Forced Folding and Structural Analysis of Metastable Proteins

Ronald W. Peterson, Karthik Anbalagan, Cecilia Tommos, and A. Joshua Wand*

Johnson Research Foundation and Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059, and Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden SE-106 91

Received April 12, 2004; E-mail: wand@mail.med.upenn.edu

It is becoming increasingly clear that a significant fraction of the proteins encoded by the human and other genomes are likely to be significantly unfolded in vitro and, often, in vivo. This will undoubtedly hamper attempts to characterize their structure by classical crystallographic or solution NMR methods. There are at least two possible reasons for this relative instability of the folded state. On one hand, some proteins are "intrinsically unfolded" in the absence of target ligands, i.e. their biologically relevant structure need not be the lowest free energy state of the protein in the absence of ligand. Indeed, there is at least one example of "synergistic folding" where two effectively unstructured co-activators of transcription bind and cooperatively form a structured heterodimer. On the other hand, some proteins must simply be considered unstable in the absence of stabilizing excluded volume effects arising from the dense packing conditions of the cellular milieu. In that case, the folded native state may indeed be the lowest-energy structure in solution, but it is not sufficiently removed in free energy from nonnative structures to be the dominant species in vitro.

In an effort to develop a means of stabilizing the folded compact state of a marginally stable protein, we have employed an approach based on the effects of a confined space on the ensemble of states occupied by a polypeptide chain. It has long been known, on theoretical grounds, that confining a linked chain within a restrictive volume will favor compact states over more extended states. A variety of means based on a confined space concept have been employed to stabilize folded proteins. These include confinement within acrylamide gels, reverse micelles, and other nanostructures.

Here we take advantage of reverse micelle encapsulation to provide a restricted environment of defined geometry to force the folding of a metastable protein that is largely unfolded in free solution. This is achieved by the encapsulation of the protein in the protective environment of the water core of a reverse micelle and dissolving the entire assembly in a low-viscosity fluid. Of particular interest here is the use of the regular and simple spherical shape of the reverse micelle to force fold unstable proteins in a predictable way that makes direct contact with polymer theory while allowing high-resolution NMR studies to be carried out (Figure 1). Importantly, as we show below, this approach also has the potential to contribute to the current debate about the nature of the unfolded state.

To illustrate the approach we have employed an extensively modified version of a three-helix bundle protein of de novo design. The original version, termed \( \alpha_3 \)W, is a small monomeric single-chain protein of 67 residues that forms a stable three-helix bundle. In an effort to explore the relative contributions of helix-stabilizing interactions and tertiary contacts to the global stability of the bundle, many of the helical residues not participating in tertiary contacts were changed to residues of significantly lower helical propensity. Using the AGADIR algorithm, the resulting sequence is predicted to have a substantially lower helical content relative to that of the original \( \alpha_3 \)W sequence. At concentrations necessary for traditional NMR experiments (>0.5 mM) the protein aggregates and precipitates.

Freshly prepared samples yield a \(^{15}\)N HSQC spectrum with few resolved cross-peaks indicative of an unfolded protein without a dominant folded conformation (Figure 2A).

The inherent instability of this sequence is overcome when the polypeptide chain is encapsulated in an AOT reverse micelle prepared in hexane under restricted water conditions. At high water loading (\( W_0 \)), the molar ratio of water to surfactant, the \(^{15}\)N HSQC spectrum of the encapsulated protein is indicative of an unfolded protein but lacks features arising from nonspecific aggregation (Figure 2B). In the case shown, the water loading is 40.
As the internal diameter of the reverse micelle approaches that of the dimensions of the native $\alpha$W structure, the NMR spectrum reverts to that indicative of a single dominant species with a chemical shift dispersion consistent with a three-helix bundle (Figure 2C). This particular spectrum was obtained with a sample having a water loading of $\sim$4. Because $\alpha$W is a relatively small protein, and the unfolded states are therefore not so extensive, low $W_0$ values are required to provide sufficient stabilization of the compact state of the protein.

The folded and unfolded states are in fast exchange on the NMR chemical shift time scale. This allows the fraction of the protein in the folded state to be determined. The chemical shift time scale. This allows the fraction of the protein in the folded state, steric effects limiting free sampling of main chain torsion angles,11 and the possible presence of significant residual structure.12 Further experiments are required to clarify this issue.

In summary, we have shown that encapsulation of a metastable protein within the restricted volume of a reverse micelle can be used to force fold the protein and allow its characterization by modern methods of NMR spectroscopy. This may have significant utility in the context of structural proteomics. In addition, this method may allow the distinction between “intrinsically unfolded” proteins and proteins simply having marginal stability. Finally, variation of the inner volume of the reverse micelle can be used to probe the character of the manifold of unfolded states.

Acknowledgment. We thank Robert Baigelman for preparation of labeled protein. This work was supported by NIH Research Grant GM 62874 and an NSF MRSEC Award (DMR00-79909) (A.J.W.) and the Swedish Research Council (C.T.).

References

(14) The sequence of the modified $\alpha$W used in these studies is: GSRRVLVTRTQLQSRLQVTTSLGGGGGIEAKLNQWESKLNNRIESLGGGGDVTSTVEADVLANDTDITSL. The protein was recombinantly expressed as a fusion with thioredoxin, a histidine-tag and a thrombin cleavage site, purified using a nickel column, and cleaved with thrombin; the $\alpha$W derivative was isolated via a second nickel column step.