Zinc-Mediated Helix Capping in A Triple-Helical Protein[†,‡]

Jie Liu, Jiayin Dai, and Min Lu*

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

Received September 9, 2002; Revised Manuscript Received March 17, 2003

ABSTRACT: Specific sequence signals at α-helix termini can assist protein folding by punctuating and cueing secondary structural elements in the final native conformation. Here we report the crystallization of a 56-residue alanine-containing peptide, denoted Ala-10_{56}, in the presence of Zn^{2+}. The 1.7 Å crystal structure shows that Ala-10_{56} forms a parallel trimeric coiled coil with three zinc ions anchoring distinct capping conformations at the amino-terminal ends of the three helices. In each polypeptide chain, the free α-amino nitrogen and carbonyl oxygen of the amino-terminal Ser residue coordinate to a Zn^{2+} ion to form a five-membered chelate, and the *syn*-unidentate interaction of the Asp7 side chain with the Zn^{2+} cation leads to the formation of a unique docking arrangement for helix capping. Moreover, the coordination of the zinc ion involves a neighboring trimer molecule in the crystal. Consequently, the crystal contacts are stabilized by carboxylate−Zn^{2+} interactions between four Ala-10_{56} trimers in the crystal lattice. The observed synergy between the protein−zinc ion recognition and the helix-packing arrangements would contribute to the conformational specificity of the Ala-10_{56} trimer.

Accurate prediction of side-chain packing and its influence on the native tertiary conformation are important objectives in modern protein structure and design efforts. Side-chain shape in the hydrophobic cores of proteins, charged side chains on the protein surface, buried polar residues, and secondary structure punctuation motifs have all been recognized as major determinants of structure and stability in helical globular proteins (*1*−*15*). It is generally believed that stereochemical punctuation marks encoded in protein sequences play a role in the grammar and syntax of protein folding (*15*). Since the mean helix length in soluble globular proteins is only about 10 residues (*16*, *17*), the four terminal amide protons and carboxyl oxygens that present groups with unsatisfied hydrogen bonds comprise a large fraction of the total hydrogen-bonding potential (*12*, *18*). Perhaps as a direct consequence, isolated α-helices are marginally stable structures and unfold easily with temperature (*19*, *20*). Stabilizing effects involving helix start and stop signals thus can provide a strong thermodynamic driving force for α-helix formation in proteins. Formulation of guidelines for both peptide design and protein structure prediction will require a detailed understanding of local structural features that stabilize helix termini.

Characteristic sequence motifs referred to as N-cap and C-cap have been identified at the N- and C-terminal ends of helices, respectively, in many natural proteins (*12*, *13*). The residues at the amino terminus of the α-helix are termed ···N″−N′−N-cap−N_1−N_2−N_3−N_4···, where N-cap is the first residue with an *i*, *i* + 4 helical hydrogen bond. Similarly, the residues at the carboxyl terminus are termed ···C_4−C_3−C_2−C_1−C-cap−C′−C″···. An α-helix can be stabilized by a N-terminal polar residue at the N-cap position with the standard free energy change of about 2 kcal/mol, as the side chain of Asn, Asp, Thr, or Ser directly hydrogen-bonds to exposed backbone amide protons (*12*, *21*−*30*). These hydrogen-bonded interactions in a protein context are often associated with the hydrophobic interaction between an apolar residue at the N′ position and nonpolar side chains within the helix (*29*, *31*). On the other hand, an α-helix can be effectively capped at the C-terminus either by Gly at left-handed helical (α_L) conformations or by Pro at extended (E) conformations (*32*−*26*). Stabilizing C-cap effects in both the α_L- and E-terminated helices are also thought to reflect direct hydrogen-bond formation between the backbone and exposed helical carbonyl oxygens as well as the interaction between hydrophobic residues before and after C-cap (*32*, *34*, *36*, *37*).

Experimental studies aimed at understanding the role of helix-capping motifs in marking the beginning and end of secondary structure have been hampered by the tendency of helix termini to fray in model peptides. Most structural information on helix initiation and termination comes from α-helices connected to turns (*12*, *13*). Some of these difficulties are minimized in the coiled coil, a protein structural motif, which consists of two or more α-helices wrapped around each other with a superhelical twist (*38*). Coiled-coil sequences contain a characteristic seven-residue repeat designated by the letters *a* through *g* (*39*). Interactions involving hydrophobic residues at positions *a* and *d* are a dominant factor in the stabilization of coiled coils, and charged side chains at positions *e* and *g* also contribute through electrostatic interactions (*39*−*42*). The coiled-coil motif appears to offer the simplest system for studying the helix-capping interactions that govern protein folding and stability.

[†] This work was supported by National Institutes of Health Grant AI42382 and by Irma T. Hirschl Trust. J.L. was supported by a postdoctoral fellowship from the Norman and Rosita Winston Foundation.

[‡] Atomic coordinates have been deposited in the Protein Data Bank (entry 1JCC).

* To whom correspondence should be addressed. Phone: (212) 746−6562. Fax: (212) 746−8875. E-mail: mlu@mail.med.cornell.edu.

10.1021/bi026828a CCC: $25.00
Published on Web 04/19/2003

Small ligands such as a metal ion serve chemical, structural, catalytic, and regulatory roles in a variety of different biological systems including myoglobin, carbonic anhydrase, transcription factors, and steroid receptors (*43*, *44*). Metal-binding sites are generally in the interior of the protein and are composed of a number of chelating groups that surround and coordinate simultaneously to the metal ion with a well-defined stereochemical preference (*45*, *46*). Protein−metal ion interactions therefore require proper chemical composition and proper stereochemistry of the metal−ligand environment. Protein−metal ion recognition and discrimination are thus increasingly used as tools in protein engineering and protein structure prediction (reviewed in refs *47* and *48*). Here we show that an α-helical trimeric coiled coil (named Ala-10_{56}) in which most of the buried bulky hydrophobic amino acids have been substituted by alanine contains three Zn^{2+}-binding sites at the N-terminal ends of the three helices. Each Zn^{2+} ion binds to a specific constellation of one α-amino nitrogen atom and four oxygen atoms (from a backbone carbonyl, two carboxylates, and a water molecule). As a result, the side chain of the N-terminal Ser residue is oriented toward each helix axis to form a characteristic L-shaped cap structure. Additionally, the zinc ions cross-link four Ala-10_{56} trimer molecules in the crystal lattice. Thus, Zn^{2+}-mediated helix capping provides a sensitive connector linking α-helix formation to the developing tertiary and quaternary structure of a protein.

MATERIALS AND METHODS

Protein Production and Purification. The sequence of the Ala-10_{56} peptide is S−S−N−A−K−I−D−Q−L−S−S−D−A−Q−T−A−N−A−K−A−D−Q−A−S−N−D−A−N−A−A−R−S−D−A−Q−A−A−K−D−D−A−A−R−A−N−Q−R−L−D−N−M−A−T−K−Y−R, with *a* positions singly underlined and *d* positions doubly underlined. Plasmid pAla10_{56} encoding Ala-10_{56} was generated by site-directed mutagenesis of pAla7 (refs *6* and *49*). The peptide was expressed in *Escherichia coli* strain BL21-(DE3)/pLysS (Novagen) at 20 °C for 12 h. Cells were lysed by ice-cold glacial acetic acid. The bacterial lysate was centrifuged (35 000*g* for 30 min) to separate the soluble fraction from inclusion bodies. The soluble fraction containing Ala-10_{56} was dialyzed into 5% acetic acid overnight at 4 °C. The peptide was purified from the soluble fraction to homogeneity by reverse-phase HPLC (Waters, Inc.) on a Vydac C-18 preparative column, using a water−acetonitrile gradient in the presence of 0.1% trifluoroacetic acid, and lyophilized. The identity of the peptide was confirmed by mass spectrometry. Peptide concentrations were determined by using tyrosine absorbance at 280 nm in 6 M guanidinium hydrochloride (GdmCl) (*50*).

Crystallization and Data Collection. Ala-10_{56} was crystallized at 4 °C using the hanging-drop vapor diffusion method. A 10 mg/mL protein solution containing 50 mM zinc acetate was mixed 1:1 with a reservoir solution and allowed to equilibrate against the reservoir solution. Crystals were obtained under several conditions using crystallization kits I and II (Hampton Research, Riverside, CA). These original crystals diffracted to only 3.6 Å resolution. Well-diffracting crystals were grown from 0.1 M sodium acetate, pH 5.6, 0.1 M zinc acetate, and 18% poly(ethylene glycol) 8000. Crystals grew as prisms to a maximum size of 0.1 × 0.1 × 0.2 mm^3 within 1−2 weeks. The crystals belong to the space group *C*2 with a trimer (1.7 Å resolution) in the asymmetric unit ($a = 62.39$ Å, $b = 33.32$ Å, $c = 71.98$ Å, $\alpha = \gamma = 90°$, $\beta = 102.55°$) or with two trimers (1.9 Å resolution) in the asymmetric unit ($a = 139.00$ Å, $b = 34.61$ Å, $c = 57.80$ Å, $\alpha = \gamma = 90°$, $\beta = 93.87°$). Triclinic crystals with the space group *P*1 ($a = 32.72$ Å, $b = 34.61$ Å, $c = 71.50$ Å, $\alpha = 76.21°$, $\gamma = 86.90°$, $\beta = 61.91°$) were also obtained under identical conditions. In comparison to the *C*2 form, the triclinic crystals diffracted poorly (between 2.7 and 2.9 Å resolution). For data collection, crystals were transferred to a drop containing 15% (v/v) glycerol in the corresponding mother liquor, mounted in nylon loops, and frozen in liquid nitrogen. Diffraction data were recorded at 100 K on an ADSC Quantum-4 CCD detector at beamline X26C at the Brookhaven National Laboratory National Synchrotron Light Source. Diffraction intensities were integrated by using DENZO and SCALEPACK software (*51*) and reduced to structure factors with the program TRUNCATE from the CCP4 program suite (*52*).

Structure Determination and Refinement. Initial phases were determined by molecular replacement with the program AMoRe (*53*) using the Lpp-56 trimer as a search model (*30*). A combined rotation-translation search (with 15−2.5 Å data) produced an unambiguous solution for Ala-10_{56} (correlation coefficiency = 57.0%; *R*-factor = 52.8%). Density interpretation and model building were done with the program O (*54*). Ten percent of the reflections were excluded from refinement and used as a cross-validation set. Crystallographic refinement of the structure was done using the program CNS 1.0 (ref *55*). Refinement steps involved rigid body refinement and group *B*-factor refinement, followed by iterative rounds of simulated annealing, positional refinement, and individual *B*-factor refinement, as well as model rebuilding using O. When the R_{free} reached approximately 35%, water molecules were added manually. The resolution of the data used in refinement was gradually extended to 1.7 Å. Throughout the refinement, bulk-solvent and overall anisotropic *B*-factor corrections were utilized, and low-resolution data were included. Noncrystallographic symmetry restraints were not used in the final refinement. Water molecules were chosen using distance and geometry criteria from main-chain carbonyl groups, amide nitrogen atoms, and side-chain polar groups. Water molecules were not positionally restrained during refinement. The final structure was confirmed using simulated annealing omit maps, omitting one residue at a time and heating the structure to 1000 K before cooling. The quality of coordinates was examined by PROCHECK (*56*). All the amino acids but a few at the helix termini were in the most favored regions of the Ramanchandran space. Asn50 of chain A lies in the additionally allowed region of the Ramanchandran plot. Ala52, Thr53, Lys54, Tyr55, Arg56 of chain A, Thr53, Lys54, Tyr55, Arg56 of chain B, and Lys54, Tyr55, Arg56 of chain C were left out of the model, as they could not be seen clearly in the electron density maps.

Circular Dichroism Spectroscopy. Circular dichroism experiments were performed on an Aviv 62A DS circular dichroism spectrometer (Aviv Associates, Lakewood, NJ) as described previously (*57*). The wavelength dependence of molar ellipticity, [θ], was monitored at 0 °C from 200 to

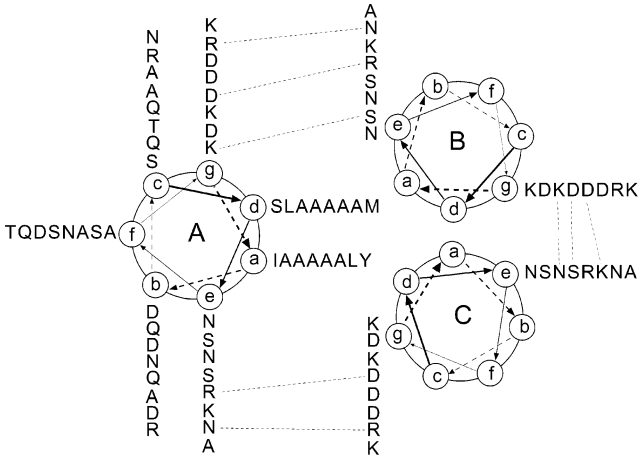

FIGURE 1: Helical wheel representation of the Ala-10_{56} trimer sequence. The positions in the 4-3 hydrophobic repeat are labeled *a* through *g*. Each helix begins at a *c* position. View is from the amino terminus. Dashed lines between helices indicate residues that are within hydrogen-bonding and ion-pairing distance in the crystal structure.

260 nm in a 0.1-cm path length cuvette with a 5-s averaging time. Thermal stability was determined by monitoring the change in the ellipticity at 222 nm as a function of temperature. Thermal melts were performed in 2 °C increments with an equilibration time of 2 min at the desired temperature and an integration time of 30 s.

RESULTS

Crystallization of Ala-10_{56} in the Presence of Zn^{2+}. We have engineered a 56-residue alanine-containing peptide, denoted Ala-10_{56}, based on the sequence template of the Lpp-56 trimerization domain as a model coiled coil for studying how interfacial interactions between α-helices affect the structure and stability of small helical proteins. This peptide contains 10 alanine residues at the central *a* and *d* positions and two bulky hydrophobic amino acids each at the most N- or C-terminal *a* and *d* positions (Figure 1). It is reasoned that Ala-10_{56} should have low inherent thermodynamic stability, because the -CH_3 side chain of the core alanines has relatively low hydrophobicity. Indeed, circular dichroism studies showed that the peptide is predominantly unfolded in aqueous solution (data not shown). To explore the effects of metal ions on the crystallization of the Ala-10_{56} peptide, we have performed extensive crystallization trials using Ala-10_{56} in the presence of Zn^{2+}. Zinc was chosen because its coordination chemistry is versatile, exhibiting variation in number, charge, structure, and amino acid composition (*58*). Several crystal forms were obtained, but the best diffracting crystals were grown by using 18% polyethylene glycol 8000, 0.1 M sodium acetate, pH 5.6, and 0.1 M zinc acetate at low temperature. The Ala-10_{56} peptide appears to require Zn^{2+} for satisfactory crystallization.

Crystal Structure of the Ala-10_{56} Trimer. Self-rotation function calculations and a 35.1% solvent content in the Ala-10_{56} crystals suggested one trimer in the asymmetric unit (*59*), wherein monomers are related by an approximate noncrystallographic 3-fold axis. The crystal structure was determined by molecular replacement. The current model has been refined to a crystallographic *R*-factor of 19.2% and a free *R*-factor of 22.3% with excellent geometry using data between 50 and 1.7 Å resolution. Details of the data collection and refinement statistics are given in Table 1. The model consists of 156 amino acid residues and incorporates 234 water molecules and three Zn^{2+} ions. Residues 52−56 of helix A, 53−56 of helix B, and 54−56 of helix C are not visible in our density maps and therefore must be disordered. A portion of the $2F_o - F_c$ electron density map and the refined atomic model are shown in Figure 2a.

Table 1: Crystallographic Data Collection and Refinement Statistics

Data Collection	
resolution (Å)	50−1.7
observed reflections	69970
unique reflections	15821
redundancy	2.5 (2.5)[a]
$I/\sigma(I)$	7.7 (3.3)[a]
completeness (%)	97.6
R_{merge} (%)[b]	7.9
Refinement	
resolution (Å)	50−1.7
reflections	14227
R_{cryst} (%)[c]	19.2
R_{free} (%)[c]	22.3
protein nonhydrogen atoms	1331
water molecules	234
zinc ions	3
RMS Deviations from Ideal Geometry	
bond lengths (Å)	0.006
bond angles (deg)	0.9
torsion angles (deg)	15.5
average temperature factors (Å^2)	
all atoms	22.1
protein atoms	19.0
solvent atoms	35.7

[a] Values in parentheses correspond to the highest resolution shell 1.76−1.70 Å. [b] $R_{merge} = \sum\sum_j|I_j(hkl) - \langle I(hkl)\rangle|/\sum|\langle I(hkl)\rangle|$, where I_j is the intensity measurement for reflection *j* and $\langle I\rangle$ is the mean intensity over *j* reflections. [c] R_{cryst} (R_{free}) $= \sum||F_o(hkl)| - |F_c(hkl)||/\sum|F_o(hkl)|$, where F_o and F_c are observed and calculated structure factors, respectively. No sigma-cutoff was applied. Ten percent of the reflections were excluded from refinement and used to calculate R_{free}.

The Ala-10_{56} peptide forms a parallel triple-helical structure with the same hallmark characteristics observed in the structures of conventional three-stranded coiled coils (Figure 2b). The left-handed superhelix creates a cylinder that is ∼78 Å long and ∼20−24 Å wide. Ten alanines, two leucines, one isoleucine, and one methionine at positions *a* and *d* from each monomer are buried at the center of the trimer. The side chains of these core residues show the characteristic "acute knobs-into-holes" packing, as seen in the crystal structure of the GCN4-pII isoleucine-zipper trimer (*1*, *60*). On the basis of distance criteria, residues on the surface of the Ala-10_{56} trimer appear to form interhelical *e* to *g* electrostatic interactions, a trait common to coiled-coil structures (see Figure 1). Because of asymmetric crystal contacts, the three individual helices in the Ala-10_{56} structure have slight differences in conformation and degree of order, with the B helix having a higher overall temperature factor (20.8 Å^2) than the A and C helices (18.3 and 17.8 Å^2, respectively). The root-mean-square (rms) deviations between the individual helices in the noncrystallographic trimer vary from 0.085 Å (helices A and B) to 0.49 Å (helices A and C).

Protein-Zn^{2+} Interactions in the Ala-10_{56} Structure. During the refinement of the Ala-10_{56} structure, 40.8, 34.9, and 47.3 σ peaks appeared at the amino termini of helices A, B, and

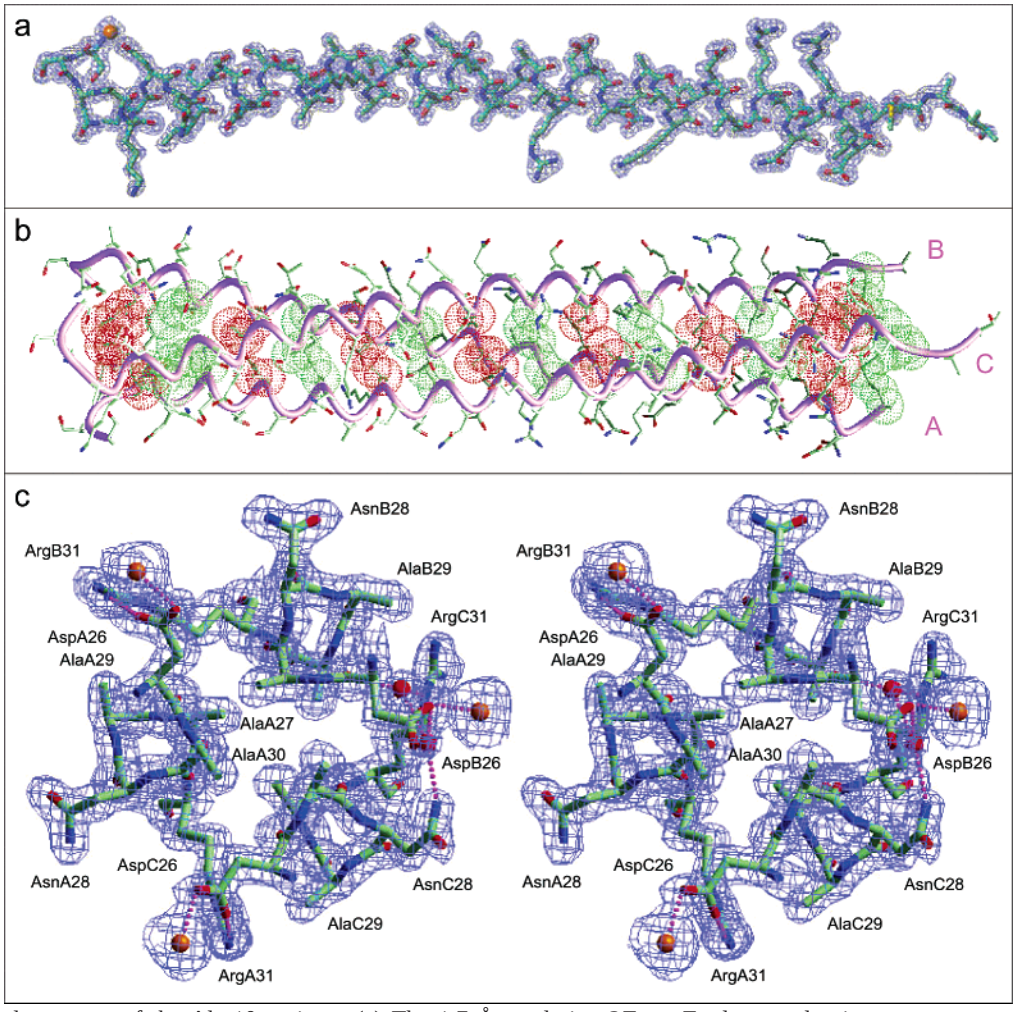

FIGURE 2: Crystal structure of the Ala-10_{56} trimer. (a) The 1.7 Å resolution $2F_o - F_c$ electron density map superposed onto the final refined structure. A side view of the C chain (residues 1−53) is shown. A Zn^{2+} ion at the amino terminus is depicted as a yellow ball. The map was contoured at 1.2 σ. (b) Side view of the trimer. The van der Waals surfaces are colored red for residues at *a* positions and green for residues at *d* positions. The helix chains are labeled A, B, and C at the carboxyl termini. (c) The cross section of the trimer at the Ala27 and Ala30 layers as visualized in a stereoview of the final $2F_o - F_c$ electron density map contoured at 1.2 σ. Zn^{2+} ions and water molecules are represented as yellow and red spheres, respectively. Figures were generated by using the Turbo-Frodo, GRASP, and SETOR programs (*78*−*80*).

C, respectively, in our difference Fourier maps which could not be adequately accounted for by water molecules. On the basis of the size of each peak and the components of the crystallization conditions, these electron densities were modeled as Zn^{2+} ions. The individual *B*-factor refinements of the modeled zinc ions converge at 20.2, 24.2, and 16.9 $Å^2$ for the A, B, and C helices, respectively (Figure 2c). These values are comparable with the average *B*-factors of the five interacting nitrogen and oxygen atoms of protein ligands and water (25.5 $Å^2$ for helix A, 32.4 $Å^2$ for helix B, and 16.7 $Å^2$ for helix C) (Table 2). Moreover, the Zn−O distances are in the range of 2.3 to 2.7 Å (Table 2), as compared to the expected value of 2.5 Å (*61*). Thus, we identified three clear Zn^{2+}-binding sites on the N-terminal ends of the three helices in the Ala-10_{56} structure.

Figure 3 shows the detailed atomic structure around the zinc ion. In each Ala-10_{56} helix, both the free α-amino group and the backbone carbonyl of the N-terminal Ser residue coordinate to a Zn^{2+} cation to form a five-membered chelate (Figure 3b) (*62*), and the bound Zn^{2+} is coordinated by the *syn*-oriented carbonyl oxygen lone electron pair of the Asp7 side chain. As a consequence of this ligation effect, the amino terminus of the helix forms an L-shaped bend perpendicular to the helix axis. In addition, Asp26* (from an adjacent symmetry-related trimer in the unit cell) and a water molecule complete the coordination polyhedron of pentacoordinate zinc ion (Figures 3 and 4). The zinc complexation is close to ideal, forming a slightly distorted trigonal bipyramid with the main-chain C═O (Ser1) → Zn ← OH (water) angles ranging from 150° to 160°, close to the expected value of 180° **(**Table 2). Zinc binding therefore clearly plays a significant role in capping the amino termini of the Ala-10_{56} helices. In the present case, the protein−zinc ion interactions may also help define the geometric relationships of the α-helices in the crystalline state and hence lead to a stable crystal lattice.

It is remarkable that the geometry and coordination number of Zn^{2+} can vary greatly to accommodate particular protein ligands (*46*). The coordination number 4 is most commonly found in protein crystal structures, but the inner-sphere coordination number of catalytic zinc may increase to 5 as the chemistry of catalysis proceeds (*44*, *63*). Interestingly, the noncatalytic, or structural, Zn^{2+} ions in the Ala-10_{56} structure possess five direct protein ligands, with a unique coordination stereochemistry. The four ligand atoms contributed by the protein lie on separate polypeptide chains and form an unusual zinc-binding site positioned between

Table 2: Characteristics of Three Zinc-Binding Sites in the Ala-10_{56} Structure

helix	ligand	distance to Zn^{2+} (Å)	*B*-factor (Å^2)	angles (deg)	
A	N (Ser1)	2.7	30.8	N−Zn−OH	87
				N−Zn−OD1(7)	132
A	O (Ser1)	2.4	25.7	O−Zn−N	68
				O−Zn−OD1(7)	85
A	OD1 (Asp7)	2.4	24.6	OD1(7)−Zn−OH	125
				OD1(7)−Zn−OD1(C26)	91
A	OH (Wat)	2.5	28.1	OH−Zn−O	150
				OH−Zn−OD1(C26)	90
A	OD1 (AspC26)	2.3	18.5	OD1(C26)−Zn−O	93
				OD1(C26)−Zn−N	132
A	Zinc		20.2		
B	N (Ser1)	2.7	37.5	N−Zn−OH	99
				N−Zn−OD1(7)	131
B	O (Ser1)	2.3	32.9	O−Zn−N	70
				O−Zn−OD1(7)	79
B	OD1 (Asp7)	2.6	34.4	OD1(7)−Zn−OH	119
				OD1(7)−Zn−OD1(B26)	119
B	OH (Wat)	2.7	38.6	OH−Zn−O	160
				OH−Zn−OD1(B26)	83
B	OD1 (AspB26)	2.3	18.8	OD1(B26)−Zn−O	82
				OD1(B26)−Zn−N	93
B	zinc		24.2		
C	N (Ser1)	2.4	17.3	N−Zn−OH	91
				N−Zn−OD1(7)	127
C	O (Ser1)	2.3	15.1	O−Zn−N	75
				O−Zn−OD1(7)	95
C	OD1 (Asp7)	2.3	16.8	OD1(7)−Zn−OH	105
				OD1(7)−Zn−OD1(A26)	90
C	OH (Wat)	2.3	15.8	OH−Zn−O	159
				OH−Zn−OD1(A26)	89
C	OD1 (AspA26)	2.3	18.3	OD1(A26)−Zn−O	93
				OD1(A26)−Zn−N	142
C	zinc		16.9		

crystal-packing-contact trimer molecules. The N-terminal serine chelate involves a special bidentate carbonyl−zinc ion coordination; the C═O → Zn^{2+} angles lie within the range of 115−118°. This binding mode, albeit not common in Zn^{2+}-binding proteins, has been observed in zinc proteases where the Zn^{2+} atom is involved in controlling the positions of substrate-binding groups in the active site (*62*, *64*). In addition, as seen in Figure 3a, the zinc ions display a preference for in-plane interactions with the carboxylate groups of Asp7 and Asp26* via *syn*-coordination stereochemistry. This geometry is thought to optimize its interaction with the oxygen sp^2 lone electron pair and typically occurs when the zinc−oxygen distance is in the range of 2.3−2.6 Å (*61*). It should also be noted that water acts as a zinc ligand in the Ala-10_{56} structure, as the powerful nucleophilic activity of Zn^{2+}-bound water is observed in catalysis by enzymes such as carbonic anhydrase and carboxypeptidase (*65*).

Ala-10_{56} is not Stabilized by Zn^{2+} in Aqueous Solution. It is possible that the protein−Zn^{2+} interactions observed crystallographically here are utilized toward the stabilization of the labile Ala-10_{56} trimer. If so, one would anticipate that the addition of Zn^{2+} would stabilize the Ala-10_{56} peptide under native folding conditions. To address this issue, circular dichroism experiments were used to assess the effects of Zn^{2+} on the folding and stability of Ala-10_{56} in 50 mM Tris-HCl, pH 8.0, 100 mM NaCl as well as in 100 mM sodium acetate, pH 5.2, 100 mM NaCl. There is no apparent increase in the α-helical content and/or the thermal stability of Ala-10_{56} with added zinc acetate under both acidic pH and pH 8 conditions (data not shown), indicating that the peptide in solution does not bind zinc ions. Given that the binding of three zinc ions at the N-terminal helical ends in the Ala-10_{56} trimer structure necessitates a crystal packing contact (Figure 4), this result is not surprising and may well be expected.

DISCUSSION

The crystallographic view of the Ala-10_{56} peptide demonstrates that the formation of three zinc-binding sites can serve a dual role in its folding: they cap the free N-terminal ends of three helices, and promote an intermolecular packing arrangement for coiled-coil trimer formation. These results expand our structural knowledge of helix capping to the relationships among protein−zinc recognition, molecular conformation, and intermolecular association in a triple-helical context. Each helix in the 1.7 Å Ala-10_{56} crystal structure forms a distinct N-terminal bend characteristic of capping structures seen in proteins and peptides (*12*, *27*, *29*, *30*). The Zn^{2+} cation within each cap anchors a well-defined hydrogen-bonding network in a trigonal bipyramid fashion: the peptide carbonyl oxygen and the α-amino nitrogen of Ser1 complex zinc through a five-membered chelate structure, and the zinc coordination by the Asp7 side chain projects the chain in a direction orthogonal to the α-helical axis. As a result, the zinc ions shield the exposed amide main-chain groups efficiently and direct the folding of single dominant capping conformations. Moreover, the ligation of each bound Zn^{2+} to the Asp26 side chain on the surface of the neighboring molecule cross-links four Ala-10_{56} trimers in crystal contacts. It would appear that the binding of Zn^{2+} to the Ala-10_{56} peptide may greatly reduce conformational space during its search for the native fold and therefore help specify the parallel triple-helical conformation. The coupled helix-capping and crystal-packing interactions may underlie a unique mechanism for the folding and association of the metastable Ala-10_{56} trimer under crystallization conditions.

The structure of the Ala-10_{56} peptide provides an opportunity to investigate several factors to be considered in the protein engineering of zinc-binding sites. First, zinc coordination polyhedra in proteins are typically tetrahedral or distorted tetrahedral, although zinc-binding sites are varied in their numbers and geometries, depending on the structural, catalytic, or regulatory role of the metal ion. Our results show that structural zinc readily accepts a distorted pentacoordinate geometry which is required for the binding of three Zn^{2+} ions to the Ala-10_{56} trimer in the solid state. This contrasts with the presumed change of catalytic zinc from four- to five-coordination during the course of enzymatic turnover (*44*, *63*). Second, metal ions are commonly sequestered from solvent by their molecular environments. The analysis of molecular structure databases by Eisenberg and co-workers reveals the tendency of metals to bind at centers of high hydrophobic contrast (*45*). They find that hydrophilic atomic groups (containing nitrogen, oxygen, or sulfur atoms) which directly ligate the metal are embedded within a large shell of carbon-containing groups. The hydrophobic outer sphere around the metal ion is also thought to favor zinc binding since the hydrophobic environment restricts flexibility of the binding site, thereby minimizing the conformational entropy loss incurred on metal binding (*66*, *67*). The zinc ions in the Ala-10_{56} structure are exposed to bulk solvent and utilize a

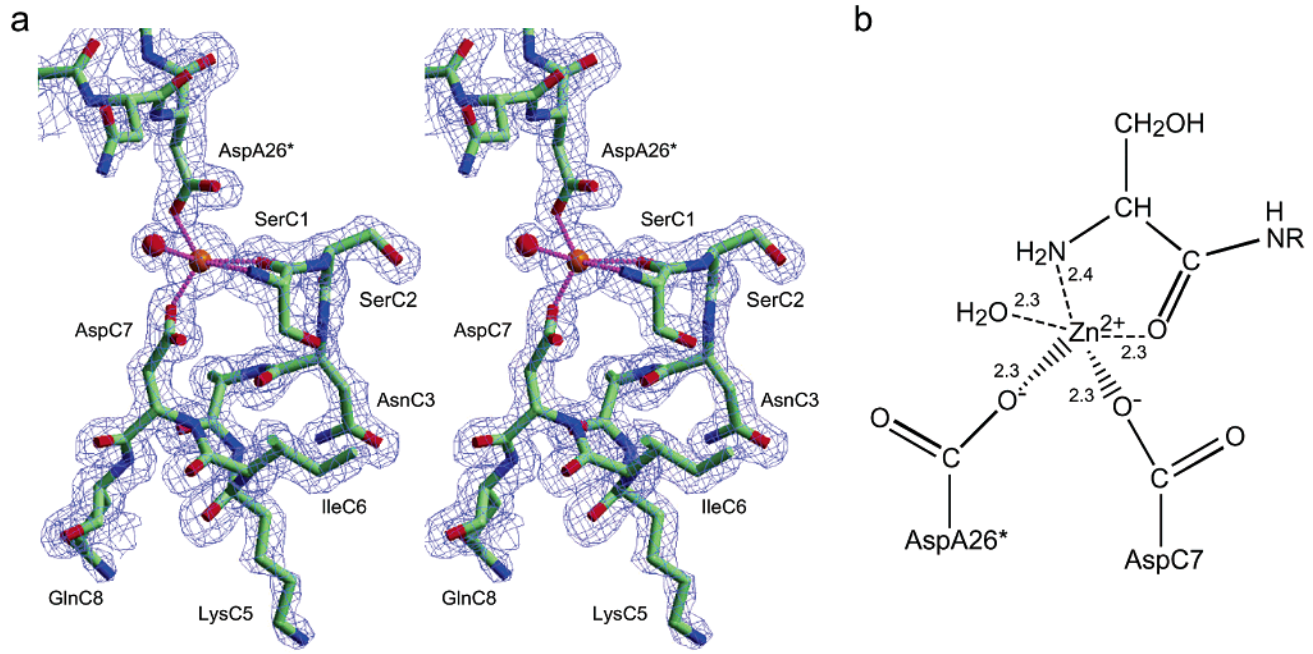

FIGURE 3: Zn^{2+}−protein ligand interactions in the Ala-10_{56} structure. (a) A portion of the 1.7 Å $2F_o - F_c$ electron density map contoured at 1.5 σ is superposed on the final model, showing stereoview of the amino terminus of the C chain (residues 1−8), a Zn^{2+} atom (yellow sphere), a water molecule (red sphere), and AspA26* (from an adjacent trimer in the crystal lattice). Figure was prepared with the program SETOR (*80*). (b) A schematic of the zinc-binding site of the C chain is shown. The amino-terminal serine engages zinc via a five-membered chelate. Bond distances are given in Å.

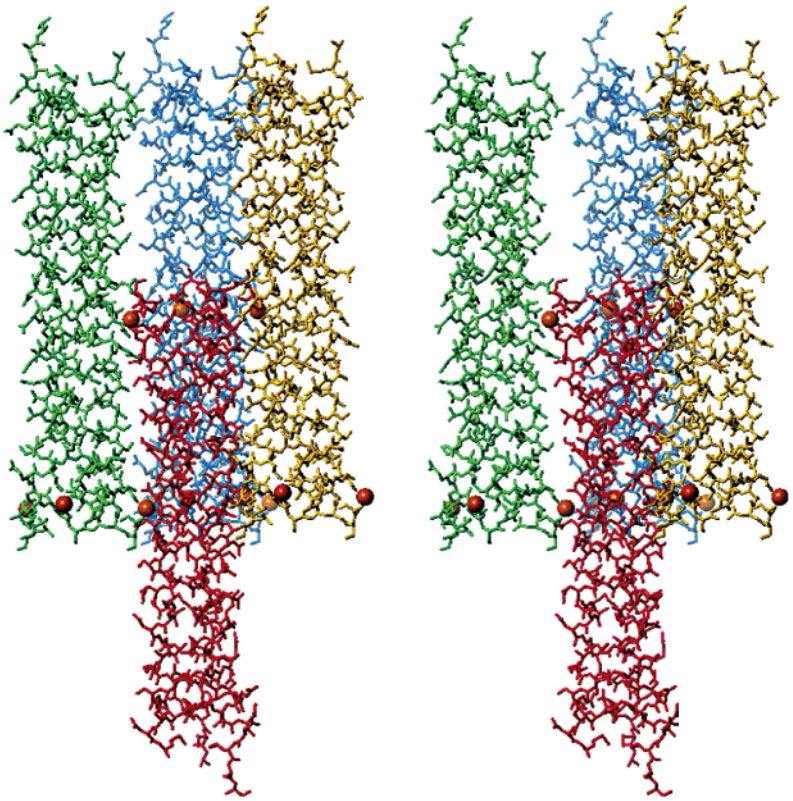

FIGURE 4: Stereoview of the Cα backbones of four Ala-10_{56} trimers in the crystal lattice. The zinc ions are shown as yellow balls in the amino termini of the α-helices. Figure was created with the program Turbo-Frodo (*78*).

water molecule as their direct ligands. In this case, it is possible that the high ionic strength of the crystallization conditions may decrease the solvation potential of the chelating groups as well as their electrostatic interactions with the zinc ion. Third, the zinc ligation within the Ala-10^{56} structure is finely tuned through the secondary, tertiary, and quaternary effects of the protein matrix. Our results are consistent with the notion that long-range protein−metal ion interactions through space (either solvent or the protein milieu) or through hydrogen-bond Coulombic interactions are important for modulating metal ion affinity, specificity, and function (reviewed in refs *46* and *48*).

Early zinc-binding studies of peptides and proteins reveal an enthalpic contribution to zinc−ligand association (*68*, *69*). Moreover, the loss of the conformational entropy of the polypeptide chain as it becomes bound to a zinc ion is not expected to surpass the entropy gain of solvent release from metal-binding site organization (*70*). Hence, the thermodynamics of protein−zinc interactions are generally thought to be dominated by substantial entropy effects, at least as zinc binds to a small peptide. The engineering of zinc-binding sites into small peptides or large proteins is thus being used to stabilize both single α-helices and protein tertiary structure (*48*, *71*−*75*); for example, 3-His carbonic anhydrase-like Zn^{2+}-binding sites were incorporated into a protein composed of four helices connected by three short loop sequences (α_4) (*73*) and a redesigned antibody-like protein (the minibody) (*75*). While the Ala-10_{56} trimer structure exhibits extensive fixed tertiary interactions that are characteristic of naturally occurring proteins, the peptide is largely unfolded under physiological conditions and is not affected by Zn^{2+} when added. Given the fact that the geometric preferences of the zinc-binding sites in the Ala-10_{56} structure are highly specific, it remains possible that the trimer formation is a consequence of crystal packing contacts. Nonetheless, our results show that zinc-mediated helix capping can contribute to the exquisite conformational specificity of the Ala-10^{56} coiled-coil trimer.

There is growing evidence that many protein domains or even full-length proteins are intrinsically unstructured. Such disordered proteins appear to play an important but poorly understood role in the pathogenesis of a wide variety of human diseases, ranging from Alzheimer's disease to spongiform encephalopathies (for review, see refs *76* and *77*). However, the conformational determinants of the unstructured proteins have not yet been defined. Understanding their unstructured nature and its role in biological processes presents an important future challenge. The Ala-10_{56} and Ala-14 peptides that undergo a folding transition only during crystallization could prove to be useful for investigating the role of quarternary interactions in determining protein folding and stability (*81*).

ACKNOWLEDGMENT

We thank Dieter Schneider for help using beamline X26C at the Brookhaven National Laboratory.

REFERENCES

1. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science 262*, 1401−1407.
2. Zhou, N. E., Kay, C. M., and Hodges, R. S. (1992) *J. Biol. Chem. 267*, 2664−2670.
3. Munson, M., O'Brien, R., Sturtevant J. M., and Regan, L. (1994) *Protein Sci. 3*, 2015−2022.
4. 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.
5. Gonzalez, L., Jr., Brown, R. A., Richardson, D., and Alber, T. (1996) *Nature Struct. Biol. 3*, 1002−1010.
6. Liu, J., Cao, W., and Lu, M. (2002) *J. Mol. Biol. 318*, 877−888.
7. O'Shea, E. K., Rutowski, R., and Kim, P. S. (1992) *Cell 68*, 699−708.
8. Betz, S., Fairman, R., O'Neil, K. T., Lear, J., and Degrado, W. F. (1995) *Philos. Trans. R. Soc. London B. Biol. Sci. 348*, 81−88.
9. Lumb, K. J., and Kim, P. S. (1995) *Biochemistry 34*, 8642−8648.
10. Gonzalez, L., Jr., Woolfson, D. N., and Alber, T. (1996) *Nat. Struct. Biol. 3*, 1011−1018.
11. Oakley, M. G., and Kim, P. S. (1998) *Biochemistry 37*, 12603−12610.
12. Presta, L. G., and Rose, G. D. (1988) *Science 240*, 1632−1641.
13. Richardson, J. S., and Richardson, D. C. (1988) *Science 240*, 1648−1652.
14. Matthews, B. W. (1995) *Adv. Protein Chem. 46*, 249−278.
15. Aurora, R., and Rose, G. D. (1998) *Protein Sci. 7*, 21−38.
16. Barlow, D. J., and Thornton, J. M. (1988) *J. Mol. Biol. 201*, 601−619.
17. Penel, S., Morrison, R. G., Mortishire-Smith, R. J., and Doig, A. J. (1999) *J. Mol. Biol. 293*, 1211−1219.
18. Pauling, L., Corey, R. B., and Branson, H. R. (1951) *Proc. Natl. Acad. Sci. U.S.A. 37*, 205−210.
19. Marqusee, S., and Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898−8902.
20. Lyu, P. C., Liff, M., Marky, L. A., and Kallenbach, N. R. (1990) *Science 250*, 669−673.
21. Serrano, L., and Fersht, A. R. (1989) *Nature 342*, 296−299.
22. Bell, J. A., Beckel, W. J., Sauer, U., Baase, W. A., and Matthews, B. M. (1992) *Biochemistry 31*, 3590−3596.
23. Serrano, L., Neira, J.-L., Sancho, J., and Fersht, A. R. (1992) *Nature 356*, 453−455.
24. Lyu, P. C., Wemmer, D. E., Zhou, H. X., Pinker, R. J., and Kallenbach, N. R. (1993) *Biochemistry 32*, 421−425.
25. Forood, B., Feliciano, E. J., and Nambiar, K. P. (1994) *Proc. Natl. Acad. Sci. U.S.A. 90*, 838−842.
26. Doig, A. J., Chakrabartty, A., Klingler, T. M., and Baldwin, R. L. (1994) *Biochemistry 33*, 3396−3403.
27. Gong, Y., Zhou, H. X., Guo, M., and Kallenbach, N. R. (1995) *Protein Sci. 4*, 1446−1456.
28. Doig, A. J., and Baldwin, R. L. (1995) *Protein Sci. 4*, 1325−1336.
29. Lu, M., Shu, W., Ji, H., Spek, E., Wang, L., and Kallenbach, N. R. (1999) *J. Mol. Biol. 288*, 743−752.
30. Shu, W., Liu, J., Ji, H., and Lu, M. (2000) *J. Mol. Biol. 299*, 1101−1112.
31. Munoz, V., Blanco, F. J., and Serrano, L. (1995) *Nat. Struct. Biol. 2*, 380−385.
32. Schellman, C. (1980) in *Protein Folding* (Jaenicke, R., Ed.) pp 53−61, Elsevier/North-Holland, New York.
33. Milner-White, E. J. (1988) *J. Mol. Biol. 199*, 503−511.
34. Aurora, R., Srinivasan, R., and Rose, G. D. (1994) *Science 264*, 1126−1130.
35. Prieto, J., and Serrano, L. (1997) *J. Mol. Biol. 274*, 276−288.
36. Gunasekaran, K., Nagarajaram, H. A., Ramakrishnan, C., and Balaram, P. (1998) *J. Mol. Biol. 275*, 917−932.
37. Kumar, S., and Bansal, M. (1998) *Proteins: Struct. Funct. Genet. 31*, 460−476.
38. Cohen, C., and Parry, D. A. D. (1990) *Proteins: Struct. Funct. Genet. 7*, 1−15.
39. McLachlan, A. D., and Stewart, M. (1975) *J. Mol. Biol. 98*, 293−304.
40. Hodges, R. S., Sodek, J., Smillie, L. B., and Jurasek, L. (1972) *Cold Spring Harb. Symp. Quant. Biol. 37*, 299−310.
41. Parry, D. A. D. (1982) *Biosci. Rep. 2*, 1017−1024.
42. Lupas, L., Van Dyke, M., and Stock, J. (1991) *Science 252*, 1162−1164.
43. Williams, R. J. P. (1985) *Eur. J. Biochem. 150*, 231−248.
44. Matthews, B. W. (1988) *Acc. Chem. Res. 21*, 333−340.
45. Yamashita, M. M., Wesson, L., Eisenman, G., and Eisenberg, D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5648−5652.
46. Christianson, D. W. (1991) *Adv. Protein Chem. 42*, 281−355.
47. Regan, L. (1995) *Trends Biochem. Sci. 20*, 280−285.
48. DeGrado, W. F., Summa, C. M., Favone, V., Nastri, F., and Lombardi, A. (1999) *Annu. Rev. Biochem. 68*, 779−819.
49. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
50. Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.
51. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
52. CCP4. (1994) *Acta Crystallogr. Sect. D 50*, 760−763.
53. Navaza, J. (1994) *Acta Crystallogr. Sect. A 50*, 157−163.
54. Jones, T. A., Zou, J.-Y., and Cowan, S. W. (1991) *Acta Crystallogr. Sect. A 47*, 110−119.
55. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D 54*, 905−921.

56. Laskowski, R. A., MacArthur, M. V., Moss, D. D., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
57. Shu, W., Ji, H., and Lu, M. (1999) *Biochemistry 38*, 5378−5385.
58. Cotton, F. A., and Wilkinson, G. (1980) *Advanced Inorganic Chemistry*, 4th ed., p 14, Wiley, New York.
59. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
60. Crick, F. H. C. (1953) *Acta Crystallogr. 6*, 689−697.
61. Carrell, C. J., Carrell, H. L., Erlebacher, J., and Glusker, J. P. (1988) *J. Am. Chem. Soc. 110*, 8651−8656.
62. Holmes, M. A., and Matthews, B. W. (1981) *Biochemistry 20*, 6912−6920.
63. Christianson, D. W., and Lipscomb, W. N. (1989) *Acc. Chem. Res. 22*, 62−69.
64. Christianson, D. W., and Lipscomb, W. N. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7568−7572.
65. Merz, K. M. (1990) *J. Mol. Biol. 214*, 799−802.
66. Cram, D. (1986) *Angew. Chem., Int. Ed. Engl. 25*, 1039−1057.
67. Serpersu, E. H., Shortle, D., and Mildvan, A. S. (1986) *Biochemistry 25*, 68−77.
68. Henkens, R. W., and Sturtevant, J. M. (1968) *J. Am. Chem. Soc. 90*, 2669−2676.
69. Henkens, R. W., Watt, G. D., and Sturtevant, J. M. (1969) *Biochemistry 8*, 1874−1878.
70. Berg, J. M. (1989) *Prog. Inorg. Chem. 37*, 143−185.
71. Ghadiri, M. R., and Choi, C. (1990) *J. Am. Chem. Soc. 112*, 1630−1632.
72. Ruan, F. Q., Chen, Y. Q., and Hopkins, P. B. (1990) *J. Am. Chem. Soc. 112*, 9403−9404.
73. Handel, T., and DeGrado, W. F. (1990) *J. Am. Chem. Soc. 112*, 6710−6711.
74. Regan, L., and Clarke, N. D. (1990) *Biochemistry 29*, 19878−10883.
75. Pessi, A., Bianchi, E., Crameri, S., Venturni, S., Tramontano, A., and Sollazzo, M. (1993) *Nature 362*, 367−369.
76. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321−331.
77. Dobson, C. M. (2002) *Philos. Trans. R. Soc. London Sect. B 356*, 133−145.
78. Jones, T. A. (1978) *J. Appl. Cryst. 11*, 268−272.
79. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins: Struct. Funct. Genet. 11*, 281−296.
80. Evans, S. V. (1993) *J. Mol. Graph. 11*, 134−138.
81. Liu, J., and Lu, M. (2002) *J. Biol. Chem. 277*, 48708−48713.

BI026828A