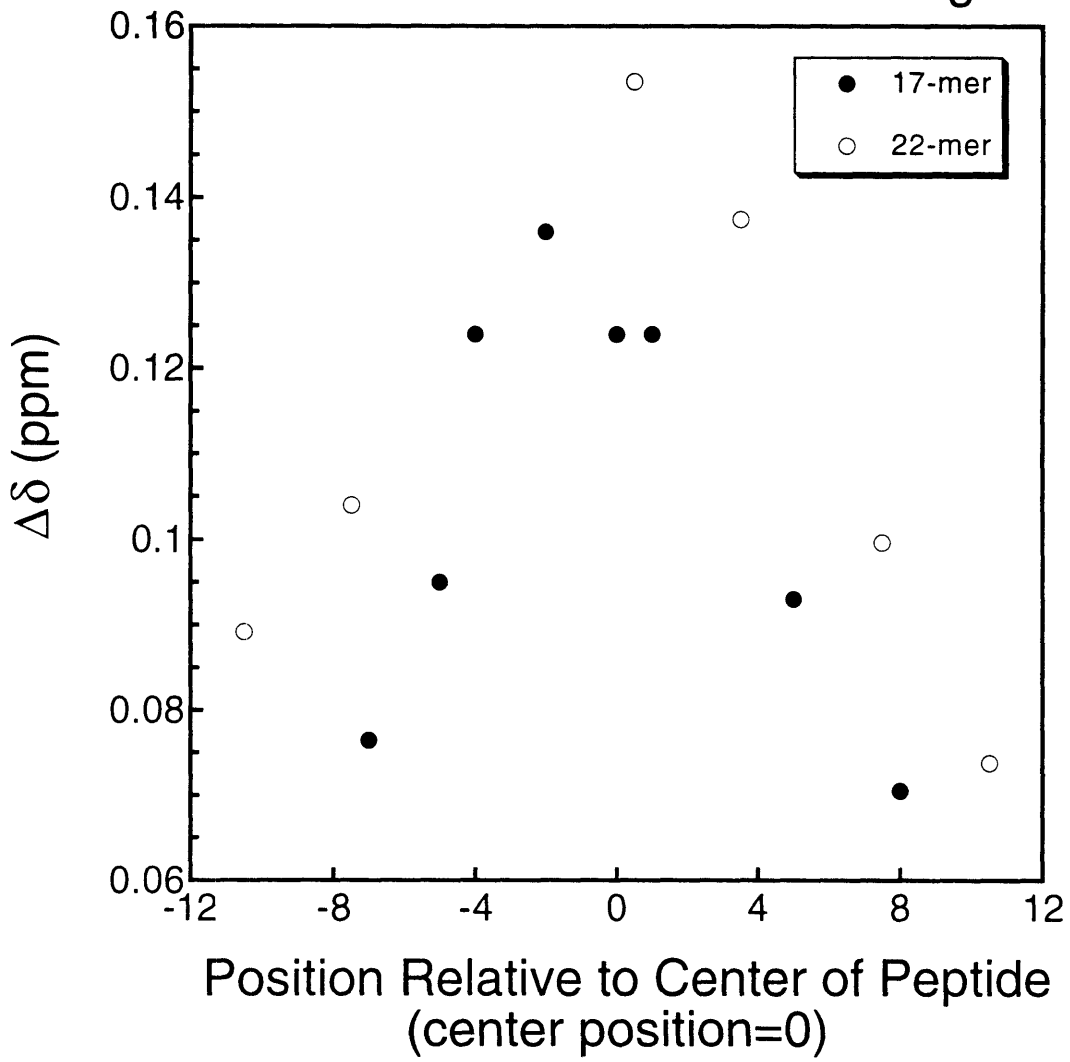

0.16
0.14
0.12
0.1
0.08
0.06
Δδ (ppm)
-12
-8
-4
0
4
8
12
17-mer
22-mer
Position Relative to Center of Peptide
(center position=0)

Fig. 4

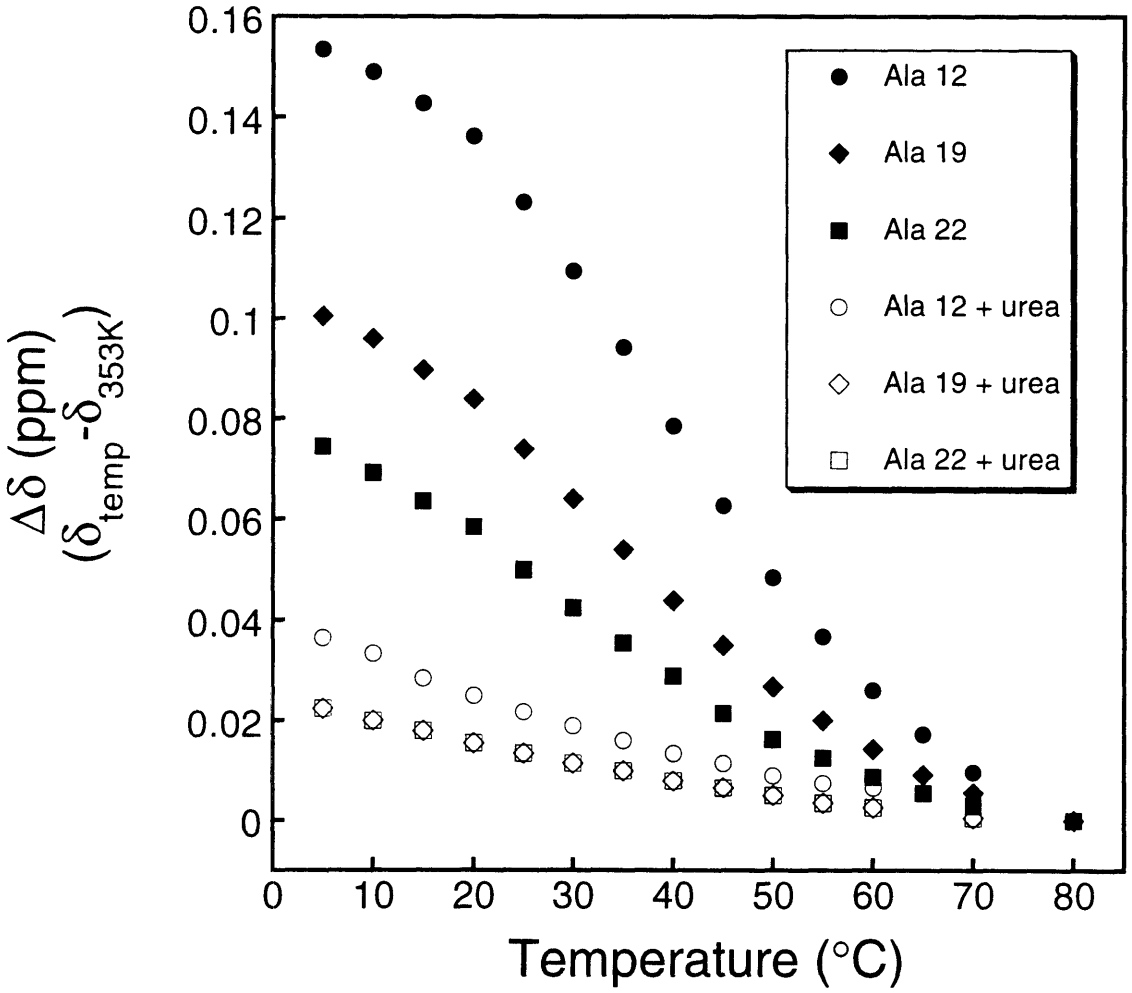
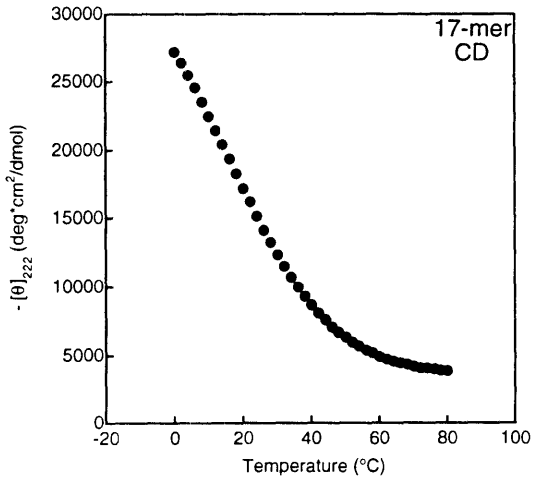
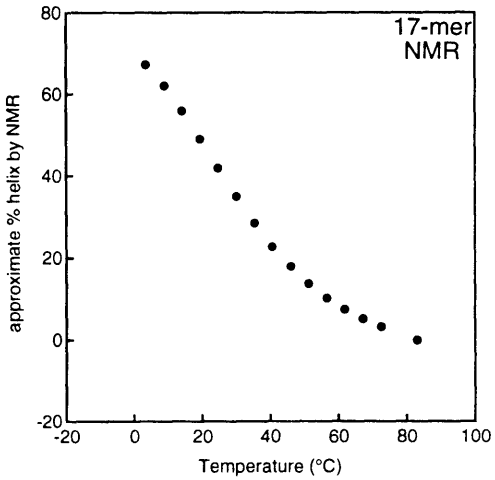
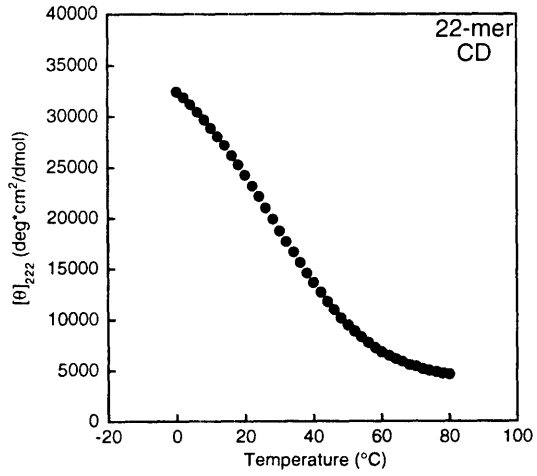

0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
10
20
30
40
50
60
70
80
Temperature (°C)
Δδ (ppm)
(δtemp-δ353K)
Ala 12
Ala 19
Ala 22
Ala 12 + urea
Ala 19 + urea
Ala 22 + urea

Fig. 5


17-mer
CD
- [θ]222 (deg*cm2/dmol)
Temperature (°C)


17-mer
NMR
approximate % helix by NMR
Temperature (°C)


22-mer
CD
[θ]222 (deg*cm2/dmol)
Temperature (°C)

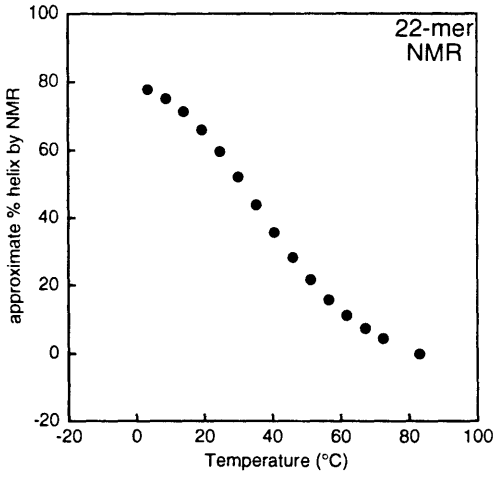

22-mer
NMR
approximate % helix by NMR
Temperature (°C)

Fig. 6

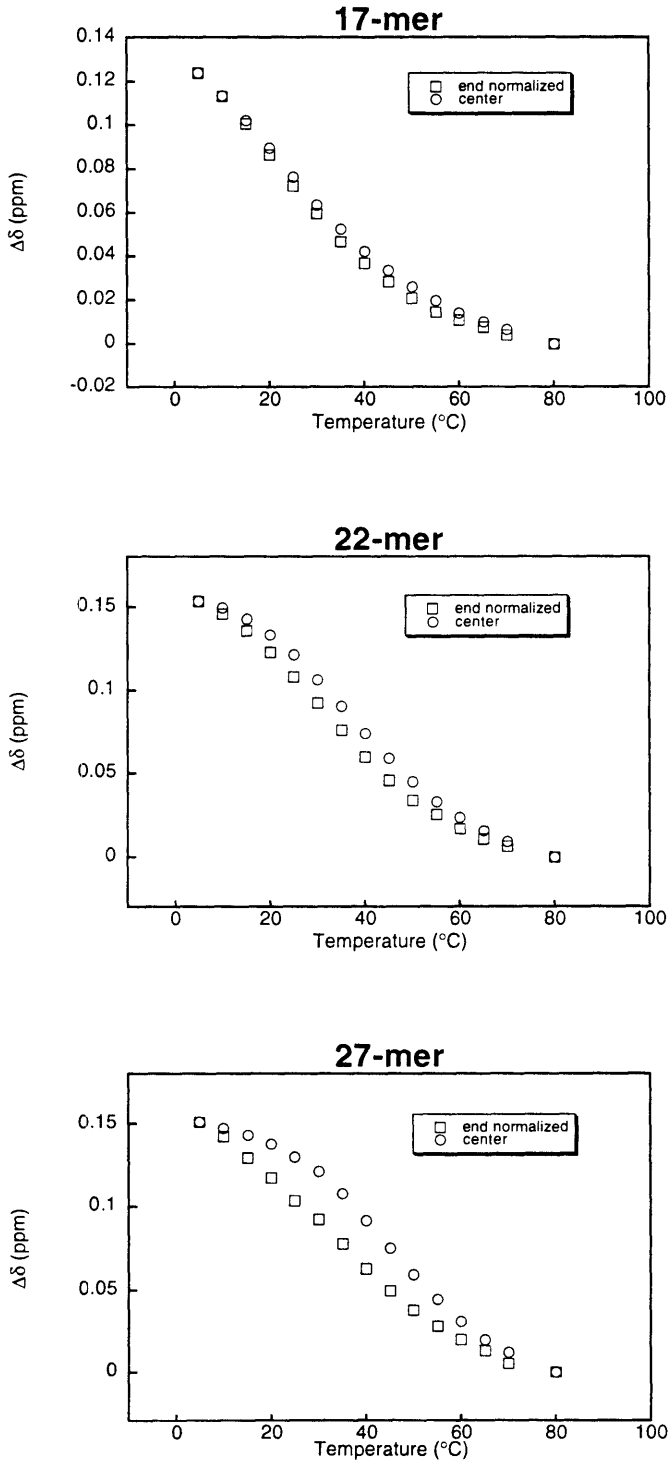
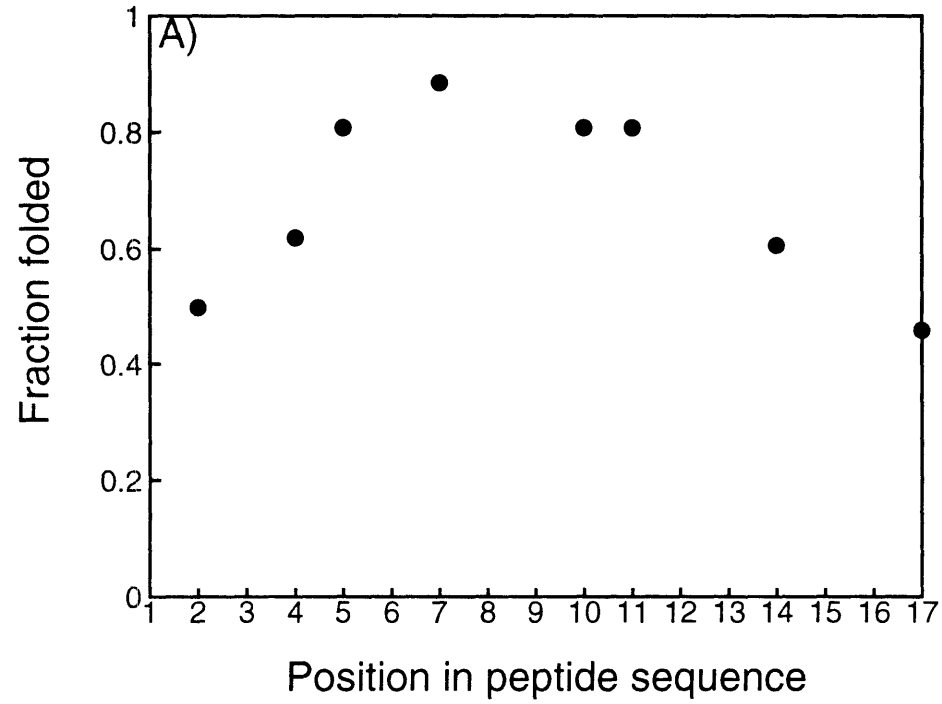

17-mer
end normalized
center
Δδ (ppm)
Temperature (°C)
22-mer
end normalized
center
Δδ (ppm)
Temperature (°C)
27-mer
end normalized
center
Δδ (ppm)
Temperature (°C)

Fig. 7


A)
Fraction folded
0
0.2
0.4
0.6
0.8
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Position in peptide sequence

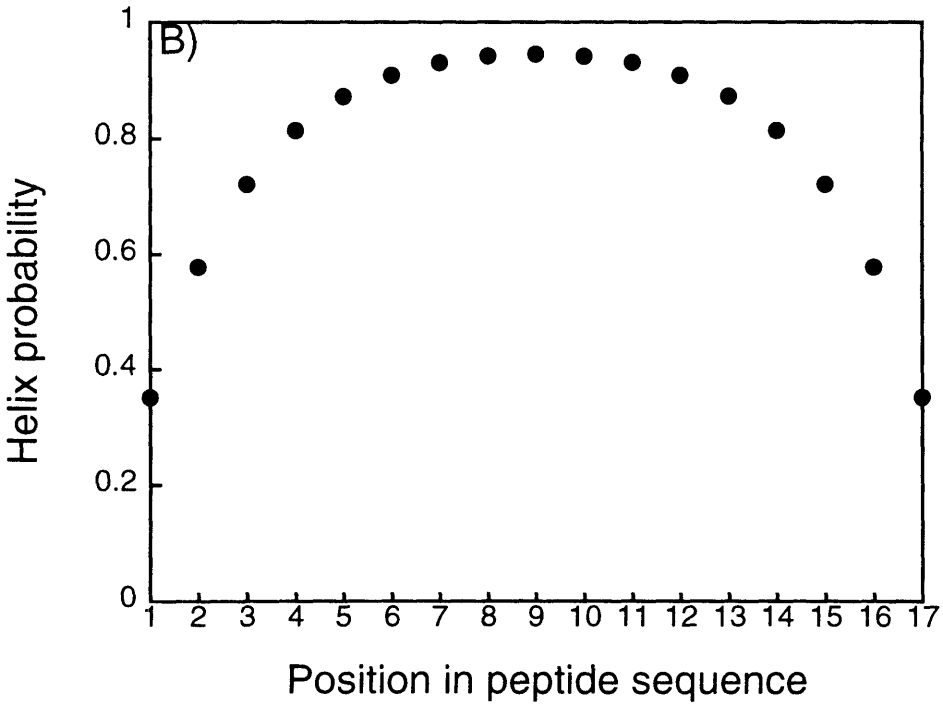
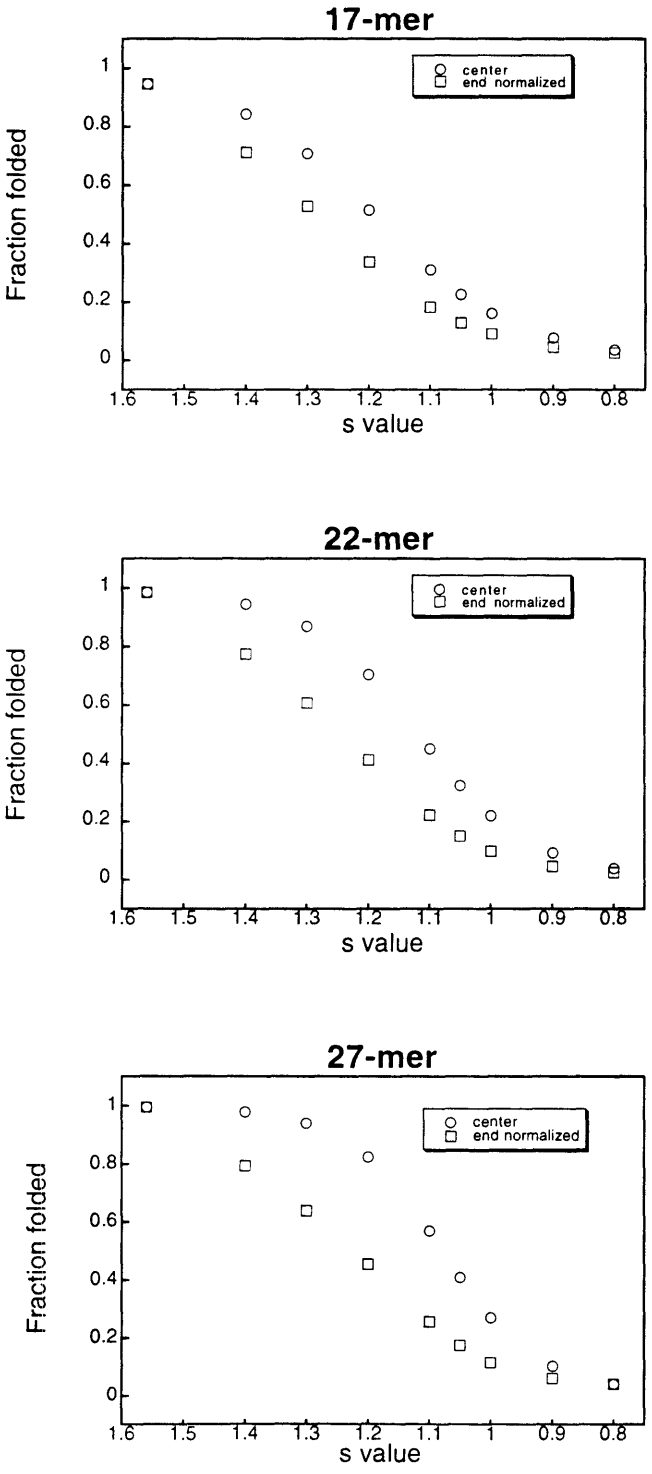

B)
Helix probability
0
0.2
0.4
0.6
0.8
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Position in peptide sequence

Fig. 8


17-mer
center
end normalized
Fraction folded
1
0.8
0.6
0.4
0.2
0
1.6 1.5 1.4 1.3 1.2 1.1 1 0.9 0.8
s value
22-mer
center
end normalized
Fraction folded
1
0.8
0.6
0.4
0.2
0
1.6 1.5 1.4 1.3 1.2 1.1 1 0.9 0.8
s value
27-mer
center
end normalized
Fraction folded
1
0.8
0.6
0.4
0.2
0
1.6 1.5 1.4 1.3 1.2 1.1 1 0.9 0.8
s value

Biographical Note

Brenda Arlene Schulman

Education

1989	The Johns Hopkins University, Baltimore, MD B. A., Biology, General Honors and Biology Honors
1989 - present	Doctoral candidate Massachussetts Institute of Technology Department of Biology Thesis Advisor: Professor Peter S. Kim, Ph.D.

Awards

Beneficial-Hodson Scholar, 1985-1989
Phi Beta Kappa, 1988
National Science Foundation Undergraduate Research Fellow, summer 1988
McElroy Award in Biology, The Johns Hopkins University, 1989
National Science Foundation Predoctoral Fellow, 1989-1992

Publications

1. Haussler, M.R., Terpening, C.M., Jurutka, P.W., Meyer, J., **Schulman, B.A.**, Haussler, C.A., Whitfield, G.K., and B.S. Komm (1988). Vitamin D hormone receptors: structure, regulation and molecular function. Progress in Endocrinology, 763-770.

2. Pihl, T.D., Schicho, R.N., Black, L.K., **Schulman, B.A.**, Maier, R.J. and R.M. Kelly (1990). Hydrogen-sulfur autotrophy in the hyperthermophilic archaebacterium *Pyrodictium brockii*. Biotechnology and Genetic Engineering Reviews **8**, 345-377.

3. Pihl, T.D., Black, L.K., **Schulman, B.A.** and R.J. Maier (1992). Hydrogen-oxidizing electron transport components in the hyperthermophilic archaebacterium *Pyrodictium brockii*. Journal of Bacteriology **174**, 137-143.

4. **Schulman, B.A.** and P.S. Kim (1994). Hydrogen exchange in BPTI variants that do not share a common disulfide bond. Protein Science **3**, 2226-2232.

5. Peng, Z-y., Wu, L.C., **Schulman, B.A.**, and P.S. Kim (1995). Does the molten globule have a native-like tertiary fold? Philosophical Transactions of the Royal Society B **348**, 43-47.

6. **Schulman, B.A.**, Redfield, C., Peng, Z-y., Dobson, C.M. and P.S. Kim (1995). Different subdomains are most protected from hydrogen exchange in the molten globule and native states of α-lactalbumin. Journal of Molecular Biology **253**, 651-657.

7. Wu, L.C., **Schulman, B.A.**, Peng, Z-y., and P.S. Kim (1996). Disulfide determinants of calcium-induced packing in α-lactalbumin, Biochemistry, in press.

8. **Schulman, B.A.** and P.S. Kim. Non-cooperative protein folding, Science, submitted.

Personal Born June 18, 1967, Tucson, AZ
U. S. Citizen