PROTEIN STRUCTURE

Darwin among Enzymes

from our Molecular Biology Correspondent

X-RAY crystallography is a many-splendoured science, which is now being enlisted in the service of molecular evolution. The evolutionary divergence of protein molecules has been studied primarily in terms of primary structure. The biggest effort has been mounted on cytochrome *c*, which being small and a constituent of the most rudimentary as well as the most advanced forms of life, can scarcely be bettered for the purpose. The cytochrome *c*s of a great variety of species, from the kangaroo to the castor bean, have been sequenced, to the point that the irreducible core of invariant features, which are presumed to be indispensable for biological function, can now be defined with a fair degree of confidence. So far, thirty-five of the hundred odd amino-acids remain invariant, including one run of eleven, and almost as many again are subject only to conservative substitutions. In all cases there is a high lysine content, and the protein is very basic. The basic, acidic and hydrophobic residues are strikingly clustered, and glycines, hydroxylic and aromatic side chains, and the haem-linked groups are strongly conserved.

Dickerson *et al.* (*J. Biol. Chem.*, **246**, 1511 ; 1971), with the aid of 2.8 Å resolution structures of ferricytochrome *c* from two sources—horse heart and bonito—have now set out to throw light on the path of evolution in terms of the stereochemistry. The molecule gives the appearance of being folded around the haem group, which is lodged in a crevice, attached on one side to two cysteines and a histidine, on the other to methionine (thus resolving a venerable argument concerning the nature of the sixth haem ligand). One of the propionic acid side chains of the haem group is secured in the interior of the molecule by a criss-cross of hydrogen bonds. Across the crevice is the only α-helical segment of the chain, and as Dickerson *et al.* note, the various rules purporting to relate helicity to sequence fail prodigiously when applied to cytochrome *c*. Bends in the chain, in the form of 3_{10}-helices are present, and are now beginning to appear as one of the generalities of protein structure. Cytochrome *c* has the form of a coating of predominantly polar side chains over a shell of hydrophobic side chains on the inside ("oil drop"), a large number of the latter packing closely round the haem. The invariant residues are distributed as follows: besides the four linked to the haem group, the tract 70–80 constitutes one side of the haem crevice (containing the methionine ligand) ; three invariant aromatic residues reside in a channel to one side of the haem, which, it is surmised, gives access to the surface for electron transfer ; two invariant prolines are found to one side ; a channel on the opposite side contains two aromatic side chains. Lysines are clustered round the outlets of these channels, and probably represent an operational binding site. On the other side of the molecule from the haem crevice is a cluster of nine acidic groups, which are surmised to be implicated in binding functions. Further generalizations that emerge from inspection are that prolines and glycines, which tend to be invariant, exert geometric control over bends in the chain, and the hydroxylic side chains are involved in internal hydrogen bonds. The highly variable positions occur at outside corners, where the side chain can float freely in the solvent.

Two related proteins, widely separated in evolutionary terms, are mammalian myoglobin and the monomeric insect haemoglobin, erythrocruorin. The structure of the latter was determined by Huber and his colleagues, who now survey the differences and similarities between the two molecules (Huber *et al.*, *Europ. J. Biochem.*, **19**, 42 ; 1971). The structures are clearly broadly similar, though only some 20 per cent of the residues are conserved. The situation is somewhat complicated by some incompatibilities between the structure and the published sequence, some parts of which Huber *et al.* have tentatively revised. Relative to the myoglobin chain, that of erythrocruorin shows several deletions, a phenomenon which has considerable structural consequences, in particular a bodily displacement of 2 Å of one of the long helices (B helix). There are other displacements and distortions of α-helical segments, but the most striking feature of the comparison undoubtedly concerns the environment of the haem group. Only five of the eighteen residues packed round the haem are conserved. The functionally important C-terminal tyrosine, common to all mammalian haemoglobins, is replaced by erythrocruorin by a methionine. A large number of phenylalanines occupy the interior, no less than seven in the haem pocket, and a glutamic acid replaces the distal histidine, long supposed indispensable for oxygenation (though recently found to be replaced by arginine in an abnormal human haemoglobin variant).

Robertus, Alden and Kraut (*Biochem. Biophys. Res. Commun.*, **42**, 334 ; 1971) have used X-ray crystallography to settle a long-standing conjecture that subtilisin BPN′ ('Nagarse'), as produced in Japan, is the self-same enzyme as subtilisin B, or 'Novo', from Denmark. Peptide mapping reveals no differences, but cannot exclude some conservative substitutions, and the enzymatic characteristics are also indistinguishable. The nature of the parent organisms is veiled in commercial secrecy, for they are used in the manufacture of enzyme detergents. The 2.5 Å difference map of Robertus *et al.* settles the issue ; it is essentially featureless, but for small aberrations, thought to be differences in bound solvent distribution, resulting from differences in crystallization conditions. Evidently the products will wash equally white.

A New View of Cosmic Infrared Sources

COSMIC infrared sources are usually explained by shells of dust grains around ultraviolet sources, the dust converting the energy to infrared wavelengths. An alternative view is put forward in next Monday's *Nature Physical Science* in which N. C. Wickramasinghe (Institute of Theoretical Astronomy, Cambridge) suggests that the absorption of cosmic ray energy rather than ultraviolet energy by dust grains, and its re-emission in the infrared, can better account for the observational facts. What has inspired Wickramasinghe to develop this viewpoint is that the short term variability of cosmic infrared sources indicates that they have dimensions of the order of a parsec or less, yet using conventional dust models and ultraviolet sources it is hard to see how the dimensions could be less than ten to a hundred parsec.

Wickramasinghe has therefore considered in detail what happens when a cosmic ray nucleon traverses a dust grain, and points out that cosmic ray heating could be the dominant source of grain energy near objects such as supernovae shells or other exploding objects. He shows that intensity variations on time scales of days or months, as have been observed, are compatible with this model, and the amounts of dust required seem reasonable. In other words, the infrared emission can be generated in sources that are an order of magnitude smaller than would be required if ultraviolet photons were the energy source.

Wickramasinghe has also been able to bring into his account other features of the infrared emission from galaxies, notably a turnover in the spectrum which occurs at 2.2 μm which he explains in terms of the effect on the grains of sputtering. It will be interesting to see whether the conventional explanations of the infrared sources can be improved to deal with the difficulty highlighted by Wickramasinghe's article, or whether his idea will come out on top.