

# Hierarchical folding of cytochrome *c*

Syun-Ru Yeh and Denis L. Rousseau

**Formation of secondary structure during protein folding can now be probed on the submillisecond time scale. By monitoring the formation of  $\alpha$ -helix formation during the initial collapse of cytochrome *c*, its folding mechanism is delineated.**

To catch a glimpse of the complete molecular picture of a protein folding reaction is one of the major challenges for protein chemists. A central problem lies in clarification of the early folding events. Several lines of evidence suggest that the initial step in protein folding involves the collapse of a polypeptide chain<sup>1</sup>. However, it is unclear whether the collapse is accompanied by or preceded by any secondary and/or tertiary structure formation — does secondary structure form first and guide the formation of the tertiary structure? If a structure forms early in the folding process, is it always native-like or could it contain non-native elements (misfolded structures)? If wrong structural elements form, would they retard or accelerate the subsequent folding events? Are these processes different for each protein or are there general rules that are common to all proteins?

On page 514 of this issue, Akiyama *et al.*<sup>2</sup> address these questions by probing the formation of secondary structure in the folding reaction of cytochrome *c* in the submillisecond time scale with circular dichroism (CD). These new results show that the early folding intermediate of cytochrome *c* consists of ~20% of the native  $\alpha$ -helical structure and is insensitive to the refolding conditions. The authors also show that the majority of the secondary structure is established after the early collapse of the polypeptide chain with a time constant of ~500  $\mu$ s. These results together with previous observations in the literature suggest that the bottleneck for the folding reaction of cytochrome *c* originates from the final search for the native structure rather than from the initial condensation of the polypeptide chain. This elegant work reported by Akiyama *et al.*<sup>2</sup> not only advances our understanding of the early folding events of cytochrome *c*, but also provides a new tool for future investigations of other protein systems.

## Probing the early folding events

The problem in addressing the early folding events in the past has resulted from the



**Fig. 1** The cartoon illustrates that a blind man develops a very wrong impression of an elephant depending on which body parts he touches, demonstrating the importance of the examination of a folding reaction by as many probes as possible prior to drawing conclusions.

difficulties in finding appropriate techniques for initiating folding reactions and for probing the molecular structure of a protein as folding proceeds. For decades reversible folding reactions have been studied with stopped-flow instrumentation by mixing unfolded proteins with buffer thereby initiating the folding reaction by diluting out the denaturant. However, the typical mixing dead time for stopped-flow instrumentation is on the order of a few milliseconds during which a large portion of the reaction has gone to completion for many proteins. (The lost information is often referred to as the 'burst phase'). To overcome this limitation, several techniques have been explored to initiate the folding reactions without a mixing event, such as laser-induced CO-photodissociation<sup>3</sup>, photo-induced electron transfer<sup>4</sup> and laser temperature jump<sup>5,6</sup>. Although each of these techniques has had considerable success, the first two are only applicable to proteins such as cytochrome *c* that can have a photolabile ligand and a redox center and the third technique is limited to thermally-induced unfolding or refolding.

In an effort to make measurements that could be applied to all protein systems, submillisecond solution mixers have been developed over the past five years<sup>7-9</sup>. With

these new mixers the submillisecond folding kinetics have been probed by various spectroscopic techniques, including resonance Raman scattering<sup>7,10-12</sup>, tryptophan fluorescence<sup>8,9</sup> and small angle X-ray scattering<sup>13</sup>. Cytochrome *c*, a small  $\alpha$ -helical protein, has been one of the model proteins for folding studies because it has a heme group that is axially coordinated by a histidine (His 18) and a methionine (Met 80) that is covalently linked to the polypeptide chain through two cysteine residues (14 and 17). The environment of the heme can be monitored during folding by optical absorption and resonance Raman scattering. In addition, there is a single tryptophan residue in the protein with a strong fluorescence in the unfolded structure. This fluorescence is fully quenched in the native state due to Förster energy transfer to the heme. Thus it provides useful information concerning the change in molecular size during folding. Finally, two potential non-native heme ligands (His 26 and His 33) can give rise to misfolded structures that give additional information concerning the peptide chain dynamics during folding<sup>10-12</sup>.

## A controversial issue

With the development of new technology, the details of the folding mechanism

## news and views

of cytochrome *c* are beginning to emerge. However, controversy exists concerning the properties of the structures of the collapsed state. Roder and colleagues argue that the collapse results in a highly structured ensemble of states that are on the folding pathway. On the other hand Englander and coworkers argue that the collapsed state lacks any secondary or tertiary structure and is simply a response to the change in solvent conditions — that is, the consequence of transferring the polypeptide from a high concentration of denaturant to a low concentration of denaturant (good solvent to poor solvent).

The conclusions of the Roder group are based upon their tryptophan fluorescence studies<sup>14</sup>. They followed the folding reaction from 45 to 750  $\mu$ s and detected two distinct exponential phases, with time constants of  $\sim$ 50 and  $\sim$ 500  $\mu$ s. The rapid phase represented a loss of  $\sim$ 60% of the fluorescence and the second phase accounted for an additional 30%. The temperature dependence yielded activation enthalpies of 7.5 and 1.0 kcal mol<sup>-1</sup> for the rapid and the slow phases, respectively. The presence of activation barriers is consistent with the proposal that the early condensation of the polypeptide chain produces specific folding intermediate(s) rather than it being a simple response to a change in solvent conditions (a barrier-free process). Subsequent hydrogen exchange (H/D exchange) studies by Roder and coworkers revealed that within 2 ms of the initiation of folding there is moderate protection in the three helices that are present in the native structure (the N- and C-terminal helices and the 60s helix) and they postulated that the initial stages of folding involve the formation of partially organized productive folding intermediates<sup>15</sup>.

To assess the nature of the early events, Englander and coworkers took a different approach<sup>16</sup>. They compared the folding of the native protein with that of the truncated proteins fragments, F1–65 and F1–80, in which the C-terminal residues are deleted from the 104-residue native protein. From CD and H/D exchange measurements, they proposed that the truncated proteins contain no native secondary structure, even under native conditions, and thus are good models for the unfolded protein. Based on the observations that the burst phase kinetics monitored by CD and tryptophan fluorescence were the same in the truncated proteins as in the native protein one millisecond after mixing, they concluded that this early phase involved a collapse of the protein

due to the change in solvent conditions and did not represent the formation of an intermediate with defined secondary or tertiary structural elements. The dilemma is that the assumption that the truncated proteins contains no secondary structure is inconsistent with the conclusions drawn several years ago by Stellwagen and colleagues who reported native-like features in the same truncated proteins in aqueous solvents at neutral pH<sup>17,18</sup>.

### Chain condensation and secondary structure formation

The results reported by Akiyama *et al.*<sup>2</sup> shed new light on this controversial issue. Combined with previously reported data, an integrated view of cytochrome *c* folding is emerging. First, an intermediate, formed in  $<$ 100  $\mu$ s, is associated with a major condensation of the polypeptide chain<sup>12</sup>. This is confirmed by an exponential decrease ( $\tau = \sim$ 50  $\mu$ s) in the tryptophan fluorescence intensity, which indicates that there is a significant reduction in the overall chain dimension on this early time scale<sup>14</sup>. The compaction of the polypeptide chain during this nascent phase is also reflected by rapid changes in the heme ligation states<sup>12</sup> because the reduction in the polypeptide chain dimension brings potential native and non-native heme ligands, His 18 and His 33, respectively, into proximity of the heme. The results reported by Akiyama *et al.*<sup>2</sup> demonstrate that the intermediate formed during this phase attains  $\sim$ 20% of the native  $\alpha$ -helical structure. The secondary structure formation during this phase is insensitive to refolding conditions, confirming that the collapse is a kinetically controlled process.

Additional changes occurring over the next several hundred microseconds are characterized by a further collapse of the polypeptide chain and the development of  $\sim$ 70% of the native secondary structure. The fluorescence changes reported by Roder and coworkers<sup>14</sup> and the changes in far UV CD observed by Akiyama *et al.*<sup>2</sup> both display exponential decay rates of  $\sim$ 500  $\mu$ s. A collapsed protein was confirmed by small angle X-ray scattering studies in which a radius of gyration not very different from that of the native structure was observed in the 150–500  $\mu$ s time scale<sup>13</sup>. Furthermore, H/D exchange experiments show that the major helices are formed within a few milliseconds after the initiation of the folding, but the protection factors are several orders of magnitude smaller than those in the native state<sup>15</sup>. This suggests that the helical structures in this time domain are loosely

packed, perhaps resembling the solvated helices observed during the folding reaction of apo-myoglobin<sup>6,19</sup>.

Finally, the native structure evolves from the collapsed protein with a much slower rate ranging from tens to hundreds of milliseconds. The rate strongly depends on the refolding conditions. At neutral and high pH, misfolding due to the coordination of His 33 (ref. 20) to the heme results in a slow folding reaction involving a heterogeneous population. In contrast, under conditions disfavoring the coordination of His 33 (for example, at pH lower than 5 when His 33 is protonated<sup>21,22</sup>, or in the presence of exogenous imidazole that prevents the mis-coordination of His 33 by binding to the heme<sup>8</sup>) or in histidine mutants<sup>23</sup>, the native structure is reached within a few milliseconds. The changes in the heme coordination states during this folding phase follow a thermodynamic equilibrium<sup>12</sup>, indicating that the polypeptide chain retains significant structural flexibility at this stage. This is consistent with the proposal that the secondary structural elements formed early in the reaction are only loosely packed and are localized in a few specific polypeptide segments<sup>15</sup>, imposing a minimal effect on modulating the heme coordination states. Although the intrinsic reactivity of histidine towards the heme iron is much higher than that of the native ligand (Met 80), the equilibrium is shifted towards the native methionine coordination state, because it is biased by the native tertiary polypeptide fold as the folding progresses.

From an energetic point of view, the major reduction in entropy occurs during the early chain condensation phase in which the conformational space of the polypeptide is dramatically reduced without the development of extensive tertiary interactions. It is during the subsequent ligand exchange phase that the energy is minimized through structural fluctuations biased toward the native conformation. In essence, the high efficiency of the folding reaction is made possible by the rapid condensation of the polypeptide chain into a collapsed structure and the high fidelity of the folding reaction is achieved through the subsequent thermodynamically-controlled search for the global minimum.

### Closing remarks

There is no universal probe that can provide snapshots of the conformational changes of a protein along the folding coordinate with molecular resolution. It is through the integrated view obtained by various techniques that an accurate picture

can be obtained without falling into traps of illusions, as depicted by the blind man's view of the elephant (Fig. 1), adapted from an old Chinese fable. The results reported by Akiyama *et al.*<sup>2</sup> provide secondary structure information in the submillisecond time scale; they have added a new dimension to the current view of cytochrome *c* folding. Newly developed submillisecond techniques are opening new avenues leading to the understanding of the earliest events of folding in many other protein systems as well. They reveal that the development of the secondary structural elements early in the folding reaction varies from one protein to another. For example, a much larger fraction of the native secondary structure of apo-myoglobin<sup>6,19</sup> is established during the early phase of folding as compared to cytochrome *c*. To generate a general view of protein folding, it is important to investigate as many protein systems as possible by the new methods.

#### Acknowledgments

We thank E. Stellwagen of the University of Iowa for helpful discussions. This work was supported by a grant from the National Institutes of Health.

*Syun-Ru Yeh and Denis L. Rousseau are in the Department of Physiology and Biophysics, Albert Einstein College of Medicine, Yeshiva University, Resnick Campus, 1300 Morris Park Avenue, Bronx, New York 10461, USA. Correspondence should be addressed to D.L.R. email: rousseau@aecom.yu.edu*

1. Dobson, C. M., Sali, A. & Karplus, M. *Angew. Chem. Int. Ed.* **37**, 868–893 (1998).
2. Akiyama, S., Takahashi, S., Ishimori, K. & Morishima, I. *Nature Struct. Biol.* **7**, 514–520 (2000).
3. Jones, C. M. *et al. Proc. Nat. Acad. Sci. USA* **90**, 11860–11864 (1993).
4. Pascher, T., Chesick, J. P., Winkler, J. R. & Gray, B. H. *Science*, **271**, 1558–1560 (1996).
5. Gruebele, M., Sabelko, J., Ballew, R. & Ervin, J. *Acc. Chem. Res.* **31**, 699–707 (1998).
6. Dyer, R. B., Gai, F., Woodruff, W. H., Gilmanshin, R. & Callender, R. H. *Acc. Chem. Res.* **31**, 709–716 (1998).
7. Takahashi, S. *et al. Nature Struct. Biol.* **4**, 44–50 (1997).
8. Chan, C.-K. *et al. Proc. Nat. Acad. Sci. USA* **94**, 1779–1784 (1997).
9. Shastry, M. C. R., Luck, S. D. & Roder, H. *Biophys. J.* **74**, 2714–2721 (1998).
10. Yeh, S.-R., Takahashi, S., Fan, B. & Rousseau, D. L. *Nature Struct. Biol.* **4**, 51–56 (1997).
11. Yeh, S.-R. & Rousseau, D. L. *Nature Struct. Biol.* **5**, 222–228 (1998).
12. Yeh, S. R., Han, S., and Rousseau, D. L. *Acc. Chem. Res.* **31**, 727–736 (1998).
13. Pollack, L. *et al. Proc. Natl. Acad. Sci. USA* **96**, 10115–10117 (1999).
14. Shastry, M. C. R. & Roder, H. *Nature Struct. Biol.* **5**, 222–228 (1998).
15. Sauder, J. M. & Roder, H. *Folding Des.* **3**, 293–301 (1998).
16. Sosnick, T. R., Shtilerman, M. D., Mayne, L. & Englander, S. W. *Proc. Nat. Acad. Sci. USA* **94**, 8545–8550 (1997).
17. Babul, J., McGowan, E. B. & Stellwagen, E. *Arch. Biochem. Biophys.* **148**, 141–147 (1972).
18. Bryant, C. & Stellwagen, E. *J. Biol. Chem.* **260**, 332–336 (1985).
19. Callender, R. H., Gilmanshin, R. Dyer, R. B. & Woodruff, W. H. *Annu. Rev. Phys. Chem.* **49**, 173–202 (1998).
20. Colon, W., Wakem, L. P., Sherman, F. & Roder, H. *Biochemistry* **36**, 12535–12541 (1997).
21. Elove, G. A., Bhuyan, A. K. & Roder, H. *Biochemistry* **33**, 6925–6935 (1994).
22. Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. *Nature Struct. Biol.* **1**, 149–156 (1994).
23. Pierce, M. M. & Nall, B. T. *Protein Sci.* **6**, 618–627 (1997).

## StARTing to understand cholesterol transfer

Douglas M. Stocco

**The crystal structure of a lipid transfer START domain has been solved and shown to contain a hydrophobic tunnel that likely binds cholesterol. These results have implications for understanding the mechanism of action of the steroidogenic acute regulatory protein (StAR), which is indispensable in steroid biosynthesis.**

The steroid hormones — which are required for the homeostatic maintenance of blood pressure, carbohydrate metabolism and reproductive function — represent some of the most important molecules found in the body. Their synthesis is regulated by signals from the anterior pituitary gland that act on specific steroidogenic cells found mainly in the adrenal glands and gonads. In response to these signals the steroidogenic acute regulatory protein (StAR) is synthesized. This protein is required for the rapid increase in steroid hormone production and mediates the delivery of cholesterol to the enzyme — cytochrome P450 side chain cleavage (P450 scc) enzyme — that converts cholesterol to pregnenolone, the first steroid formed<sup>1</sup> in the steroidogenic pathway. This enzyme resides in the inner mitochondrial membrane, and it has long been known that the transfer of chole-

sterol from the outer mitochondrial membrane to this enzyme is the rate-limiting step in steroidogenesis<sup>2</sup>.

To date, the details of how the biosynthesis of steroid hormones is regulated have remained obscure. Now, the paper by Tsujishita and Hurley<sup>3</sup> in the May issue of *Nature Structural Biology* reports a high resolution tertiary structure of a critical region, called the START domain (for StAR related lipid transfer domains), of the MLN64 protein, a 50 kDa protein of unknown function that is highly and specifically expressed in the malignant cells of breast carcinomas. This START domain contains a hydrophobic tunnel that appears capable of binding a single cholesterol molecule. Thus, this finding is a first step toward understanding cholesterol transfer.

START domains are found in a wide variety of proteins including the phos-

phatidylcholine transfer protein, acyl-CoA thioesterase, p122-RhoGAP, the Goodpasture antigen binding protein and, most importantly for the work described here, MLN64 and StAR<sup>4</sup>. They are ~200-residue lipid-binding motifs<sup>4</sup>, and the importance of the new structure is that the START domain in MLN64 is highly homologous to the START domain in the StAR protein that is required for the regulation of steroid biosynthesis.

Many studies have demonstrated excellent correlations between StAR expression and steroid hormone biosynthesis. For example, expression of StAR through transfection results in increases in the transfer of cholesterol to the inner mitochondrial membrane and in steroid biosynthesis by both steroidogenic and nonsteroidogenic cells<sup>5–7</sup>. Therefore it is clear that StAR can somehow mediate cholesterol transfer to the inner mito-