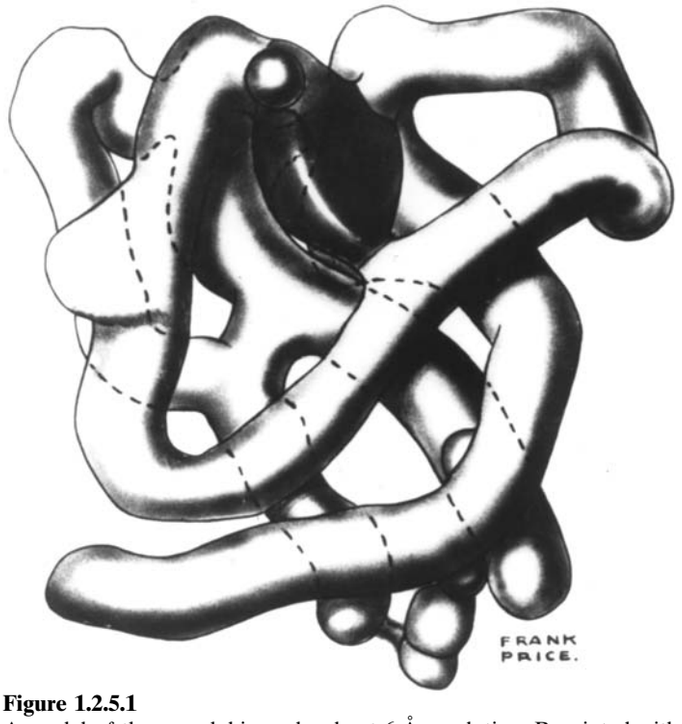

Figure 1.2.5.1
A model of the myoglobin molecule at 6 Å resolution. Reprinted with permission from Bodo *et al.* (1959). Copyright (1959) Royal Society of London.

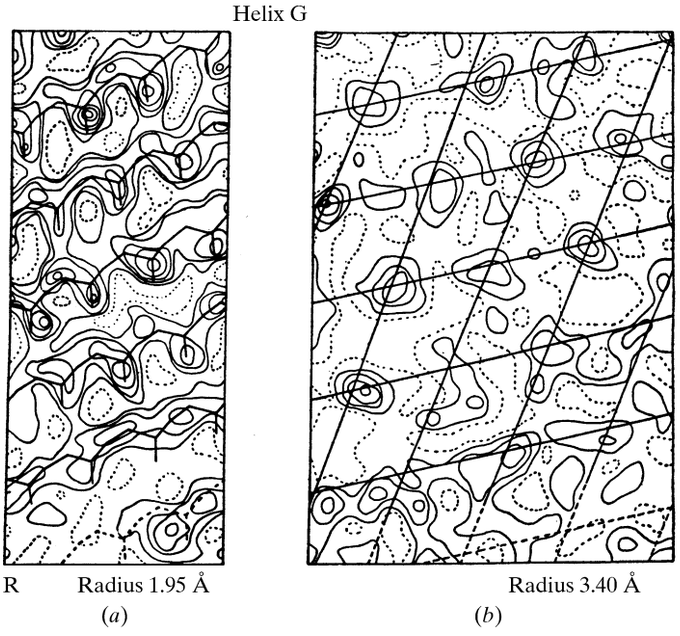

Figure 1.2.5.2
Cylindrical sections through a helical segment of a myoglobin polypeptide chain. (*a*) The density in a cylindrical mantle of 1.95 Å radius, corresponding to the mean radius of the main-chain atoms in an α-helix. The calculated atomic positions of the α-helix are superimposed and roughly correspond to the density peaks. (*b*) The density at the radius of the β-carbon atoms; the positions of the β-carbon atoms calculated for a right-handed α-helix are marked by the superimposed grid (Kendrew & Watson, unpublished). Reprinted with permission from Perutz (1962). Copyright (1962) Elsevier Publishing Co.

One problem with the isomorphous replacement method was the determination of accurate parameters that described the heavy-atom replacements. Centric projections were a means of directly determining the coordinates, but no satisfactory method was available to determine the relative positions of atoms in different derivatives when there were no centric projections. In particular, it was necessary to establish the relative y coordinates for the heavy-atom sites in the various derivatives of monoclinic myoglobin and in monoclinic horse haemoglobin. Perutz (1956) and Bragg (1958) had each proposed solutions to this problem, but these were not entirely satisfactory. Consequently, it was necessary to average the results of different methods to determine the 6 Å phases for myoglobin. However, this problem was solved satisfactorily in the structure determination of haemoglobin by using an $(F_{\rm H1} - F_{\rm H2})^2$ Patterson-like synthesis in which the vectors between atoms in the two heavy-atom compounds, H1 and H2, produce negative peaks (Rossmann, 1960). This technique also gave rise to the first least-squares refinement procedure to determine the parameters that define each heavy atom.

Perutz used punched cards to compute the first three-dimensional Patterson map of haemoglobin. This was a tremendous computational undertaking. However, the first digital electronic computers started to appear in the early to mid-1950s. The EDSAC1 and EDSAC2 machines were built in the Mathematical Laboratory of Cambridge University. EDSAC1 was used by John Kendrew for the 6 Å-resolution map of myoglobin (Bluhm *et al.*, 1958). EDSAC2 came online in 1958 and was the computer on which all the calculations were made for the 5.5 Å map of haemoglobin (Cullis *et al.*, 1962) and the 2.0 Å map of myoglobin. It was the tool on which many of the now well established crystallographic techniques were initially developed. By about 1960, the home-built, one-of-a-kind machines were starting to be replaced by commercial machines. Large mainframe IBM computers (704, 709 *etc.*), together with FORTRAN as a symbolic language, became available.

1.2.5. The first protein structures (1957 to the 1970s)

By the time three-dimensional structures of proteins were being solved, Linderström-Lang (Linderström-Lang & Schellman, 1959) had introduced the concepts of 'primary', 'secondary' and 'tertiary' structures, providing a basis for the interpretation of electron-density maps. The first three-dimensional protein structure to be solved was that of myoglobin at 6 Å resolution (Fig. 1.2.5.1) in 1957 (Kendrew *et al.*, 1958). It clearly showed sausage-like features which were assumed to be α-helices. The iron-containing haem group was identified as a somewhat larger electron-density feature. The structure determination of haemoglobin at 5.5 Å resolution in 1959 (Cullis *et al.*, 1962) showed that each of its two independent chains, α and β, had a fold similar to that of myoglobin and, thus, suggested a divergent evolutionary process for oxygen transport molecules. These first protein structures were mostly helical, features that could be recognized readily at low resolution. Had the first structures been of mostly β structure, as is the case for pepsin or chymotrypsin, history might have been different.

The absolute hand of the haemoglobin structure was determined using anomalous dispersion (Cullis *et al.*, 1962) in a manner similar to that used by Bijvoet. This was confirmed almost immediately when a 2 Å-resolution map of myoglobin was calculated in 1959 (Kendrew *et al.*, 1960). By plotting the electron density of the α-helices on cylindrical sections (Fig. 1.2.5.2), it was possible to see not only that the Pauling prediction of the α-helix structure was accurately obeyed, but also that the C_β atoms were consistent with *laevo* amino acids and that all eight helices were right-handed on account of the steric hindrance that would occur between the C_β atom and carbonyl oxygen in left-handed helices.

The first enzyme structure to be solved was that of lysozyme in 1965 (Blake *et al.*, 1965), following a gap of six years after the excitement caused by the discovery of the globin structures. Diffusion of substrates into crystals of lysozyme showed how substrates bound to the enzyme (Blake, Johnson *et al.*, 1967), which in turn suggested a catalytic mechanism and identified the essential catalytic residues.

From 1965 onwards, the rate of protein-structure determinations gradually increased to about one a year: carboxypeptidase (Reeke *et al.*, 1967), chymotrypsin (Matthews *et al.*, 1967), ribonuclease (Kartha *et al.*, 1967; Wyckoff *et al.*, 1967), papain (Drenth *et al.*, 1968), insulin (Adams *et al.*, 1969), lactate dehydrogenase (Adams *et al.*, 1970) and cytochrome *c* (Dickerson *et al.*, 1971) were early examples. Every new structure was a major event. These structures laid the foundation for structural biology. From a crystallographic point of view, Drenth's structure determination of papain was particularly significant in that he was able to show an amino-acid sequencing error where 13 residues had to be inserted between Phe28 and Arg31, and he showed that a 38-residue peptide that had been assigned to position 138 to 176 needed to be transposed to a position between the extra 13 residues and Arg31.

The structures of the globins had suggested that proteins with similar functions were likely to have evolved from a common precursor and, hence, that there might be a limited number of protein-folding motifs. Comparison of the active centres of chymotrypsin and subtilisin showed that convergent evolutionary pathways could exist (Drenth *et al.*, 1972; Kraut *et al.*, 1972).

The variety of structures that were being studied increased rapidly. The first tRNA structures were determined in the 1960s (Kim *et al.*, 1973; Robertus *et al.*, 1974), the first spherical virus structure was published in 1978 (Harrison *et al.*, 1978) and the photoreaction centre membrane protein structure appeared in 1985 (Deisenhofer *et al.*, 1985). The rate of new structure determinations has continued to increase exponentially. In 1996, about one new structure was published every day. Partly in anticipation and partly to assure the availability of results, the Brookhaven Protein Data Bank (PDB) was brought to life at the 1971 Cold Spring Harbor Meeting (H. Berman & J. L. Sussman, private communication). Initially, it was difficult to persuade authors to submit their coordinates, but gradually this situation changed to one where most journals require coordinate submission to the PDB, resulting in today's access to structural results *via* the World Wide Web.

The growth of structural information permitted generalizations, such as that β-sheets have a left-handed twist when going from one strand to the next (Chothia, 1973) and that 'cross-over' β-α-β turns were almost invariably right-handed (Richardson, 1977). These observations and the growth of the PDB have opened up a new field of science. Among the many important results that have emerged from this wealth of data is a careful measurement of the main-chain dihedral angles, confirming the predictions of Ramachandran (Ramachandran & Sasisekharan, 1968), and of side-chain rotamers (Ponder & Richards, 1987). Furthermore, it is now possible to determine whether the folds of domains in a new structure relate to any previous results quite conveniently (Murzin *et al.*, 1995; Holm & Sander, 1997).

1.2.6. Technological developments (1958 to the 1980s)

In the early 1960s, there were very few who had experience in solving a protein structure. Thus, almost a decade passed after the success with the globins before there was a noticeable surge of new structure reports. In the meantime, there were persistent attempts to find alternative methods to determine protein structure.

Blow & Rossmann (1961) demonstrated the power of the single isomorphous replacement method. While previously it had been thought that it was necessary to have at least two heavy-atom compounds, if not many more, they showed that a good representation of the structure of haemoglobin could have been made by using only one good derivative. There were also early attempts at exploiting anomalous dispersion for phase determination. Rossmann (1961) showed that anomalous differences could be used to calculate a 'Bijvoet Patterson' from which the sites of the anomalous scatterers (and, hence, heavy-atom sites) could be determined. Blow & Rossmann (1961), North (1965) and Matthews (1966) used anomalous-dispersion data to help in phase determination. Hendrickson stimulated further interest by using Cu $K\alpha$ radiation and employing the anomalous effect of sulfur atoms in cysteines to solve the entire structure of the crambin molecule (Hendrickson & Teeter, 1981). With today's availability of synchrotrons, and hence the ability to tune to absorption edges, these early attempts to utilize anomalous data have been vastly extended to the powerful multiple-wavelength anomalous dispersion (MAD) method (Hendrickson, 1991). More recently, the generality of the MAD technique has been greatly expanded by using proteins in which methionine residues have been replaced by selenomethionine, thus introducing selenium atoms as anomalous scatterers.

Another advance was the introduction of the 'molecular replacement' technique (Rossmann, 1972). The inspiration for this method arose out of the observation that many larger proteins (*e.g.* haemoglobin) are oligomers of identical subunits and that many proteins can crystallize in numerous different forms. Rossmann & Blow (1962) recognized that an obvious application of the technique would be to viruses with their icosahedral symmetry. They pointed out that the symmetry of the biological oligomer can often be, and sometimes must be, 'noncrystallographic' or 'local', as opposed to being true for the whole infinite crystal lattice. Although the conservation of folds had become apparent in the study of the globins and a little later in the study of dehydrogenases (Rossmann *et al.*, 1974), in the 1960s the early development of the molecular replacement technique was aimed primarily at *ab initio* phase determination (Rossmann & Blow, 1963; Main & Rossmann, 1966; Crowther, 1969). It was only in the 1970s, when more structures became available, that it was possible to use the technique to solve homologous structures with suitable search models. Initially, there was a good deal of resistance to the use of the molecular replacement technique. Results from the rotation function were often treated with scepticism, the translation problem was thought to have no definitive answer, and there were excellent reasons to consider that phasing was impossible except for centric reflections (Rossmann, 1972). It took 25 years before the full power of all aspects of the molecular replacement technique was fully utilized and accepted (Rossmann *et al.*, 1985).

The first real success of the rotation function was in finding the rotational relationship between the two independent insulin monomers in the *P*3 unit cell (Dodson *et al.*, 1966). Crowther produced the fast rotation function, which reduced the computational times to manageable proportions (Crowther, 1972). Crowther (1969) and Main & Rossmann (1966) were able to formulate the problem of phasing in the presence of noncrystallographic symmetry in terms of a simple set of simultaneous