Acta Cryst. (1995). D**51**, 33–38

Crambin: a Direct Solution for a 400-Atom Structure

By Charles M. Weeks, Herbert A. Hauptman, G. David Smith and Robert H. Blessing

Medical Foundation of Buffalo, 73 High Street, Buffalo, NY 14203, USA

Martha M. Teeter

Department of Chemistry, Boston College, Chestnut Hill, MA 02167, USA

and Russ Miller

Department of Computer Science, State University of New York at Buffalo, Buffalo, NY 14260, USA

(*Received* 23 *December* 1993; *accepted* 8 *August* 1994)

Abstract

The crystal structure of crambin, a 46-residue protein containing the equivalent of approximately 400 fully occupied non-H-atom positions, was originally solved at 1.5 Å by exploiting the anomalous scattering of its six S atoms at a single wavelength far removed from the absorption edge of sulfur. The crambin structure has now been resolved without the use of any anomalous-dispersion measurements. The technique employed was an *ab initio* 'shake-and-bake' method, consisting of a phase-refinement procedure based on the minimal function alternated with Fourier refinement. This method has successfully yielded solutions for a smaller molecule (28 atoms) using 1.2 Å data, and a crambin solution was obtained at 1.1 Å.

Introduction

Crambin is a small (M_r = 4716 Da), hydrophobic protein found in the seeds of the Abyssinian cabbage and homologous to a family of membrane-active plant toxins (Teeter, Mazer & L'Italien, 1981). It crystallizes in space group $P2_1$. There are three disulfide bridges among its 46 amino-acid residues, and anomalous scattering by the six S atoms provided the means for the original solution of the structure at 1.5 Å resolution (Hendrickson & Teeter, 1981). The diffraction data have since been recollected to 0.83 Å at 130 K without a capillary tube using the method of Hope (Teeter & Hope, 1986; Hope, 1988). The final protein model, including disordered atoms but excluding H atoms, contains 377 protein atoms and 147 solvent atoms (water plus one ethanol molecule) (Teeter, Roe & Heo, 1993). Previous attempts to solve crambin *ab initio* by conventional multisolution direct methods have been unsuccessful (Sheldrick, Dauter, Wilson, Hope & Sieker, 1993).

The shake-and-bake technique provides a powerful new formulation of direct methods. It has recently been applied successfully to a variety of known structures ranging in size up to the 317 non-H atom gramicidin A dimer (Langs, 1988). It has also been used to solve two previously unknown peptide structures which contain 105 and 110 independent non-H atoms and which had proven intractable to other methods (Miller, DeTitta, Jones, Langs, Weeks & Hauptman, 1993). Shake-and-bake consists of generating trial structures, comprised of randomly positioned atoms, each of which is then refined alternately in reciprocal and direct space as illustrated in Fig. 1. The percentage of such trial structures that converge to solution is a function of, among other things, size and complexity of the structure, resolution and quality of data, and space group.

Reciprocal-space phase refinement is based on a simple parameter-shift procedure (Weeks, DeTitta, Hauptman, Thuman & Miller, 1994), which reduces the value of the minimal function,

$$R(\varphi) = \Bigg(\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} \{ \cos(\varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}}) - [I_1(A_{\mathbf{HK}})/I_0(A_{\mathbf{HK}})] \}^2 + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}| \times \{ \cos(\varphi_{\mathbf{L}} + \varphi_{\mathbf{M}} + \varphi_{\mathbf{N}} + \varphi_{-\mathbf{L}-\mathbf{M}-\mathbf{N}}) - [I_1(B_{\mathbf{LMN}})/I_0(B_{\mathbf{LMN}})] \}^2 \Bigg) \times \Bigg(\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}| \Bigg)^{-1}, \quad (1)$$

(Hauptman, 1988; Hauptman, 1991; DeTitta, Weeks, Thuman, Miller & Hauptman, 1994). The minimal function expresses a relationship among phases related by triplet and negative quartet invariants,

which have the associated parameters (or weights),

$$A_{\mathbf{HK}} = [2/(N^{1/2})]|E_{\mathbf{H}}E_{\mathbf{K}}E_{\mathbf{H+K}}|, \quad (2)$$

and

$$B_{\mathbf{LMN}} = (2/N)|E_{\mathbf{L}}E_{\mathbf{M}}E_{\mathbf{N}}E_{\mathbf{L+M+N}}|[(|E_{\mathbf{L+M}}|^2 + |E_{\mathbf{M+N}}|^2 + |E_{\mathbf{N+L}}|^2) - 2], \quad (3)$$

respectively, where the $|E|$'s are the normalized structure-factor magnitudes and N is the number of atoms, assumed identical, in the unit cell. $R(\varphi)$ is a measure of the mean-square difference between the calculated structure invariants and their expected values, as given by the ratio of Bessel functions, and is expected to have a minimum, R_T, when the phases are equal to their correct values for some choice of origin and enantiomorph. The formula for R_T,

$$R_T = \tfrac{1}{2} + 1\Big/\left(\sum_{\mathbf{H,K}} A_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}|\right) \times \left[\sum_{\mathbf{H,K}} A_{\mathbf{HK}}\left(\{[I_2(A_{\mathbf{HK}})]/[2I_0(A_{\mathbf{HK}})]\} - \{[I_1(A_{\mathbf{HK}})]/[I_0(A_{\mathbf{HK}})]\}^2\right) + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}|\left(\{[I_2(B_{\mathbf{LMN}})]/[2I_0(B_{\mathbf{LMN}})]\} - \{[I_1(B_{\mathbf{LMN}})]/[I_0(B_{\mathbf{LMN}})]\}^2\right)\right] < \tfrac{1}{2}, \quad (4)$$

does not require prior knowledge of the phases and, therefore, can be calculated *ab initio*.

Optimization in real space requires a density-modification algorithm which, when atomic or near-atomic resolution data are available, consists simply of picking an appropriate number of the highest peaks from the Fourier map (conventional Fourier refinement). If markedly unequal atoms are present, appropriate numbers of peaks (atoms) can be weighted by the proper atomic numbers during transformation back to reciprocal space. Thus, *a priori* knowledge concerning the chemical composition of the crystal is utilized, but no knowledge of constitution is required or used during peak selection. Much of the power of the method appears to derive from this objective, unconditional, and automatic alternation of phase and Fourier refinement where the predetermined number of cycles is based on structure size.

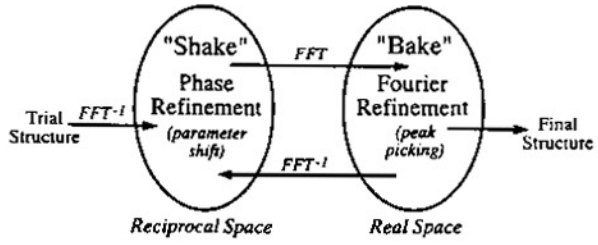


Fig. 1. The shake-and-bake algorithm.

Solutions are trial structures having a close match between peak positions and the true atomic positions for some choice of origin and enantiomorph. Solutions typically have mean phase errors of 30° or less. In space groups such as $P2_12_12_1$, where there are only a few possible discrete origin positions, shake-and-bake trials for known structures can be rapidly screened for solutions by examining the mean phase error or average absolute value of the deviations of the phases from their known values calculated using final refined coordinates and thermal parameters. In all space groups, similar judgements can be made by examining the cosine invariant figure-of-merit,

$$\mathrm{COSFOM} = \left[\sum_{\mathbf{H,K}} A_{\mathbf{HK}}|\cos(\varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}}) - \cos(\varphi^T_{\mathbf{H}} + \varphi^T_{\mathbf{K}} + \varphi^T_{-\mathbf{H}-\mathbf{K}})| + \sum_{\mathbf{L,M,N}} B_{\mathbf{LMN}} \times |\cos(\varphi_{\mathbf{L}} + \varphi_{\mathbf{M}} + \varphi_{\mathbf{N}} + \varphi_{-\mathbf{L}-\mathbf{M}-\mathbf{N}}) - \cos(\varphi^T_{\mathbf{L}} + \varphi^T_{\mathbf{M}} + \varphi^T_{\mathbf{N}} + \varphi^T_{-\mathbf{L}-\mathbf{M}-\mathbf{N}})|\right] \times \left(\sum_{\mathbf{H,K}} A_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}|\right)^{-1}, \quad (5)$$

which measures the average weighted absolute value of the difference between the values of the invariants computed using the trial (φ) and known phases (φ^T). Although the values of the individual phases depend on the choice of origin and enantiomorph, the cosine invariants are independent of these choices. Therefore, cosine invariants can be compared without first referring two phase sets to a common origin and enantiomorph. For the 0.83 Å crambin data, typical COSFOM values were 0.32 for solutions and 0.73 for non-solutions.

Results

The 1.5 Å data used by Hendrickson & Teeter (1981) in the original solution of crambin using anomalous dispersion were measured on a Picker FACS-1 diffractometer at room temperature. The data used in this investigation were remeasured on a Nicolet $P2_1$ diffractometer equipped with a graphite monochromator and a Nicolet LT-1 low-temperature attachment. A crambin crystal was first coated with oil, then mounted on a glass fiber and rapidly transferred into a cold N_2 stream at 130 K (Hope, 1988).

In all, 29443 useable diffraction intensities to a minimum interplanar spacing of 0.83 Å were recorded from Cu $K\alpha$ radiation using ω scans, and the structure was refined using this new data (Teeter, Roe & Heo, 1993).

Normalized structure-factor magnitudes ($|E|$'s) were computed using programs (Blessing, 1989) based on the method of Levy, Thiessen & Brown (1970), and the reflections were ordered by decreasing magnitude. All equivalent reflections were averaged, ignoring differences due to anomalous dispersion. The 4000 reflections with the largest $|E|$'s were then used to generate the corresponding 40000 triplets having the largest values of $A_{\mathbf{HK}}$. Use of negative quartets in shake-and-bake is optional, and although their inclusion can improve the success rate, more computing time is required. Consequently, negative quartets were omitted in this experiment.

One thousand trial structures, each consisting of two randomly positioned atoms per asymmetric unit, were then generated without geometrical restrictions and subjected to 200 cycles of the shake-and-bake procedure using the computer program *SnB* (Miller, Gallo, Khalak & Weeks, 1994). In space group $P2_1$, at least two atoms should be included in an initial trial structure, otherwise all the initial phase values of the substructure will be calculated as 0 or 180°. The crambin molecule contains 327 unique non-H atoms, and it was known from experimental density measurements that the equivalent of 91 fully occupied water molecules was present in the asymmetric unit. Therefore, since a total of approximately 400 atoms was expected, the 400 largest peaks on each Fourier map were used in subsequent structure-factor calculations with the six largest peaks being treated as S atoms and the remainder being considered as C atoms. Absolutely no geometrical constraints were imposed on the peaks.

Phase refinement was carried out using a parameter-shift procedure in which $R(\varphi)$ is reduced in value. In each shake-and-bake cycle, $R(\varphi)$ is first computed based on the set of phase values obtained from the preceding structure-factor calculation. The value of the first phase is then incremented by 90° and $R(\varphi)$ is recalculated. If $R(\varphi)$ decreases, the first phase is again incremented by 90° and $R(\varphi)$ again recalculated. On the other hand, if $R(\varphi)$ increases after the initial incrementation, the first phase is decremented once or twice by 90°. The phase value resulting in the smallest value of $R(\varphi)$ is then selected for subsequent calculations. After completing the permutation of the first phase, the parameter-shift routine considers the remaining phases, in sequence, in an analogous fashion. Note that, when varying the ith phase value, the new values determined for the previous $i-1$ phases are used immediately in the calculation of $R(\varphi)$. Centrosymmetric phases were treated as general phases without regard to the restrictions space-group symmetry places on their values. This is because previous experiments with smaller structures had shown this procedure to yield a higher success rate than restricting centrosymmetric phases only to their permissible values.

Following 200 or fewer shake-and-bake cycles, 36 of the 1000 trial structures yielded solutions, giving a success rate of *ca* 4% for the complete 0.83 Å crambin data set. Fig. 2 shows a histogram of the $R(\varphi)$ values for these trials as produced by the *SnB* program, and the bimodal nature of the distribution indicates that solutions are present. The correlation between actual solutions identified by COSFOM (values ≈ 0.32) and probable solutions indicated here by final $R(\varphi)$ values in the range 0.412–0.423 is perfect. Thus, even if the crambin structure had not been previously known, the 36 solutions would have been readily identified.

One of the successful trials was then examined in detail. A convenient measure of the quality of the map is the number of peaks that lie within 0.5 Å of correct atomic positions. Table 1 displays this information for E maps computed using 4000 phases. Data are given for the case where the structure was modeled in the structure-factor calculation as C_{400} and then for the case where the structure was modeled as S_6C_{394}. The coordinates obtained from the final shake-and-bake cycle were then used to perform one final cycle of Fourier refinement alone using a fast Fourier program written at the Medical Foundation of Buffalo by Stephen Potter. Fourier refinement applying the conditions $F_{obs} > 2\sigma_F$ and $F_{calc}/F_{obs} > 0.3$ was carried out using all measured reflections as well as just the 4000 reflections used in the shake-and-bake procedure itself. The number of the largest peaks from the last shake-and-bake cycle

R(φ) Range	Number of Trials	
0.412-0.416	20	***
0.417-0.421	15	**
0.422-0.426	1	*
0.427-0.431	0	
0.432-0.436	0	
0.437-0.441	0	
0.442-0.446	0	
0.447-0.451	2	*
0.452-0.456	5	*
0.457-0.461	13	**
0.462-0.466	42	*****
0.467-0.471	139	**********
0.472-0.476	252	**********
0.477-0.481	317	**********
0.482-0.486	155	**********
0.487-0.491	35	*****
0.492-0.496	2	*
0.497-0.501	1	*

Fig. 2. A histogram of the final $R(\varphi)$ values for 1000 crambin trials at 0.83 Å resolution as displayed by the *SnB* program. The bimodal distribution clearly separates the 36 solutions from the non-solutions.

used in the structure-factor calculation to generate phases for the final Fourier refinement was also varied (200, 250, 300, 350, 400 peaks) and both E maps and F maps were computed. F-map Fourier refinement based on phases for the full data set (27337 'based' reflections) phased by a 300-peak structure-factor calculation ($R_{cryst} = 0.38$) gave the largest number of matches, and these results are also included in Table 1. Fig. 3 shows a histogram of the 400 largest peaks on this map for which the distances to the matching refined protein atomic positions were in the ranges indicated. Most of these distances are between 0.05 and 0.2 Å, and there are very few in the range 0.5–1.0 Å. The missing atoms are scattered throughout the molecule. Most, but not all, of these atoms occur in the larger side chains and have either relatively high thermal motion or are disordered.

As noted earlier (Table 1), a comparison of the final protein atomic positions with that of the Fourier refined E map shows that 86% are correct, but the quality of the shake-and-bake maps is more apparent from their appearance on a graphics workstation. The majority of the E-map peaks are located within the boundary of the crambin molecule, while the solvent region contains relatively few large peaks. Only 26 water molecules in the crambin structure were refined as fully occupied and of these only 15 have a B_{iso} value of less than 8.0 Å^2. Both E- and F_o-map peaks are present which correspond to these 15 water molecules and although F_o density was present for some of the fully occupied waters with B's in excess of 20.0 Å^2, no E-map peaks were present for these atoms. Many of the water molecules which were refined at 0.80 occupancy and with B values of 8.0 Å^2 or less were represented by both E- and F-map peaks.

E-map peaks are typically very sharp and are surrounded by regions of very low density. Thus, it is usually necessary to contour E maps at much higher levels (3–4 standard deviations) than are usually used for F_o maps. We have found it convenient to contour the crambin E maps at 3σ, while F_o maps have been contoured between 1/2 and $2/3\sigma$. The E maps provide positions for atoms, while the F_o maps show connectivity and the positions of additional atom centers. For example, Fig. 4(*a*) illustrates the E and F_o density in the vicinity of the Tyr44 side chain. E-map peaks clearly show the positions of C^α, C^β, C^γ and one C^δ atom; however, the F_o map reveals the positions of all the atoms of the tyrosine side chain. Fig. 4(*b*), which is fairly typical of the quality of the maps in general, shows the maps in the vicinity of Pro5. Both the E and F_o maps are unambiguous in tracing the main chain and for providing the positions for the side-chain atoms. The crambin sequence is heterogeneous at position 22, which is either a proline or a serine residue. While the E map strongly suggests the presence of serine at this position, the F_o maps also provide some indication for a proline residue (Fig. 4*c*), in agreement with the refined structure. It is noteworthy that the phases from this procedure are accurate enough to provide some indication of this heterogeneity.

Table 1. *Percentages of the largest* 400 *peaks having distances* < 0.5 Å *to matching true atomic positions with occupancy* > 0.5

	Protein (%)	Solvent (%)
Trial structure C_{400}	74	29
Trial structure S_6C_{394}	83	34
Trial structure S_6C_{394} with E-map Fourier refinement	86	38
Trial structure S_6C_{394} with F-map Fourier refinement	92	51

Not all the large E or F_o peaks are correct. An incorrect peak is located in the vicinity of the C^α—C^β bond of Thr21 and could cause some confusion, were it not for the fact that all atoms of this residue are very well defined. One region of the maps which is somewhat ambiguous is the segment between Pro36 and Thr39, which possesses the highest average main-chain thermal motion. The positions of only seven of 16 possible backbone atoms are indicated by the E map, although Pro36 and Ala38 are best defined by the E map. The interpretation of the maps at Gly37 would be ambiguous and some difficulty could be encountered at Thr39. At position 40 (cysteine) the map is again quite clear, but at Pro41 there are no E-map peaks which correspond to the pyrrolidine ring. Again, the den-

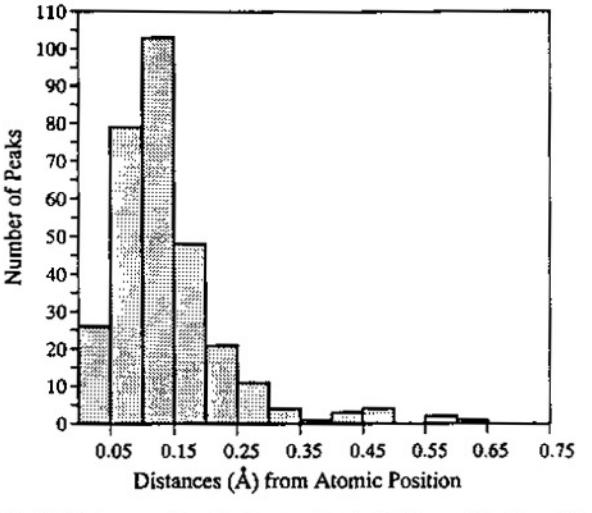
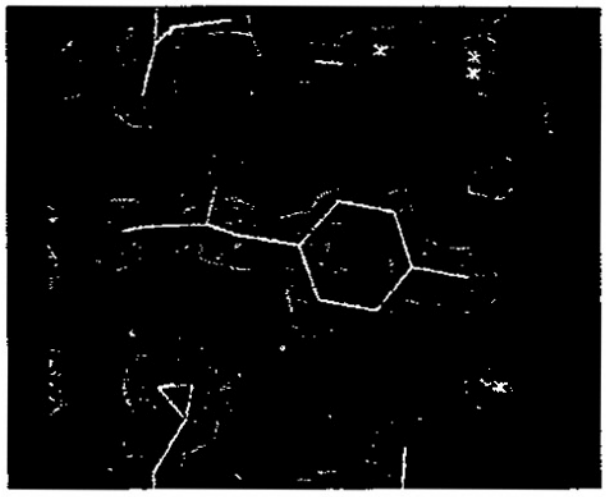
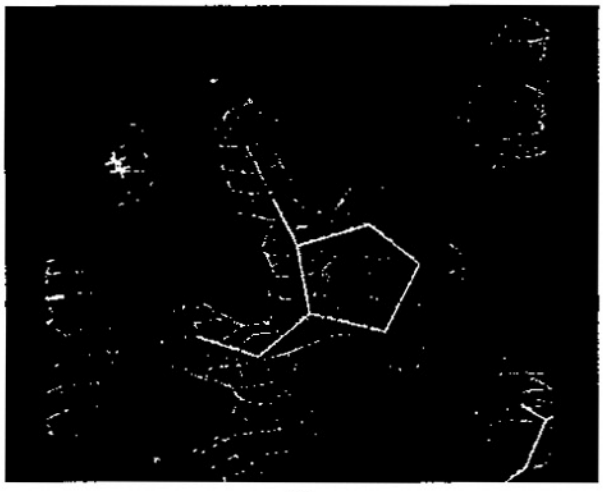


Fig. 3. Distances of peaks from refined atomic positions (protein atoms only).



(a)



(b)

(c)

Fig. 4. E (green) and F_o (blue) density for (a) Tyr44, (b) Pro5 and (c) residue 22.

sity from the F_o map provides a relatively clear indication of the positions of the five atoms.

The success rate for crambin using the shake-and-bake procedure is high relative to that for gramicidin A, a structure of comparable size. In the latter case, it was shown by prescreening trials based on the known initial mean phase error that solutions exist (Weeks, DeTitta, Miller & Hauptman, 1993), but no solutions were found among the first 2000 random trials processed. In order to determine if this discrepancy might in some way be related to the presence of the six S atoms in crambin, normalized structure-factor magnitudes were computed for the isomorphous structure which would be obtained by replacing the six cysteine residues in crambin by alanine residues. The contribution from the six S atoms was vectorially subtracted from the scaled experimental magnitudes $|F_o|$ and, from the adjusted $|F_o|$, new values of the magnitudes $|E_o|$ were established by the same procedure used with the original experimental $|F_o|$. The data for this imaginary protein were then subjected to the same shake-and-bake process as were the original native crambin data. After 200 cycles, eight of 1000 trial structures had converged to solution, yielding a success rate of about 1%. Thus, although the presence of the six S atoms did serve to increase the success rate, this did not make the difference between success and failure.

In order to determine the minimum-resolution crambin data to which the shake-and-bake procedure, as presently formulated, can be applied successfully, the experimental data were truncated at 1.0 and 1.1 Å resolution, and the entire process beginning with the computation of the normalized structure-factor magnitudes was repeated for each subset. The results of these experiments are presented in Table 2. There is an abrupt decline in success rate between 1.0 and 1.1 Å with only one trial structure leading to a solution at 1.1 Å. The experiment was not repeated at 1.2 Å because of the anticipated low success rate and large amount of computing time that might be required to unequivocally determine if a solution could be obtained. Previous experimentation using similarly truncated data for a 28-atom steroid, 9-α-methoxycortisol, has shown that, although the success rate begins to decrease substantially at resolutions less than 1.1 Å, the method is still successful at 1.2 Å but does not produce stable solutions at 1.3 Å. The results are consistent with previous observations which indicate that classical small-molecule direct methods fail if fewer than half the reflections in the range 1.1–1.2 Å are observed (Sheldrick, Dauter, Wilson, Hope & Sieker, 1993). The clear separation at each resolution between the entries in column 4 and column 5 of Table 2 shows that the final value of $R(\varphi)$ correctly identifies solutions.

Table 2. *Solutions per* 1000 *shake-and-bake trials as a function of resolution*

Resolution (Å)	No. of solutions	R_T [equation (4)]	$R(\varphi)$ range of solutions	$R(\varphi)$ range of non-solutions	COSFOM range of solutions	COSFOM range of non-solutions
0.83	36	0.422	0.412–0.423	0.450–0.499	0.30–0.35	0.70–0.75
1.0	9	0.429	0.423–0.429	0.453–0.494	0.30–0.36	0.70–0.76
1.1	1	0.439	0.450	0.466–0.499	0.41	0.71–0.77

Discussion

The results presented above for crambin show that *ab initio* direct methods, as implemented in the shake-and-bake procedure, can solve structures containing as many as 400 non-H atoms in the asymmetric unit. Solutions are identified on the basis of the calculated value of the minimal function, which results in maps that are easily interpreted. Furthermore, these maps are improved by additional Fourier refinement. The presence of S atoms in a structure of this size increases the success rate (*i.e.* 4 *versus* 1% solutions), but such atoms are not essential. Based on the success rate observed, the expected time to achieve a solution, if crambin were an unknown, is 2–3 d on a dedicated Silicon Graphics R4000 workstation with the present version of the *SnB* program. Other successful applications of this method to relatively large known peptide structures include a 174-atom valinomycin analog (1.1 Å resolution data) and a 176-atom valinomycin–dioxane complex (full Cu $K\alpha$ data) (Langs, Blessing & Duax, 1992).

So far, it has not proven possible to solve structures using the shake-and-bake procedure if the resolution of the data is less than *ca* 1.2 Å. Crambin can be solved at 1.1 Å resolution using truncated data (one of 1000 trials), and a solution might be obtained at 1.2 Å if several thousand trials were examined. The replacement of the peak picking procedure with a density-modification scheme more appropriate for lower resolution data as well as alternate phase-refinement techniques are under investigation.

This research was supported by NIH grant GM-46733 and by NSF grants IRI-9108288 and MCB-9219857. The authors also thank Häkon Hope for measuring the crambin data used in these experiments, Hanif Khalak and Steve Gallo for assisting with many of the calculations, and Thinking Machines Corporation for making time available on their CM-5 computers.

References

BLESSING, R. H. (1989). *J. Appl. Cryst.* **22**, 396–397.

DETITTA, G. T., HAUPTMAN, H. A., MILLER, R., PAGELS, M., SABIN, T., THUMAN, P. & WEEKS, C. M. (1991). *Proceedings of the Sixth Distributed Memory Computing Conference*, pp. 587–594. New York: IEEE Computer Society Press.

DETITTA, G. T., WEEKS, C. M., THUMAN, P., MILLER, R. & HAUPTMAN, H. A. (1994). *Acta Cryst.* A**50**, 203–210.

HAUPTMAN, H. A. (1988). Am. Crystallogr. Assoc. Meet. Abstract R4. Philadelphia, USA.

HAUPTMAN, H. A. (1991). In *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 324–332. Oxford: IUCr/Oxford Univ. Press.

HAUPTMAN, H. A., WEEKS, C. M., SMITH, G. D., TEETER, M. M. & MILLER, R. (1993). Am. Crystallogr. Assoc. Meet. Abstract No. PI18. Albuquerque, USA.

HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107–113.

HOPE, H. (1988). *Acta Cryst.* B**44**, 22–26.

LANGS, D. A. (1988). *Science*, **241**, 188–191.

LANGS, D. A., BLESSING, R. H. & DUAX, W. L. (1992). *Int. J. Pept. Protein Res.* **39**, 291–299.

LEVY, H. A., THIESSEN, W. E. & BROWN, G. M. (1970). Am. Crystallogr. Assoc. Meet. Abstract No. B6. New Orleans, USA.

MILLER, R., DETITTA, G. T., JONES, R., LANGS, D. A., WEEKS, C. M. & HAUPTMAN, H. A. (1993). *Science*, **259**, 1430–1433.

MILLER, R., GALLO, S. M., KHALAK, H. G. & WEEKS, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.

SHELDRICK, G. M., DAUTER, Z., WILSON, K. S., HOPE, H. & SIEKER, L. C. (1993). *Acta Cryst.* D**49**, 18–23.

TEETER, M. M. & HOPE, H. (1986). *Trans. NY Acad. Sci.* **482**, 163–165.

TEETER, M. M., MAZER, J. A. & L'ITALIEN, J. J. (1981). *Biochemistry*, **20**, 5437–5443.

TEETER, M. M., ROE, S. M. & HEO, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.

WEEKS, C. M., DETITTA, G. T., HAUPTMAN, H. A., THUMAN, P. & MILLER, R. (1994). *Acta Cryst.* A**50**, 210–220.

WEEKS, C. M., DETITTA, G. T., MILLER, R. & HAUPTMAN, H. A. (1993). *Acta Cryst.* D**49**, 179–181.