Association-induced folding of globular proteins

VLADIMIR N. UVERSKY*†‡, DANIEL J. SEGEL§, SEBASTIAN DONIACH§ AND ANTHONY L. FINK†¶

*Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064; †Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia; and §Department of Physics, Stanford University, Stanford, CA 94305

ABSTRACT It has generally been assumed that the aggregation of partially folded intermediates during protein refolding results in the termination of further protein folding. We show here, however, that under some conditions the association of partially folded intermediates can induce additional structure leading to soluble aggregates with many native-like properties. The amount of secondary structure in a monomeric, partially folded intermediate of staphylococcal nuclease was found to double on formation of soluble aggregates at high protein or salt concentrations. In addition, more globularity, as determined from Kratky plots of small-angle x-ray scattering data, was also noted in the associated states.

Protein association (aggregation) is a major biomedical and biotechnological problem. Diseases such as the amyloidoses, prion diseases, and cataracts are caused by protein association, as are several diseases involving inclusion bodies or other amorphous deposits (e. g. inclusion body myositis, light chain deposition disease, Huntington's disease) (1, 2). Formation of inclusion bodies is a common problem in the production of recombinant proteins (3–5). The storage and delivery of protein drugs are also often complicated by the association process. Protein refolding is often also accompanied by aggregation, especially at higher protein concentrations, which has been attributed to the association of partially folded intermediates (6–11).

Partially folded intermediates have been detected under both transient and equilibrium protein folding conditions (12–15). The structural properties of such intermediates are diverse, ranging from substantially unfolded to almost as structured as the native state (15). We have previously established that acid-unfolded staphylococcal nuclease (pH 2.5) can be transformed into one of three partially folded intermediates, depending on the nature and concentration of anions added to neutralize the repulsive effect from the net positive charges in the acid-unfolded polypeptide chain (16–17). Such partially folded species have a strong propensity to aggregate: for both apomyoglobin and staphylococcal nuclease (SNase) we have observed that the propensity to associate decreased with increasing structural content, and the least structured intermediates were monomeric only at rather dilute concentrations.

We present here results of the effect of association on the structural properties of a partially folded intermediate of SNase.

MATERIALS AND METHODS

SNase. SNase was grown and purified from a cloned gene kindly supplied by D. Shortle (Johns Hopkins Univ. School of Medicine). The homogeneity of the protein samples was checked electrophoretically by using the PhastSystem (Phar-macia). Protein concentrations were determined from the published molar extinction coefficient. Salt titrations at low pH were carried out by making a series of solutions of desired salt concentration and adjusting the pH value with HCl or NaOH. Samples were incubated overnight before measurements. Sodium sulfate was from Sigma. The pH values were measured with a microcombination glass electrode [Microelectrodes (Londonderry, NH), model MI-410].

Circular Dichroism (CD). CD data were collected on an Aviv 60DS CD spectrometer at 23°C. Far-UV spectra from 185–260 nm were collected with 5 sec per point signal averaging; θ_{222} measurements were taken in kinetics mode with the signal averaged over 120 sec. A 0.1-cm fixed pathlength cell was used. The fraction of native secondary structure was estimated as $N_{ss}\% = ([\theta]_{222} - [\theta]_{222}^U)/([\theta]_{222}^N - [\theta]_{222}^U)$, where $[\theta]_{222}$ is the molar ellipticity value at given conditions, while $[\theta]_{222}^U$ and $[\theta]_{222}^N$ are its values in completely unfolded and native states, respectively.

Small-Angle X-Ray Scattering (SAXS). Solution x-ray scattering experiments were carried out at the Stanford Synchrotron Radiation Laboratory on Beam Line 4–2. A flow-cell with 10-μm thick mica windows, a pathlength of 1.3 mm, and a sample capacity of 45 μl was used to reduce extended exposure of the sample to radiation. The sample cell was thermostatted and maintained at 20°C. For Kratky plots the scattering data were plotted as $I(Q) \times Q^2$ vs. Q, where $I(Q)$ is the x-ray scattering intensity, $Q = 4\pi \sin\theta/\lambda$ is the momentum transfer, and 2θ and λ are the scattering angle and the wavelength of the x-rays, respectively.

RESULTS

When a salt-free solution of SNase is titrated from neutral pH to low pH the protein unfolds in a very cooperative process with a midpoint of pH 4. The addition of various anions at pH 2.5 leads to the formation of one of three possible partially folded intermediates (16–17). In the present work we have focused on the least structured partially folded nonglobular intermediate, A_1, which is induced by chloride or sulfate ions. At pH 2.5 and 250 mM Na_2SO_4, and low protein concentrations, SNase is monomeric and ≈50% folded, based on the amount of secondary structure and the radius of gyration (ref. 16 and manuscript in preparation). The inset to Fig. 1 shows the θ_{222} as the acid-unfolded state (U_A) is titrated with sodium sulfate: the initial transition, complete by 100 mM sulfate, corresponds to the conversion of U_A to monomeric A_1 (see also Fig. 2B); the second transition, complete by 0.8 M sulfate, demonstrates the effect of increasing ionic strength, and corresponds to salt-induced dimerization (see below).
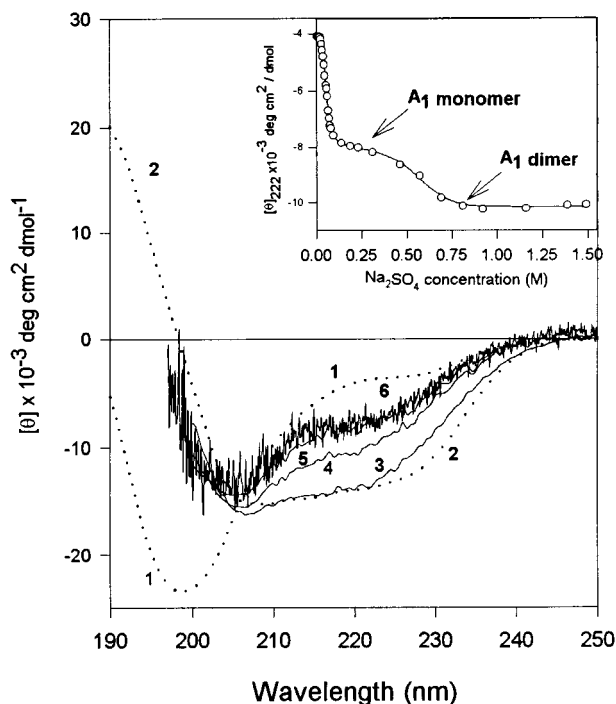
FIG. 1.   The effect of protein and salt concentration on the secondary structure content of the A_1 intermediate of SNase. The inset shows the sulfate titration of U_A at 0.15 mg/ml protein (see text). The CD spectra show the increase in secondary structure as protein concentration is increased. Spectrum 1 is the U_A, pH 2.5, no salt; spectrum 2 is the native state, pH 7.5. Spectra 3–6 of the A_1 state were collected under the same conditions (pH 2.5 and 250 mM Na_2SO_4) but at different protein concentrations: spectrum 6, the noisy curve, at 0.08 mg/ml; spectrum 5 at 0.21 mg/ml; spectrum 4 at 0.52 mg/ml; and spectrum 3 at 3.92 mg/ml.

The effect of increasing protein concentration on the secondary structure of the partially folded intermediate is also shown in Fig. 1: the far-UV CD spectra of the A_1 state of SNase (pH 2.5, 250 mM sodium sulfate) at protein concentrations below 0.2 mg/ml (spectra 5 and 6) where the species is monomeric, indicate the conformation is substantially unfolded ($\approx50\%$ as folded as the native state, as measured by θ_{222}), compared with the native state. SAXS studies showed that under these conditions (≤0.2 mg/ml) the SNase intermediates are monomeric.
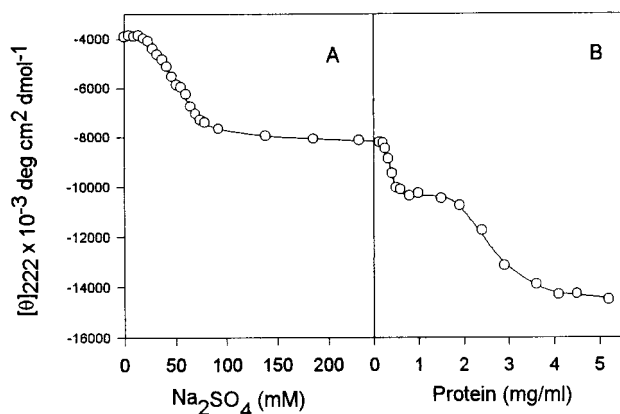

FIG. 2.   Anion- and protein-induced secondary structure in acid-unfolded SNase (pH 2.5). (*A*) the effect of increasing Na_2SO_4 at low protein concentration (0.15 mg/ml). (*B*) The effect of increasing protein concentration at fixed sulfate (250 mM), leading to association. All measurements were carried out at 23°C.

At higher concentrations of protein two transitions were observed by far-UV CD (Fig. 2*B*). The first of these transitions was complete by a protein concentration of ≈0.5 mg/ml, and led to the formation of a new species (spectrum 4) with considerably more secondary structure, $\approx70\%$ of the θ_{222} of the native structure. SAXS data (see below) and size-exclusion HPLC showed that under these conditions the SNase was predominantly dimeric. Further increase in the protein concentration induced a second transition leading to formation of larger soluble aggregates with more secondary structure content. Above 4 mg/ml protein the CD spectrum (spectrum 3) is almost identical to that of the native protein!

Fig. 2 depicts the formation of secondary structure in staphylococcal nuclease at acid pH. As with many other proteins the addition of anions to U_A leads to increased secondary structure due to screening of the positive charges (18–19). Increasing the sulfate concentration from ≈25 to ≈75 mM transforms the acid-unfolded state into the least structured partially folded conformation, A_1, while further increase in anion content up to 250 mM sulfate had little effect (Fig. 2*A*). Increasing the protein concentration at fixed sulfate concentration (250 mM) induces additional secondary structure (Fig. 2*B*). This figure clearly shows that the process of association-induced secondary structure formation has sequential (biphasic) character. First, increasing the protein concentration from ≈0.2 to ≈0.5 mg/ml leads to the transition to an intermediate state with more secondary structure. This intermediate exists at moderate protein concentrations (up to 1.5 mg/ml) and then it, in turn, is transformed to a new species with native-like secondary structure. Based on SAXS (see below) and size-exclusion HPLC data (not shown), the species at the end of the sulfate titration is monomeric, whereas the species at moderately high protein concentrations is dimeric, and at high protein concentrations is multimeric. The transition in Fig. 2A, and the second transition in Fig. 2B, were fully reversible, whereas the first transition in Fig. 2B was not (although this is probably a kinetic effect).

There are two particularly noteworthy points. (*i*) The apparently native-like conformation (as well as the other less structured intermediates) lacks the unique tertiary structure characteristic of native SNase, i.e., its near-UV CD spectrum is that of the unfolded molecule (data not shown). (*ii*) The dimeric intermediate described here is rather close in its structural properties to the monomeric species observed for acid-unfolded SNase in the presence of trifluoroacetate, the A_2 state (17).

As the intensity of the far UV CD signal is sensitive to each type of ordered secondary structure, it is difficult to separate the contributions of inter- and intramolecular interactions in the process of the association-induced secondary structure formation. To aid in clarifying this point we have used SAXS that can give information about the size, compactness and shape of the scattering molecule (20). In addition it has been shown that analysis of the scattering data in the form of a Kratky plot can provide information about the globularity of the molecule (20–23). The Kratky plot for the native protein will show a characteristic maximum, while for the unfolded polypeptide there will be no maximum (21–23). Partially folded conformations will exhibit a Kratky plot with a maximum whose magnitude depends on their degree of compactness (22).

Fig. 3 shows Kratky plots for staphylococcal nuclease under different experimental conditions. Curves 1 and 5 represent the scattering profiles for acid-unfolded (pH 2.5, no salt) and native (pH 7.5) protein, respectively. The native protein shows the characteristic maximum, while the acid-unfolded SNase scatters as a Gaussian chain, consistent with the lack of a globular core. Curve 2 is for the monomeric form of the A_1 partially folded intermediate [0.2 mg/ml (pH 2.5) and 250 mM Na_2SO_4], and shows that in this conformational state the
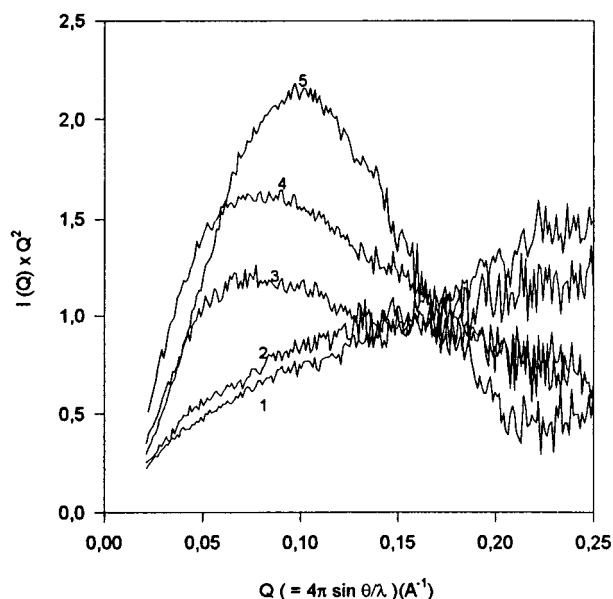
FIG. 3. Kratky plots for SNase in the different conformational states at 20°C: *1,* pH 2.5, no salt (acid-unfolded state); *2,* pH 2.5, 250 mM Na_2SO_4, protein concentration is 0.2 mg/ml (A_1 state); *3,* pH 2.5, 250 mM Na_2SO_4, protein concentration is 0.55 mg/ml (dimerized A_1 state); *4,* pH 2.5, 250 mM Na_2SO_4, protein concentration is 3.55 mg/ml (oligomeric A_1 state); *5,* pH 7.5 (native state).

molecule has no significant globular structure, i.e., the Kratky plot is very similar to that for the unfolded protein. The lack of globular structure attributed to the monomeric intermediate correlates well with recent work by using poly-L-lysine and poly-L-glutamic acid as models in which it was established that formation of intramolecular α-helical elements joined by flexible coils without formation of a hydrophobic core (globular structure) does not affect the shape and intensity of the Kratky plot (23). At the same time Fig. 3 clearly shows that an increase in protein concentration under these conditions leads to the appearance of globular structure (curves 3 and 4), with the amount of globular structure increasing with increasing protein concentration.

This effect cannot be explained just by the association-induced formation of intermolecular interactions: such intermolecular associations of proteins usually lead to the formation of β-sheet structure (11). However, it has been shown that for poly-L-lysine in the β-sheet conformation, the Kratky plot is quantitatively similar to the plot for a Gaussian chain and notably different from the globular one (23). It is known that the β-sheet conformation of poly-L-lysine is formed mostly due to intermolecular interactions (24) (as a consequence, polypeptides in this conformation have no globular core). Therefore, the increase in concentration of the partially folded intermediate induces not only formation of intermolecular structure but leads also to the appearance of new intramolecular structure within an individual protein molecule, as manifested by both increased secondary structure and globular core.

SAXS data are very sensitive to intermolecular protein association. The value of forward-scattered intensity ($I(Q)$ as $Q \rightarrow 0$) is proportional to the square of the molecular weight of the molecule (20). I(0) for a pure dimer sample will therefore be twice that for a sample with the same number of monomers because each dimer will scatter four times as strongly, but there will be half as many as in the pure monomer sample. Our results are consistent with the conclusion that SNase molecules are predominantly monomeric in all three anion-induced intermediates: A_1; A_2 (acid-unfolded SNase in the presence of trifluoroacetate); and A_3 (formed in the

presence of trichloroacetate at low protein concentrations, ≤ 0.2 mg/ml); as well as in the native and acid-unfolded states. The increase in protein concentration (0.5–2.0 mg/ml) at moderate sulfate (0.25 M) or Cl^- (0.9 M) leads to a 2-fold increase in $I(0)$ (from ≈ 3.6 to ≈ 7.1). Increasing concentrations of these anions at high protein concentrations results in the much larger increases in $I(0)$. For example, at the end of second association-induced transition (0.25 M SO_4^{2-}, 4.0 mg/ml) $I(0)$ is 31.3, which would imply the formation of octamers if the sample is monodisperse. Further increases in protein concentration lead to further increases in the value of $I(0)$, and ultimately the protein precipitates (data not shown).

The results can be summarized, Fig. 4, in the form of a protein and sulfate-dependent conformational "phase" diagram for the association-induced conformational transitions in SNase at low pH. The solid-line boundaries between different conformations are constructed on the basis of the midpoint values of the corresponding Na_2SO_4-induced transitions at fixed protein concentrations, or those of transitions induced by the increasing in protein concentration at fixed sulfate content. Unfolded monomers exist at low sulfate concentration in a wide range of protein concentrations. Partially folded, loosely packed (nonglobular) monomers (A_1), appear in dilute protein solutions when the sulfate concentration increases. Globular (more tightly packed) dimers with larger secondary structure are formed in more concentrated protein solutions with sulfate concentrations >150 mM. Higher multimers with the largest structural order exist in a limited range of high protein and sulfate concentrations. Finally, the shaded area represents the conditions where precipitation occurs. As noted, most of these boundaries reflect reversible transitions.

DISCUSSION

We believe that the monomeric A_1 intermediate corresponds to molecules containing a core of native-like secondary struc-
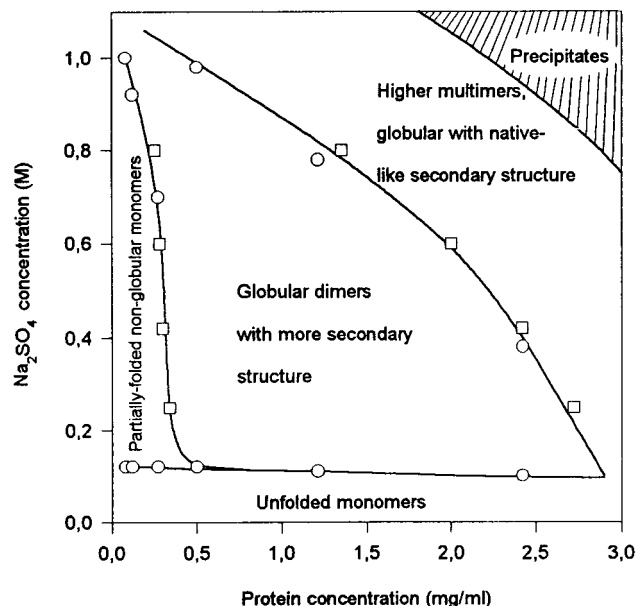

FIG. 4. Sulfate- and protein concentration-dependent conformational "phase" diagram for the conformational states of staphylococcal nuclease at 23°C and pH 2.5. The conformational space consists of U_A, A_1, globular dimers with more secondary structure than the monomer, and globular oligomers with native-like secondary structure content. The continuous lines show the boundaries between these states. Boundaries were determined on the basis of the midpoint (C_m) values of the corresponding Na_2SO_4-induced transitions at fixed protein concentrations (squares), or midpoint (C_m) values of transitions induced by the increasing in protein concentration at fixed sulfate content (circles). The shaded area represents the conditions at which precipitation occurs.

ture with the remainder of the polypeptide chain in a relatively disordered state (which may have regions flickering in and out of native-like secondary structure). The lack of globularity (a tightly packed core) shown by the Kratky plots for A_1 indicate that this core is rather small and loosely packed. Two regions in SNase have been suggested to be the initially formed core, the $\alpha 2$-helix and the $\beta 2$- and $\beta 3$-sheets (25): these are thus logical regions of structure in the proposed core in A_1.

The data show that a partially folded intermediate, with only 50% of the native-like secondary structure in its monomeric form, gains additional secondary structure on association to form dimers or higher multimers, as the protein concentration is increased. Although the multimeric aggregates have native-like secondary structure they lack the characteristic tertiary structure of the native conformation, as revealed by their lack of near-UV CD signal. The Kratky plots indicate that the structure is nevertheless tightly packed.

There are two possible models to explain these observations: (*i*) association of the monomers occurs first and leads to the induction of the additional structure, or (*ii*) the higher concentration of protein or salt increases the population of more structured intermediates that then associate. Although the data do not allow an unambiguous choice, we favor the former for several reasons. In a sense these models represent extremes of a continuum, and the question boils down to whether the hydrophobic surfaces involved in association reflect transient or stable regions of the monomeric intermediate structure.

It is most likely that the intermolecular association reflects specific interactions between hydrophobic surfaces of one partially folded molecule with those of another, where these specific interactions are ones that normally occur intramolecularly and lead to the formation of the native state, i.e., in a sense the association arises from nonproductive (intermolecular) interactions that normally are formed (intramolecularly) on the pathway to the native state. In other words if we consider the monomeric native state to arise by the coalescence of structural building blocks (subdomains) in an intramolecular fashion, the aggregates arise by these same interactions but in an intermolecular fashion. Thus the dimers may represent a form of domain swapping (26) in which regions of one A_1 intermediate molecule interact with a second molecule in a specific interaction that results in additional secondary structure formation. In this regard it is interesting to note that a dimeric species of SNase has been reported (27). It is noteworthy that the A_1 dimer is particularly stable.

Although it is well known that peptides with high helical propensity will associate and thus stabilize the helical structure at high peptide concentrations (28, 29), this is the first report that the association or aggregation of a partially folded intermediate of a protein containing a variety of types of secondary structure leads to the formation of additional secondary structure. It is conceivable that the association arises from interactions of amphiphilic helices in the intermediate (possibly only present transiently), analogous to the association of helical peptides, but the CD difference spectra show no evidence of significant increased helix content on association. In addition, the starting point in the present case is a stable partially folded intermediate with substantial structure, in contrast to helical peptide case.

It is not clear how widespread the phenomenon of association-induced secondary structure formation may be, nor if it is physiologically significant. However, the fact that inclusion bodies may have substantial native-like secondary structure (30) could be one consequence. On the other hand, one of the characteristics of aggregated proteins (such as inclusion bod-

ies, folding aggregates, and amyloid) is the presence of β-structure (11), and the association-induced structure in the case of SNase does not seem to involve significant nonnative β-sheet structure. The CD difference spectrum between monomeric A_1 and its dimer is consistent with an increase in β-structure, but the transition from the dimer to higher multimers is more complex, and as noted, results in formation of a CD spectrum similar to that of the native state. Thus it is possible that the association of SNase reported here represents a more specific type of interaction than that typically found in protein aggregation.

1. Carrell, R. W. & Lomas, D. A. (1997) *Lancet* **350,** 134–138.
2. Thomas, P. J., Qu, B. H. & Pedersen, P. L. (1995) *Trends Biochem. Sci.* **20,** 456–459.
3. Marston, F. A. O. (1986) *Biochem. J.* **240,** 1–12.
4. Schein, C. H. (1989) *Bio/Technology* **7,** 1141–1149.
5. Wetzel, R. (1992) in *Stability of Protein Pharmaceuticals, Part B:* In Vivo *Pathways of Degradation and Strategies for Protein Stabilization*, eds. Ahern, T. J. & Manning, M. C. (Plenum, New York), Vol. 3, pp. 43–48.
6. London, J., Skrzynia, C. & Goldberg, M. E. (1974) *Eur. J. Biochem.* **47,** 409–415.
7. Zettlmeissl, G., Rudolph, R. & Jaenicke, R. (1979) *Biochemistry* **18,** 5567–5571.
8. Jaenicke, R. (1995) *Phil. Trans. R. Soc. London B* **348,** 97–105.
9. DeYoung, L. R., Fink, A. L. & Dill, K. A. (1993) *Acc. Chem. Res.* **26,** 614–620.
10. Mitraki, A. & King, J. (1989) *Bio/Technology* **7,** 690–697.
11. Fink, A. L. (1998) *Folding Design* **3,** R9–R15.
12. Kuwajima, K. (1989) *Proteins* **6,** 87–103.
13. Kim, P. S. & Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59,** 631–660.
14. Ptitsyn O. B. (1995) *Adv. Protein Chem.* **47,** 83–229.
15. Fink, A. L. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24,** 495–522.
16. Fink, A. L., Calciano, L. J., Goto, Y., Nishimura, M. & Swedberg, S. A. (1993) *Protein Sci.* **2,** 1155–1160.
17. Uversky, V. N. Karnoup, A. S., Segel, D. J., Seshadri, S., Doniach, S. & Fink, A. L. (1998) *J. Mol. Biol.*, in press.
18. Goto, Y., Takahashi, N. & Fink, A. L. (1990) *Biochemistry* **29,** 3480–3488.
19. Fink, A. L., Calciano, L. J., Goto, Y., Kurotsu, T. & Palleros, D. R. (1994) *Biochemistry* **33,** 12504–12511.
20. Glatter, O. & Kratky, O. (1982) *Small Angle X-Ray Scattering* (Academic, New York), p. 515.
21. Kataoka, M., Hagihara, Y., Mihara, K. & Goto, Y. (1993) *J. Mol. Biol.* **229,** 591–596.
22. Doniach, S., Bascle, J., Garel, T. & Orland, H. (1995) *J. Mol. Biol.* **254,** 960–967.
23. Semisotnov, G. V., Kihara, H., Kotova, N. V., Kimura, K., Amemiya, Y., Wakabayashi, K., Serdynk, I. N., Timchenko, A. A., Chiba, K., Nikaido, K., Ikura, T. & Kuwajima, K. (1996) *J. Mol. Biol.* **262,** 559–574.
24. Adler, A. J., Greenfield, N. J. & Fasman, G. D. (1973) *Methods Enzymol.* **27,** 675–735.
25. Wang Y. & Shortle D. (1995) *Biochemistry* **34,** 15895–15905.
26. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. (1995) *Protein Sci.* **4,** 2455–2468.
27. Green, S. M., Gittis, A. G., Meeker, A. K. & Lattman, E. E. (1995) *Nat. Struct. Biol.* **2,** 746–751.
28. Yoshida, K., Shibata, T., Masai, J., Sato, K., Noguti, T., Go, M. & Yanagawa, H. (1993) *Biochemistry* **32,** 2162–2166.
29. Kaiser, E. T. & Kezdy, F. J. (1984) *Science* **223,** 249–55.
30. Oberg, K., Chrunyk, B. A., Wetzel, R. & Fink, A. L. (1994) *Biochemistry* **33,** 2628–2634.