Cooperative Folding Units of *Escherichia coli* Tryptophan Repressor

A. Wallqvist,* T. A. Lavoie,# J. A. Chanatry,# D. G. Covell,* and J. Carey#
#Department of Chemistry, Princeton University, Princeton, New Jersey 08544-1009, and *Frederick Cancer Research and Development Center, National Cancer Institute, Science Applications International Corporation, Frederick, Maryland 21702 USA

ABSTRACT A previously published computational procedure was used to identify cooperative folding units within tryptophan repressor. The theoretical results predict the existence of distinct stable substructures in the protein chain for the monomer and the dimer. The predictions were compared with experimental data on structure and folding of the repressor and its proteolytic fragments and show excellent agreement for the dimeric form of the protein. The results suggest that the monomer, the structure of which is currently unknown, is likely to have a structure different from the one it has within the context of the highly intertwined dimer. Application of this method to the repressor monomer represents an extension of the computations into the realm of evaluating hypothetical structures such as those produced by threading.

INTRODUCTION

Information encoded in a protein's sequence directs its folding, yet the relationship between sequence and folding is complicated and incompletely understood at the molecular level (for a review see Dill and Chan, 1997). However, many proteins incorporate cooperative folding units into their three-dimensional structures. A cooperative folding unit may be defined operationally by its amide-hydrogen exchange behavior. A tertiary substructure of a protein that is resistant to amide-hydrogen exchange unless the protein unfolds globally was originally described as the slow-exchange core of a protein and was proposed to correspond to the so-called folding core, an early structured intermediate on the folding pathway (Woodward, 1993). In fact, amide-proton exchange from the native state identifies the same folding core as do quenched-flow kinetic studies for cytochrome *c* (Roder et al., 1988; Bai et al., 1995) and for ribonuclease H (Chamberlain et al., 1996; Raschke and Marqusee, 1997), indicating a correspondence between cooperative folding units of the native protein and stable substructures that form early in folding. In several other cases, amide-hydrogen exchange from the native state identifies additional substructures that unfold cooperatively (e.g., Hiller et al., 1997; Fuentes and Wand, 1998), suggesting that partially folded, native-like substructures can exist independently of the folding of the entire chain.

The existence of stable substructures and their correspondence with folding units have motivated a number of theoretical and experimental approaches to the identification of such units in the native structures of proteins. Although the identification of a stable substructure does not necessarily imply that it is part of the folding pathway, early theoretical attempts were made to delineate protein folding pathways through interactions of various substructures (Miyazawa and Jernigan, 1982; Moult and Unger, 1991; Chelvanayagam et al., 1992). Recent attempts to identify native-like substructures have focused on hydrophobicity, solvent exposure, and packing criteria (Hilser and Freire, 1996; Tsai and Nussinov, 1997). In what has been termed the new view of protein folding (Bryngelson et al., 1995), a possible role for these stable substructures is to function as nucleation sites (Panchenko et al., 1996; Dill and Chan, 1997).

We and others have recently developed computational tools for identifying stable substructures of a protein (Hilser and Freire, 1996; Wallqvist et al., 1997). These approaches are based on the known structure of the protein in the native state, derived from either x-ray diffraction or NMR data, and use empirical free-energy scoring functions to evaluate an ensemble of hypothetical unfolded states. In our approach, the scoring function defines the probability, for each residue, of being in the native or unfolded state, and cooperative folding units are delineated by considering these probabilities for each amino acid in the protein: residues that have similar probabilities for the native configuration and are in proximity in the native three-dimensional structure are considered a cooperative folding unit. The application of this method to a number of globular proteins (Wallqvist et al., 1997) identified cooperative folding units that showed excellent correspondence to the slow-exchange cores defined experimentally. The fact that energy calculations based on native-state models can identify substructures similar to those observed by hydrogen exchange implies that information about substructure definition is encoded in the native fold.

In the present paper we extend this method by using it to predict the independent folding units within the tryptophan repressor (TrpR) in both dimeric and monomeric states. Distinct stable substructures are predicted, depending upon whether the monomeric chain is evaluated in isolation or in the context of the extensively intertwined homodimer. Because the monomer structure is unknown, these predictions serve as a test of the hypothetical structure of the folded

Received for publication 24 March 1999 and in final form 10 June 1999.

Address reprint requests to Dr. Jannette Carey, Department of Chemistry, Princeton University, Princeton, NJ 08544-1009. Tel.: 609-258-1631; Fax: 609-258-6746; carey@chemvax.princeton.edu.

Dr. Wallqvist's present address is Chemistry Department, Rutgers University, Piscataway, NJ 08854-8087.

0006-3495/99/09/1619/08 $2.00

monomer used in the calculations. In the case of dimeric TrpR, the independent folding units identified computationally show good correspondence with the pattern of experimental evidence on local structure and stability within TrpR and its fragments. In particular, proteolytic dissection of the TrpR dimer appears to closely reflect the identified stable substructures. For monomeric TrpR, limited availability of experimental data permits only partial evaluation of predicted independent folding units, but several lines of evidence suggest a monomer structure quite different from its structure within the context of the native TrpR dimer. Because of the success of proteolysis in evaluating the predicted substructures of the TrpR dimer, we suggest that proteolysis of a monomeric mutant form of TrpR may be a useful experimental tool for identifying the stable substructures present in this protein. The results obtained for the TrpR monomer and dimer, taken together, reinforce the utility of the computational method for identifying cooperative units within proteins of known structure, and they suggest a combination of experimental and computational approaches that may be useful in evaluating hypothetical structures.

MATERIALS AND METHODS

Theoretical domain identification

All theoretical calculations were performed according to procedures reported previously (Wallqvist et al., 1997). Initial calculations indicated that with a segment window size of 10 residues, the unfolding penalties converged to values that did not vary appreciably compared to results with smaller window sizes. Although the resolution between unfolding penalties of two amino acids close in sequence was improved in calculations employing segment window sizes of six and three, such calculations on the *trp* repressor dimer would have required excessive CPU time. The calculation of unfolding penalties of *trp* repressor and its fragments employed the crystal coordinate set denoted 3wrp (Lawson and Sigler, 1988) obtained from the Protein Data Bank. The native dimeric *trp* aporepressor was generated according to instructions in the PDB file. Calculations for the monomer unit utilized $10^{6.2}$ configurations in the ensemble of denatured states for window size = 6 and $10^{4.4}$ for window size = 10. The dimer calculations used $10^{7.3}$ configurations and a window size = 10.

Production of TrpR fragments

A large-scale (2L) culture of pTAL/CY15071(λDE3) expressing TrpR fragment 52–108 was grown to an OD_{600} of ~0.5 at 37°C and induced with isopropylthiogalactoside for 4 h. Cells were harvested by centrifugation. The wet cell paste was resuspended in 10 mM Tris-HCl (pH 7.6) (5 ml/g wet weight cells); EDTA was added to a final concentration of 1 mM; phenylmethylsulfonyl fluoride was added to a final concentration of ~0.1 mM to inhibit serine protease activity. Cell membranes were ruptured by passing the resuspended cell paste through a French pressure cell three times. The mixture was centrifuged at 10K rpm for 30 min at 4°C. Streptomycin sulfate (20% w/v) was added to precipitate DNA. The mixture was centrifuged at 10K rpm for 30 min at 4°C; solid ammonium sulfate was added to the supernatant to a final concentration of 40% (w/v), and the mixture was stirred at 4°C for 2 h. The resulting protein pellet was resuspended and dialyzed extensively against P11 buffer (10 mM sodium phosphate, pH 7.6; 100 mM NaCl; 0.1 mM EDTA) at 4°C. The resulting solution was applied to a P-11 phosphocellulose column. The column was washed with ~100 ml of P-11 buffer, and the fragment was then eluted with 450 ml of a linear salt gradient from 0.15 M NaCl to 0.8 M NaCl (in P-11 buffer). Fractions containing protein were concentrated using Amicon Centricon-3 or Amicon Centriprep-1 concentrators. The concentrated sample was loaded onto a G-50 Sephadex gel filtration column equilibrated in 1× P-11 buffer eluted with the same buffer. Fragment 72–108 was purified from a chymotryptic digest of trp aporepressor according to published procedures (Carey, 1989; Tasayco and Carey, 1992). Protein concentrations were determined from extinction coefficients measured in 6 M guanidine-HCl according to the method of Gill and von Hippel (1989): 5690 M^{-1} cm^{-1} for fragment 52–108 and 4729 M^{-1} cm^{-1} for fragment 72–108.

Circular dichroism

Experiments were performed on an AVIV 62DS circular dichroism spectrometer equipped with thermoelectric control of cell temperature. Spectra were recorded at 4°C with 0.2-nm step size and 1-s averaging time.

Proteolysis

Reactions contained 20 μM fragment and 1.3 μg/ml chymotrypsin at room temperature in 10 mM sodium phosphate (pH 7.6). Aliquots were removed at timed intervals, stopped by boiling in sodium dodecyl sulfate gel sample buffer, and resolved by electrophoresis through an 18% acrylamide sodium dodecyl sulfate gel. N-terminal sequencing of fragments excised from a blot of a duplicate gel was carried out by the Princeton University Synthesis/Sequencing facility with 10 rounds of automated Edman degradation.

RESULTS AND DISCUSSION

Theoretical results

The calculations employed in this work determine whether a segment of a protein can maintain a native-like conformation and derive favorable interactions within an ensemble of hypothetical, partially denatured states, as previously described in detail by Wallqvist et al. (1997). The results are expressed as free energy per residue in units of kT. The free energies are arbitrarily given a positive sign, indicating that the native-like state is at lower energy and that its disruption requires energy input, the so-called unfolding penalty. Thus peaks in this plot identify regions predicted to participate in native-like substructures.

The *trp* repressor protein is a highly intertwined dimer (Schevitz et al., 1985), in which the tertiary structure is formed almost entirely by intermolecular interactions between the two monomers (Fig. 1). Thus we may expect different results from unfolding penalty calculations on the dimer and on the isolated monomer with the same structure. The topology of this hypothetical monomer, derived by deleting one subunit from the native TrpR dimer, is very similar to the x-ray crystal structures of calmodulin and troponin C (Sundaralingam et al., 1985; Babu et al., 1985; Herzberg and James, 1985). L-Tryptophan (not shown) binds to the dimer in two symmetrical pockets formed by residues from helices C and E (or c and e) of one subunit and the b-c (or B-C) turn of the other subunit. Calculations were performed on both apo- and holo-repressor forms with similar results; the present method is not sufficiently sensi-

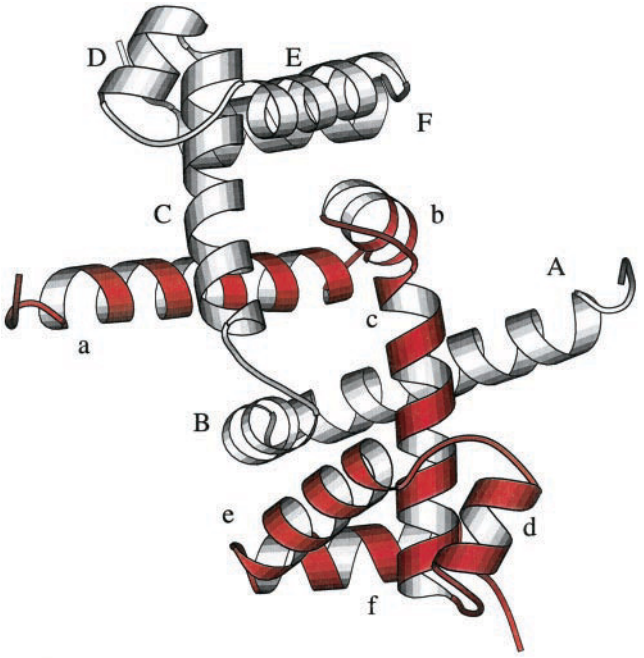

FIGURE 1 TrpR structure. Coordinates of the TrpR aporepressor dimer structure were obtained from PDB file 3wrp containing residues 5–108. The schematic protein structure is rendered with the Molscript graphics program (Kraulis, 1991). One subunit is shown in red to reveal the intertwined nature of the dimer. Helices are marked by uppercase letters in one subunit and by lowercase letters in the other.

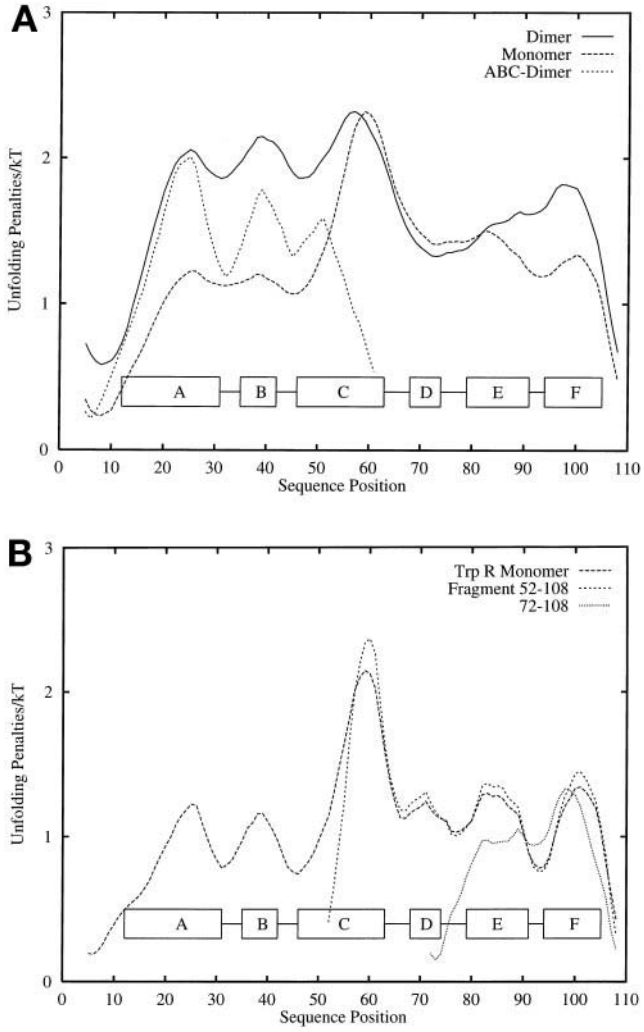

FIGURE 2 Unfolding penalties for *trp* aporepressor and its fragments. Unfolding penalties were calculated as described in the text and are displayed on the *y* axis in units of *kT*. Lettered boxes above the *x* axis represent α-helical segments from the x-ray structure of the TrpR dimer shown in Fig. 1 *A*. (*A*) Comparison of monomer, dimer, and dimer fragment 5–62 (ABC dimer). Results with dimeric chains are shown for one subunit only. The window size for these calculations was 10 residues. (*B*) Comparison of monomer with fragments 52–108 and 72–108. Window size: six residues. The differences in peak shape for the monomer in *A* and *B* are due to the change in window size.

tive to detect differences due to single amino acids or residues.

Fig. 2 *A* shows the results of unfolding penalty calculations for both the dimer and the hypothetical monomer of TrpR. In both cases, helix C scores as the most stable segment of the identified substructures, but its most favored tertiary partners differ, depending on the presence of a second monomer. When a second chain is present, the calculations indicate that the A and B helices from each chain form the most stable substructure, together with helix C. Helix F also shows significant stabilization associated with the predicted dimeric substructure involving helices A-C. In the absence of a second chain, the calculations identify a substructure comprising the C-terminal parts of helix C and nearby regions of helices D and E, with diminished relative importance of helix F. The large differences in unfolding penalty for helices A-B in the monomer and dimer are likely related to the fact that in the dimer, most of the contacts made by helices A, B, and the N-terminal part of helix C are intermolecular. Unlike the dimer case, the unfolding penalty in the monomer for helices A-B is very similar to that for D-E-F, suggesting that these two protein regions may compete with each other for interactions with helix C. The quantitative agreement of unfolding penalties for the helix C-D region in the monomer and dimer suggests that this region may form a stable substructure independently of other chain segments.

In an effort to illuminate the role of helix C, calculations were also carried out for various chain segments. Three examples relevant for comparison with available experimental data are shown in Fig. 2. Calculations on the ABC segment of the dimer (Fig. 2 *A*) show a pattern similar to that found in the intact dimer, except that in the absence of the distal parts of the chain, the unfolding penalty peak for helix C is shifted to residue 51, and the C-terminal segment of helix C is not predicted to be part of a stable substructure. The unfolding penalties for fragment 52–108 (Fig. 2 B) are in quantitative agreement with those calculated for the intact monomer, supporting the suggestion that this chain segment may be structurally independent of the N-terminal parts of the monomer. Taken together, these results suggest that, in the presence of a second chain, helices A, B, and C form an

intermolecular folding unit together with helix F, whereas in the monomer, helix C may instead fold with adjacent C-terminal portions of the chain. The unfolding penalty for helix F is quantitatively similar regardless of the presence of helix C (fragment 72–108, Fig. 2 *B*); similar results are observed for helix A in the dimer in Fig. 2 *A*. For helix F, these results suggest the possibility that this helix may represent an independent folding unit.

Comparison with experiment

The predictions from the present calculations can be evaluated by comparison with available experimental evidence on the structure and stability of TrpR and its fragments. A large body of data by now supports a view of the dimeric TrpR apoprotein with extensive dynamics in the amino-terminal residues and in the helix-turn-helix motif, but with a well-ordered helical structure throughout the rest of the molecule (for recent reviews see Luisi and Sigler, 1990; Lavoie and Carey, 1994)). The flexibility of the helix-turn-helix region in both apo- and holorepressor forms has been studied extensively by NMR as well as x-ray crystallography. Lawson et al. (1988) showed that two crystal forms of the holorepressor grown under identical conditions have distinctly different conformations in the DNA-binding region (helix D–turn–helix E) in that the C-D and the D-E interhelical turns and the second turn of the D helix are shifted in one crystal form compared to the other. This result suggests that the D-E region is quite flexible and can adopt different (though not necessarily isoenergetic; Jin et al., 1999) conformations under the influence of crystal packing forces.

NMR studies also reveal that the DE region of TrpR has unusual dynamics. Although the chemical shift ranges of protons in helices D and E indicate some helical character (Zhao et al., 1993), the majority of the nuclear Overhauser effect (NOE) connectivities expected for helices D and E of aporepressor and for helix D of holorepressor are not observed (Arrowsmith et al., 1991b; Czaplicki et al., 1991; Zhao et al., 1993). Furthermore, the exchange rates of backbone amide protons of helices D and E in apo- and holorepressor are at least two orders of magnitude faster than those in the hydrophobic core of the protein (Arrowsmith et al., 1991a; Czaplicki et al., 1991). Combined with data on ^{15}N relaxation rates (Zheng et al., 1995), the results indicate that aporepressor helices D and E are stable as helices on a nanosecond, but not a millisecond, time scale. A highly flexible helix-turn-helix region in the aporepressor is also consistent with molecular dynamics simulations (Komeiji et al., 1991, 1994; Howard and Kollman, 1992; Guenot and Kollman, 1992). Furthermore, a very large heat capacity change upon binding of L-trp has been interpreted as reflecting the burial of protein surface area due to folding in the helix-turn-helix region to form the ligand-binding pocket (Jin et al., 1993).

Proteolytic sensitivity of the peptide backbone is also consistent with this pattern of segmental flexibility. Despite the relatively low sequence specificity of chymotrypsin and the large number of its substrate residues present in TrpR, this enzyme cleaves wildtype apo- and holorepressor at only two major sites, peptide bonds 7–8 in the flexible N-terminus (Carey, 1989) and 71–72 in helix D (Tsapakos et al., 1985; Carey, 1989), giving rise to fragments 8–71 and 72–108. The same cleavage kinetics and product identities are also observed for both apo- and holorepressor forms of the TrpR mutant L75F (Jin et al., 1999), in which Phe replaces Leu at position 75, the last residue of helix D. The lack of cleavage at position 75 is particularly surprising in the mutant protein because chymotrypsin prefers Phe over Leu as the residue at the site of the scissile bond (Schellenberger et al., 1991). The fact that Leu^{71} within helix D is cleaved, whereas the nearby residue 75 adjacent to the interhelical turn is not cleaved, suggests that the helix itself may be more dynamic than the interhelical turn. Helical segments of proteins are generally protease-resistant (Fontana et al., 1993, 1997).

Isolated fragments 8–71 and 72–108 of wild-type TrpR can reassemble to form a complex with circular dichroism (CD) and NMR spectra closely resembling the native aporepressor dimer (Tasayco and Carey, 1992). The reassembly reaction follows an obligately ordered series of steps that is postulated to reflect some of the steps on the TrpR folding pathway, a suggestion that is supported by recent folding data (Gloss and Matthews, 1997). In the first step, fragment 8–71 undergoes a concentration-dependent increase in helix content according to CD results, and the 1D ^{1}H NMR spectrum of the resulting complex contains a subset of the signature aromatic and upfield methyl resonances of the native dimer, consistent with a nativelike tertiary environment for many residues in the fragment assembly. A shorter fragment encompassing only residues 17–51 (approximately the middle of helix A to the middle of helix C) also displays concentration-dependent helix content according to CD analysis (Tasayco and Carey, unpublished observations).

Fragment 72–108 in isolation is only marginally helical at any concentration, but upon the addition of 8–71 a net increase in helix content is observed in CD mixing experiments, and the NMR signature of the native TrpR dimer is restored, despite the lack of covalent connection between residues 71 and 72 in each subunit (Tasayco and Carey, 1992). Furthermore, a 13-residue peptide corresponding to helix F, although not helical in isolation, can also combine with fragment 8–71 with a net increase in helix content. These results are consistent with the extensive evidence of flexibility of the helix-turn-helix domain, and they strongly support the notion that helices A, B, C, and F form a cooperative folding unit of the TrpR dimer, as also suggested by the present theoretical results. Recent studies on the flexibility and dynamics of RNase S, with a single proteolytic cleavage between peptide bonds 20 and 21, suggest that this complex has dynamics very similar to those of intact RNase A, despite the covalent break (Nadig et al., 1996). It is not yet known if this behavior will be general to fragment complementation systems, but results on TrpR to

date suggest a similar picture. Indeed, the extensive dynamics of the helix-turn-helix region of TrpR demonstrated by recent NMR evidence suggests that the reconstituted complex of TrpR fragments is a better model for the intact native protein than could be appreciated previously by comparison with the crystal structures. Nevertheless, crystallographic evidence also indicates that helices ABCF from each subunit form the dimeric core of TrpR, whereas helices D and E protrude from the dimer surface (Schevitz et al., 1985).

Because of the high stability of the intertwined TrpR dimer, the isolated monomer becomes significantly populated only at subnanomolar concentrations in the absence of denaturants (C. A. Royer, personal communication). In the presence of denaturants, TrpR undergoes a two-state, cooperative unfolding transition between folded dimer and unfolded monomer (Gittelman and Matthews, 1990). Thus, for the TrpR monomer, comparable information on local structures is not available. However, early in the refolding reaction of the TrpR dimer, an intermediate is populated that has the properties of a partially folded monomer (Mann and Matthews, 1993). Recently a TrpR mutant has been engineered with a dimer-disrupting interface (Shao et al., 1997). Monomers of this mutant persist up to micromolar concentrations in native conditions, and spectroscopic data indicate a molten globule-like form with secondary and tertiary structure but lacking a cooperative unfolding transition. The two tryptophan residues of this monomer are in ordered hydrophobic environments, and they contribute to the stabilization of a nonnative tertiary structure for the monomer (Shao and Matthews, 1998). Detailed structural information is not yet available for the monomer, however.

Some experimental data are available for two fragments of TrpR, encompassing residues 52–108 and 72–108, that may be relevant for evaluating the predicted monomer substructures. Fragment 52–108 was designed to incorporate all of the residues important for binding of L-trp and DNA (Lavoie, 1996). Inspection of the x-ray crystal structures of TrpR suggests that the C-terminal half of helix C must be present for that purpose, and because residue 52 in helix C is Gly, this residue was chosen as the N-terminus of the designed fragment. In the crystal structure of TrpR there is a bend in helix C at the central residue, Gly^{52} (Schevitz et al., 1985). Interestingly, the peak of the calculated unfolding penalty in Fig. 2 *A* for the ABC dimer occurs at the adjacent residue Leu^{51}, suggesting the possibility that this region of the protein marks a naturally occurring boundary between structural units.

Both fragments 52–108 and 72–108 are monomeric over wide concentration ranges, as judged by size exclusion chromatography and CD for 52–108 (Lavoie, 1996) and by CD, NMR, and gel filtration for 72–108 (Tasayco and Carey, 1992). Both fragments exhibit only slight helicity in aqueous buffer, with 19% of residues in helical conformation for 52–108 and 16% for 72–108 (Fig. 3 *A*). TFE titration indicates that fragment 52–108 has higher helix content than 72–108 at all trifluoroethanol concentrations at which solubility permits direct comparison (Fig. 3 *B*). Fragment 52–108 shows some resistance to chymotryptic proteolysis. Over a ~90-min time course (Fig. 3 *C*), digestion produced three slightly shorter fragments with N-termini at residue 52 (Lavoie, 1996), demonstrating that all shortening occurs from the C-terminal end of the fragment. The electrophoretic mobility of the fragments indicates removal of up to 20 residues from the C-terminus. These results are consistent with the clustering of residues in the predicted folding unit and with the finding that fragment 72–108 shows no protection from proteolysis (Carey, 1989; Jin et al., 1999; and data not shown). The chymotryptic resistance at the N-terminus of fragment 52–108 is not trivially due to the absence of suitable cleavage sites in the segment 52–71, as numerous suitable substrate residues (Schellenberger et al., 1991) are located throughout the segment.

Titration of fragment 52–108 with 8–71, analyzed by fluorescence and CD signal changes (Lavoie, 1996), yielded a binding constant of at least 10^7 M^{-1}, indicating that the two fragments can form a stable complex despite the redundant chain segment between residues 52 and 70. Complementation of fragments containing redundancies is not unexpected and was demonstrated originally for fragments of cytochrome *c* (Taniuchi et al., 1986). Numerous attempts to concentrate fragment 52–108 to the levels required for NMR analysis (i.e., 1–2 mM) were unsuccessful because of precipitation and aggregation of the protein under a wide range of conditions unless the TrpR operator target DNA was present, implying interaction with the DNA (Lavoie, 1996). However, even at low temperature (4°C), the NMR spectrum revealed no shifts or broadening in the characteristic DNA imino or methyl resonances, and the protein resonances were very sharp and were not well dispersed, indicating a highly dynamic interaction. No other evidence could be found that indicated that fragment 52–108 binds to either DNA or L-trp, or is independently folded.

Relation of theory to experiment

The agreement between unfolding penalty predictions and the large body of experimental data for the TrpR dimer reinforces the utility of the theoretical calculations in identifying stable substructures within native proteins of known structure, as has been already shown for a number of other native proteins (Wallqvist et al., 1997). The close agreement between the predicted substructures of the TrpR dimer and the results of hydrogen exchange, proteolytic dissection, and fragment reassembly for that protein suggests that proteolysis may also idenitfy tertiary substructures or cooperative folding units of the TrpR monomer, if this species could be isolated. Proteolysis using enzymes of low intrinsic sequence specificity, coupled with careful analysis of fragmentation patterns and kinetics, can often identify labile chain segments between structured regions (Fontana et al., 1993, 1997; Hubbard, 1998; Carey, 1999) and may identify autonomous subdomains comprising associating but nonco-

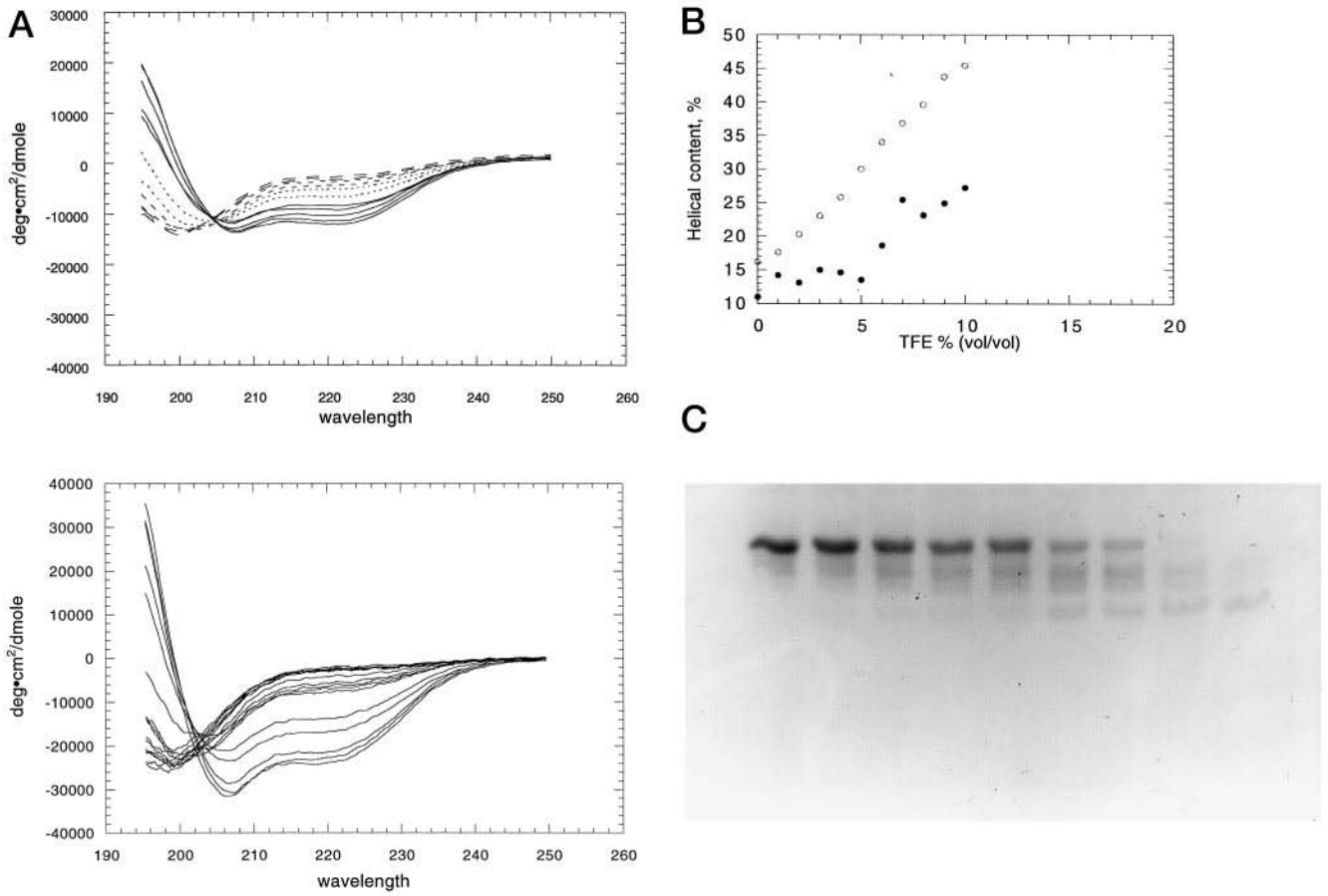

FIGURE 3 Structure in monomeric TrpR fragments. (*A*) CD spectra of 20 μM fragments in 20 mM sodium phosphate, pH 7.6, 0.1 mM EDTA (*top trace*) or with TFE added at 0–10% (v/v) in 1% increments (*top to bottom traces*, respectively). *Top*: Fragment 52–108. *Bottom*: Fragment 72–108. (*B*) Effect of TFE on helix content for fragment 52–108 (○) and 72–108 (●). The percentage helicity was calculated by the method of Chen et al. (1974). (*C*) Proteolysis of fragment 52–108. Lanes from left to right represent digestion for 0, 1, 2, 4, 10, 20, 30, 45, and 90 min under the reaction conditions given in Materials and Methods. Bands labeled A, B, and C were excised from a blot of a duplicate gel and subjected to N-terminal sequencing by Edman degradation. The molecular mass of the starting material is ~6100 Da.

valent chain segments (Wu et al., 1994; Peng and Wu, 1999). By analyzing the structures of proteolytic fragments and assemblies with low-resolution spectroscopic methods, such as CD, UV absorbance, or fluorescence, which report on secondary and tertiary structure content but which measure only the average signals in a population of molecules, one can obtain a view of the structured parts of the protein at a resolution equivalent to that at which the protein is cleaved, in effect increasing the molecular resolution of those methods.

The present application of unfolding penalty calculations to monomeric TrpR represents an attempt to use these methods to predict substructures, using a hypothetical structure as a template. If successful, this application could be extended to evaluate other hypothetical protein structures, such as those produced by threading algorithms or structure homology models. Just as in the case of the TrpR dimer, the calculations can be evaluated in comparison with experimental data on protein fragments. In the present case, experimental evidence is limited to existing fragments 52–108 and 72–108 of TrpR, permitting only partial evaluation of the hypothetical monomer structure. The available data offer some weak correlations with the unfolding penalty calculations for this monomer but do not strongly or uniquely support the hypothetical structure. Moreover, the calculations based on this structure do not predict additional fragments that could be designed to provide unique tests of the model structure. Several additional lines of evidence argue against this structure as a good model for the TrpR monomer. Although the topology of the model structure is shared by calmodulin and troponin C, the latter proteins differ in having highly polar surfaces, unlike the large exposed hydrophobic surface areas of the TrpR monomer, which become buried only upon dimer formation. In addition, calmodulin has a more globular fold in solution according to NMR data (Kuboniwa et al., 1995; Zhang et al., 1995). Furthermore, recent evidence on a monomeric mutant of TrpR (Shao and Matthews, 1998) suggests a more collapsed, although probably molten-globule-like, structure for the mutant monomer.

Monomeric TrpR typifies a large number of proteins for which direct structural analysis is elusive for one reason or

another. Although the recently described TrpR mutant is monomeric at concentrations up to tens of mM, it aggregates at the higher concentrations that would be required for structural determination or hydrogen exchange analysis (Shao and Matthews, 1997). The ability of proteolysis to identify native-like substructures of dimeric TrpR suggests that proteolytic dissection of the monomer, combined with characterization of the resulting protein fragments by low-resolution biophysical methods, might provide an objective means of evaluating the monomer structure. In conjunction with the unfolding penalty calculations employed here, this kind of data might permit the rejection of an incorrect model structure and should give clues to the identity of correct substructures. Proteolysis appears to be one of the very few means of studying the residual structure of monomeric TrpR at a resolution higher than could be achieved by the spectroscopic methods that have already been applied to the intact monomeric mutant (Shao and Matthews, 1997, 1998). The recent availability of this mutant protein provides a new avenue for the characterization of TrpR monomer structure by proteolytic dissection. This system may therefore provide a favorable case in which to test the suggestion that proteolysis can be a useful adjunct to unfolding penalty calculations for evaluating model structures.

Determining the fold of a protein based solely on the primary structure is a daunting theoretical task. Identification of a folding core, as in the present approach, can help researchers to sketch out essential parts of the native fold; indeed, the predicted existence of a folding core could be one indication of a foldable protein. This analysis requires a fold template upon which the sequence can be threaded. In the current work we have focused on a template derived from a known dimeric form of the protein. In the dimer structure the calculated unfolding penalties correlate very well with stable substructures of the protein identified via hydrogen exchange, NMR dynamics analysis, and proteolytic digestion. The monomeric form of the protein poses a greater challenge both theoretically and experimentally, because information on the possible folds (if any) of this structure is missing and is not readily accessible experimentally. Thus we based the theoretical calculations for the monomer form on the structure it displays within the context of the dimer. This ensures that we use a structural configuration that is represented by the knowledge-based potentials, even though it still is a hypothetical structure. The properties of the monomer and its fragments were characterized experimentally and compared with the theoretical calculations. Matching experimental data with the predicted properties in the dimer case showed that the theoretical calculations can give a plausible picture of folding cores. In the monomer case no such matchup was possible among currently available data, indicating that the hypothetical structure for the monomer is incorrect and suggesting a route for acquiring additional relevant experimental data. The limiting factor in applying the theoretical model is the requirement for input of coordinates, either from NMR or x-ray measurements or from a plausible hypothetical structure. Here we have employed a very conservative approach in not generating coordinate sets from the potential, instead relying only on the experimentally known coordinate set. Additional evidence in the form of experimental data, such as those presented here, will be required to validate the theoretical predictions.

The authors thank the staff of the Biomedical Supercomputing Center, Frederick Cancer Research and Development Center, for assistance and access to computer facilities.

This work was supported by National Institutes of Health grant GM 43558 to JC, and by the National Institutes of Health Intramural Targeted Antiviral AIDS Program and the National Cancer Institute, under contract with the Science Applications International Corporation (AW and DGC).

The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

Arrowsmith, C. H., J. Czaplicki, S. B. Iyer, and O. Jardetzky. 1991b. Unusual dynamical features of the *trp* repressor from *E. coli*. *J. Am. Chem. Soc.* 113:4020–4022.

Arrowsmith, C., R. Pachter, R. Altman, and O. Jardetzky. 1991a. The solution structures of *Escherichia coli trp* repressor and *trp* aporepressor at an intermediate resolution. *Eur. J. Biochem.* 202:53–66.

Babu, Y. S., J. S. Sack, T. J. Greenhough, C. E. Bugg, A. R. Means, and W. J. Cook. 1985. Three-dimensional structure of calmodulin. *Nature.* 315:37–40.

Bai, Y., T. R. Sosnick, L. Mayne, and S. W. Englander. 1995. Protein folding intermediates: native-state hydrogen exchange. *Science*. 269: 192–197.

Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins Struct. Funct. Genet.* 21:167–195.

Carey, J. 1989. Trp repressor arms contribute binding energy without occupying unique locations on DNA. *J. Biol. Chem.* 264:1941–1945.

Carey, J. 1999. Identification of protein domains and subdomains by systematic proteolysis. *Methods Enzymol.* (in press).

Chamberlain, A. K., T. M. Handel, and S. Marqusee. 1996. Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nature Struct. Biol.* 3:782–787.

Chelvanayagam, G., Z. Reich, R. Bringas, and P. Argos. 1992. Prediction of protein folding pathways. *J. Mol. Biol.* 227:901–916.

Chen, Y.-H., J. T. Yang, and K. H. Chau. 1974. Determination of helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry.* 13:3350–3359.

Czaplicki, J., C. Arrowsmith, and O. Jardetzky. 1991. Segmental differences in the stability of the *trp* repressor peptide backbone. *J. Biomol. NMR.* 1:349–361.

Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nature Struct. Biol.* 4:10–19.

Fontana, A., P. Polverino de Laurento, and V. de Filippis. 1993. Molecular aspects of proteolysis of globular proteins. *In* Protein Stability and Stabilization. W. Van den Tweel, A. Harder, and M. Buitelaar, editor. Elsevier Science, Amsterdam. 101–110.

Fontana, A., P. Polverino deLaureto, V. DeFilippis, E. Scaramella, and M. Zambonin. 1997. Probing the partly folded states of proteins by limited proteolysis. *Folding Design.* 2:R17–R26.

Fuentes, E. J., and A. J. Wand. 1998. Local dynamics and stability of apocytochrme b562 examined by hydrogen exchange. *Biochemistry.* 37:3687–3698.

Gill, S. C., and P. H. von Hippel. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182: 319–326.

Gittelman, M. S., and C. R. Matthews. 1990. Folding and stability of *trp* aporepressor from *Escherichia coli. Biochemistry.* 29:7011–7020.

Gloss, L. M., and C. R. Matthews. 1997. Urea and thermal equilibrium denaturation studies on the dimerization domain of *Escherichia coli trp* repressor. *Biochemistry.* 36:5612–5623.

Guenot, J., and P. A. Kollman. 1992. Molecular dynamics studies of a DNA-binding protein. 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the *E. coli* trp repressor. *Protein Sci.* 1:1185–1205.

Herzberg, O., and M. N. James. 1985. Structure of the calcium regulatory muscle protein troponin-C at 2.8 Å resolution. *Nature.* 313:653–659.

Hiller, R. Z. H. Zhou, M. W. W. Adams, and S. W. Englander. 1997. Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200 degrees C. *Proc. Natl. Acad. Sci. USA.* 94: 11329–11332.

Hilser, V. J., and E. Freire. 1996. Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection. *J. Mol. Biol.* 262:756–771.

Howard, A. E., and P. A. Kollman. 1992. Molecular dynamics studies of a DNA-binding protein. 1. A comparison of the trp repressor and trp aporepressor aqueous simluations. *Protein Sci.* 1:1173–1184.

Hubbard, S. J. 1998. The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta* 1382:191–206.

Jin, L., J. W. Fukayama, I. Pelczer, and J. Carey. 1999. Long-range effects on dynamics in a temperature-sensitive mutant of trp repressor. *J. Mol. Biol.* 285:361–378.

Jin, L., J. Yang, and J. Carey. 1993. Thermodynamics of ligand binding to *trp* repressor. *Biochemistry.* 32:7302–7309.

Komeiji, Y., M. Uebayashi, J.-i. Someya, and I. Yamato. 1991. Molecular dynamics simulation of trp aporepressor in a solvent. *Protein Eng.* 4:871–875.

Komeiji, Y., M. Uebayashi, and I. Yamato. 1994. Molecular dynamics simulations of trp apo- and holorepressors: domain structure and ligand-protein interactions. *Protein Struct. Funct. Genet.* 20:248–258.

Kraulis, P. 1991. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24:946–950.

Kuboniwa, H., N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee, and A. Bax. 1995. Solution structure of calcium-free calmodulin. *Nature Struct. Biol.* 2:768–776.

Lavoie, T. A. 1996. Molecular origin of specificity and cooperativity in the *trp* repressor system. Ph.D. thesis. Princeton University.

Lavoie, T. A., and J. Carey. 1994. Adaptability and specificity in DNA binding by *trp* repressor. *Nucleic Acids Mol. Biol.* 8:184–195.

Lawson, C. L., R.-g. Zhang, R. W. Schevitz, Z. Otwinowski, A. Joachimiak, and P. B. Sigler. 1988. Flexibility of the DNA-binding domains of *trp* repressor. *Proteins Struct. Funct. Genet.* 3:18–31.

Luisi, B., and P. B. Sigler. 1990. The stereochemistry and biochemistry of the trp repressor-operator complex. *Biochim. Biophys. Acta.* 1048: 113–126.

Mann, C. J., and C. R. Matthews. 1993. Structure and stability of an early folding intermediate of *Escherichia coli trp* aporepressor measured by far-UV stopped-flow circular dichroism and 8-anilino-1-naphthalene sulfonate binding. *Biochemistry.* 32:5282–5290.

Miyazawa, S., and R. L. Jernigan. 1982. Equilibrium folding and unfolding pathways for a model protein. *Biopolymers.* 21:1333–1363.

Moult, J., and R. Unger. 1991. An analysis of protein folding pathways. *Biochemistry.* 30:3816–3824.

Nadig, G., G. S. Ratnaparkhi, R. Varadarajan, and S. Vishveshwara. 1996. Dynamics of ribonuclease A and ribonuclease S: computational and experimental studies. *Protein Sci.* 5:2104–2114.

Panchenko, A. N., A. Luthey-Schulten, and P. G. Wolynes. 1996. Foldons, protein structural modules, and exons. *Proc. Natl. Acad. Sci. USA.* 93:2008–2013.

Peng, Z.-y., and L. C. Wu. 1999. Autonomous protein folding units. *Adv. Protein Chem.* (in press).

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

Roder, H., G. A. Elöve, and S. W. Englander. 1988. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labeling and proton NMR. *Nature.* 225:700–704.

Schellenberger, V., K. Braune, H. J. Hofman, and H. D. Jakubke. 1991. The specificity of chymotrypsin. A statistical analysis of hydrolysis data. *Eur. J. Biochem.* 199:623–636.

Schevitz, R. W., Z. Otwinowski, A. Joachimiak, C. L. Lawson, and P. B. Sigler. 1985. The three-dimensional structure of *trp* repressor. *Nature.* 317:782–786.

Shao, X., P. Hensley, and C. R. Matthews. 1997. Construction and characterization of monomeric tryptophan repressor: a model for an early intermediate in the folding of a dimeric protein. *Biochemistry.* 36: 9941–9949.

Shao, X., and C. R. Matthews. 1998. Single-tryptophan mutants of monomeric tryptophan repressor: optical spectroscopy reveals nonnative structure in a model for an early folding intermediate. *Biochemistry.* 37:7850–7858.

Sundaralingam, M., R. Bergstrom, G. Strasburg, S. T. Rao, P. Roychowdury, M. Greaser, and B. C. Wang. 1985. Molecular structure of troponin C from chicken skeletal muscle at 3-angstrom resolution. *Science.* 227: 945–948.

Taniuchi, H., G. R. Parr, and M. A. Juillerat. 1986. Complementation in folding and fragment exchange. *Methods Enzymol.* 131:185–217.

Tasayco, M. L., and J. Carey. 1992. Ordered self-assembly of polypeptide fragments to form nativelike dimeric *trp* repressor. *Science.* 255: 594–597.

Tsai, C., and R. Nussinov. 1997. Hydrophobic folding units derived from dissimilar monomer structures and their interactions. *Protein Sci.* 6:24–42.

Tsapakos, M. J., P. V. Haydock, M. Hermodson, and R. L. Somerville. 1985. Ligand-mediated conformational changes in *trp* repressor protein of *E. coli* probed through limited proteolysis and the use of specific antibodies. *J. Biol. Chem.* 260:16383–16394.

Wallqvist, A., G. W. Smythers, and D. G. Covell. 1997. Identification of cooperative folding units in a set of native proteins. *Protein Sci.* 6:1627–1642.

Woodward, C. 1993. Is the slow-exchanging core the protein folding core? *Trends Biochem. Sci.* 18:359–360.

Wu, L. C., R. Grandori, and J. Carey. 1994. Autonomous subdomains in protein folding. *Protein Sci.* 3:369–371.

Zhang, M., T. Tanaka, and M. Ikura. 1995. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nature Struct. Biol.* 2:758–767.

Zhao, D., C. H. Arrowsmith, X. Jia, and O. Jardetzky. 1993. Refined solution structures of the *Escherichia coli trp* holo- and aporepressor. *J. Mol. Biol.* 229:735–746.

Zheng, Z., J. Czaplicki, and O. Jardetzky. 1995. Backbone dynamics of *trp* repressor studied by ^{15}N NMR relaxation. *Biochemistry.* 34:5212–5223.