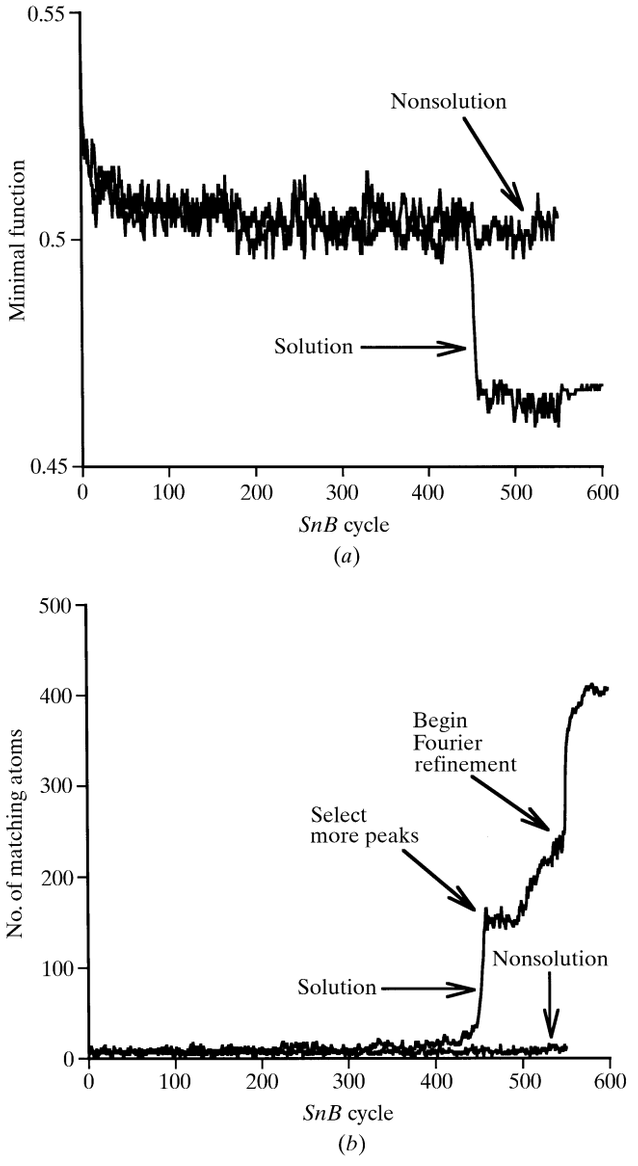

Figure 16.1.9.3
Tracing the history of a solution and a nonsolution trial for scorpion toxin II as a function *of Shake-and-Bake* cycle. (*a*) Minimal-function figure of merit, and (*b*) number of peaks closer than 0.5 Å to true atomic positions. Simple peak picking (200 or $0.4N_u$ peaks) was used for 500 (N_u) cycles, and 500 peaks (N_u) were then selected for an additional 50 ($0.1N_u$) dual-space cycles. The solution (which had the lowest minimal-function value) was then subjected to 50 cycles of Fourier refinement.

this example, a second abrupt increase in correct peaks occurs when Fourier refinement is started.

Since the correlation coefficient is a relatively absolute figure of merit (given atomic resolution, values greater than 65% almost invariably correspond to correct solutions), it is usually clear when *SHELXD* has solved a structure, although when the data do not extend to atomic resolution the CC values are less informative, and for a substructure they depend strongly on the data quality.

16.1.10. Applying dual-space programs successfully

The solution of the (known) structure of triclinic lysozyme by *SHELXD* and shortly afterwards by *SnB* (Deacon *et al.*, 1998) finally broke the 1000-atom barrier for direct methods (there happen to be 1001 protein atoms in this structure!). Both programs have also solved a large number of previously unsolved structures that had defeated conventional direct methods; some examples are listed in Table 16.1.10.1. The overall quality of solutions is generally very good, especially if appropriate action is taken during the Fourier-refinement stage. Most of the time, the *Shake-and-Bake* method works remarkably well, even for rather large structures. However, in problematic situations, the user needs to be aware of options that can increase the chance of success.

16.1.10.1. Avoiding false minima

The frequent imposition of real-space constraints appears to keep dual-space methods from producing most of the false minima that plague practitioners of conventional direct methods. Translated molecules have not been observed (so far), and traditionally problematic structures with polycyclic ring systems and long aliphatic chains are readily solved (McCourt *et al.*, 1996, 1997). False minima of the type that occur primarily in space groups lacking translational symmetry and are characterized by a single large 'uranium' peak do occur frequently in *P*1 and occasionally in other space groups. Triclinic hen egg-white lysozyme exhibits this phenomenon regardless of whether parameter-shift or tangent-formula phase refinement is employed. An example from another space group (*C*222) is provided by the Se substructure data for AdoHcy hydrolase (Turner *et al.*, 1998). In this case, many trials converge to false minima if the feature in the *SnB* program that eliminates peaks at special positions is not utilized.

The problem with false minima is most serious if they have a 'better' value of the figure of merit being used for diagnostic purposes than do the true solutions. Fortunately, this is not the case with the uranium 'solutions', which can be distinguished on the basis of the minimal function [equation (16.1.4.2)] or the correlation coefficient [equation (16.1.6.1)]. However, it would be inefficient to compute the latter in each dual-space cycle since it requires that essentially all reflections be used. To be an effective discriminator, the figure of merit must be computed using the phases calculated from the point-atom model, not from the phases directly after refinement. Phase refinement can and does produce sets of phases, such as the uranium phases, which do not correspond to physical reality. Hence, it should not be surprising that such phase sets might appear 'better' than the true phases and could lead to an erroneous choice for the best trial. Peak picking, followed by a structure-factor calculation in which the peaks are sensibly weighted, converts the phase set back to physically allowed values. If the value of the minimal function computed from the refined or *unconstrained* phases is denoted by $R_{\rm unc}$ and the value of the minimal function computed using the *constrained* phases resulting from the atomic model is denoted by $R_{\rm con}$, then a function defined by

$$R \text{ ratio} = (R_{\rm con} - R_{\rm unc})/(R_{\rm con} + R_{\rm unc}) \qquad (16.1.10.1)$$

can be used to distinguish false minima from other nonsolutions as well as the true solutions (Xu *et al.*, 2000). Once a trial falls into a false minimum, it never escapes. Therefore, the *R* ratio can be used, within *SnB*, as a criterion for early termination of unproductive trials. Based on data for several *P*1 structures, it appears that termination of trials with *R* ratio values exceeding 0.2 will eliminate most false minima without risking rejection of any potential solutions. In the case of triclinic lysozyme, false minima can be recognized, on average, by cycle 25. Since the default recommendation would be for 1000 cycles, a substantial saving in CPU time is realized by using the *R* ratio early-termination test.