3. _____, in *Endocrinological Frontiers in Physiological Insect Ecology*, F. Sehnal, A. Zabza, D. L. Denlinger, Eds. (Wroclaw Technical Univ. Press, Wroclaw, Poland 1988).
4. L. M. Riddiford and C. M. Williams, *Science* **156**, 541 (1967).
5. L. M. Riddiford, *ibid.* **158**, 139 (1967).
6. R. T. Cardé and E. F. Taschenberg, *J. Insect Physiol.* **30**, 109 (1984).
7. A. K. Mattoo and J. C. Suttle, *The Plant Hormone Ethylene* (CRC Press, Boca Raton, FL, 1991).
8. Larvae were collected in June and July of 1988 through 1990 from corn fields in Mississippi and shipped as larvae to Beltsville, MD, by overnight delivery.
9. A. K. Raina, J. A. Klun, E. A. Stadelbacher, *Ann. Entomol. Soc. Am.* **79**, 128 (1986).
10. A. K. Raina, E. A. Stadelbacher, J. D. Warthen, Jr., in *Proceedings of a Conference on Insect Chemical Ecology*, I. Hrdy, Ed. (Academia Praha and SPB Academic Publishing, The Hague, in press).
11. S. P. Burg and E. A. Burg, *Plant Physiol.* **42**, 144 (1967).
12. R. A. Flath, R. R. Forrey, J. O. John, B. G. Chan, *J. Agric. Food Chem.* **26**, 1290 (1978).
13. W. W. Cantelo and M. Jacobson, *J. Environ. Sci. Health* **A14**, 695 (1979).
14. C. D. Reid and R. L. Lampman, *J. Chem. Ecol.* **15**, 1109 (1989).
15. E. M. Beyer, Jr., *Plant Physiol.* **58**, 268 (1976).
16. T. G. Kingan, A. K. Raina, M. B. Blackburn, M. Ma, *Soc. Neurosci. Abstr.* **16**, 856 (1990).
17. T. G. Kingan, *Analyt. Biochem.* **183**, 283 (1989). The antiserum used is highly specific for the complete 33-amino acid peptide; recognition of NH_2- or COOH-terminal 14– or 15–amino acid fragments is less than 0.1% that of the full-length peptide.
18. E. A. Stadelbacher, *Environ. Entomol.* **10**, 766 (1981).
19. R. G. Vogt, G. D. Prestwich, M. R. Lerner, *J. Neurobiol.* **22**, 74 (1991).
20. J. Huignard, E. Thibout, J.-C. Biermont, in *Advances in Invertebrate Reproduction*, M. Prochet, J.-C. Andries, A. Dhainaut, Eds. (Elsevier, Amsterdam, 1986), vol. 4, pp. 425–432.
21. We thank J. Meinwald, L. M. Riddiford, M. Edelman, J. Anderson, and J. Suttle for critical reading of an early version of the manuscript. We also thank E. A. Stadelbacher for collecting field insects and P. Thomas Laemont for technical assistance.

9 September 1991; accepted 6 November 1991

Ordered Self-Assembly of Polypeptide Fragments to Form Nativelike Dimeric *trp* Repressor

MARIA LUISA TASAYCO AND JANNETTE CAREY*

Subdomain-size proteolytic fragments of *Escherichia coli trp* repressor have been produced that assemble in defined order to regenerate fully native dimers. By characterization of the secondary and tertiary structures of isolated and recombined fragments, the structure of assembly intermediates can be correlated with the kinetic folding pathway of the intact repressor deduced from spectroscopic measurement of folding rates. The nativelike structure of these intermediates provides further evidence that protein folding pathways reflect the stabilities of secondary structural units and assemblies found in the native state. The proteolytic method should be generally useful in adding structural detail to spectroscopically determined folding mechanisms.

A MAJOR GOAL OF STUDYING PROtein folding mechanisms is to describe the structures of intermediates along the folding pathway [for review, see (*1*)]. Only one method has the required temporal and structural resolution to follow the folding process: nuclear magnetic resonance (NMR) analysis of hydrogen exchange rates for individual protons (*2*). Although this method is powerful, its application is limited to the rather small set of proteins under ~20 kD for which complete resonance assignments are available. More widely applicable kinetic studies that use global spectroscopic probes, such as near-ultraviolet (UV) extinction coefficient, far-UV ellipticity, or fluorescence intensity, yield only formal mechanistic intermediates that lack structural definition. It would be generally useful if such abstract mechanisms could be fleshed out in structural detail. We report (i) an approach that uses biochemical dissection of a complex protein into subdomain-size fragments and (ii) structural characterization of the ordered assembly process that these fragments undergo (*3*).

The *trp* aporepressor of *Escherichia coli* (TrpR) has an unusually complex subunit architecture (*4*). The dimer is built from two highly intertwined chains (Fig. 1), each of which has an overall dumbbell-shaped fold similar to that of calmodulin or troponin (*5*). Helices A, B, and part of C, together with A′, B′, and part of C′, constitute most of the subunit interface. These six helices contain many of the hydrophobic residues of the protein, are tightly packed, and show little flexibility (*6–8*). The amphipathic COOH-terminal helix, F, from one subunit packs against helices A′ and B′ from the other subunit to complete the dimer interface, creating a complex topology. Together, helices A, B, F, and part of C from each subunit form a rigid core domain (*4, 6*), whereas helices D and E are involved in DNA binding (*9*). The intertwined nature of the TrpR dimer implies that association

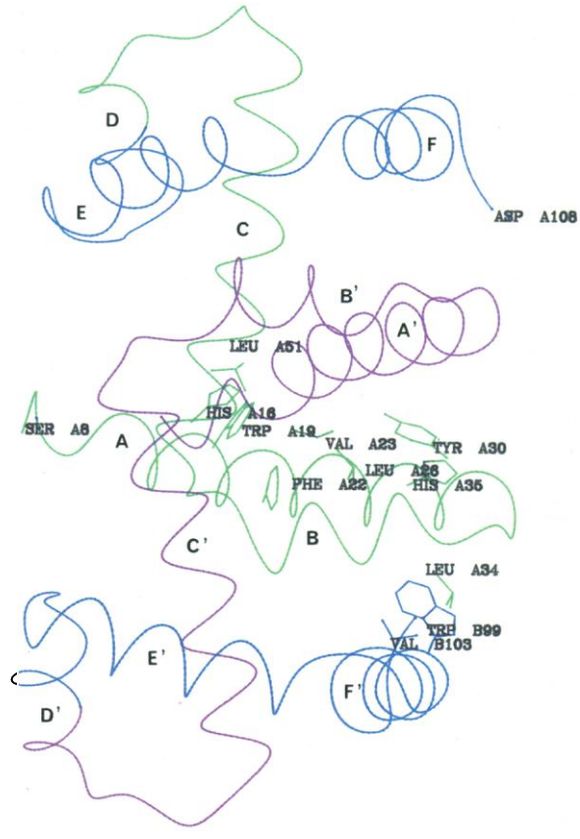

Fig. 1. Structure of the *trp* repressor dimer. The polypeptide backbone is represented by a string from residues 8 to 108. Residues 8 to 71 (N-fragment) of subunit A are shown in green, and those from subunit B are in purple; residues 72 to 108 (C-fragment) from each subunit are shown in blue. Helices are labeled A to F in subunit A and A′ to F′ in subunit B. Nativelike packing of certain hydrophobic residues is detectable by NMR. The side chains of these residues are shown in skeletal form and identified by three-letter code, residue number, and subunit (A or B); symmetry-related residues from the two subunits are indicated only once. The model was constructed with coordinates from file 3WRP [TrpR holorepressor (*4, 6*)] from the Brookhaven Protein Data Bank with deletion of the exogenous Trp ligand.

Chemistry Department, Princeton University, Princeton, NJ 08544.

*To whom correspondence should be addressed.

of the two polypeptide chains must be a fairly early step on the folding pathway. Consistent with this topological constraint are spectroscopic studies (*10*) of refolding, which reveal complex multiphase kinetics including bimolecular folding steps, intramolecular folding steps, and nonfolding isomerization steps. Despite the complexity of both structure and kinetics, equilibrium folding (*10*) of TrpR is two-state, precluding detailed structural studies of intermediates.

We generated proteolytic fragments (*11*) of the repressor that illuminate folding of the intact protein. Chymotrypsin cleaves the repressor after Tyr^7 in the NH_2-terminal "arm" and after Leu^{71} in the first turn of helix D (*11, 12*). The NMR (*7, 8, 12*) and x-ray (*4, 6, 9*) structures of various forms of the repressor show that the NH_2 terminus and the DNA-binding helices are the most flexible parts of the protein. These dynamic features of the protein structure are reflected in the proteolytic pattern: under native conditions, only 2 of the more than 50 potential chymotrypsin cleavage sites in each subunit are attacked in 6 hours (*11*). The two cleavage events are kinetically distinct and can be followed by circular dichroism (CD) spectropolarimetry (*11*) for preparative isolation of NMR quantities of fragments. Because of the architecture of TrpR, cleavage of the dimer at position 71 of each monomer separates the rigid core domain (*4, 6*) into two pairs of subdomain fragments, N (residues 8 to 71) and C (residues 72 to 108). These fragments were readily purified from one another by molecular-sieve chromatography in presence of urea and were renatured by dilution or dialysis or both.

The 500-MHz NMR spectrum of TrpR that lacks the seven NH_2-terminal residues ["armless" TrpR, residues 8 to 108 (*12*)] is shown in Fig. 2A. Although broad and severely overlapping NMR lines are characteristic for a protein of this molecular weight (~25 kD), complete sequence-specific resonance assignments have been achieved by Arrowsmith and co-workers (*7, 8, 12*). Features in the armless TrpR spectrum diagnostic for the folded native state (*7, 12*) include resonances from five methyl groups shifted upfield as a result of packing against aromatic residues in the hydrophobic interior (see Fig. 1) and characteristic resonance frequencies in the aromatic region, which also reflect packing (*7*).

Many of the features expected for nativelike protein are present in the NMR spectrum of the renatured N-fragment (Fig. 2B). This fragment contains five of the six aromatic residues of armless TrpR. Despite the tendency of this fragment to aggregate at NMR concentrations, with concomitant broadening of peaks, most of the resonances in the aromatic region of the N-fragment spectrum coincide with resonances from the native protein (*13*). These results suggest that packing of most aromatic groups in this fragment is nativelike, implying that at least helix A, which contains most of these residues, is formed in the N-fragment. The high helix content found by CD [(*14*); see Fig. 2E], combined with the NMR spectrum, suggests that the folded portion of N-fragment includes most of helices A, B, and C. The overall similarity of the N-fragment spectrum

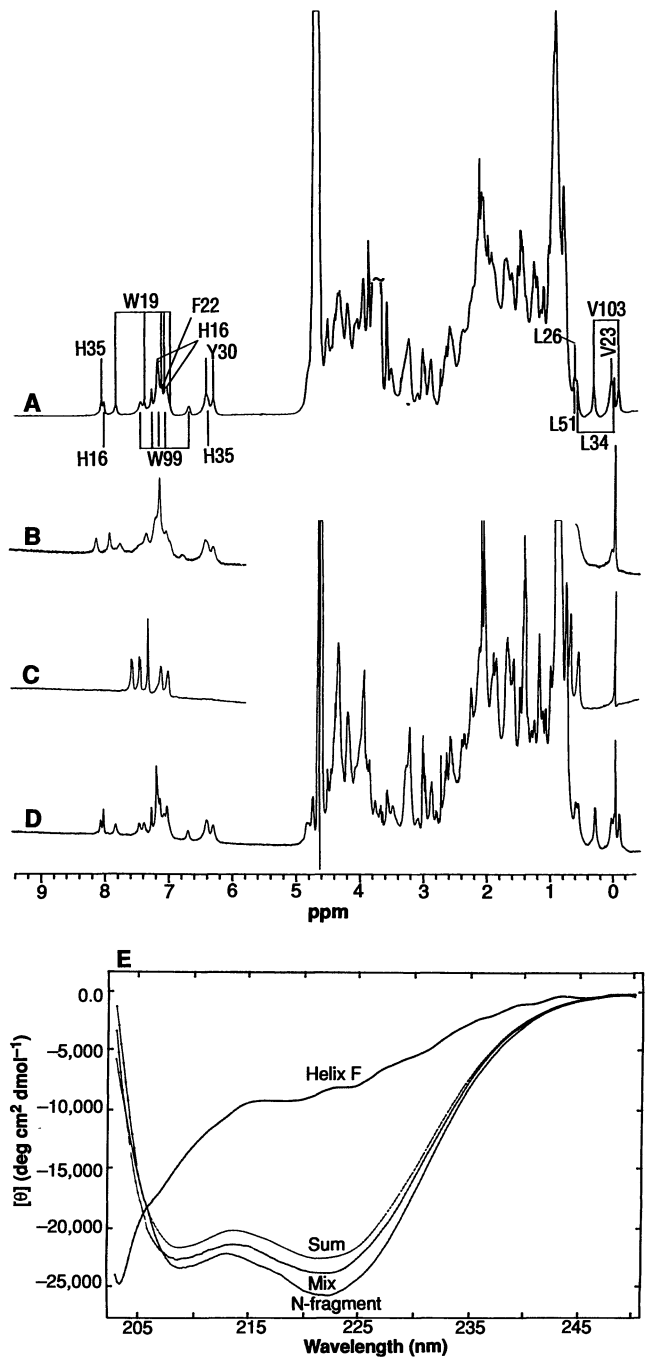

Fig. 2. Structure of isolated and reconstituted fragments. Protein concentrations were determined by absorption spectroscopy in 6 M guanidine hydrochloride (*23*) and are reported in terms of monomer. (**A** through **D**) NMR spectra. Spectra were acquired on a 500-MHz JEOL spectrometer at 38°C with a 1.5-s pulse delay and presaturation of the water peak before data acquisition. Final conditions were approximately 1 mM protein, 10 mM sodium phosphate, pD 7.6 (uncorrected), 0.5 M NaCl, and 99.9% D_2O containing a trace of sodium trimethylsilylpropionate as a reference standard. Abbreviations for the amino acid residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (A) Armless repressor (intact protein containing residues 8 to 108). Assignments are from (*7*). The large peak at ~3.6 ppm is due to polyethyleneglycol contamination. (B) N-fragment. Conditions were as for armless protein, except that 19% acetonitrile was added to the N-fragment sample to reduce aggregation. This amount of cosolvent had no effect on helicity of the sample, as assessed from the CD spectrum taken at a protein concentration of ~0.7 mM with a cell (13-μm path length) at the synchrotron light source, Brookhaven National Laboratory. (C) C-fragment. Conditions were as for armless protein. (D) Reconstituted mixture. Equimolar amounts of N- and C-fragments were combined to give a final protein concentration of ~1 mM each. Other conditions were as for armless protein; that is, no acetonitrile was added. (**E**) Circular dichroism spectra of N-fragment and helix F peptide. We obtained spectra on an Aviv 62DS spectropolarimeter at 25°C using a tandem cell (2-cm by 0.4375-cm path length) with a bandwidth of 1.5 nm and step size of 0.25 nm. Buffer was 10 mM sodium phosphate, pH 7.6, and 50 mM NaCl. Helix F peptide (residues 94 to 106 of intact TrpR; sequence $CH_3COValGluLeu$-$ArgGlnTrpLeuGluGluValLeuLeuLysNH_2$) was synthesized by an automated method and was purified on a 3 cm by 25 cm column containing C_4 flash-chromatography resin with a shallow gradient of acetonitrile in 0.1% trifluoroacetic acid (*21*). Spectra of individual fragments were taken at protein concentrations of 7.8 μM. "Sum" spectrum was acquired by having one fragment on each side of the tandem cell at equal concentrations (7.8 μM) and volumes. "Mix" spectrum was taken after thorough mixing by repeated inversion of the tandem cell.

to that of intact protein, as well as several kinds of biochemical evidence (*15*), strongly suggest that this fragment is dimeric. Thus, many of the residues of the N-fragment participate in forming a nativelike dimeric structure.

The NMR spectrum of the C-fragment is shown in Fig. 2C. The only aromatic residue in this fragment, Trp^{99}, gives rise to five resonances in positions intermediate between those found for Trp in a random-coil conformation (*16*) and in the intact repressor (*7*). Trp^{99} is in the center of helix F (Fig. 1), one of the most highly ordered regions in the native protein (*4, 6–9, 12*). In the intact protein, folding of helix F causes a strong upfield shift of the Val^{103} methyl group due to proximity of Trp^{99} and strong *i,i*+1 NH–NH nuclear Overhauser effects (NOEs) (*7*), both of which are absent from the C-fragment spectrum (Fig. 2C). The NMR spectrum suggests that secondary and tertiary structure are weak or absent in the isolated C-fragment; this conclusion is also supported by the CD spectrum, which shows ~0% of the residues in helical conformation.

An equimolar mixture of the N- and C-fragments gave an NMR spectrum (Fig. 2D) that is essentially indistinguishable from that of the intact protein. All of the diagnostic features of the intact protein are present in the spectrum of the reconstituted mixture; there are only minor differences in a small number of chemical shifts, mostly in the Cβ region. Thus, it appears that upon binding to the N-fragment dimer, the C-fragment adopts a folded conformation similar to that in native aporepressor (*17*). The binding energy of the N- and C-fragments therefore must be sufficient to drive the folding of C-fragment into a nativelike structure.

To begin probing the structural basis for the reconstitution of the N- and C-fragments and to further confirm the nativelike structure of the reconstituted mixture, we used CD to analyze the ability of a synthetic helix-F peptide to bind to the N-fragment (Fig. 2E). Though the helix F peptide by itself shows no clear helical structure, there is a net increase in helix content upon mixing relative to the sum of the two spectra (*18*). Thus, it appears that reconstitution of C-fragment with the N-fragment dimer is mediated primarily by helix F, as inspection of the crystal structure (*4, 6*) suggests. Taken together, our results suggest an ordered assembly process in which two N-fragments first associate to form an autonomous dimeric subdomain that provides a surface for the folding of helix F.

The ordered assembly of TrpR fragments suggests a tentative and testable mechanism for some steps in the folding of the intact repressor. The NH_2-terminal segment of each monomer first adopts a nascent structure. An early step of this kind is formally required to provide a recognition surface for dimerization. Association of two polypeptide chains follows, forming a dimeric interface comprising most of helices A, A′, B, B′, C, and C′. At this stage of the folding process, helices D, E, and F from both chains are incorrectly packed and may be disordered. In a subsequent step these distal portions of each chain cross the subunit interface, allowing helix F to pack correctly against helices A, B, and C of the opposite subunit (*19, 20*). This model for the folding pathway is consistent with the kinetic mechanism deduced (*10*) from spectroscopic studies. In particular, our results support folding through dimeric but not fully native intermediates [D* in (*10*)]. Tests of the model would require kinetic measurements of folding for the isolated and reconstituted fragments. Comparison with rate constants for folding of intact repressor (*10*) may allow these putative intermediates to be placed on the folding pathway.

The approach presented here has the potential to go beyond reconstitution systems (*3*) that have aimed to equate protein fragments with structural intermediates on the folding pathway. This success appears to arise from dissection into fragments sufficiently small that they are capable of folding only upon recombination in defined order. In this sense, our approach is similar and complementary to that of Kim and coworkers (*21*), who have used medium-size synthetic peptides to mimic structured intermediates in the folding pathway of bovine pancreatic trypsin inhibitor (BPTI). An advantage of the proteolytic approach is that the structure and dynamics of the protein direct cleavage to sites potentially relevant for folding, with no requirement for prior knowledge of intermediate structures, such as is available in the case of only a few proteins like BPTI. This approach may therefore provide a useful early step in characterizing intermediate structures, even for proteins or domains with folding pathways that are quite complex.

Our results provide additional support for a model in which the pathway of protein folding is populated by intermediates assembled from nativelike secondary structural elements (*1, 2, 21*). A key feature of this model is the stability of secondary structure units and assemblies: though marginal, the stability of isolated structural elements may influence the choice of pathway and is enhanced by mutual interaction during the folding process. If such a model can be generalized, protein folding may be illuminated by understanding the subdomain hierarchies (*22*) of native proteins.

REFERENCES AND NOTES

1. P. S. Kim and R. L. Baldwin, *Annu. Rev. Biochem.* **59**, 631 (1990).
2. H. Roder, G. A. Elöve, S. W. Englander, *Nature* **335**, 700 (1988); J. B. Udgaonkar and R. L. Baldwin, *ibid.*, p. 694.
3. Although there have been numerous studies of protein fragments, their structures, and their reconstitution, most have been aimed at elucidating the assembly of prefolded domains [for review, see R. Jaenicke, *Biochemistry* **30**, 3147 (1991)] rather than the folding steps leading to domains (*22*). In other cases, emphasis has been placed on reconstitution of function. Some of the most detailed fragment studies, on staphylococcal nuclease and cytochrome c [cytc; for review, see H. Taniuchi, G. R. Parr, M. A. Juillerat, *Methods Enzymol.* **131**, 185 (1986)], have been interpreted in terms of features that stabilize the native state rather than in terms of correlation with folding pathways. In part, this has been due to lack of kinetic detail in the folding reactions of these proteins, a limitation only recently overcome for cytc (*2*). For the cytc system, preliminary data indicate that fragments containing the two early-folding NH_2- and COOH-terminal helical regions (*2*) can associate (L. Wu and J. Carey, unpublished results). In the cytc case, proteolysis is the only way to isolate heme-containing fragments.
4. R. W. Schevitz, Z. Otwinowski, A. Joachimiak, C. L. Lawson, P. B. Sigler, *Nature* **317**, 782 (1985); R.-G. Zhang *et al.*, *ibid.* **327**, 591 (1987).
5. O. Herzberg and M. N. G. James, *ibid.* **313**, 653 (1985); M. Sundaralingam *et al.*, *Science* **227**, 945 (1985); Y. S. Babu *et al.*, *Nature* **315**, 37 (1985).
6. C. L. Lawson *et al.*, *Proteins* **3**, 18 (1988).
7. C. H. Arrowsmith, R. Pachter, R. B. Altman, S. B. Iyer, O. Jardetzky, *Biochemistry* **29**, 6332 (1990); E. I. Hyde *et al.*, *Eur. J. Biochem.* **183**, 545 (1989).
8. C. H. Arrowsmith, J. Czaplicki, S. B. Iyer, O. Jardetzky, *J. Am. Chem. Soc.* **113**, 4020 (1991).
9. Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
10. M. S. Gittelman and C. R. Matthews, *Biochemistry* **29**, 7011 (1990).
11. J. Carey, *J. Biol. Chem.* **264**, 1941 (1989). Cleavage at Tyr^7 is nearly complete when cleavage at Leu^{71} begins.
12. C. H. Arrowsmith, J. Carey, L. Treat-Clemons, O. Jardetzky, *Biochemistry* **28**, 3875 (1989).
13. Three upfield methyl peaks are expected in a nativelike N-fragment spectrum (7): Val^{23} (which in the intact protein is ring-current shifted by Phe^{22} and Trp^{19}), Leu^{26} (because of nearby Phe^{22} and Tyr^{30}), and Leu^{51} (shifted by Trp^{19}). Upfield-shifted resonances for Leu^{34} are not expected in N-fragment, because this fragment lacks Trp^{99}, the side chain responsible for ring-current effects on its methyl groups in intact repressor (7). The presence of a peak very close to the chemical shift position of Val^{23} methyl (0 to 0.1 ppm) and the broad lines in the region of Leu^{26} and Leu^{51} methyls (0.6 to 0.7 ppm) suggest that the environment of these residues in the N-fragment is nativelike. The peak at ~6.8 ppm arises from a partly folded portion of N-fragment and not from contaminating Trp^{99}, as shown by urea denaturation of the reconstituted mixture. In the absence of urea (Fig. 2D), the 6.8-ppm peak is missing from the mixture spectrum, and the Trp^{99} peak is present at ~6.7 ppm, as in armless TrpR. At 2 M urea the 6.8-ppm peak begins to grow in next to the Trp^{99} resonance, which is unaffected at this urea concentration.
14. W. C. Johnson, *Proteins* 7, 205 (1990). From analysis of x-ray data, intact TrpR should have ~75% helical residues, whereas by CD it has ~55%. The NMR results of Arrowsmith and co-workers (*7, 8, 12*) agree with the CD value for helix content of intact TrpR, suggesting that the solution structure of TrpR has lower helix content than the crystal structure. CD analysis shows that N-fragment has ~56% of its residues in helical conformation; the fraction expected for a fully folded nativelike N-fragment dimer, on the basis of NMR data, is 63%.
15. Genetic studies [R. L. Kelley and C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 483 (1985)] had shown that an amber mutation at codon 68 in the repressor gene produced a protein fragment that was able to

inactivate wild-type protein produced in the same cell, presumably by formation of stably folded but inactive heterodimers. Thus, the first 68 residues, which form helices A, B, and C in the intact protein, are sufficient for heterodimer formation. The following lines of evidence further support a dimeric structure for the nativelike N-fragment: mobility in native polyacrylamide gel electrophoresis (*11*); concentration dependence of sedimentation velocity; concentration dependence of the fluorescence spectrum and the fluorescence-detected urea denaturation midpoint; and glutaraldehyde cross-linking. In addition, N-fragment can form heterodimers with both intact and armless TrpR.

16. A. Bundi and K. Wüthrich, *Biopolymers* **18**, 285 (1979).
17. Other lines of evidence support the conclusion that reconstituted *trp* repressor has nativelike structure. First, the only Trp residues in TrpR are 19 and 99. Fluorescence spectra show nearly identical maximum wavelengths (λ_{max}) and total integrated intensities for intact TrpR, armless TrpR, intact $Trp^{99} \rightarrow$ Phe mutant TrpR (C. Royer, personal communication), N-fragment, and the reconstituted mixture of N- and C-fragments. Time-resolved fluorescence spectra show a 4-ns lifetime component for both N-fragment and the $Trp^{99} \rightarrow$ Phe mutant proteins. These results suggest that the fluorescence spectrum of intact protein is due almost entirely to Trp^{19} and that the environment of this residue in N-fragment is also nativelike. Second, we (in collaboration with C. L. Lawson) have obtained crystals of the reconstituted mixture in the presence of L-tryptophan under the same conditions that yield tetragonal crystals (*4, 6*) of intact *trp* holorepressor. In addition, reconstitution detected by CD shows that maximal increase in helicity occurs at a 1:1 ratio of N- and C-fragments at ~8 μM each; addition of up to sixfold excess C-fragment causes no further increase in helicity.
18. Helix F represents ~32% of the residues of C-fragment (12 of 37 residues). The fraction of residues in helical conformation went down slightly after reconstitution of N-fragment with C-fragment, from ~56% to ~47%, which is consistent with formation of helical structure in only the helix F portion of C-fragment, with little or no change either in the remaining residues of C-fragment or in the N-fragment portion of the reconstituted molecule.
19. The intertwined topology of TrpR is strikingly similar to that of the all-α-helical interferon dimer (*20*), even though the two proteins share no apparent homology. Interferons occur in both monomeric and dimeric forms, and x-ray structures are known for each. Comparison of these two forms reveals a remarkable conservation of chain folds (*20*). The NH_2-terminal helices are arranged almost identically in both monomeric and dimeric interferons, whereas the position of the two COOH-terminal helices differs in the two forms: in the monomer, these helices are packed internally among the NH_2-terminal helices, whereas in the dimer the distal segment of each polypeptide crosses the subunit interface to pack with the opposite subunit, as in *trp* repressor. These observations suggest the possibility that an early step in assembly of dimeric interferon involves folding of the NH_2-terminal part of the protein into a nativelike structure common to both monomeric and dimeric forms. This fold provides a recognition surface for dimerization, whereas the distal chain segments are incorrectly or incompletely folded until a later stage of the folding process.
20. S. E. Ealick *et al.*, *Science* **252**, 698 (1991).
21. T. G. Oas and P. S. Kim, *Nature* **336**, 42 (1988); J. P. Staley and P. S. Kim, *ibid.* **344**, 685 (1990).
22. A. M. Lesk and G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4304 (1981); J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
23. R. S. Mulvey, R. J. Gualtieri, S. Beychok, *Biochemistry* **13**, 782 (1974).
24. Supported by grant GM43558 to J.C. We thank K. Valentine for assistance in acquiring some of the NMR spectra and R. Punch and F. Figueirido for assistance with computer modeling. Discussions with D. Barrick and P. Kim are acknowledged.

3 September 1991; accepted 25 November 1991

Toward a Dynamical Structure of DNA: Comparison of Theoretical and Experimental NOE Intensities

JANE M. WITHKA, S. SWAMINATHAN,* JAYASHREE SRINIVASAN, DAVID L. BEVERIDGE, PHILIP H. BOLTON

Comparisons of experimental and calculated interproton nuclear Overhauser effect (NOE) buildup curves for duplex d(CGCGAATTCGCG)$_2$ have been made. The calculated NOEs are based on molecular dynamics simulations including counterions and water and on the single-structure canonical A, B, and crystal forms. The calculated NOE effects include consideration of the motions of individual interproton vectors and the anisotropic tumbling of the DNA. The effects due to inclusion of anisotropic tumbling are much larger than those due to the local motion, and both improve the agreement between calculated and experimental results. The predictions based on the dynamical models agree significantly better with experiment than those based on either of the canonical forms or the crystal structure.

THE FIRST DETERMINATION OF THE structure of DNA was based on diffraction data on fibers (*1*); subsequently, higher resolution structural information was obtained by crystallography on oligonucleotides (*2*). A number of crystal structures in the A, B, and Z families have been reported (*2–5*). The analysis of these structures has revealed further evidence for sequence-dependent fine structure and bending. However, the applicability of these observations to the structure of DNA in solution has not been determined, because the solution structure per se is not accurately known.

Studies of DNA in solution based on extensive molecular dynamics (MD) simulations including counterions and water (*6*) have recently been reported from this laboratory. A theoretical prediction on the structure of duplex d(CGCGAATTCGCG)$_2$ in solution (Fig. 1) exhibits qualitatively, if not quantitatively, the propeller twist in the base pairs and local axis deformations at or near the GC/AT interface seen in the crystal structure of Dickerson and Drew (*2*). However, the calculations involve approximations at the level of the intermolecular force field and assumptions in the simulation protocol (*7–9*), and the accuracy of the predictions based on MD simulation of DNA structure in solution has not yet been fully validated. A detailed comparison of theoretical and experimental results for DNA in solution is required to determine the quality of the MD results that are obtained in the absence of any artificial energy constraints.

NOE data provide, in principle, extensive information on interproton distances in molecules (*10–12*). However, structures cannot be determined directly from the NOE data because of a number of factors, which include information on the short-range interproton distance (≤5 Å), spin diffusion effects, and the internal motion and anisotropic tumbling of the DNA molecules (*13–16*). We have recently described a procedure for obtaining the NOE intensities directly from the trajectories of MD simulation (*15*). This method includes consideration of anisotropic tumbling of the duplex and the orientation of the ^1H–^1H vectors to the symmetry axis of the duplex and the extent of local motion of the ^1H–^1H vectors. These features along with the interproton distances can be used to determine the experimentally observable ^1H–^1H NOEs (*13–15*).

The orientation of the ^1H–^1H vectors to the helical symmetry axis is important because a duplex dodecamer DNA, an asymmetric molecule with a length about twice that of the diameter, undergoes anisotropic tumbling in solution (*13–16*). Therefore, a ^1H–^1H vector pointing along the symmetry axis will have an effective correlation time longer than one perpendicular to the helical axis because the tumbling of the DNA helix in solution will be faster about the symmetry axis than about the axis perpendicular to the symmetry axis. The local motion of a ^1H–^1H vector, of extent θ, is about the average orientation, β, of the ^1H–^1H vector to the helical axis (Fig. 2).

Kaptein and co-workers recently included some aspects of internal motion in conjunction with internuclear distance constraints in

Chemistry Department, Wesleyan University, Middletown, CT 06459.

*Present address: Gilead Sciences, Foster City, CA 94404.