REVIEW

Autonomous subdomains in protein folding

LAWREN C. WU,¹ RITA GRANDORI, AND JANNETTE CAREY

Chemistry Department, Princeton University, Princeton, New Jersey 08544-1009

(RECEIVED October 7, 1993; ACCEPTED October 29, 1993)

Abstract

Proteolytic dissection of native *trp* repressor and horse heart cytochrome *c* has been used to infer some of the steps in the folding pathways of the intact proteins. For both proteins, small fragments are capable of undergoing spontaneous noncovalent association to form subdomains with native-like secondary and/or tertiary structural features, suggesting that dissection/reassembly may be a general method to gain insight into the structures of folding intermediates. The importance of this approach is its simplicity and potential applicability to studying the folding pathways of a wide range of proteins. The proteases report on the structure and dynamics of the native state, circumventing the need for prior knowledge of the structures of folding intermediates. The observation that small fragments of proteins can associate noncovalently suggests that protein folding can be viewed as an intramolecular “recognition” process. The results imply that substantial information about protein structure and folding is encoded at the level of subdomains, and that chain connectivity has only a minor role in determining the fold.

Keywords: proteolysis; recognition; reconstitution; self-assembly

A major goal of studying protein folding is to identify the structures of intermediates along the folding pathway, in the hope that this knowledge will illuminate the mechanism of protein folding. The folding pathway (Fig. 1) can be viewed as consisting of a series of steps (reviewed by Kim & Baldwin, 1990; Dyson & Wright, 1993) in which the unstructured polypeptide chain first forms local, often regular structures, which assemble through increasingly nonlocal interactions into the native tertiary structure. Efforts to describe the structures of intermediates have met with widely varied success along this pathway, for rather different reasons. The earliest steps can occur too fast to be observed directly by experiment. Nevertheless, the sequence-dependent aspects of these early steps are rather well understood, especially in the case of helices (Shoemaker et al., 1987; Marqusee et al., 1989) and turns (Dyson et al., 1988), because synthetic peptides have been useful as models for these strictly local interactions. In contrast, the late steps in the pathway are slow enough to be observed directly, and they have been characterized in detail for a number of proteins (Kim & Baldwin, 1990). These late steps are usually minor rearrangements that occur when the protein is nearly fully folded, so a large number of structural probes is available that can report on these steps.

The middle steps, in which local structured regions must coalesce, are not generally well described. These steps involve rapid

and cooperative folding of partially ordered chains, which precludes simple direct observation and limits the number of structural probes available to monitor the process. The only method that has been successful in identifying the structures of these intermediates has been hydrogen exchange pulse-labeling of backbone amides during the folding reaction, followed by NMR detection in the native state (Roder et al., 1988; Udgaonkar & Baldwin, 1988). Application of this method, however, is restricted to a narrow set of proteins that have complete NMR assignments and favorable kinetics for both the folding and exchange reactions. Because few proteins meet these criteria, new and more generally applicable methods of probing these steps are desirable.

We showed recently that a simple biochemical approach may yield information on the order of folding steps and the structures of presumptive intermediates for the *trp* repressor of *Escherichia coli* (Tasayco & Carey, 1992). This protein has an unusual topologically constrained dimer architecture (Schevitz et al., 1985), which makes its folding pathway of interest while at the same time introducing significant complication to studies of its folding pathway. Gittelman and Matthews (1990) have shown that several kinetic phases can be detected by absorbance, fluorescence, or CD, and they have established the molecularity of each phase from its concentration dependence. Yet, as for so many other proteins, complete NMR assignments for the repressor (Arrowsmith et al., 1990) and detailed spectroscopic kinetic analysis do not yet permit a structural interpretation of its folding pathway.

Mild proteolysis of the native repressor dimer with chymotrypsin yielded a very limited number of fragments (2 per sub-

Reprint requests to: Jannette Carey, Chemistry Department, Princeton University, Princeton, NJ 08544-1009; e-mail: carey@chemvax.princeton.edu.

¹ Present address: Whitehead Institute, 9 Cambridge Center, Cambridge, Massachusetts 02142.

Fig. 1. Schematic illustration of the structural pathway of protein folding (for recent review, see Kim & Baldwin, 1990). This picture of the pathway is used merely as a conceptual guide for the present discussion and is not meant to imply primacy of secondary structure formation over global collapse, nor to favor kinetic over thermodynamic control of the folding process. Modified, with permission, from Richardson (1981).

unit; Carey, 1989). These fragments could be denatured, purified away from each other, renatured, and analyzed by CD and NMR to determine their secondary and tertiary structures alone and in combination (Tasayco & Carey, 1992). The results showed that these 4 fragments refold spontaneously in an ordered series of steps to yield native-like dimeric repressor. Subsequent experiments showed that synthetic peptides as short as 13–35 residues could participate in some assembly steps (Tasayco & Carey, 1992, and unpubl. results). We postulated that the self-assembly reactions mimic some of the folding steps of the intact protein. On this basis, we proposed a detailed structural model that makes testable kinetic predictions for the folding of the intact repressor dimer. It is now necessary to determine the folding kinetics of these fragments and their combinations, and thus, we hope, to equate assembly intermediates with folding intermediates.

These results appear to be significant for several reasons. First, the approach makes use of extraordinarily simple biochemical methods that can be used on any protein. For the all-helical *trp* repressor, we used CD to follow the time course of preparative chymotryptic cleavage and analyzed samples from the CD cuvette during digestion by using low molecular weight denaturing polyacrylamide gel electrophoresis (Carey, 1989). It is noteworthy that the repressor is more than 50 mol percent of hydrophobic residues that could provide suitable target sites for chymotryptic cleavage, yet in a period of over 6 h only 2 frag-

ments are generated, containing nearly the entire protein sequence. This result shows that protease accessibility is severely restricted by the structure and dynamics of the native state. Thus, protease digestion has some advantages over *de novo* design of synthetic peptide mimics of folding intermediates (Oas & Kim, 1988) because there is no need for prior knowledge of intermediate structures.

To evaluate the generality of these features of our approach, we have begun to apply it to additional proteins and have recently demonstrated its applicability to horse heart cytochrome *c* (cytc). We chose this protein because its structural folding pathway had been elucidated using hydrogen exchange and NMR by Roder et al. (1988). They found that amino- and carboxy-terminal helices are protected from exchange with similar kinetics early in the folding process, when the rest of the chain is unprotected, suggesting that the 2 helices may interact early in folding. In the native state of cytc these 2 helices are packed together at an angle of about 90°. However, synthetic peptides corresponding to these regions of the cytc sequence were not helical and showed no evidence of interaction (D. Barrick, S. Marqusee, & R.L. Baldwin, pers. comm.) unless covalently connected by a disulfide bond (Kuroda, 1993). It seemed likely that the heme and/or nearby residues might be essential for the interaction; the only way to obtain fragments containing the heme is by proteolysis of the native protein.

Fragment 1–38H containing residues 1–38 of cytc with the

heme covalently attached via thioether linkages to cysteines 14 and 17 was purified from a trypsin digest of citraconylated cytc according to published methods (Fisher et al., 1973; Taniuchi et al., 1986) and titrated with a synthetic peptide representing the C-terminal helix. CD difference spectra (mixture of the 2 peptides minus sum of individual peptide spectra) are markedly helical through the titration range (Wu et al., 1993); this is a stringent criterion for an increase in helix content and de facto evidence for interaction of the 2 peptides with concomitant increase in helix content. The titration data suggest an association constant in the range of $\sim 10^{-5}$ M and formation of a specific, heme-dependent 1:1 complex. NMR measurements (Wu et al., 1993) show that the T_1 relaxation times of protons in the C-terminal peptide are differentially affected by the heme iron ion, with those closest to the helix-pairing site being most enhanced. These results support the hypothesis (Roder et al., 1988) that the terminal helical segments interact in the early stages of cytc folding. Further work is needed to determine the minimum fragment size capable of reconstitution, the role of the heme, and the nature of tertiary interactions in the complex.

The demonstration that short peptides from 2 rather different proteins can undergo noncovalent association to form subdomains with some native-like structural features has a number of noteworthy implications. First, it suggests that this approach might be generally useful in studying protein folding reactions, including those that are not amenable to NMR exchange studies, although additional examples should be demonstrated. Second, the results indicate that the folding reaction can be viewed as a special case of molecular recognition in which the interacting partners are intramolecular rather than intermolecular. This insight could simplify both conceptualization and experimentation on the protein folding problem. Third, empirical definition of protein domains and subdomains by proteolysis may be a useful adjunct to sequence homology analysis, as it potentially identifies units that may have evolved under independent, local structural and/or functional constraints. Fourth, cytc is one of many proteins that have their termini nearby in the native state (Thornton & Sibanda, 1983). Although this result is expected for random-flight chains (Flory, 1953), the fact that many folded proteins exhibit this feature could indicate that for these proteins, the termini may come together before substantial structure is formed in the remainder of the chain.

Finally, dissection to small fragments that can refold upon noncovalent recombination implies that substantial information about protein structure and folding is encoded at the level of subdomains: portions of proteins that are larger than individual secondary structure elements but smaller than domains (Rose, 1979). The recent results reviewed here suggest that the definition of subdomain structure in proteins can be relaxed to include chain segments that are not covalently connected, i.e., that are not necessarily contiguous in the primary sequence. This consideration may be important for current efforts to develop algorithms (Nussinov & Wolfson, 1991; Zehfus, 1993) that identify subdomains in protein structures. More importantly, the results imply that chain connectivity (i.e., topology) is not fold-determining. Rather, the role of chain connectivity may be to combat entropy during folding by increasing the effective local concentrations of individual secondary structural elements, with the fold being defined by the interactions among these elements. Further support for this concept comes from the observation

that a TIM-barrel protein folds correctly even when circularly permuted (Luger et al., 1989), implying that specific tertiary interactions determine the end product of the folding process. Orengo and Thornton (1993) have also recently pointed out that internal packing, as judged by vector comparison of structural environments, can be quite similar in proteins of different topologies. Thus, protein topology may not be a definitive characteristic of the relationships among proteins.

Acknowledgments

This work was supported by grant GM43558 from NIH. J.C. is a Clare Boothe Luce Preceptor.

References

- Arrowsmith CH, Pachter R, Altman RB, Iyer SB, Jardetzky O. 1990. Sequence-specific ^1H NMR assignments and secondary structure in solution of *Escherichia coli* trp repressor. *Biochemistry* 29:6332–6341.
- Carey J. 1989. Trp repressor arms contribute binding energy without occupying unique locations on DNA. *J Biol Chem* 264:1941–1945.
- Dyson HJ, Rance M, Houghten RA, Lerner RA, Wright PE. 1988. Folding of immunogenic peptide fragments of proteins in water solution I. Sequence requirements for the formation of a reverse turn. *J Mol Biol* 201:161–200.
- Dyson HJ, Wright PE. 1993. Peptide conformation and protein folding. *Curr Opin Struct Biol* 3:60–65.
- Fisher WR, Taniuchi H, Anfinsen CB. 1973. On the role of heme in the formation of the structure of cytochrome c. *J Biol Chem* 248:3188–3195.
- Flory PJ. 1953. *Principles of polymer chemistry*. Ithaca, New York: Cornell University Press.
- Gittelman MS, Matthews CR. 1990. Folding and stability of trp aporepressor from *Escherichia coli*. *Biochemistry* 29:7011–7020.
- Kim PS, Baldwin RL. 1990. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* 59:631–660.
- Kuroda Y. 1993. Residual helical structure in the C-terminal fragment of cytochrome c. *Biochemistry* 32:1219–1224.
- Luger K, Hommel U, Herold M, Hofsteenge J, Kirschner K. 1989. Correct folding of circularly permuted variants of a $\beta\alpha$ barrel enzyme in vivo. *Science* 243:206–210.
- Marqusee S, Robbins VH, Baldwin RL. 1989. Unusually stable helix formation in short alanine-based peptides. *Proc Natl Acad Sci USA* 86:5286–5290.
- Nussinov R, Wolfson HJ. 1991. Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques. *Proc Natl Acad Sci USA* 88:10495–10499.
- Oas TG, Kim PS. 1988. A peptide model of a protein folding intermediate. *Nature* 336:42–48.
- Orengo CA, Thornton JM. 1993. Alpha plus beta folds revisited: Some favored motifs. *Structure* 1:105–120.
- Richardson JS. 1981. Protein anatomy. *Adv Protein Chem* 34:167–339.
- Roder H, Elöve GA, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335:700–704.
- Rose GD. 1979. Hierarchic organization of domains in globular proteins. *J Mol Biol* 134:447–470.
- Schevitz RW, Otwinowski Z, Joachimiak A, Lawson CL, Sigler PB. 1985. The three-dimensional structure of trp repressor. *Nature* 317:782–786.
- Shoemaker KR, Kim PS, York E, Stewart JM, Baldwin RL. 1987. Tests of the helix dipole model for stabilization of α -helices. *Nature* 326:563–567.
- Taniuchi H, Parr GR, Juillerat MA. 1986. Complementation in folding and fragment exchange. *Methods Enzymol* 131:185–217.
- Tasayco ML, Carey J. 1992. Ordered self-assembly of polypeptide fragments to form native-like dimeric trp repressor. *Science* 255:594–597.
- Thornton JM, Sibanda BL. 1983. Amino and carboxy-terminal regions in globular proteins. *J Mol Biol* 167:443–460.
- Udgaonkar JB, Baldwin RL. 1988. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335:694–699.
- Wu LC, Laub PB, Elöve GA, Carey J, Roder H. 1993. A noncovalent peptide complex as a model for an early folding intermediate of cytochrome c. *Biochemistry* 32:10271–10276.