10 Interpretation of the electron density map

10.1 The Graphics Model Building Procedure

In the process of building an atomic model, the polypeptide sequence is fitted to the electron density using computer graphics. The experimental electron density map rarely shows continuous density from the N- to the C- terminus due to errors in the phases used in the map calculation as well as disorder or flexibility of the protein. By recognising some parts of the sequence in the electron density map from their shape, a number of sequence fragments is produced (Bränden & Jones, 1990). The sequence assignments are then extended from the fragments until the backbone (N, C, C\(_\alpha\), O) cannot be traced further. Obtaining correct chain density depends on finding appropriate sequence fragments for which the density is readily recognizable. If high resolution data are available, the chain direction can be determined by recognising the electron-dense features representing the carbonyl oxygens. Characteristic orientation of side chains in helices also indicates chain direction.

The electron density can be skeletonised by connecting high values in the electron density map (Greer, 1974). Side chains are defined as a skeletal piece between a tip and a branch point. Editing of the skeleton connectivity may lead to a correct continuous trace. The map interpretation is facilitated by the fact that proteins contain significant amounts of regular structures, such as right-handed \(\alpha\)-helices, \(\beta\)-strands and standard turns (Jones & Thirup, 1986). During the chain trace, chemical and physical information about the
proteins can be helpful. For example, the heavy-atom positions used to determine the structure should have chemically sensible protein ligands. By using a protein database of the best refined protein structures an initial model can be built from fragments of known structures which match portions of the skeletonised electron density. Certain skeleton "atoms" can be assigned as $C_{\alpha}$ positions. This provides an initial model which can be used as a framework for the location of peptide fragments by allowing the program O (Jones et al., 1991) to align the fragments along the skeleton at suitable intervals. Once the $C_{\alpha}$ coordinates of the protein have been identified, the main chain can be constructed by linking together peptide fragments taken from a library of refined protein structures and enforcing the use of stereochemically correct pieces in the polypeptide model (Jones et al., 1991). As most amino acids adopt only a few preferred side chain conformations (rotamers), the initial model can be built fully automatically from the $C_{\alpha}$ trace using main and side chain conformational databases. The fit of each residue can then be further improved either manually (Jones, 1978) or automatically by real space methods (Jones & Liljas, 1984). Errors can occur, but these can usually be corrected during refinement. For example, the chain direction of the correctly identified secondary structure element may be wrong; incorrect connections between secondary structure elements may result in wrong main chain connections; the sequence could be out of register for some residues and there could be incorrect conformations for side chains or in the main chains (e.g. peptide flip; Bränden & Jones, 1990).
10.2 Interpretation of the Neu5Ac Lyase Electron Density Map

The final averaged phases at 2.5 Å resolution were used with the observed intensities to compute electron density maps in space group $P3_121$. Inspection of the electron density map showed that helical segments were left-handed coils and that the correct space group is therefore $P3_221$. The map was skeletonised using the program O (Jones et al., 1991). The skeletons were calculated with a base level and increment of $1.3\sigma$ and $1\sigma$ respectively. The base level is the electron density level below which skeletons are not formed. The increment allows the skeleton to favour higher density areas. Main chain skeletons were defined as skeleton fragments consisting of at least ten connected vertices. The electron density was also displayed as a three-dimensional surface contoured at one standard deviation level.

The protein fold determination and the isolation of a single molecule within the tetramer from the neighbouring molecules was performed from contoured mini-maps drawn on transparent plastic sheets using a 1 Å grid and contouring at $1\sigma$ with $0.5\sigma$ increments.

The observed domain structure was eight parallel $\beta$-strands surrounded by eight $\alpha$-helices with three additional helices at one terminus. These helices and $\beta$-strands were labelled $A$ to $H$ and $a$ to $h$, respectively and the topology tentatively discerned to be that of an $\alpha/\beta$ barrel. As a guide to model building on the graphics, the $C_\alpha$ positions were traced on contoured mini-maps on plastic sheets and confirmed with skeletons using computer graphics and the program
(Jones et al., 1991). $C_\alpha$ positions were placed about 3.8 Å apart from each other, which is the distance between adjacent $\alpha$-carbons in all polypeptide chains independent of the $\phi, \psi$ angles of the residues involved. The model was then built by fitting polyalanine to the electron density using database fragments of five consecutive $\alpha$-carbon positions. These pentapeptides were constructed by searching a library containing thirty well refined protein structures representing these residues. The pentapeptide with the lowest $r.m.s.$ deviation between its $C_\alpha$ positions and the selected $C_\alpha$ positions obtained from the skeletonisation was accepted. Adjacent pentapeptides overlapped by one residue. The quality of the electron density was very high throughout the whole molecule with clear side chain density appearing for most residues (Figure 22). The only ambiguous connectivity occurred at the active site, where the main chain seemed to have two possible branches (Figure 23). This was due to an electron dense feature found in the active site which was later modelled as a sulfate ion bound to the catalytic lysine–165. The chain direction was consistent with the connectivity. Continuous electron density was observed throughout the whole molecule and a preliminary polyalanine model of 291 residues was built (out of the 297 residues present in the protein). It was assumed there was missing electron density at both termini due to their flexibility. It was shown later that electron density was missing for the first three and the last three amino acids. The program O (Jones et al., 1991) was used to build the side chains by using the side chain direction of the alanine polypeptide when extending the side chain beyond $C_\beta$. The hydrophobic tryptophan (Trp–215) was identified to align the amino acid sequence to the observed electron density.
Figure 22: Electron density calculated from the final averaged phases at 2.5 Å resolution looking into the active site of the barrel. The main chain electron density is continuous with clear side chain electron density. The yellow line represents the skeleton “atoms” identified as $C_\alpha$ trace fitted into the density. This Figure was produced using TURBO-FRODO (Roussel & Cambillau, 1989).
Figure 23: Electron density calculated from the final averaged phases at 2.5 Å resolution. The electron dense feature branching near the Cα of Thr–48 caused the only ambiguity in the main chain connectivity during tracing of the initial main chain and was modelled later as a sulfate ion. The yellow line represents the skeleton "atoms" identified as Cα trace fitted into the density. This Figure was produced using TURBO-FRODO (Roussel & Cambillau, 1989).
Figure 24: Electron density calculated from the final averaged phases at 2.5 Å resolution. The Trp–215 density was identified to align the amino acid sequence to the observed electron density initiating the side chain model building at this residue. The yellow line represents the skeleton “atoms” identified as $C_\alpha$ trace fitted into the density. This Figure was produced using TURBO-FRODO (Roussel & Cambillau, 1989).
(Figure 24). The large size, characteristic shape and very ordered nature of this side chain and its position in the amino acid sequence (being followed by Arg–216 and Tyr–217 and preceded by Met–208 and Tyr–210) allowed easy recognition of the corresponding electron density for this side chain. The side chain model building was initiated at this residue and consistency was found between the density and the amino acid sequence extending in both directions from Trp–215 using the program O. Electron density beyond C$_\beta$ was missing for sixteen long side chains located at the surface of each protomer (Arg–6, Lys–24, Glu–36, Lys–71, Lys–73, Lys–75, Lys–96, Glu–116, Lys–146, Asp–150, Asp–183, Lys–222, Lys–225, Lys–234, Glu–279 and Lys–280) presumably due to dynamic disorder of these side chains. Electron density was available for the rest of the side chains and 291 backbone residues (typically one – three standard deviations above the noise). The atomic model for the other three protomers was generated by applying the non-crystallographic operators to the traced model of the protomer.