

## 16.1. AB INITIO PHASING

**Table 16.1.10.1**Some large structures solved by the *Shake-and-Bake* method

Previously known test data sets are indicated by an asterisk (\*). When two numbers are given in the resolution column, the second indicates the lowest resolution at which truncated data have yielded a solution. The program codes are *SnB* (S) and *SHELXD* (D). The largest substructures solved by these two programs are mentioned in the text.

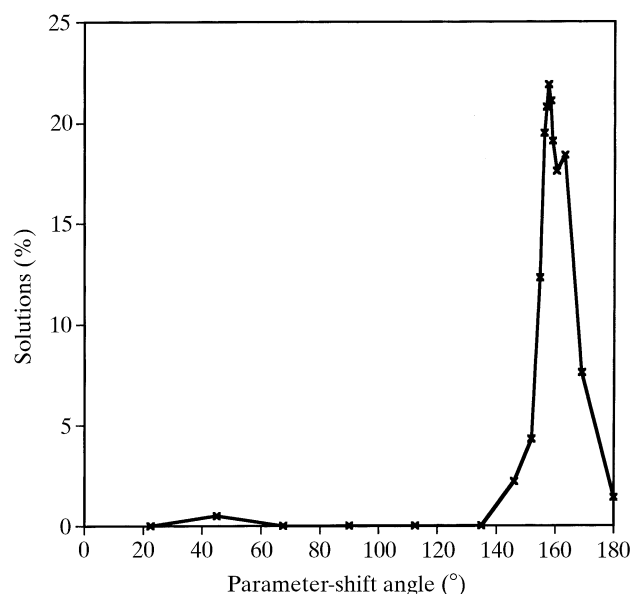
Compound	Space group	$N_u$ (molecule)	$N_u$ + solvent	$N_u$ (heavy)	Resolution (Å)	Program	Reference
Hirustasin	$P4_32_12$	402	467	10S	1.2–1.55	D	[1]
Cyclodextrin derivative	$P2_1$	448	467	—	0.88	D	[2]
Alpha-1 peptide	$P1$	408	471	Cl	0.92	S	[3]
Rubredoxin*	$P2_1$	395	497	Fe, 6S	1.0–1.1	S, D	[4]
Vancomycin	$P1$	404	547	12Cl	0.97	S	[5]
BPTI*	$P2_12_12_1$	453	561	7S	1.08	D	[6]
Cyclodextrin derivative	$P2_1$	504	562	28S	1.00	D	[7]
Balhimycin*	$P2_1$	408	598	8Cl	0.96	D	[8]
Mg-complex*	$P1$	576	608	8Mg	0.87	D	[9]
Scorpion toxin II*	$P2_12_12_1$	508	624	8S	0.96–1.2	S	[10]
Bucandin	$C2$	516	634	10S	1.05	D	[11]
Decaplanin	$P2_1$	448	635	4Cl	1.00	D	[12]
Amylose-CA26	$P1$	624	771	—	1.10	D	[13]
Viscotoxin B2	$P2_12_12_1$	722	818	12S	1.05	D	[14]
Mersacidin	$P3_2$	750	826	24S	1.04	D	[15]
Cv HiPIP H42Q*	$P2_12_12_1$	631	837	4Fe	0.93	D	[16]
Feglymycin	$P6_3$	828	1026	—	1.10	D	[17]
Acutohaemolysin	$C2_1$	1010	1242	17S	0.8	S	[18]
Tsuchimycin	$P1$	1069	1293	24Ca	1.00	D	[19]
HEW lysozyme*	$P1$	1001	1295	10S	0.85	S, D	[20], [21]
rc-WT Cv HiPIP	$P2_12_12_1$	1264	1599	8Fe	1.20	D	[16]
Cytochrome c3	$P3_1$	2024	2208	8Fe	1.20	D	[22]

References: [1] Usón *et al.* (1999); [2] Aree *et al.* (1999); [3] Privé *et al.* (1999); [4] Dauter *et al.* (1992); [5] Loll *et al.* (1998); [6] Schneider (1998); [7] Reibenspiess *et al.* (2000); [8] Schäfer *et al.* (1998); [9] Teichert (1998); [10] Smith *et al.* (1997); [11] Kuhn *et al.* (2000); [12] Lehmann *et al.* (2003); [13] Gessler *et al.* (1999); [14] Pal *et al.* (2008); [15] Schneider *et al.* (2000); [16] Parisini *et al.* (1999); [17] Bunkóczi *et al.* (2005); [18] Liu *et al.* (2003); [19] Bunkóczi (2004); [20] Deacon *et al.* (1998); [21] Walsh *et al.* (1998); [22] Frazão *et al.* (1999).

It should be noted that *SHELXD* optionally deletes the highest peak if the second peak is less than a specified fraction (*e.g.* 40%) of the height of the first, in an attempt to ‘kick’ the structure out of a false minimum.

Recognizing false minima is, of course, only part of the battle. It is also necessary to find a real solution, and essentially 100% of the triclinic lysozyme trials were found to be false minima when the standard parameter-shift conditions of two 90° shifts were used. In fact, significant numbers of solutions occur only when single-shift angles in the range 140–170° are used (Fig. 16.1.10.1), and there is a surprisingly high *success rate* (percentage of trial

structures that go to solutions) over a narrow range of angles centred about 157.5°. It is also not surprising that there is a correlated decrease in the percentage of false minima in the range 140–150°. This suggests that a fruitful strategy for structures that exhibit a large percentage of false minima would be the following. Run 100 or so trials at each of several shift angles in the range 90–180°, find the smallest angle which gives nearly zero false minima, and then use this angle as a single shift for many trials. Balhimycin (Schäfer *et al.*, 1998) is an example of a large non- $P1$  structure that also requires a parameter shift of around 154° to obtain a solution using the minimal function.

**Figure 16.1.10.1**

Success rates for triclinic lysozyme are strongly influenced by the size of the parameter-shift angle. Each point represents a minimum of 256 trials.

**16.1.10.2. Choosing a refinement strategy**

Variations in the computational details of the dual-space loop can make major differences in the efficacy of *SnB* and *SHELXD*. Fig. 16.1.10.2 shows the results of different strategies tested on a 148-atom  $P1$  structure (Karle *et al.*, 1989) while developing *SHELXD*. The CPU time requirements of parameter-shift (PS) and tangent-formula expansion (TE) are similar, both being slower than no phase refinement (NR). In real space, the random-omit-map strategy (RO) was slightly faster than simple peak picking (PP) because fewer atoms were used in the structure-factor calculations. Both of these procedures were much faster than iterative peaklist optimization (PO). The original *SHELXD* algorithm (TE + PO) performs quite well in comparison with the *SnB* algorithm (PS + PP) in terms of the percentage of correct solutions, but less well when the efficiency is compared in terms of CPU time per solution. Surprising, the two strategies involving random omit maps (PS + RO and TE + RO), which had been included in the test as placebos, are much more effective than the other algorithms, especially in terms of CPU efficiency. Indeed these two runs appear to approach a 100% success rate as the number of cycles becomes large. The combination of random omit maps and Karle-type tangent