Fixing register errors and filling-in missing loops

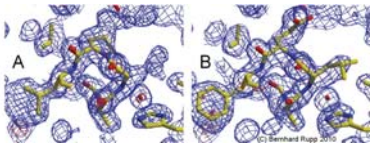


Figure 12-26 A three-residue register shift error. A shift of the sequence assignment with correct backbone conformation is a register error. In the depicted bias-minimized 1.8 Å electron density the incorrect residues Gly-Pro-Leu (A) are easily replaced with the correct residues Leu-Glu-Phe (B). Register shift is easy to detect in high resolution electron density, but not so in low resolution structures, where register shifts present one of the more insidious problems.

In register errors the backbone fits, but the residues are shifted out of sequence. Loops often need to be manually corrected or completed.

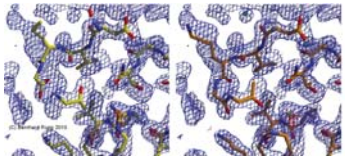


Figure 12-35 Automated loop rebuilding by *ARP/wARP*. Our starting model had a gap in a loop region (left panel). *ARP/wARP* rebuilt the loop and joined the ends properly. This worked well because we had good 1.7 Å data. The view is clipped and the loop actually continues out of the paper plane after the second glycine. Note the incorrect lysine torsion typically for the minor errors that need to be manually adjusted during polishing of the final model.

Building the solvent structure – discrete water molecules and networks

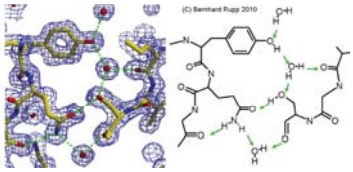
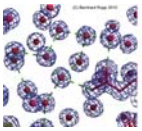


Figure 12-27 An extended hydrogen bond network containing a number of typical hydrogen bonds. The green arrows in the schematic point toward the hydrogen accepting electron lone pairs. Note that (i) each bond has a proper donor–acceptor pair (which clearly defines the correct orientation of the Gln residue), and (ii) the different types of interactions: direct backbone-side chain, side chain–side chain, and water-mediated interactions between residues. Typical X–O–H angles are 120° for the sp2 hybridized orbitals of the -OH groups, and 104.5° for H–O–H in the nearly tetrahedrally coordinated water oxygen atom. The covalent O–H distance is 0.96 Å. The map is an omit map calculated without water molecules in the model.



The water network needs to be chemically plausible. Water molecules without plausible connection to protein or other solvent atoms are probably noise and should not be placed.

How many water molecules can we expect?

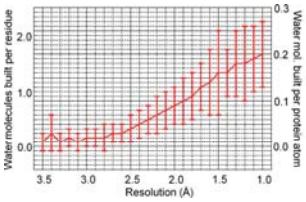
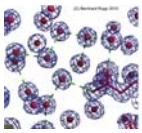


Figure 12-29 Number of discrete water solvent molecules built in protein structures. The graph shows the plausible trend that more discrete waters can be built in high-resolution structures than in low-resolution structures. Note that the linear part of the graph extrapolates to zero at about 2.9 Å, making one wonder what the water molecules in low resolution structures really might be. The graph has been redrawn with kind permission from a figure by Clemens Vonrhein, Global Phasing, Ltd.



The water network needs to be chemically plausible. Water molecules without plausible connection to protein or other solvent atoms are probably noise and should not be placed.

Building the solvent structure – discrete solvent molecules

Many solvent components from the crystallization cocktails can be discrete

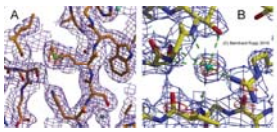


Figure 12-30 Solvent molecules in a protein structure. A metal ion (Mg^{2+}, left panel) and organic molecule (2-methyl-2,4-pentanediol, MPD, right panel) are bound to surface residues of concanavalin A. Note the distorted octahedral environment, which clearly indicates, together with shorter bond lengths (2.0–2.2 Å) that the coordinated central atom is not just another water molecule. MPD is a commonly used crystallization additive. Note the pentagonal arrangement formed by the hydrogen bonded atoms, similar to frequently observed arrangements involving ordered water molecules at protein surfaces. PDB entries 1gkb[107] and 2nk1[108] (the MPD is modeled only as discrete water molecules in the deposited structure). Both maps are *Shake&wARP* omit maps computed without the solvent molecules.

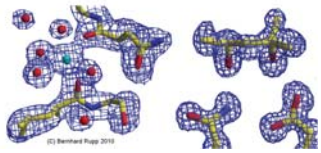


Figure 12-32 Se-Met residues and chloride ion in twelfl structure. The sequence of the *ARP/wARP* auto-built 2.5 Å twelf model is properly assigned for 97% of the residues; the remainder are missing residues in disordered loops. Distinctly visible (A) is the high density for the Se-Met residues. Four "heavy waters" in the structure (B) are chloride ions based on the coordination distances > 3 Å, the environment, and the refinement results as detailed in the text.

Building the solvent structure – discrete ligand molecules

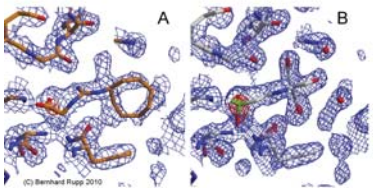
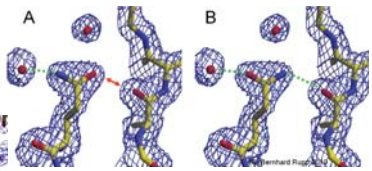


Figure 12-31 The sweetener that was bitter. The panels show 2.1 Å maps contoured at 1σ (blue) and 5σ (red). (A) A super-sweetener presumed to be bound to a taste receptor analog built into electron density map generated from *CNS* ML $2mF_o - DF_c$ coefficients. Poor density fit and a questionable vdW contact prompted critical re-examination of the model, and the fit and local chemistry proved to be perfect for TES buffer (2-(2-hydroxy-1,1-dihydroxymethyl-ethylamino)-ethanesulfonic acid), a component from the crystallization cocktail (B). The *Shake&wARP* map in (B) has slightly less noise and cleaner connectivity and clearly reveals the true nature of the ligand. PDB entry 1yni.[109]

Binding sites are by nature designed to bind something. They will grab anything in the crystallization cocktail – and high concentration will further increase occupancy. Chemical plausibility and excellent electron density are a must for ligand studies!

Polishing the model: NQH flips and correlated split side chains

Figure 12-32 Ambiguity of glutamine side chain placement. Example of an incorrectly placed Gln side chain (A). The hydrogen bonding pattern is much more plausible in the correct, flipped position (B), where each hydrogen bond has a donor–acceptor pair.



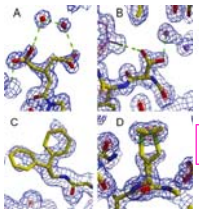


Figure 12-33 Split side chain conformations. Panel A shows the acidic residue Glu participating in a solvent hydrogen bond network, and the split serine residue (B) participates in intermolecular crystal contacts with its neighbor (purple Gln residue) in one conformer, and correlated hydrogen-bonds to a water molecule in the alternate conformation. The bottom panel row shows the hydrophobic residues Phe (C) and Met (D) located on the surface of a protein in split conformations, with only partial electron density visible. The resolution of the electron density maps is about 1.7 Å.

NQH normally cannot be positioned based on electron density alone. Programs like Molprobity return flipped residues based on chemical plausibility.