# A HIERARCHIC MODEL FOR THE SELF-ASSEMBLY OF GLOBULAR PROTEINS

George D. Rose, *University of Delaware, Department of Chemistry, Newark, Delaware 19711 U.S.A.*

In the past years many globular proteins have been shown to be neatly separable into large, spatially compact, continuous-chain regions called domains. Often protein crystallographers have dissected their molecules into two or three of these domains and observed that biologically relevant binding sites are situated between them. Domains are now regarded as commonplace components in proteins, and proposals have been made that a given domain is a separately organized, independently assembled, and functionally distinct unit within the whole protein monomer (1–3). These ideas have been stimulated by recent findings that "genes in pieces" can code for "proteins in pieces" (4, 5).

We have been interested in understanding the role of domains in protein structures, and to that end we developed a computer program to identify the domains from x-ray coordinates (6). In this analysis, the polypeptide chain is viewed conceptually as a tangle of string in space. A single planar cut is made after which the chain is examined, as illustrated in Fig. 1. Domains are said to exist whenever some cut divides the tangle into two linear segments, and not into many smaller segments. A ball of yarn, for example, has no domains.

Using this computational tool, we found that the large qualitatively described domains mentioned in the literature are only a special case of a far more general phenomenon: namely, big domains are, in turn, composed of smaller domains. In greater detail, a whole protein monomer can be successively subdivided into subdomains of decreasing size until finally individual units of secondary structure are achieved, as shown in Fig. 2. We consistently observed this unified architecture in a test set of 22 x-ray elucidated proteins. A similar result has also been reported by Crippen (8).

A hierarchy is a kind of ordering in which each component of some structure is wholly contained within a parent component, as suggested by the nested boxes in Fig. 3. The domains of a protein are found to be organized in a hierarchy.

As an example, the organization of hexokinase monomer is shown schematically in Fig. 4. In this representation, each branch point describes the way that a hexokinase domain is subdivided into its component subdomains. In this analysis, no part of the chain is excluded after the protein is partitioned into its domains; every residue belongs to one domain or another.

---

Dr. Rose's current address is Pennsylvania State University, Milton S. Hershey Medical Center, Department of Biological Chemistry, Hershey, Pa. 17033.

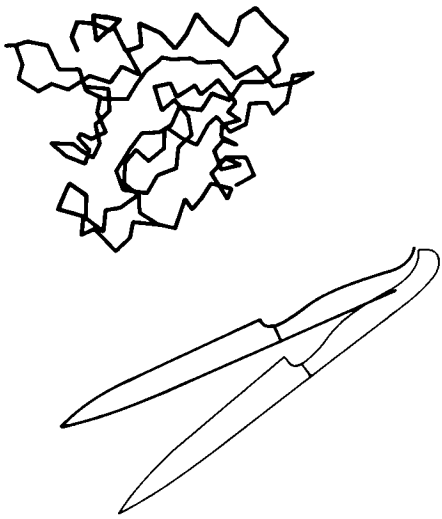

Figure 1 Finding the domains in a protein. A planar cut is chosen that optimizes division of the molecule into continuous-chain fragments.

The hierarchic pattern of domain organization is experimental evidence in favor of a self-assembly process that proceeds by hierarchic condensation. This folding model can be represented by a tree structure, such as shown in Fig. 4, but with the arrows leading from bottom to top. Any folding pathway (9) leading to the native state can be described by such a tree, and each tree is completely specified by a stepwise sequence of folding events.

This sequential model for folding is appealing because only a computationally small number of trees will exist for any given protein. Steps in the assembly dynamics of a given tree are as follows:

(*a*) Nearby hydrophobic chain elements interact to form small folding intermediates of marginal stability (10).

↓

(*b*) Neighboring intermediates coalesce, leading to successively larger intermediate structures.

↓

.

.

.

↓

(*c*) Upon completion, the native structure and hydrophobic core of the whole protein monomer are achieved (11).

Every step in this cascade is characterized by a condensation of continuous-chain intermediates formed at the previous step, with each step enhancing the compactness and conformational stability of the molecule.

The number of conceivable folding trees for each protein depends upon the possible interactions between folding intermediates. The small equilibrium intermediates, near the base of a folding tree, will be rapidly interconvertible. Only as these smaller modules merge with each other to form successively larger intermediates will conformations of persisting stability begin to emerge. However, most folding trees are expected to terminate prematurely

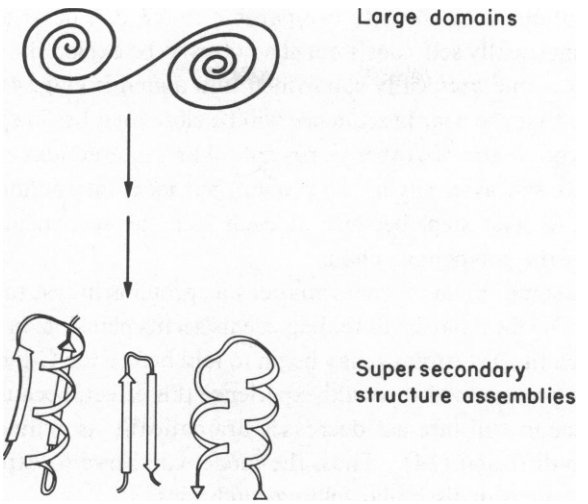


Figure 2 Levels of structure discovered by the domains' program. The whole protein monomer is divided into large domains; these are further subdivided in successive steps until supersecondary structures (7) and then individual helices and strands are found.

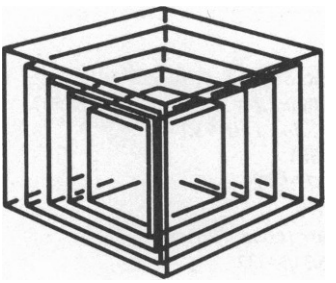

Figure 3 Chinese boxes form a structural hierarchy. Each box is wholly contained within the next larger box.

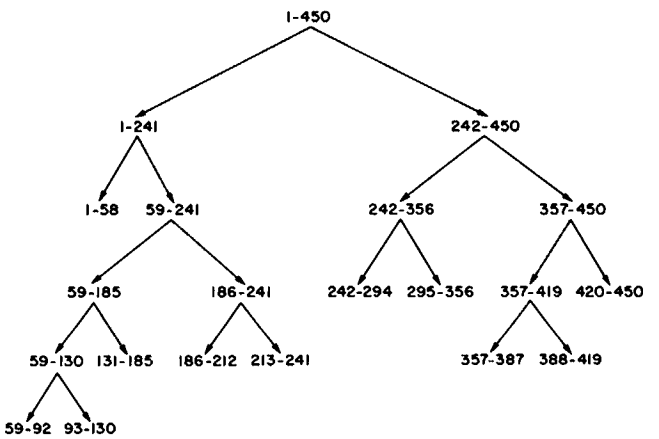


Figure 4 The domain structure of hexokinase monomer taken from our computer analysis of 22 proteins (6). Each branch point shows how a discovered hexokinase domain is divided into subdomains. The abbreviated tree shown here is terminated at the point where the domains no longer form closed structures.

(12) before conformations with stability comparable to $kT$ can be attained, leaving only a minor residue of geometrically self-consistent structures to be explored.

That proteins can be hierarchically subdivided into domains (Fig. 4) shows that parts of the polypeptide chain that are near in sequence will be close together in space as the hierarchy is traversed from bottom to top. In other words, all folding interactions are local ones at some step in the cooperative self-assembly of the protein; but local interactions at the $i$th step are not yet local at the $(i-1)$st step, because at each step the intermediates will include an increasing measure of the polypeptide chain.

Proteins in cells assume the same conformations as proteins in test tubes (13), in many, if not in all, instances. The local nature of folding events seems plausible in the case of an in vivo nascent chain in which the N-terminus may begin to fold before the C-terminus is completed. But even a denatured protein in vitro would experience this effect, because the likelihood that two sites along the chain will interact decreases dramatically as a function of the distance between them due to diffusion (14). Thus, the model can serve to explain the existence of similar folded end-products in dissimilar folding conditions.

This work was supported by National Institutes of Health grant GM27370.

*Received for publication 24 January 1980.*

## REFERENCES

1. Wetlaufer, D. B. 1973. *Proc. Natl. Acad. Sci. U.S.A.* **70**:697–701.
2. Liljas, A., and M. G. Rossman. 1974. *Annu. Rev. Biochem.* **43**:475–507.
3. Ptitsyn, O. B. 1978. *Fed. Eur. Biochem. Soc. Lett.* **93**:1–4.
4. Gilbert, W. 1978. *Nature (Lond.).* **271**:501.
5. Blake, C. C. F. 1979. *Nature (Lond.).* **277**:598.
6. Rose, G. 1979. *J. Mol. Biol.* **134**:447–470.
7. Levitt, M., and C. Chothia. 1976. *Nature (Lond.).* **261**:552–558.
8. Crippen, G. M. 1978. *J. Mol. Biol.* **126**:315–332.
9. Levinthal, C. 1968. *J. Chim. Phys.* **65**:44–45.
10. Rose, G. 1978. *Nature (Lond.).* **272**:586–590.
11. Kauzmann, W. 1959. *Adv. Protein Chem.* **14**:1–63.
12. Cohen, F. E., T. J. Richmond, and F. M. Richards. 1979. *J. Mol. Biol.* **132**:275–288.
13. Anfinsen, C. B. 1973. *Science (Wash. D. C.).* **181**:223–230.
14. Karplus, M., and D. L. Weaver. 1979. *Biopolymers* **18**:1421–1437.

# LOCAL CONFORMATIONAL RELAXATIONS AND PROTEIN FOLDING-UNFOLDING TRANSITION

Tian Yow Tsong and Minoru I. Kanehisa, *Department of Physiological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205 U.S.A.*

A protein molecule is a small system; as such its conformation constantly fluctuates (1). These events cover a broad time range: picoseconds for bond vibrations, oscillations and rotations, nanoseconds to microseconds for chain motions at a local level, and milliseconds to seconds for folding-unfolding transition. In a recent study we examined how events associated with rapid intramolecular motions or local conformational relaxations might affect the dynamics of the chain folding. We have proposed a cluster model (2–5) which postulates that