An Alanine-Zipper Structure Determined by Long Range Intermolecular Interactions*

Received for publication, August 27, 2002, and in revised form, September 19, 2002
Published, JBC Papers in Press, October 3, 2002, DOI 10.1074/jbc.M208773200

Jie Liu‡ and Min Lu§

From the Department of Biochemistry, Weill Medical College of Cornell University, New York, New York 10021

A major challenge in protein folding is to identify and quantify specific structural determinants that allow native proteins to acquire their unique folded structures. Here we report the engineering of a 52-residue protein (Ala-14) that contains exclusively alanine residues at the hydrophobic a and d positions of a natural heptad-repeat sequence. Ala-14 is unfolded under normal solution conditions yet forms a parallel three-stranded a-helical coiled coil in crystals. Ala-14 trimers in the solid state associate with each other through the pairing of polar side chains and formation of an extended network of water-mediated hydrogen bonds. In contrast to the classical view that local intramolecular tertiary interactions dictate the three-dimensional structure of small single-domain proteins, Ala-14 shows that long range intermolecular interactions can be essential in determining the metastable alanine-zipper structure. A similar interplay between short range local and longer range global forces may underlie the conformational properties of the growing class of natively unstructured proteins in biological processes.

The classical view of protein folding posits that local secondary and tertiary interactions promote the association of cooperative structural elements in a process of sequential stabilization to form a limited set of transient intermediates and finally the native state (1–12). The current theory emphasizes the role of the free energy landscape of a polypeptide chain in driving rapid and efficient folding into a unique well defined structure (13–17). In any description of protein folding, a major challenge is to identify specific structural determinants that guide the folding process. Although considerable progress has been achieved toward understanding how compact single-domain proteins fold (1–12), the factors that specify and stabilize larger and more extended protein complexes remain largely undefined despite their critical importance in the complex environment inside a living cell. We report here a new alanine-zipper structure that undergoes a native folding transition only during crystallization. A network of long range solvent-mediated interactions is observed in the 1.3-Å x-ray crystal structure of this alanine-zipper trimer that appears to directly control its folding into a helical trimer structure.

EXPERIMENTAL PROCEDURES

Protein Production—Plasmid pAla14 encoding Ala-14 was derived from plp56 (Ref. 18). The sequence Cys-Gly-Gly was added to the N terminus of Ala-14 to produce Ala-14C. Similarly, Gly-Gly-Cys was added to the C terminus of Ala-14 to produce Ala-14C. The sequences Cys-Gly-Gly and Gly-Gly-Cys were added at the N and C termini, respectively, of Ala-14 to generate Ala-14CC. The proteins were expressed in Escherichia coli strain BL21(DE3)/pLYsS (Novagen). Cells were grown at 20 °C in LB media to an optical density of 0.8 at 600 nm and induced with isopropyl-β-D-thiogalactoside for 10 h. Cells were lysed by glacial acetic acid on ice. The bacterial lysate was centrifuged (35,000 × g for 30 min) to separate the soluble fraction from inclusion bodies. The soluble fraction containing denaturing Ala-14 was dialyzed into 5% acetic acid overnight at 4 °C. Proteins from the soluble fraction were purified to homogeneity by reverse-phase high performance liquid chromatography (Waters, Inc.) on a Vydac C-18 preparative column (Hesperia, CA) using a water-acetonitrile gradient in the presence of 0.1% trifluoroacetic acid and 0.01% trifluoroacetic acid and 0.01% heptfluorobutyric acid. The molecular weights of each peptide were confirmed by using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (PerSeptive Biosystems, Framingham, MA).

To produce the disulfide-linked Ala-14SOS protein (see Fig. 2C), the reduced Ala-14NC, Ala-14C, and Ala-14CC peptides were mixed and incubated in reduct buffer (1 mM oxidized glutathione, 1 mM reduced glutathione, 50 mM NaCl, 0.1 mM Tris-HCl, pH 8.7, 1 mM EDTA) for 3 days at 4 °C to a total protein concentration of 40 mg/ml. This treatment was followed by high pressure liquid chromatography purification as described above. The designed Ala-14SOS protein with disulfides between Ala-14NC and the N terminus of Ala-14CC and between Ala-14CC and the C terminus of Ala-14CC was analyzed by proteolysis with trypsin and mass spectrometry. In the oxidized sample, the disulfide-bonded homodimers of the CGGSSNAK and ADNAAGGC fragments were observed. In the reducing sample, the masses of these homodimers disappear and new masses appear, corresponding to the individual fragments.

Crystallization, Structure Determination, and Refinement—A 25-mg/ml stock of Ala-14 was prepared in water. The best diffracting crystals grew at 4 °C from 1 μl of the stock added to 1 μl of the reservoir buffer (0.1 mM sodium acetate, pH 4.0, 1.3 mM sodium citrate) and allowed to equilibrate against the reservoir buffer. The crystals belong to the space group P21 (a = 21.84 Å, b = 26.94 Å, c = 45.80 Å, α = 88.05°, β = 84.95°, γ = 87.56°) and contain three monomers in the asymmetric unit. Prior to data collection, crystals were transferred into cryosolution containing the reservoir buffer and 15% glycerol, harvested, and frozen in liquid nitrogen. Diffraction data were collected at the X25 beamline at the Brookhaven National Laboratory using a Brandeis B4 CCD detector. Reflection intensities were integrated and scaled with the program AMoRe (21). Initial phases were determined by molecular replacement using the Ala-7 trimer (20) as a search model in the program AMoRe (21). Electron density map interpretation and model building were done with the program O (22). The structure of Ala-14 was refined at 1.3-Å resolution using the program CNS 1.0 (Ref. 20). The isotropic CNS model was further refined at 1.3-Å resolution using restrained anisotropic displacement parameters for all atoms with the program SHELX97 (Ref. 24). The final model was verified using simulated annealing omit maps. All main-chain torsional angles
Crystal Structure of an Alanine Zipper

**RESULTS AND DISCUSSION**

**Alanine Zipper**—The outer membrane lipoprotein is the most abundant protein of *E. coli* containing a trimeric \( \alpha \)-helical coiled-coil domain (Lpp-56) that is embedded between the outer membrane and the periplasmic peptidoglycan (18). Classical coiled-coil proteins share a characteristic seven-amino acid repeat containing bulky hydrophobic side chains at the first (a) and fourth (d) positions (30–32). The crystal structure of Lpp-56 reveals that this coiled-coil trimer has an unusual feature: three alanines occupying successive a and d positions form a local structure we refer to as an alanine zipper (18). In previous studies, we showed that four additional alanine layers can be stably incorporated into the triple-helical structure despite the fact that each alanine layer is destabilizing significantly relative to large hydrophobic groups (20). Because alanine has relatively low hydrophobicity, the effect generally considered to be an important determinant of protein stability (33–35), we surmised that van der Waals packing interactions within the –CH\(_3\) groups are insufficient to specify and sustain an extended alanine-zipper conformation under physiological conditions. To explore the limits of three-dimensional structure formation in the absence of strong hydrophobic stabilization, we engineered an alanine-zipper protein called Ala-14 by the complete substitution of each bulky hydrophobic amino acid at the a and d positions of the Lpp-56 heptad-repeat sequence with alanine (Fig. 1A).

**Crystal Structure of the Ala-14 Trimer**—To evaluate the high resolution features of the interfacial alanine side-chain packing, we determined the x-ray crystal structure of the Ala-14 peptide at 1.3-Å resolution (Fig. 1B). The Ala-14 structure was refined to a conventional R-factor of 15.0% with a free R-factor of 20.9% and root mean square deviations from ideal bond lengths and bond angles of 0.010 Å and 2.0°, respectively (Table 1). Ala-14 forms a parallel triple-stranded \( \alpha \)-helical structure with novel and unique features (Fig. 1, C and D). The superhelix creates a cylinder that is \( \sim \)18-Å wide and \( \sim \)74-Å long. An
approximate 3-fold axis of symmetry coincides with the superhelical axis. The alanine side chains at the $a$ and $d$ positions point into the center of the trimer, forming a nonpolar core at the interface of the three $\alpha$-helices. Cross-sectional layers containing alanine at the $a$ positions alternate with layers containing alanine at the $d$ positions. Residues at positions $e$ and $g$ pack against alanines at $a$ and $d$, respectively, to complete the hydrophobic core. Interhelical salt bridges and charge-stabilized hydrogen bonds form with high frequency between charged or polar side chains at the $g$ position of one heptad and at the $e$ position of a preceding or succeeding heptad.

The Ala-14 trimer shows a classical acute knobs-into-holes packing characteristic of parallel triple-stranded coiled coils (25, 36). At both the $a$ and $d$ layers, the $C_{\alpha}-C_{\alpha}$ bond of each alanine knob makes an acute angle ($\sim 60^\circ$) with the $C_{\alpha}-C_{\alpha}$ vector at the base of the recipient hole among four residues on a neighboring helix. This similar geometric packing at positions $a$ and $d$ is a distinguishing structural feature of trimer formation (36). In 13 of 14 $a$ and $d$ layers, the dihedral angles $\phi$ and $\psi$ of the alanine backbones are approximately $-40$ and $-70$, corresponding to the $\alpha$-helix region of Ramachandran space (37). The alanine residues in the C-terminal most $d$ layer assume dihedral angles near 160 and $-70$ and adopt an extended conformation. These end effects cause the superhelix to be locally unwound, and five coordinated water molecules are observed to be anchored to the Ala-48B backbone carbonyl in the core structure.

An interplay between alanine side-chain packing and backbone conformation can be discerned in the trimer structure as dramatic changes in the twist of individual helices and the helix-helix spacing. The superhelical parameters (radius $R_\phi$, pitch $\omega_\phi$, and $\alpha$-position $C_{\alpha}$ phase angle $\phi$) for the Ala-14 and Lpp-56 trimers are as follows: $R_\phi = 5.1 \text{ Å}, \omega_\phi = 99$ residues/turn, and $\phi = 13.1^\circ$ for Ala-14; $R_\phi = 6.1 \text{ Å}, \omega_\phi = 115$ residues/turn, and $\phi = 16.8^\circ$ for Lpp-56. The alanine backbone deviation from classical coiled coils can be described as a localized unwinding of the superhelix (decrease in $\omega_\phi$), decrease in the phase angle $\phi$, and decrease in the supercoil radius $R_\phi$. As a result, the backbones of the three helices in the Ala-14 trimer are more highly curved and wrap more tightly around the superhelical axis, producing a very close spacing between adjacent helices (8.8 Å). This type of helix contact may indeed correspond to an optimal geometry for a novel class of tight helix-helix association in proteins (38).

**Solution Properties of Ala-14**—Surprisingly, Ala-14 has only $\sim 20\%$ helix content at 3 mg/ml protein concentration in neutral pH phosphate-buffered saline at 0 °C as determined by CD (Fig. 3A). Under these conditions, Ala-14 lacks a “folded” base line at low temperatures and displays a broad noncooperative thermal unfolding transition (Fig. 2B). Sedimentation equilibrium measurements indicate that Ala-14 exists as a monomer, even at a concentration of 5 mg/ml. Strikingly, there is no apparent increase in the $\alpha$-helical content and/or the thermal stability of Ala-14 in the presence of 0.1 m sodium acetate, pH 4.0, 1.3 m sodium citrate (crystallization conditions) relative to phosphate-buffered saline, pH 7.0. Taken together, these data indicate that the Ala-14 peptide lacks intermolecular association and forms a mix of intramolecular helix and coil under native folding conditions in solution.

To assess the role of the monomer-trimer equilibrium of Ala-14 on its folding in aqueous solution, we generated a covalently linked Ala-14 trimer whose stability is concentration-independent. In this single-chain version of the protein denoted Ala-14SOS, the three Ala-14 strands are connected by short disulfide-bonded peptide linkers (Fig. 2C). The helical CD signal observed in Ala-14SOS is comparable, and the thermal stability is identical to that found in the isolated Ala-14 peptide (Fig. 2). In contrast to what is commonly found in other coiled-coil models, the stability of Ala-14 is insensitive to its molecular identity. Thus, the Ala-14 peptide is largely unfolded in solution, notwithstanding that the high resolution crystal structure of the Ala-14 trimer shows specific tertiary packing among helices. This conclusion is in accord with the predicted loss of $\sim 15.6$ kcal/mol in net hydrophobic stabilization energy of the triple-helical structure associated with alanine substitutions (26) and the notion that the $–CH_2$ side chain is too short to form stable nonpolar clusters (39).

**Intertrimer Packing Interactions**—What interactions might then explain the precise native-like assembly of Ala-14 helical trimers in the solid state? One clue comes from the observation that crystals of the Ala-14 peptide have exceptionally low values of solvent content (26.5%) and crystal volume/unit of protein molecular mass (1.70 Å³/dalton). The crystal structure displays a distinctive packing arrangement in which layers of parallel triple helices are situated closely along the helical axis with respect to each other (Fig. 3A). In the direction of the helical axis, there is little space between the C-terminal end of one triple-helical molecule and the N-terminal end of the next (Fig. 3B). This compact arrangement makes it possible to create sufficient intertrimer contacts to stabilize the crystal lattice. In effect, the lattice provides helix-terminating interactions that anchor the ends of successive triple helices in the crystal. Residues Ser-1B, Ser-2B, Asn-3A, Asn-3B, Asn-3C, Ala-4A, Lys-5A, Asp-7B, Gln-14B, Lys-19B, Asn-28A, Ser-32A, Asp-33A, Lys-38B, Asp-39B, Arg-47A, Asp-49A, Asp-49B, Asn-50A, Asn-50C, Ala-51A, and Ala-51C are positioned such that the $–CH_2$ side chain is too short to form stable nonpolar clusters (39).
partners for exposed terminal amide protons and carboxyl oxygens (40). The helix capping and solvation of the ends presumably account in part for the presence of a high resolution molecular structure in the Ala-14 trimer under crystallization conditions.

A second unique feature of Ala-14 in the crystalline state is the presence of continuous strips of intertrimer interactions along the helical axis. These appear to be critical in determining the compact lateral assembly of Ala-14 trimers (Fig. 5A).

The network of extended interstrand connections among peptide atoms and ordered water molecules includes the following: (i) 2 salt bridges and 16 charge-stabilized hydrogen bonds between the adjacent triple helices/trimer (Fig. 5, B–E, Asp-12C/Arg-31C* and Arg-43C/Ser-11B*) and (ii) water-mediated hydrogen-bonding interactions between polar or charged side chains and backbone carbonyl or amide groups also occur frequently (Fig. 5, A and C–E). Of the 186 water molecules added in the crystal structure, 154 (82%) are involved in direct interactions (within 3.25 Å) with the peptide. The average B-factor value of the structured water molecules, 28.2 Å², is not much higher than the 21.3 Å² B-factor of protein side-chain atoms. The hydration of the Ala-14 trimer seems to be dominated by its involvement in water-mediated interchain hydrogen bonds. The combination of specific intertrimer electrostatic interactions with capping effects at the helix termini suffices to lock a labile Ala-14 solution structure into a defined native state in the crystal.

Polar Interactions—Proteins acquire their spatial structures under the influence of a manifold of noncovalent interactions among backbone and side-chain atoms including the hydrophobic effect, van der Waals interactions, and hydrogen bonds and other electrostatic interactions. Whereas the relative importance, selectivity, and context dependence of these forces remain imperfectly understood, the burial of hydrophobic surfaces is thought to play a major role in governing protein architecture and folding, at least in relatively small protein folds (3, 33–35, 41–44). Ala-14 provides a striking example of how a native structure can still form once this localized driving force has been severely attenuated. In contrast to proposals that solvent-exposed electrostatic interactions contribute only marginally to protein stability, we find that an array of weak polar interactions distributed along the chain can impose sufficient conformational constraints to stabilize the alanine-zipper trimer.

Buried alanine side chains provide a small van der Waals component to the thermodynamic stability of the alanine-zipper trimer. The residual enthalpic interactions among the core Ala side chains are not sufficient to maintain the alanine-zipper fold in solution. Our finding that the covalently tethered Ala-14 trimer analog fails to form a triple-helical structure in solution indicates that alanine-zipper formation is not solely opposed by the loss of translational conformational entropy. It would appear that favorable enthalpy contributions from residual van der Waals packing, backbone hydrogen-bonding, and...
the network of solvent interactions guide and stabilize the formation of the alanine-zipper structure during crystallization. The precision of this process is reflected by the high resolution of the crystal structure (1.3 Å) with low temperature factors (16.6 and 28.9 Å² for protein and solvent atoms, respectively).

Biological Implications—This process of triple-helical assembly in the crystalline state is to our knowledge the first example of stabilization of a native state directed by long range intermolecular rather than short range intramolecular interactions. It is tempting to speculate that extended solvent-mediated interactions such as what we describe here mediate the assembly of the coiled-coil rod of myosin to form filaments in muscle. There is growing evidence that a class of proteins referred to as intrinsically unstructured proteins are unfolded and devoid of native tertiary structure under normal solution conditions yet can self-assemble into highly organized aggregates that play a causative role in the pathogenesis of human diseases including Alzheimer’s and the transmissible spongiform encephalopathies (45, 46). Understanding the fundamental nature of the mechanism of such large-scale assembly processes is a critical step in the development of strategies to prevent and treat these incurable diseases. By analogy with Ala-14, long range intermolecular interactions might entail formation of solvent-mediated interactions that promote assembly of the proteins into oligomeric fibrils.

Acknowledgments—We thank Dr. Michael Becker for assistance at the Synchrotron Beamline, Xiuwen Ma for peptide purification, and Dr. Neville Kallenbach for critical reading of the paper.

REFERENCES
1. Kim, P. S., and Baldwin, R. L. (1982) Annu. Rev. Biochem. 59, 631–660
2. Matthews, C. R. (1993) Annu. Rev. Biochem. 62, 653–683
3. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O’Neil, K. T., and DeGrado, W. F. (1995) Science 270, 935–941
4. Munoz, V., and Serrano, L. (1996) Fold. Des. 1, R71–R77
5. Dobson, C. M., and Karplus, M. (1999) Curr. Opin. Struct. Biol. 9, 92–101
6. Persh, A. R. (1999) Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding, W. H. Freeman & Co., New York
7. Baldwin, R. L., and Rose, G. D. (1999) Trends Biochem. Sci. 24, 26–33
8. Plaxco, K. W., Simons, K. T., Ruczinski, I., and Baker, D. (2000) Biochemistry 39, 11177–11183
9. Rumbley, J., Hoang, L., Mayne, L., and Englander, S. W. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 105–112
10. Guerius, R., and Serrano, L. (2001) Curr. Opin. Struct. Biol. 11, 101–106
11. Bonneau, R., and Baker, D. (2001) Annu. Rev. Biophys. Biomol. Struct. 30, 173–189
12. Persh, A. R., and Daggett, V. (2002) Cell 108, 573–582
13. Bryngelson, J. D., Onuchic, J. N., Soxet, N. D., and Wolynes, P. G. (1995) Proteins Struct. Funct. Genet. 21, 167–185
14. Dill, K. A., and Chan, H. S. (1997) Nat. Struct. Biol. 4, 10–19
15. Pande, V. S., Grosberg, A., Tanaka, T., and Rokhsar, D. S. (1998) Curr. Opin.
