# The specificity of protein aggregation

*Jeannine M. Yon*

Unraveling the mechanisms of protein folding represents one of the most challenging problems in biology today. The reams of amino acid sequence arising from large-scale genome sequencing projects makes the solution of this problem all the more urgent—progress in the prediction of three-dimensional structure from primary protein sequence is reliant on greater knowledge of the protein folding process. Currently, there are two major impediments to solving the folding problem: First, the high cooperativity of the process, and second, its rapidity.

Protein aggregation is a widespread phenomenon that can occur during protein folding in vivo as well as in vitro. The likelihood of intermediates in a folding pathway generating misfolded species is dependent on such conditions as protein concentration, pH, temperature, ionic strength, and redox environment: There is kinetic competition between the correct fold and misfolded forms, which may or may not result in the formation of aggregates. Aggregation in the form of insoluble inclusion bodies is frequently encountered during the overexpression of recombinant proteins in foreign hosts, and this represents a serious problem for the biotechnology industry. The pathogenic mechanisms of certain genetic disorders, such as Creutzfeldt–Jakob and Alzheimer's disease, are also known to involve protein aggregation[1].

The formation of aggregates has often been considered a nonspecific association of partially folded polypeptide chains through hydrophobic interactions. However, several analyses suggest that aggregation occurs by specific intermolecular associations[2]. In this issue (pp. 1283–1287), Speed et al.[3] provide an elegant demonstration of the specificity of aggregation. They show that during the in vitro refolding of a mixture of two proteins, folding intermediates do not coaggregate with each other, but only with themselves. Two proteins were studied: Tail spike endorhamnosidase and coat protein from phage P22. The first of these is a thermostable trimer whose folding intermediates are thermolabile and either undergo productive refolding or form multimeric aggregate intermediates. The P22 coat protein, which comprises the capsid shell of phage P22, yields either a productive fold or "off-pathway" aggregates upon refolding. Both proteins have been intensively studied in King's laboratory.

In the present work[3], the authors first denatured the proteins in urea and then chose refolding conditions such that aggregation competes with correct folding. Folding and soluble aggregation intermediates of the two proteins were characterized by western blotting after refolding, either separately or mixed together. No heterogeneous aggregates were observed, clearly indicating that only self-association of transient refolding molecules occurs in the formation of soluble multimers. This result suggests the possibility of obtaining relatively pure protein from the inclusion body state and underlines the importance of kinetic factors.

The control of kinetic factors is indeed very important for refolding in vitro. The formation of transient multimeric species during the fast refolding phase of a monomeric protein, yeast phosphoglycerate kinase, and their dissociation during the slow folding phase to yield the monomeric active protein, has been observed for protein concentrations as low as 0.05 μM[4]. It is a specific process involving a defined region of the N-terminal domain[4]. Transient oligomeric forms have also been reported for bovine growth hormone[5] and for tryptophan synthase[6]. It is feasible that several other proteins undergo reversible self-association during the fast phase of their refolding process; however, this multimeric species may be difficult to detect.

One mechanism to account for aggregate formation during refolding of oligomeric proteins is domain swapping. This was first suggested by Monod and later proposed by London et al.[7] to account for the formation of aggregates during the refolding of tryptophanase. The mechanism involves the replacement of one domain of a monomeric protein by the same domain of an identical polypeptide chain, thus resulting in an intertwined oligomer[8].

Recent results have demonstrated that the in vitro refolding of several proteins proceeds through multiple pathways[9] at least in the early steps. Some of them could result from self-association of molecules either partially or incorrectly refolded, which further yields the correctly folded protein or aggregates, depending on the conditions that affect the kinetic parameters. The same parts of the polypeptide chain might be specifically recognized in the intermolecular association as in the intramolecular interaction.

What is the intermediate that generates the multimeric species? With the lack of experimental information, this is difficult to determine—we can only hypothesize. Speed et al.[3] suggest that the kinetic effects observed may reflect transient associations of molten-globule intermediates. However, in the molten-globule state nearly all the secondary structure is formed, whereas the tertiary interactions are not established. It seems more likely that the intermolecular associations arise from fluctuating species that precede the molten-globule state; although these intermediates are fluctuating and very unstable, they could be capable of specific recognition. The determination of this intermediate will be very important in the control of the folding process.

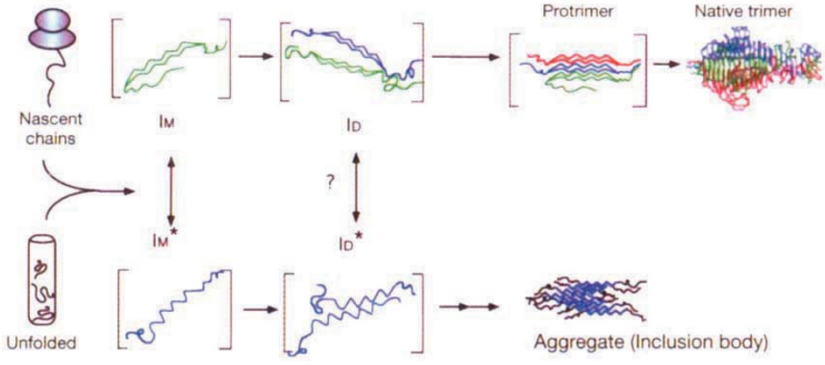


**Figure 1. Folding and aggregation pathway in vivo and in vitro.**

*Jeannine M. Yon is professor of biochemistry at Université de Paris-Sud, Laboratoire d'Enzymologie Physicochimique et Moléculaire, Unité Associée du Centre National de la Recherche Scientifique, Bâtiment 430, 91405 Orsay, Cedex, France.*

1. Kelly, J.W. 1996. *Curr. Opin. Struct.Biol.* **6**:11–17.
2. Brems, D.N. et al. 1988. *Proc.Natl.Acad.Sci. USA* **85**:3367–3371.
3. Speed, M.A. et al. 1996. *Nature Biotechnology* **14**:1283–1287.
4. Pecorari, F. et al. 1996. *J. Biol. Chem.* **271**:5270–5276.
5. Brems, D.N. 1988. *Biochemistry* **27**:4541–4546.
6. Matthews, C.R., and Crisanti, M.M. 1981. *Biochemistry* **20**:784–792.
7. London, J. et al. 1974. *Eur. J.Biochem.* **47**:409–415.
8. Bennett, M.J. et al. 1995. *Protein Sci.* **4**:2455–2468.
9. Radford, S.E. et al. 1992. *Nature* **358**:302–307.