J|A|C|S
COMMUNICATIONS

Published on Web 02/28/2004

A High-Resolution Probe of Protein Folding

Laura B. Sagle, Jörg Zimmermann, Philip E. Dawson, and Floyd E. Romesberg*

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California, 92037

Received January 8, 2004; E-mail: floyd@scripps.edu

Protein folding is a key step in the manifestation of function from genetic information, and understanding the folding process has emerged as a significant, but challenging, problem. A fundamental obstacle to studying the folding of any protein, even under equilibrium conditions, is the general absence of residue-specific probes that are sensitive to the conformational details of the protein and that also have time resolution sufficient to resolve rapidly interconverting intermediates. In principle, the folding pathway of a protein could be fully characterized by IR spectroscopy. However, the spectral congestion inherent to proteins and denaturants has prevented the direct use of IR spectroscopy.

Interestingly, proteins and denaturants have a "transparent window" in their IR spectrum that is free of absorptions, between ~1800−2700 cm^{-1}. Any sufficiently strong absorption in this region may be directly observed, despite high protein or denaturant concentrations. Thus, to develop a nonperturbative and residue-specific protein probe we have been examining the use of C−D bonds, which absorb at ~2100 cm^{-1}.[1,2] In addition to spectral resolution, these probes provide inherently high time resolution; an intermediate that gives rise to only a 1 cm^{-1} shift in the C−D stretching frequency would have to interconvert on the 10^{11} s^{-1} time scale in order to not be resolved. The inherent spectral, structural, and temporal resolution of the technique should greatly facilitate the detection and characterization of equilibrium unfolding intermediates.

For our initial application of the technique, we have chosen to study the folding of horse heart cytochrome *c* (cyt *c*). The reversible folding of cyt *c* has previously been characterized using several indigenous chromophores, including the heme cofactor and a tryptophan residue. Equilibrium unfolding induced by guanidine hydrochloride (GdnHCl) has been characterized intensively. Plots of various spectroscopic signals versus denaturant concentration typically show an apparent two-state transition with a midpoint at ~2.5 M GdnHCl that is fully shifted to the unfolded state by 3.5 M GdnHCl.[3−7] However, Raman signals from the heme and tryptophan residue show different transitions, indicating the presence of an intermediate.[8]

Equilibrium amide exchange NMR studies indicate the presence of at least two partially folded states during titration with GdnHCl.[9] At least one equilibrium unfolding intermediate has been detected by SAXS, with maximum population at ~2.9 M GdnHCl.[6] On the basis of MCD spectra, this intermediate is thought to involve bis-histidinyl coordination by His18 and His26 or His33.[7] Thus, a consensus regarding the presence and nature of equilibrium unfolding intermediates has not yet been reached due, at least in part, to the differences in techniques used. The general ability of C−D probes to detect even rapidly interconverting intermediates with residue-specific detail should help resolve the issue.

An important aspect of the C−D method is its general applicability to any part of the protein: residue, side chain, or backbone. Thus, to selectively incorporate C−D bonds at different positions throughout the protein, we semisynthesized cyt *c* using

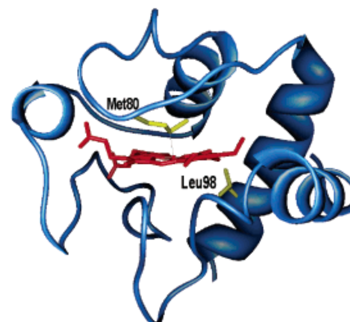

Figure 1. Cyt *c* showing Met80 and Leu98.[14]

the autocatalytic religation approach (ref 10 and Supporting Information (SI)). Briefly, a cyt *c* 1−65-homoserine lactone peptide fragment containing the covalently bound heme was generated using CNBr cleavage at Met65 of horse heart cyt *c*. This peptide was purified and then refolded at pH 7.0 in the presence of 1 equiv of chemically synthesized[11] cyt *c* 66−104 containing a specifically deuterated Leu or Met residue. Fragment association and aminolysis of the homoserine lactone resulted in high yields of semisynthetic cyt *c* Met65hSer.

Leu98 is located in the C-terminal helix of the protein (Figure 1), and is thought to unfold only at high denaturant concentration.[9] To characterize the folding of the protein at this residue, methyl group vibrations of the free leucine amino acid were first examined computationally using Gaussian98 with the LANL2DZ basis set.[12] Two asymmetric stretches at 2185 and 2175 cm^{-1} and a third approximately symmetric stretch at 2064 cm^{-1} were predicted. A spectrum of free Cδ-d_3 leucine (50 mM of 1:1 mixture of d_1- and d_2-labeled isotopomers, Cambridge Isotopes) was recorded in aqueous solution (100 mM sodium acetate, pH 5), acetic acid, ethyl acetate, and heptanol (SI). In each case, two relatively strong absorptions were observed at ~2221 and ~2210 cm^{-1}, which were assigned as asymmetric stretches, and a weaker absorption at ~2065 cm^{-1} was assigned as the symmetric stretch of the CD_3 group. The solvent studies showed that the asymmetric stretching frequencies are sensitive to the dielectric constant of the solvent, blue-shifting as the dielectric constant is increased, with no significant changes in the observed line widths (SI).

The IR spectrum of the folded, oxidized Leu98 deuterated protein (synthesized with a 1:1 mixture of d_1- and d_2-labeled isotopomers) at pH 6.2 showed two absorptions, one at ~2202 cm^{-1} and one at 2216 cm^{-1} (Figure 2). These absorptions are similar to the asymmetric stretches observed with the free amino acid, and are thus similarly assigned. The symmetric stretch observed in the free amino acid was not visible, presumably due to its smaller intensity and the lower concentrations mandated by the solubility of the protein. The single pair of asymmetric absorptions implies that the vibrations are not significantly differentiated in the two isotopomers.

With the addition of GdnHCl, the spectrum of the folded protein is replaced by two new peaks (Figure 2), assigned as the asymmetric

10.1021/ja049890z CCC: $27.50 © 2004 American Chemical Society

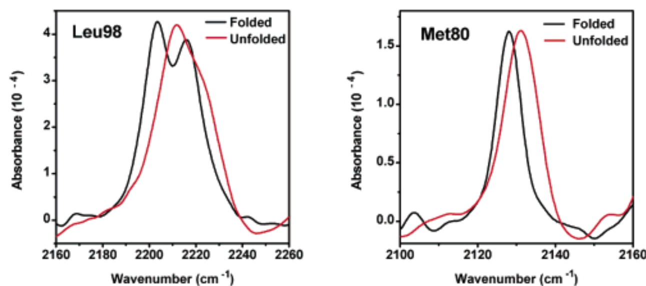

Figure 2. Folded and unfolded spectra.

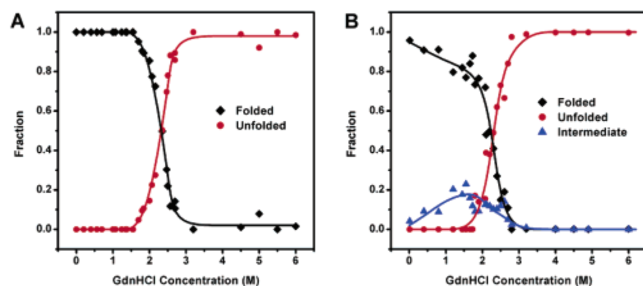

Figure 3. (A) Fraction folded and unfolded at Leu98 (based on asymmetric stretch). (B) Fraction folded and unfolded (symmetric stretch) and intermediate (single observable absorption) at Met80. The lines do not represent fits and are intended only as visual aids.

stretches of the unfolded protein, based on the invariance to additional GdnHCl as well as the UV/vis spectrum. The absorptions of the unfolded protein are blue-shifted and broadened, relative to those of the folded protein, indicating that the residue is in a more polar and disordered environment. Every spectrum at intermediate denaturant concentration is well reproduced by a combination of the folded and unfolded spectra, and after normalization a clear isosbestic point is observed (SI). The unfolding at this residue is thus two-state, without the intervention of an intermediate, and the relative weighting of each spectrum reflects the relative concentrations of folded and unfolded states. Three independent sets of data were collected, and the normalized intensity of an asymmetric stretch of the folded and unfolded proteins is plotted in Figure 3a (plotting the intensity of the other absorption yielded identical results). At Leu98, the GdnHCl induced unfolding of cyt *c* is two-state, with a transition midpoint of 2.3 ± 0.05 M GdnHCl. This is similar to, but somewhat earlier, than the transitions detected with other techniques, as discussed above.

The side chain of Met80 provides one of the two protein-based heme ligands (Figure 1), and in contrast to Leu98, its dissociation is thought to occur at low denaturant concentrations.[9] Computationally, with free methionine, two asymmetric stretches were predicted at 2234 and 2240 cm^{-1}, and a third approximately symmetric stretch was predicted at 2085 cm^{-1}. The IR spectrum of free (methyl-d_3) methionine (acquired with conditions identical to those described above for leucine) showed in each case a broad absorption at 2254 cm^{-1}, assigned as overlapping asymmetric stretches, and a stronger absorption at 2136 cm^{-1}, assigned as the approximately symmetric stretch. Both the symmetric and the asymmetric stretching frequencies are sensitive to the dielectric constant of the solvent, blue-shifting as the dielectric constant is increased, with no significant changes in the observed line widths (SI).

As with the free amino acid, two IR absorptions were apparent with (methyl-d_3)Met80 cyt c, a strong absorption at 2129 cm^{-1} and a weaker absorption at 2254 cm^{-1}, assigned as the approximately symmetric stretch and the overlapping asymmetric stretches of the CD_3 group, respectively (Figure 2). With added GdnHCl, the spectrum of the folded protein disappears and is replaced by two new peaks (Figure 2), assigned to the unfolded protein on the basis of their invariance to increasing denaturant and the corresponding UV spectrum. As at Leu98, the absorption spectrum of the unfolded protein is blue-shifted and broadened relative to that of the folded protein, indicating that the Met80 environment is more polar and disordered in the unfolded state. Interestingly, in contrast to Leu98, the spectrum of Met80 at intermediate denaturant concentrations showed an absorption which could not be assigned to either the folded or the unfolded protein. The absorption of this third species appears at ∼1 M denaturant, achieves a maximal concentration at ∼1.5 M, and then disappears by ∼2.7 M denaturant (Figure 3b). The species is likely connected to the folded protein by conformational changes related to heme ligation or misligation. While a more quantitative description of the changes Met80 will require a detailed model, the similarities in fraction unfolded at Met80 and Leu98 argue for a global unfolding event at 2.9 M GdnHCl.

Cyt *c* folding at Leu98 and Met80 is distinctly different. The C-terminal helix undergoes a simple two-state unfolding transition, while the Met80 ligand populates a third state that likely involves misligation at the heme center. To our knowledge, this is the first direct and residue-specific observation of this denaturant-induced protein intermediate and complements previous experiments that are based on observation of the cofactor[8,13,14] or amide exchange rates.[9] The details of the third state are currently being investigated. In addition, time-resolved experiments are also in development that will provide a residue specific, real time view of folding and determine if the observed third species is an on- or off-pathway folding intermediate.

Acknowledgment. Funding was provided by the La Jolla Interfaces in Science (L.J.I.S.) and the Skaggs Institute for Chemical Biology. We thank Profs. Taiha Joo and Ralph Jimenez and Dr. Robert Goldbeck for helpful conversations, and the NIH for funding (GM 59380).

Supporting Information Available: Materials, methods, and raw data (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Chin, J. K.; Jimenez, R.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 2426−2427.
(2) Chin, J. K.; Jimenez, R.; Romesberg, F. E. *J. Am. Chem. Soc.* **2002**, *124*, 1846−1847.
(3) Santucci, R.; Ascoli, F. *J. Inorg. Biochem.* **1997**, *68*, 211−214.
(4) Tsong, T. Y. *Biochemistry* **1975**, *14*, 1542−1547.
(5) Hagihara, Y.; Tan, Y.; Goto, Y. *J. Mol. Biol.* **1994**, *237*, 336−348.
(6) Segel, D. J.; Fink, A. L.; Hodgson, K. O.; Doniach, S. *Biochemistry* **1998**, *37*, 12443−12451.
(7) Thomas, Y. G.; Goldbeck, R. A.; Kligar, D. S. *Biopolymers (Biospectrosc.)* **2000**, *57*, 29−36.
(8) Rush, T. S.; Spiro, T. G. *Spectroscopic Methods in Bioinorganic Chemistry*; ACS Symposium Series 692; American Chemical Society: Washington, DC, 1992; pp 212−219.
(9) Bai, V. W.; Sosnick, T. R.; Mayne, L.; Englander, S. W. *Science* **1995**, *269*, 192−197.
(10) Wallace, C. J.; Clark-Lewis, I. *J. Biol. Chem.* **1992**, *267*, 3852−3861.
(11) Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. *Int. J. Pept. Protein Res.* **1992,** *40*, 180−193.
(12) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Adamo, C.; Jaramillo, J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.: Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
(13) Russell, B. S.; Melenkivitz, R.; Bren, K. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000***, 15*, 8312−8317.
(14) Bushnell, G. W.; Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 585−595.

JA049890Z