Prior to 1959 it was generally assumed that every protein structure would be radically different given the almost infinite possibilities in secondary structural arrangements. Today, with the advent of a relatively large catalogue of structures for water-soluble proteins, order is emerging from diversity, albeit not without controversy. Proteins can be generally classified in certain architectural categories. Within the structural divisions are often found repeating topological motifs or domains with super-secondary structures which provide specific and similar functions for various proteins. Examples can be drawn from the spatial superposition of Cα atoms in nucleotide, polysaccharide, and heme binding proteins as well as viral capsid subunits. Yet the code which relates amino acid sequence to structure is highly degenerate, permitting alteration of specific residues without loss of fold, function, or ancestral relationship (“divergent” evolution). Introns may provide the genetic mechanism to shuffle about the function-specific domains. On the other hand, structural equivalence (“convergent” evolution) was found in molecules displaying only weak or non-existent functional relationships, such as superoxide dismutase and the immunoglobulin domain or haptoglobin and the serine proteases where even primary structural homology is preserved. The concept of convergence is further enhanced by the spatially superimposable active centers of molecules bearing little topological similarity; for example, subtilisin and chymotrypsin or the zinc dependent enzymes. Quantitative attempts have been made to distinguish the two evolutionary schemes though not with complete success.

The wealth of biologically significant structures produced by X-ray crystallography seems to have narrowed their possible diversity and yet expanded the modes and etiology of their formation.

Although the α-helix and β-sheet were predicted prior to observation, the prediction of favourable tertiary structures has proved much more elusive. With the increasing data bank of protein crystal structures, observations on structures, rather than theory, form the basis of our current understanding of protein structure. In recent years the major advances have been in the area of protein topology. We now know, for example, that proteins fall into structural families, that certain super-secondary structures (e.g., the Greek key) occur frequently and that the larger proteins sub-divide into domain structures. These topological preferences can be incorporated into the prediction of protein structure by the method of generating all possible topologies for a given protein, and then attempting to identify the correct fold. To do this successfully it is necessary to develop criteria, and to understand the factors which make the native fold particularly favourable. Such criteria can only be derived by detailed analyses of the available protein structures, including not only consideration of topology but also the many other different aspects of protein structure which combine to stabilise the native state. With increasing refinement of protein coordinates, reliable data on side-chain conformation and packing between side chains are now available. This opens a new area of protein structure analysis.

The results of several detailed analyses performed in the Department of Crystallography at Birkbeck will be described. In the area of topology, a survey has been made of the ‘role’ of the amino and carboxy terminal regions in protein structures. For example, we find that the termini often form interdomain links or monomer–monomer contacts, but are rarely involved in the active