PROTEIN FOLDING: LOCAL STRUCTURES, DOMAINS AND ASSEMBLIES

R. Jaenicke

Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, D-8400 Regensburg, Federal Republic of Germany

SUMMARY: Globular proteins show the intrinsic property of acquiring their spatial structure in an autonomous way, based solely on their amino-acid sequence and their aqueous or non-aqueous environment. In order to gain insight into the mechanism of folding the essential steps in the "hierarchical condensation" from the nascent (unfolded) to the native state of a given protein have to be characterized. As taken from spectral data, short-range interactions stabilize well-defined local structures (α-helices, β-turns, loops) in independent segments of the polypeptide chain. In proceeding from elements of secondary- and supersecondary structure to subdomains and domains, the native tertiary and quaternary structure are finally generated by the merging and docking of domains and subunits. Reconstitution after preceding denaturation allows the mechanism of folding to be unravelled; at the same time, it allows "authentic" (native) protein to be recovered from "inclusion bodies".

FOLDING IN VIVO AND FOLDING IN VITRO, STRUCTURE PREDICTIONS

The structural integrity of proteins in solution depends on the solvent parameters. Accordingly, one would predict that protein folding is strongly influenced by the environment. However, a variety of experimental findings have proven that the solvent conditions upon translation and reconstitution are less critical than expected: in vitro folding and assembly may be accomplished in dilute buffer solution in the absence of components involved in cellular folding events; biologically active thermophilic proteins may be expressed in mesophilic hosts; cotranslational and posttranslational modifications such as glycosylation or processing do not necessarily interfere with the intrinsic capacity of the polypeptide chain to acquire its native 3D structure. Except for the influence of viscosity and specific ligands (coenzymes, substrates, ions etc), hardly any attempts have been made to mimic the cytoplasm in folding experiments. The fact that the nascent or refolding chain requires neither extrinsic factors nor the input of energy in order to generate the native structure has been considered sufficient evidence to postulate that the genetic code governs both translation and folding. Whether there is a unique folding code as the "second half of the genetic code" remains still to be shown (Fasman, 1989). That it cannot be colinear is trivial for the following rea-
sons: both local next-neighbor and non-local through-space interactions are involved in the minimization of potential energy; as a consequence, identical stretches of polypeptide chain may determine different 3D structures; widely differing ("homologous") sequences code for identical topologies; subdomains and domains as cooperative entities are separated by connecting peptides exhibiting anomalous configurations; extrinsic effects or effectors (not inherent in the amino-acid sequence) may play a significant role in the folding process. The latter argument has been shown to be essential in cases where cofactors or chaperones serve to stabilize intermediates of folding or assembly (Gerschitz et al., 1978; Ellis, 1990; Fischer & Schmid, 1990). Other cell-biological implications that may interfere with a general 1D → 3D algorithm of protein folding are: cellular compartmentalization, genome organization, transcription control, codon usage, amino-acid pools, kinetic competition of folding and association in overexpressing hosts, discontinuity in the rate of translation etc (Jaenicke, 1987, 1988a).

In spite of these pitfalls, there have been numerous attempts to forecast the 3D structure of proteins or their mode of folding: Search programs for sequence homologies have been successfully applied to correlate given primary structures to a limited number of protein "families". Statistical analyses of preferences for α-helices, β-strands, turns, or random structures provide secondary structure predictions with reliabilities of the order of 65%. Topological considerations and docking procedures have been developed to optimize both minimum hydrophobic surface area and maximum packing. Energy minimization and molecular dynamics calculations, as well as semi-quantum mechanical and statistical mechanical methods proved useful in reducing the number of possible conformations from an astronomically high value to only few. They have been most valuable in characterizing conformational changes with high precision. A combination of all available methods in terms of knowledge-based computer-aided structure predictions has been conceived by Blundell et al. (1987). The result of ≈ 90% correct prediction with an rms deviation < 3 Å is most satisfactory from the theoretical point of view; however, to predict the functional state, one has to be > 99% correct, so that at present all structure predictions have still to be taken with a grain of salt.

THERMODYNAMICS VS KINETICS

The acquisition of the native 3D conformation of a protein is determined by the kinetic pathway rather than random search. Based on this hypothesis, folding occurs by the fastest route available. It includes well-populated intermediates and generates the native state as the kinetically accessible