Table 16.1.9.1
Recommended parameter values for the *SnB* program

Values are expressed in terms of N_u, the number of unique non-H atoms (solvent atoms are typically ignored). Full-structure recommendations are for data sets measured to 1.1 Å resolution or better. Only heavy atoms or anomalous scatterers are counted for substructures.

Parameter	Full structures	Substructures
Phases	$10N_u$	$30N_u$
Triplet invariants	$100N_u$	$300N_u$
Peaks (with S, Cl) Peaks (no 'heavy')	$0.4N_u$ $0.8N_u$	N_u
Cycles	$N_u/2$ if $N_u < 100$ or if $N_u < 400$ with S, Cl *etc.*; N_u otherwise	$2N_u$ (minimum 20)

way but using only the weak reflections, *i.e.* those not used directly for phasing.

16.1.9.2. Parameters and procedures

All of the major parameters of the *Shake-and-Bake* procedure (*i.e.*, the numbers of refinement cycles, phases, triplet invariant relationships and peaks selected) are a function of structure size and can be expressed in terms of N_u, the number of unique non-H atoms in the asymmetric unit. These parameters have been fine-tuned in a series of tests using data for both small and large molecules (Weeks, DeTitta *et al.*, 1994; Chang *et al.*, 1997; Weeks & Miller, 1999*b*). Default (recommended) parameter values used in the *SnB* program are summarized in Table 16.1.9.1. At resolutions in the 1.1–1.4 Å range, recalcitrant data sets can sometimes be made to yield solutions if (1) the phase:invariant ratio is increased from 1:10 to values ranging between 1:20 and 1:50 or (2) the number of dual-space refinement cycles is doubled or tripled. The presence of moderately heavy atoms (*e.g.* S, C, Fe) greatly increases the probability of success at resolutions less than 1.2 Å; in general, the higher the fraction of such atoms the more the resolution requirement can be relaxed, provided that these atoms have low *B* values. Thus, disulfide bridges are much more helpful than methionine sulfur atoms because they tend to have lower *B* values. Parameter recommendations for substructures are based on an analysis of the peak-wavelength anomalous-difference data for S-adenosylhomocysteine (AdoHcy) hydrolase (Turner *et al.*, 1998). Parameter shift with a maximum of two 90° steps [indicated by the shorthand notation PS(90°, 2)] is the default phase-refinement mode. However, some structures (especially large *P*1 structures) may respond better to a single larger shift [*e.g.* PS(157.5°, 1)] (Deacon *et al.*, 1998). This seems to reduce the frequency of false minima (see Section 16.1.10.1).

In general, the parameter values used in *SHELXD* are similar to those used in *SnB*. However, the *combination* of random omit maps with tangent extension has been found to be the most effective strategy within the context of *SHELXD* for *ab initio* solution of the full stucture, and so is used as the default mode (see Section 16.1.10.2 for details). For substructure solution, especially for small substructures, it is normally faster to use starting atoms from Patterson seeding. Although both random omit and Patterson seeding can increase the success rate by an order of magnitude, combining both does not produce much further improvement. For very large substructures, and especially for very high symmetry space groups where the Patterson analysis is more time consuming, the random-omit procedure can be the more effective of the two. The largest substructure solved by *SHELXD* is probably PDB code 2pnk (to be published), solved by Qingping Xu of the Joint Center for Structural Genomics (JCSG), with 197 correct and no incorrect Se sites out of 205 (the other eight were disordered). About 1.6 million trials were needed (using multiple CPUs) to obtain one correct solution when Patterson seeding was employed, but with the random-omit method many good solutions were obtained. This example also illustrates the point that it is important not to give up too soon!

The substructure solution program *HySS* in the *PHENIX* system is more-or-less a clone of *SHELXD*. For further details see Section 16.1.12.5.

16.1.9.3. Recognizing solutions

On account of the intensive nature of the computations involved, *SnB* and *SHELXD* are designed to run unattended for long periods while also providing ways for the user to check the status of jobs in progress. The progress of current *SnB* jobs can be followed by monitoring a figure-of-merit histogram for the trial structures that have been processed (Fig. 16.1.9.2). A clear bimodal distribution of figure-of-merit values is a strong indication that a solution has, in fact, been found. However, not all solutions are so obvious, and it sometimes pays to inspect the best trial even when the histogram is unimodal. The course of a typical solution as a function of *SnB* cycle is contrasted with that of a nonsolution in Fig. 16.1.9.3. Minimal-function values for a solution usually decrease abruptly over the course of just a few cycles, and a tool is provided within *SnB* that allows the user to visually inspect the trace of minimal-function values for the best trial completed so far. Fig. 16.1.9.3 shows that the abrupt decrease in minimal-function values corresponds to a simultaneous abrupt increase in the number of peaks close to true atomic positions. In

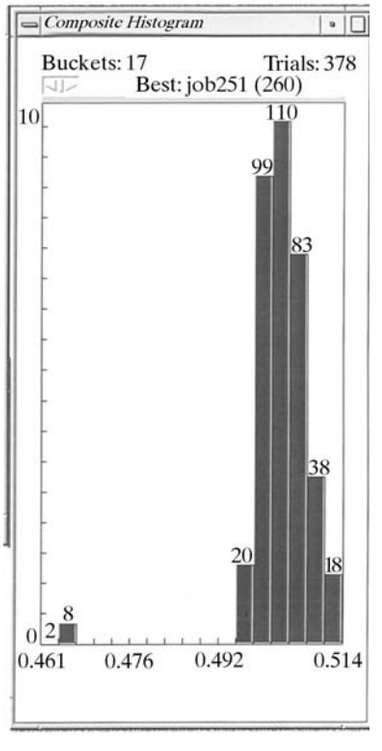

Figure 16.1.9.2
A histogram of figure-of-merit values (minimal function) for 378 scorpion toxin II trials. This bimodal histogram suggests that ten trials are solutions.