02.X-04 STRUCTURAL COMPARISON OF PROTEINS. By Patrick Argos, Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, U.S.A.

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 C_α 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, subtillisin 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.

02.X-05 THE PREDICTION OF PROTEIN STRUCTURE FROM AMINO ACID SEQUENCE. By M.J.E. Sternberg, F.E. Cohen and W.R. Taylor, Laboratory of Molecular Biophysics, Department of Zoology, South Parks Road, Oxford OX1 3PS, England.

Renaturation experiments show that in general it should be possible to predict theoretically the three-dimensional structure of a protein from its amino acid sequence. The approach of structure prediction by the minimization of an energy function will be reported and the current problems described. An alternative approach recognizes that the tertiary folds of many globular proteins involve the packing of α-helices and β-strands according to one of three motifs - the docking of α-helices to form an α/α protein, the stacking of two β-sheets (β/β), and the packing of α-helices against a predominantly parallel β-sheet (α/β). The first step is to locate the regular secondary structures and the current methods of prediction will be reported. The next step uses rules derived from analysis of the known structures, in particular the geometry of packing, the patterns of non-polar residues that mediate the interaction, and topological restrictions on the chain fold. The application of these rules in a 'combinatorial' algorithm will be reported for trials on proteins of known conformation and predictions of proteins whose structures have not been determined.

02.X-06 SOME PRINCIPLES OF PROTEIN STRUCTURE. By J. M. Thornton, Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

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 supersecondary structures (eg, 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

site. On a different aspect of structure, a study of disulphide bridges in proteins has recently been completed. Their distribution, topology, conformation and conservation were analysed. Several general patterns emerge which to some extent dictate disulphide bridge formation. For example, there is a strong preference for shorter connections, with half-cystines separated by less than 24 residues in 50% of all disulphides. This analysis of the covalent disulphide bridges led to a consideration of the weaker electrostatic salt-bridges between charged amino acid side chains. Preliminary results derived from a survey of salt bridges in high resolution protein structures will be presented.

02.X-07 WATER AND PROTEIN FOLDING. By J. M. Goodfellow, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

Solvent effects are known to play a significant role in many important aspects of protein interactions including folding. The contributions to the free energy of folding which involve water interactions include :

(i) the entropy on release of water molecules hydrogen-bonded to the unfolded conformation, and

(ii) the relative strengths of hydrogen bonds between polar-polar and polar-water groups.

Estimates of these terms are difficult to make as they depend critically on the geometry and energy of weak, not always well-characterised hydrogen bonding interactions. Finney *et al* (1980) have attempted to estimate these contributions and to compare them with other terms including the so-called hydrophobic interaction.

If more detailed calculations are to be made, we must improve our knowledge about the relevant interactions especially those between water molecules and groups on the protein. Such intermolecular potential energy functions are being developed based on the polarisable electropole model for water which allows us to incorporate the known cooperative effects in hydrogen-bonded systems. After extensive testing against experimental data on amino acid hydrate crystals, these potentials are being used to examine the state of water around biomolecules using Monte Carlo simulation techniques.

Although *potential* energies are easily extracted from these simulations *free energies* require special methods. Such methods have been used successfully on small, homogeneous systems and are being extended to look at free energy differences in the much larger and heterogeneous protein-solvent systems. In principle, it is possible to calculate the free energy difference between the native and fully extended conformations of a protein using these methods.

Finney, J. L., Gellatly, B. J., Golton, I. C. and Goodfellow, J. M., Biophysics J. 32, 17 (1980).

02.X-08 LAMBDA REPRESSOR. By Carl Pabo and Mitchell Lewis, Department of Biochemistry, Harvard University, Cambridge, Mass. USA

The crystal structure of an amino-terminal fragment of lambda repressor has been determined at 4.5 Å resolution. This fragment, which was generated by cleaving repressor with papain, contains the first 92 amino acids of repressor and binds specifically to the lambda operators. (The intact protein contains two domains. The amino-terminal domain recognizes the operators, and the carboxy-terminal domain allows the protein to dimerize.) The amino-terminus crystallized in space group $P3_121$ with cell dimensions of a=b=65 Å, c=150 Å. These crystals diffract to 2.5 Å resolution. An unusual pattern of weak reflections and the observation that related crystal forms show planar disorder allowed us to deduce that there were three layers of molecules along the c axis. (When h and k are both even, reflections that would be absent in a rhombohedral cell tend to be weak.) Each layer has six molecules, which are related by a crystallographic twofold axis and a non-crystallographic threefold axis. A single isomorphous derivative, $PtCl_4$, with anomalous measurements was used to produce a preliminary set of phases. Molecular averaging improved the quality of these phases, and a detailed image of the molecule was produced. A model is being built, and the experimental phases are being extended to 2.5 Å resolution.

We have also grown some crystals of a carboxy-terminal fragment of lambda repressor, but the current crystals do not diffract to high resolution. Further crystallization attempts are in progress.