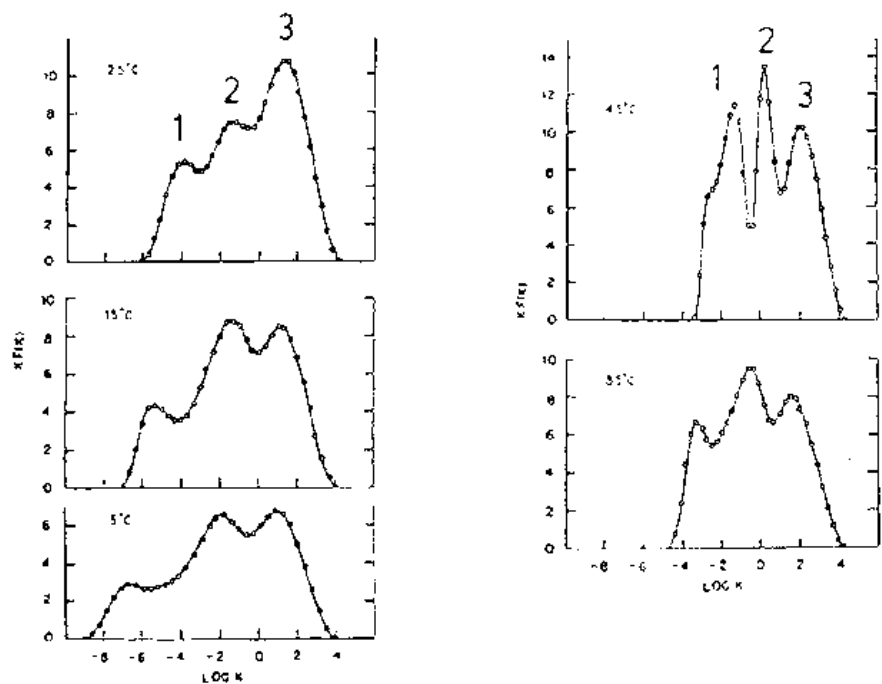


## ***Chapter 2. Substructures of proteins-Introduction***

Genome sequencing provides a wealth of information in an unknown language. One approach to surmounting this language barrier is to work backward from such universal features of protein construction and function as are known. There are several of the latter some known generally and some scarcely known at all. This monograph is a partial update of the first review article on the involvement of protein conformation in function published in 1969 by Lumry and Biltonen. It has been supplemented and updated frequently with relatively little conceptual material added until the discovery of the protein substructures in 1981 opened the way for more detailed examination of the molecular basis of protein structure and function outlined in the first review. Gregory and Lumry have described many of the consequences of that discovery. Enzymic function is used in later chapter to illustrate the possible significance of C-2 and what can be called pseudo C-2 symmetry as a source of mechanical work. The ubiquitous C-2 rotational symmetry of the catalytic domains on which enzymic catalysis depends is a major discovery. It provides a clean distinction between the popular transition-state stabilization mechanisms and the mechanical mechanism arising from destabilization of pre-transition states by conformational forces. Enzymic catalysis like protein construction itself is a consequence of the discovery in evolution of several kinds of protein substructures. The differences are not yet detectable within the current errors in the structural coordinates from protein x-ray-diffraction and nmr methodologies and as a result have generally escaped attention. However, almost all details of protein structure and function depend on the substructures so this monograph has been constructed to detail their construction and physiological importance. That can be best achieved in a limited space by using older papers to describe concepts most of which have long since been published and newer ones for

added detail and new applications. The primary vehicle making comprehensive understanding of the integration of structure and function possible is the so-called “temperature factor” determined in diffraction studies.

That qualitatively different substructures exist was implied by the finding by Linderstrøm-Lang that the rates of exchange of proton between protein sites and water fall into characteristic groups well separated along the rate-constant axis. The typical size of the slowest group was subsequently determined by Hnojwey and Reyerson. However that there are additional qualitatively and quantitatively different groups did not emerge until 1981 when the empirical probability distributions were evaluated. The probability-density distributions (pdf) for the rates of exchange of protons between protein sites, mostly backbone amides, and liquid or gaseous solvent when extracted by Gregory using Provencher’s CONTIN program for numerical evaluation of LaPlace transform consisted of three distinct, slightly overlapping peaks revealing a minimum of three kinds of protein substructures (Fig. 1.).



**Figure 2-1. Probability of an exchange site with a given rate-constant versus the rate-constant value. Found by Gregory using a numerical procedure developed by Provencher. Jaynes Max-ent procedure can also be used. The three cusps distinguish three qualitatively different kinds of substructure and there appear to be only three. Overlap of cusps is negligible since cusp I and cusp II generate different enthalpy-entropy compensation plots now found to apply to all mesophiles with paired functional domains.**

The residues in each substructure can be identified using varieties of exchange data but more detailed descriptions can be made with the Debye-Waller scattering factors known as atom “temperature” factors and usually tabulated in the Protein Database files as B values. These factors called “anisotropic displacement parameters (ADP)” by small-molecule crystallographers are related to the mean square displacement of an atom from its ideal lattice point,  $\overline{\mu^2}$  in  $\text{Å}^2$ , by the relationship for each atom  $B_i = 8\pi^2 \overline{\mu_i^2}$ . Since the quality of protein diffraction data is rarely sufficient to justify computing the three axes of the scattering ellipsoids, an isotropic model has been assumed in computing the mean square displacements. Within the accuracy of the data and refinement procedures the “isotropic” atom B value provides very useful estimates of the free volume available to the atom. Free volume has been defined in many ways. As used here it is the volume accessible for fluctuations of the center of an atom averaged over local space, packing, vibrational amplitude, conformational fluctuations and lattice disorder. B values are not corrected for neighbor-neighbor and other communal volume sharing but so long as the whole-molecule lattice fluctuations are small the errors those corrections are small. Then low B values mean low free volume and by comparison of the B pdf with the residue pdf it is found that rarely do more than a few atoms in any

residue appear in peak I, the slowest exchange-rate peak. Those atoms characterize a protein family and in contrast to their residues are tightly conserved. This distinction continues to be a source of some confusion insofar as protein “cores” are identified with whole residues rather than just the atoms in the knots. However residue conservation in this group also tends to be high with successful substitutions usually limited to the familiar like-for-like kind. Major confusion continues to arise from identifying the “slow-exchange cores” with collections of predominately aliphatic and aromatic residues without taking into account the central importance of hydrogen bonds. Slow exchange cores are confused with something called a “hydrophobic core” usually defined in terms of dispersion interactions among clusters of oily groups as the basis of stability of folding. This idea is due to misinterpretation of Kauzmann correct suggestion that the poor solubility in water of such groups is a possible major factor in folded stability. Dispersion interactions at normal densities are weak. Consider benzene crystals and petroleum jellies. So poor solubility can push polypeptides toward folding but dispersion interactions among oily residues in folded proteins can themselves make only minor contributors to stable folding. Rather residue sidechains of aromatic and aliphatic residues achieve major importance in folding in cooperation with hydrogen bonds to produce compact regions of low dielectric constant and low electrostatic potential energy. Preference for aromatic residues suggests that high polarizability may be important so it is quite likely that there are layers of sophistication in knots well beyond those obvious now.

Plots of B values versus atom or molecule are the working tool for the study of protein structure and it is important to note that atom B values must be used. As note above, only a few atoms of any residue have knot B values so conventional use of residue-average B values must be avoided. In a similar source of confusion conservation of genetic form in a protein family does not depend on exact conservation of knot residues. Rather a more reliable

description is that given by atom free volumes. The knots are the branch points that conserve the lengths and placement of the segments of the main chain and thus the zero-order structure of a protein family.