The algorithm seeks a Fourier map that is stable against repeated flipping of all density regions below δ. Obviously, a large number of missing reflections will make the algorithm less efficient, because the missing reflections are assigned a zero amplitude, which induces large termination ripples in the Fourier map. The underlying assumption of the algorithm, that the density is close to zero in large regions of the unit cell and positive in small parts of the unit cell, is no longer fulfilled and the algorithm fails. The question of incomplete data has been addressed by Palatinus *et al.* (2007). They show that the missing data can be approximated on the basis of the Patterson map of the unknown structure optimized by the maximum-entropy method. Structures that could not be solved by the original charge-flipping algorithm can be solved in this way. For small molecules, 50% or more of the reflections can be missing, and the structure can still be reconstructed by charge flipping. The situation for macromolecules is less clear.

Symmetry is an important issue. Surprisingly, in the charge-flipping method, all structures are solved in space group $P1$, and all symmetry constraints are ignored. Attempts to impose symmetry usually damage the process fatally. The disadvantage of this is that the charge density of the whole unit cell must be determined, and not just that of the asymmetric unit. Furthermore, the symmetry elements must be located once a solution has been found. A computer program, *SUPERFLIP* (Palatinus & Chapuis, 2007), and a Java applet that demonstrates the procedure in two dimensions are freely available for download at http://escher.epfl.ch/flip/.

The charge-flipping method has been adapted to proteins (Dumas & van der Lee, 2008) and applied to a $P1$ structure with 7111 atoms [*i.e.* liver alcohol dehydogenase in complex with NADH and Cd-DMSO: 5866 protein atoms, 1241 waters and 4 Cd atoms (Meijers *et al.*, 2007)]. In common with other methods described in this chapter, charge flipping is much more effective for data to very high resolution (in this case 1.0 Å) and especially for structures containing heavier atoms. The method can also, in principle, be used for substructure determination; the solution of known substructures with as many as 120 unique Se atoms is reported in the same paper.

16.1.12.7. CRUNCH2 – Karle–Hauptman determinants

The program *CRUNCH2* is quite different to the other programs mentioned in this section. With the exception of some *E*-map recycling at the end to complete a substructure, *CRUNCH2* operates entirely in reciprocal space by maximizing higher-order Karle–Hauptman determinants (Karle & Hauptman, 1950; de Gelder *et al.*, 1993). It is incorporated into the automated *CRANK* pipeline for macromolecular structure solution (Ness *et al.*, 2004). The quality of the substructure solutions obtained appears to be at least as good as those from the dual-space programs, but it may be slower for large substructures.

16.1.13. Conclusions and the grand challenge

In practice, the main use of direct methods in macromolecular crystallography is to obtain substructures using SAD and MAD data where the limitations of the method can be relaxed. There are, of course, a few structures solved *ab initio*, but they are relatively uncommon. There is a grand challenge here: to solve *ab initio* macromolecular structures using the native data alone at resolutions more typical for macromolecules without the need for specific prior structural knowledge.

The extensive (and successful) use of atomicity constraints both in real space (peak picking) and reciprocal space (tangent formula and minimal function) make it difficult to overcome the need for atomic resolution data in the *Shake-and-Bake* methods. At lower resolution, the atomicity constraint should be replaced by another based on the recurrence of model fragments that can be predicted *a priori* from the protein sequence (*e.g.* small polyalanine α-helices, β-sheets, cofactors, bases, disulfide bridges *etc.*). The effectiveness of a very small, yet accurate, fraction of the total scattering mass in the form of a fragment or heavy atoms is apparent from the results of *ACORN2* and *IL MILIONE*.

Shortly before this chapter went to press, a paper by the Usón group (Rodríguez *et al.*, 2009) showed a possible way ahead in the case of equal-atom structures, by exploiting general features of protein secondary structure. In its current form, the method requires that the protein is at least 20% α-helical and diffracts to 2.0 Å or better, requirements that would be fulfilled by at least a quarter of the protein crystal structures deposited in the PDB. The method was successfully applied to four test structures and one previously unsolved 222 amino-acid structure that diffracted to 1.95 Å and had resisted all previous attempts at solution by molecular replacement and experimental phasing. The method exploits the power of the molecular-replacement program *PHASER* (McCoy *et al.*, 2007) to search for multiple copies of (for example) 14-residue α-helices with data truncated to 2.5 Å, retaining several thousand 'best' solutions at each stage as judged by maximum-likelihood criteria. These potential multi-helix solutions are all input into a new version of the program *SHELXE* (Sheldrick, 2010) that applies density modification and main-chain tracing iteratively. At some point, depending on the size of the structure and the quality of the data, but typically for a trial structure consisting of three or four α-helices making up some 12% of the structure, the autotracing locks in and gives a relatively complete backbone trace that can be immediately recognized both by the number of connected residues traced and a correlation coefficient between the calculated and observed E values. A multiple CPU computer grid is essential for performing these numerically intensive calculations in parallel, and the whole branching and pruning operation is performed under the control of the program *ARCIMBOLDO*. This approach is still at an early stage and should benefit from fine-tuning and the inevitable future increases in computer power, but it clearly has the potential to become a main-stream *ab initio* method for the solution of protein structures.

The development, in Buffalo, of the *Shake-and-Bake* algorithm and the *SnB* program has been supported by grants GM-46733 from NIH and ACI-9721373 from NSF, and computing time from the Center for Computational Research at SUNY Buffalo. HAH, CMW and RM would also like to thank the following individuals: Chun-Shi Chang, Ashley Deacon, George DeTitta, Adam Fass, Steve Gallo, Hanif Khalak, Andrew Palumbo, Jan Pevzner, Thomas Tang and Hongliang Xu, who have aided the development of *SnB*, and Steve Ealick, P. Lynne Howell, Patrick Loll, Jennifer Martin and Gil Privé, who have generously supplied data sets. The development, in Göttingen, of *SHELXD* has been supported by the BIOXHIT Consortium and the HCM Institutional Grant ERB CHBG CT 940731 from the European Commission. GMS and IU wish to thank Thammarat Aree, Gábor Bunkóczi, Zbigniew Dauter, Judit É. Debreczeni, Judith