Symposia

S1: Theoretical Aspects of Protein Structure

1. Protein folds and protein sequences

Cyrus Chothia, Donald Bashford and Arthur M.Lesk
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The successful protein prediction of the structure of a protein from its sequence, using the known structure of an homologous protein, is of great importance for protein engineering. It is also important for testing theories or models for the sequence-dependent aspects of protein conformation. This lecture reviews the results that we have obtained in the cases where: (i) the proteins have similar folds but very different sequences; (ii) the proteins have low to moderate sequence identities; (iii) the proteins have high sequence identities. The lecture will emphasize the aspects of our results that have general application.

2. Modelling protein structure using data bases

Michael J.E.Sternberg
Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

The applications of protein sequence and protein structure data bases to protein engineering will be described. The known protein structures have been organized in a relational data base that stores the Brookhaven coordinates and computer-derived conformational features such as secondary structure, dihedral angles and solvent accessibility. The relational data base enables rapid interrogation of features. Applications include the selection of loop conformations for predicting unknown structures based on the X-ray coordinates of an homologous molecule and also to obtain conformational features (i.e. disulphide bridges) for modelling the consequences of site-specific mutations. The protein sequence data base can be used to obtain a multiple alignment of homologous sequences. The combined information from all the sequences can be used to improve secondary structure prediction by 10% and to locate probable active site residues in protein families.

3. Structure, design and modification of loop regions in proteins

J.M.Thornton, B.L.Sibanda, M.J.Edwards and D.J.Barlow
Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

In proteins the loop regions connecting the secondary structures comprise ~30% of the structure. These 'loops' are on the surface of the protein, and are often flexible. Since insertions and deletions in homologous sequences usually occur in these regions, it is expected that engineered mutations will be most easily tolerated here without destroying the three-dimensional structure. Detailed analysis of the conformations of loop regions in the proteins of known structure has revealed that for short loops there are structural families with specific sequence patterns. Such patterns can be used to guide the choice of sequences in novel protein design or in site-directed mutagenesis experiments. Loops are widely involved in recognition between molecules, such as antibody–antigen interactions and protein–receptor recognition. Peptides, excised from loop regions, can be used to elicit an immune response against the native protein. A method to identify these primary recognition loops from the structure and sequence is described.

4. Thermodynamics of electrostatic interactions in proteins

Neil K.Rogers
Laboratory of Molecular Biophysics, The Rex Richards Building, Department of Zoology, Oxford OX1 3QU, UK

There are two fundamental approaches to the thermodynamics of electrostatic interactions in proteins. The first uses a dielectric constant to describe the polarization effects of the atoms in the system: the energy so calculated has been strongly affirmed to be a free energy for many decades. This free energy may be manipulated by established relationships to yield the entropic and enthalpic components. The alternative approach is to consider all the atoms (or at least many of them) explicitly as polarizable dipoles. This removes the need to use a dielectric constant and the energy so calculated is a conformational potential, which may be manipulated by standard statistical mechanical procedures to yield the enthalpy, entropy and free energy associated with a given change. In practice these approaches are often mixed, and compromise models are used with the result that what is calculated is often difficult to assign as a thermodynamic entity. By use of the first model it will be shown that the thermodynamics of ion-pair formation in proteins is far from straightforward and that the driving force (enthalpic or entropic) depends upon the position within the globule relative to the solvent.

5. Exercise in protein design on computers

Chris Sander
Biocomputing, EMBL, D-6900 Heidelberg, FRG

During a recent 3-week workshop held at the EMBL, 24 researchers and students joined in an attempt at *de novo* protein design using available computer tools. The six groups at the workshop produced these preliminary designs: two different TIM barrels, following the topological motif of triose phosphate isomerase; two α/β proteins, i.e. proteins with an alternating sequence of the structural elements α-helix/β-strand, one of them a redesign of flavodoxin, with the original nucleotide binding site replaced by a binding site for a fluorescent dye; two α-helical bundles, with four helices each, one of them designed to bind calcium, the other designed to bind copper ions. The design strategy used at the workshop represents a departure from classical protein-folding theory. Classically, the main objective is to calculate the correct three-dimensional structure for a given amino acid sequence. Here, in contrast, the problem was restated: given a protein structure, in the form of its basic topology of secondary structure elements, design an amino acid sequence which will give rise to that structure as the protein folds. The

results are in the form of a list of the new amino acid sequence and coordinates of the corresponding protein structure. The design workshop has identified the deficiencies in our present tools and marks the beginning of an experimental/theoretical cycle of design and test of newly created proteins.

6. Correlation of coordinated amino acid substitutions with function in tobamoviruses

D.Altschuh[1], A.M.Lesk, A.C.Bloomer and A.Klug
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
[1]*Permanent address: IBMC, 15 rue Descartes 67000 Strasbourg, France*

Tobacco mosaic virus (TMV) is the best studied example of a self-aggregating system. Sequence data are available for the coat proteins of six tobamoviruses, with homologies ranging from 33 to 82%; atomic coordinates are known for tobacco mosaic virus wild-type. The constraints on the overall size and shape of the protein subunit and on the character of those regions of the subunit surface involved in quaternary structures should be reflected in the nature and pattern of acceptable amino acid substitutions. A significant spatial relationship has been found between groups of residues with identical amino acid substitution patterns. This strongly suggests that after mutation they have not become stabilized independently of each other, and that their location is linked to a particular function, at least in viruses identical to the disulphide for these residues. The most conserved feature of TMV is the RNA binding region. Core residues are conserved in all viruses or show mutations complementary in volume. The specificity of inter-subunit contacts is achieved in different ways in the three more distantly related viruses. The strategy used here for detecting coordinated substitutions has worked well within the tobamovirus family where the protein has extensive quaternary structure. If this approach can be applied equally successfully to other families of proteins, it could contribute to the understanding of protein folding and interactions.

7. mRNA translation and protein folding *in vivo*

J.C.Swaffield, I.J.Purvis and A.J.P.Brown
Biotechnology Unit, Institute of Genetics, University of Glasgow, Church Street, Glasgow G11 5JS, UK

Many highly expressed genes from *Saccharomyces cerevisiae* show a strong bias in their choice of codons for the 20 amino acids, this codon bias correlating strongly with the relative abundance of the iso-accepting tRNAs. (The set of preferred codons varies between organisms.) When 'rarely used' codons are clustered within a gene, a pause in the rate of translation is predicted that would result in the accumulation of nascent polypeptide chains of a discrete length. As proteins are synthesized from the N terminus and initial folding reactions probably occur before translation is completed, it is possible that translational pauses influence the folding of some proteins by allowing regions of the growing polypeptide chain to fold correctly before C-terminal regions are synthesized. A major problem in relating potential translational pauses to protein folding is the lack of knowledge about the tertiary/quaternary structures of many proteins for which the gene sequence is available. However, there is a tight correlation between the presence of potential translational pauses and the inter-domain regions of the *arom* multifunctional enzyme in *S. cerevisiae*. Potential translational pauses have been observed in the genes for other multifunctional enzymes in yeast (e.g. TRP3, TRP5), but in these cases the pauses do not seem to lie in inter-domain regions.

S2: Protein Structure, Stability and Dynamics

8. Flexibility and rigidity requirements for function of proteins and protein–pigment complexes

Robert Huber
Max-Planck-Institut für Biochemie, 8033 Martinsried bei München, FRG

Proteins are well designed for their functions. They may be rigid or flexible to various degrees as required for optimal performance. Flexibility at the level of amino acid side-chains occurs universally and may be important for some functions. Large-scale flexibility where large parts of a protein rearrange or move coherently are particularly interesting and will be discussed here. We may differentiate between different categories of flexibility (Bennett and Huber, 1984). Order–disorder transitions of domains and domain motions. The domains may be flexibly linked to allow rather unrestricted motion or the motion may be constrained to certain modes by hinges. The connecting segments and the hinges show characteristic structural features. The following examples will illustrate various aspects. Small proteinase inhibitors are essentially rigid molecules to provide tight complementary binding to their cognate protease (Huber and Bode, 1978). The large plasma inhibitors, however, exhibit large conformational changes upon interaction with proteases, probably for regulatory purposes (Löbermann *et al.*, 1984). The pancreatic serine proteases exhibit a disorder–order transition of their activation domain between proenzyme and enzyme forms as a means to regulate enzymic activity (Huber and Bode, 1978). Immunoglobulins show rather unrestricted and also hinged domain motions in different parts of the molecule, probably to allow cross-linking of antigens (Huber, 1984; Huber *et al.*, 1976). Citrate synthase adopts open and closed forms by a hinged domain motion to bind substrates and release products and to perform the catalytic condensation reaction respectively (Remington *et al.*, 1982). In the multi-enzyme complex riboflavin synthase in which two consecutive enzymic reactions are catalysed by two distinct enzymes restricted motions by engaging one enzyme in a capsid formed by the other, and directed substrate delivery seems to play an important role (Ladenstein *et al.*, 1986). In contrast to the previous examples, motion would be deleterious to function in the light-harvesting complexes (Schirmer *et al.*, 1985) and the reaction centres involved in the photosynthetic light reactions (Deisenhofer *et al.*, 1984, 1985). These are huge protein complexes which serve as matrixes to hold the pigments active in light absorption and electron conduction. Motion would deactivate the excited functional states of the pigments and destroy the proper geometric arrangement.