DOI: 10.1002/cphc.200700247

Collapsed But Not Folded: Looking with Advanced Optical Spectroscopy at Protein Folding

Jörg Enderlein*[a]

Protein folding, that is, the organization of proteins into a highly ordered three-dimensional structure under native conditions is one of the most interesting and fundamental topics of contemporary biophysical research. The core question is how an initially random conformation of a polypeptide chain of a protein finds its final native structure among the billions and billions of possible conformations. This is also known as the Levinthal paradox[1] originally formulated by Cyrus Levinthal in 1969: Even if each possible conformation of a typical protein is probed within only picoseconds or nanoseconds, it would take longer than the age of the universe for hitting on the correctly folded structure. To solve Levinthal's paradox it was postulated that proteins fold along specific pathways (older view), or follow one of many parallel paths down the now famous protein folding funnel that funnels a protein's conformation towards its final correct and stable fold (modern statistical mechanical view). However, the details of this funnelling remain an active area of discussion and research. Two extreme views can be distinguished: The first view assumes that the final conformation is attained by a more or less stepwise formation of secondary structure across the protein's polypeptide chain that finally folds up into the correct tertiary structure. The other extreme assumes that first the protein's polypeptide chain hydrophobically collapses into a compact but totally unstructured globule state which then converts more or less slowly into the folded structure.

Recent research has shown that the truth lies somewhere in the middle (Figure 1).[2] Hydrophobic collapse into a compact but still denatured state (the molten globule[3]) is indeed an important intermediate step in folding, but it already involves the presence of residual secondary structure. This structure, even when minimal in extent, seems to be an important aid in the further nucleation and formation of the final folded structure. Thus, an important intermediate step of folding between a completely unfolded state and the final fully formed native structure is the collapsed state of a protein, where secondary structure elements such as short α-helices and/or β-sheets are already present but no tertiary large-scale structure has yet formed.

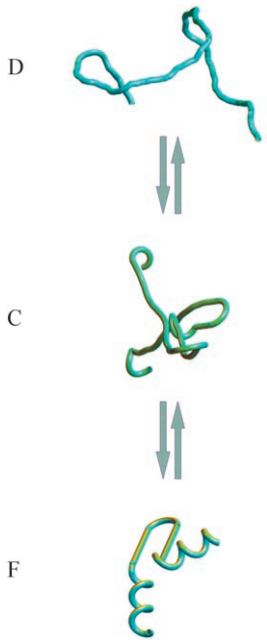

Figure 1. Pathway from the denatured, completely unfolded (D) state of a protein to the native, completely folded state (F) leads over a collapsed state (C) which still behaves like a random coil but may already contain secondary structural elements such a short α-helical or β-sheet structural elements. It is important to mention that states (D) and (C) represent *manifolds* of many different conformations, whereas state (F) is usually represented by a single well-defined conformation.

The core challenge of experimentally observing and studying this state is that under equilibrium conditions close to the native conditions of the protein the overwhelming majority of protein molecules is already in its final folded state and only a very small number of molecules will be in the pre-folded collapsed state. This makes ensemble measurements and studies of the collapsed but not yet folded state under equilibrium native conditions rather difficult. Two experimental solutions to this problem are feasible: One can either study the system far from equilibrium using stop-flow techniques, or one can use single-molecule spectroscopy for looking at individual protein molecules and thus finding rare events of collapsed but not yet folded molecules among the large number of folded protein molecules.[4]

Both approaches have been used in a recent paper by Hoffmann et al.[5] to investigate the collapsed state of the small cold shock protein Csp*Tm*. For this protein, equilibrium and time-dependent collapse have been studied at near native conditions using single molecule spectroscopy.[6] In the paper of Hoffman et al., this collapsed but unfolded state is now studied in more detail (see Figure 1).

The first method used by the authors is single-molecule Förster resonant energy transfer (smFRET).[7] In this

[a] *Prof. Dr. J. Enderlein*
Institute of Physical and Theoretical Chemistry
Eberhard Karls University Tübingen
Auf der Morgenstelle 8
72076 Tübingen (Germany)
Fax: (+49) 07071-29-5490
E-mail: joerg.enderlein@uni-tuebingen.de

method, the protein is site specifically labelled at two different positions along the peptide chain with two different fluorescent dyes. The fluorescence emission spectrum of one of both dyes (donor) is chosen in such a way that it overlaps with the absorption spectrum of the other dye (acceptor). Then, when the donor is photoexcited and finds the acceptor in close proximity, the excited state energy can be efficiently transferred from the donor to the acceptor by radiationless electromagnetic coupling. This leads to a measurable fluorescence emission signal from the acceptor which is ideally not excitable by the light source used for donor excitation. Because this so-called Förster resonance energy transfer (FRET) strongly depends on the distance between both dyes (to the sixth power), its efficiency can be used for gauging the distance between the dyes and thus the distance between the labelling sites in the protein. An important specificity of the FRET measurements by Hoffmann et al. is not to look at the average signal of a large ensemble of protein molecules but on the FRET signal of individual molecules (i.e. smFRET[7]). This allows for identifying and specifically *selecting* molecular sub-populations with different distances between labelling sites (Figure 2). This is especially important if one is interested in studying sub-populations which comprise only a small fraction of the total number of molecules, as is the case for the collapsed but not folded state of a protein under close to native conditions.

By placing FRET dye pairs at different label positions along the protein's peptide chain, the authors could map the distances between these position pairs under varying concentrations of the denaturing chemical guanidin hydrochloride (GdnHCl), using the single molecule approach for selectively looking only at molecules that are in a collapsed but unfolded state. Thus, it was possible to map this state under varying conditions (i.e. GdnHCl concentrations). The first paper applying smFRET measurements for studying the collapsed state of a protein was published by Sherman and Haran[8] where the authors successfully analyzed their experimental findings by applying models derived for the coil–globule transitions of macromolecules. Hoffman et al. extended the smFRET approach by putting acceptor–donor pairs at many different but specific sites within the Csp*Tm* molecules, thus mapping the collapsed state in a systematic way. The found behavior of the collapsed state was amazingly well described by a random Gaussian chain model, which is the simplest model for describing long polymer chains without intrinsic structural organization. In the Gaussian chain model, the protein is assumed to behave similarly to a freely jointed chain of links. The length of the links is the only free parameter of the model and is called the persistence length of the protein. The found variation of this persistence length with varying GdnHCl concentrations, and its independence on the sites of labelling suggests an isotropic collapse of the protein when switching from a completely unfolded to the collapsed state.

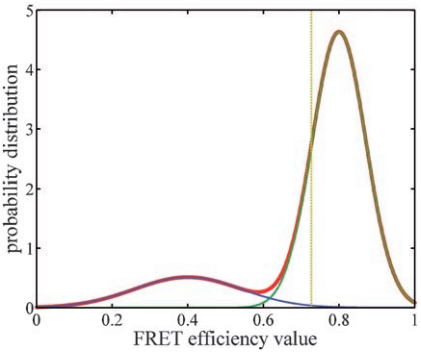

Figure 2. Principle of the single-molecule approach to FRET measurements: Instead of measuring an average FRET value of an ensemble (thin vertical line), FRET values are measured on a molecule by molecule basis and their distribution is plotted (red line). The resulting probability distribution for the FRET efficiency can then be decomposed into contributions of different sub-populations (blue and green lines). Subsequently, one can selectively study the behavior of only a single sub-population of interest (e.g. that represented by the blue line), even if this sub-population comprises only a small fraction of the total ensemble.

In a step further, the authors used single-molecule selective fluorescence-lifetime measurements for independently checking these results. As a result of the excitation energy transfer from a donor to an acceptor molecule during FRET, the donor fluorescence lifetime is reduced in a FRET-efficiency (and thus intradye distance) dependent manner. Thus, lifetime measurements are able to provide similar information on FRET efficiencies as intensity measurements of donor and acceptor fluorescence. Combining intensity with lifetime information helps to eliminate potential errors in FRET efficiency determinations, and the consistency of the results from both approaches in the study of Hoffman et al. is a strong case for the validity of the found Gaussian chain behavior of the collapsed unfolded state.

To check whether this random Gaussian chain behavior is signifying the complete absence of structure in the collapsed state of the protein, the authors applied a second powerful technique: measuring circular dichroism (CD) of the kinetically populated collapsed state. Kinetic population refers to rapid dilution of the protein from high to low GdnHCl concentration, which can be used in combination with a sufficient time-resolving spectroscopic method for studying intermediates during folding. A few microseconds to milliseconds after dilution the protein has already collapsed but is no yet folded, thus allowing for the investigation of the collapsed state on a macroscopic ensemble level. Rapid dilution is achieved by fast fluid mixing of a protein solution containing a high concentration of GdnHCl with an aqueous buffer. Mixing is performed within a specifically designed microfluidic chip, providing characteristic mixing/dilution times on a millisecond time scale. For some recent publications applying this technique for following the fast kinetics of protein collapse see, for example, ref. [9].

For studying protein secondary structure by CD, the spectral region between 190–230 nm is of special interest, where the absorption of the amide group probes the protein backbone and leads to distinctive signals for random coils, β-sheets, and α-helices. Whereas the spectral region around 220 nm is important for measuring α-helical structures, β-sheet structure shows prominent CD signal around 200 nm. Hoffmann et al. used synchrotron radiation as light source for CD measurements (SRCD), thus being able to record of CD signals across a continuous wavelength region

from 200 to 250 nm (although conventional spectrometers also allow to measure within this spectral region). Measuring at different positions of the microfluidic mixer, they recorded CD spectra at different times after collapse initiation.

What they found for the collapsed but still unfolded state of Csp*Tm* was the presence of a significant content of β-sheet structure (ca. 20% of that of the folded state) but only a negligible amount of α-helical structure (below 3% of that of the native state). Thus, the remarkable result of their combined smFRET and SRCD measurements is that the collapsed state is, on one hand, well described by a simple random Gaussian chain model, but has, on the other hand, already a high content of secondary structure. Similar results had been observed before for proteins under strongly denaturing conditions, but have not yet been seen for the collapsed state under close to native conditions. Although the study does not yield a more detailed picture of the particular organization of the β-sheets within the collapsed unfolded state, the applicability of the Gaussian chain model means that the average length of the β-sheet structural elements has to be shorter than the determined persistence length (ca. 0.39 nm under native conditions).

Two other recent publications nicely complement the paper of Hoffman et al.: In ref. [10], Merchant et al. combine smFRET measurements with molecular simulations to study the collapsed but still unfolded state of the same Csp*Tm* protein. And in ref. [11], Mukhopadhyay et al. investigate with smFRET the collapsed state of the prion protein Sup35. All these studies represent important steps towards a better understanding of the fleeting but important collapsed unfolded state of proteins under close to native conditions, which is the crucial intermediate between the completely unfolded state and the completely folded structure.

Keywords: FRET (fluorescence resonance energy transfer) · protein collapsed state · protein folding · single-molecule spectroscopy · synchrotron radiation circular dichroism

[1] C. Levinthal in *Mössbauer Spectroscopy in Biological Systems* (Eds.: I. L. Monticello, P. Debrunner, J. C. M. Tsibris, E. Munck), University of Illinois Press, Urbana, **1969**, pp. 22–24; R. Zwanzig, A. Szabo, B. Bagchi, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 20–22.

[2] V. Daggett, A. Fersht, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497–502.

[3] O. B. Ptitsyn, *Adv. Protein Chem.* **1995**, *48*, 83–229.

[4] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, S. Weiss, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5179–5184; A. A. Deniz, T. A. Laurence, M. Dahan, D. S. Chemla, P. G. Schultz, S. Weiss, *Annu. Rev. Phys. Chem.* **2001**, *52*, 233–253.

[5] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, B. Schuler, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 105–110.

[6] B. Schuler, E. A. Lipman, W. A. Eaton, *Nature* **2002**, *419*, 743–747; E. A. Lipman, B. Schuler, O. Bakajin, W. A. Eaton, *Science* **2003**, *301*, 1233–1235; C. Magg, F. X. Schmid, *J. Mol. Biol.* **2004**, *335*, 1309–1323.

[7] S. Weiss, *Science* **1999**, *283*, 1676–1683.

[8] E. Sherman, G. Haran, *Proc. Nat. Acad. Sci. USA* **2006**, *103*, 11539-11543..

[9] S. Akiyama, S. Takahashi, K. Ishimori, I. Morishima, *Nat. Struct. Biol.* **2000**, *7*, 514–520; T. Uzawa, S. Akiyama, T. Kimura, S. Takahashi, K. Ishimori, I. Morishima, T. Fujisawa, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1171–1176.

[10] K. A. Merchant, R. B. Best, J. M. Louis, I. V. Gopich, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1528–1533.

[11] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, A. A. Deniz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2649–2654.

Received: April 12, 2007
Published online on June 19, 2007

