Acta Cryst. (1993). D**49**, 179–181

Applications of the Minimal Principle to Peptide Structures

BY CHARLES M. WEEKS, GEORGE T. DETITTA, RUSS MILLER* AND HERBERT A. HAUPTMAN

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

(*Received* 11 *June* 1992; *accepted* 13 *August* 1992)

Abstract

A new direct-methods procedure has been devised which consists of phase refinement *via* the minimal function, $R(\varphi)$, alternated with Fourier summation and real space filtering. All phases are initially assigned values by computing structure factors for a randomly positioned set of atoms. These phases are then refined by using a parameter shift method to minimize $R(\varphi)$. The refined phases are Fourier transformed, and a specified number of the largest peaks in the electron-density function are found and used as a new trial structure. The probability of a trial structure converging to a solution appears to depend on structural complexity and a number of refinement parameters. This procedure shows potential for providing fully automatic routine solutions for structures in the 200–400 atom range.

Introduction

The minimal function

$$R(\varphi) = \left\{ \sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} \left[\cos T_{\mathbf{HK}} - \frac{I_1(A_{\mathbf{HK}})}{I_0(A_{\mathbf{HK}})} \right]^2 + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}| \left[\cos Q_{\mathbf{LMN}} - \frac{I_1(B_{\mathbf{LMN}})}{I_0(B_{\mathbf{LMN}})} \right]^2 \right\} \Big/ \left(\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}| \right) \quad (1)$$

(Hauptman, 1988) is a powerful relationship between the phases which has been used to form the basis of a new computer-intensive fully automatic direct-methods procedure. The double sum is taken over all reciprocal lattice vectors **H**, **K** associated with triplets

$$T_{\mathbf{HK}} = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}} \quad (2)$$

which are generated by a specified basis set of phases $\{\varphi\}$, and the triple sum is taken over all reciprocal lattice vectors **L**, **M**, **N** corresponding to the negative quartets

$$Q_{\mathbf{LMN}} = \varphi_{\mathbf{L}} + \varphi_{\mathbf{M}} + \varphi_{\mathbf{N}} + \varphi_{-\mathbf{L}-\mathbf{M}-\mathbf{N}} \quad (3)$$

generated by the set $\{\varphi\}$. The parameters A and B are defined by

$$A_{\mathbf{HK}} = (2/N^{1/2})|E_{\mathbf{H}}E_{\mathbf{K}}E_{\mathbf{H}+\mathbf{K}}| \quad (4)$$

and

$$B_{\mathbf{LMN}} = (2/N)|E_{\mathbf{L}}E_{\mathbf{M}}E_{\mathbf{N}}E_{\mathbf{L}+\mathbf{M}+\mathbf{N}}|[(|E_{\mathbf{L}+\mathbf{M}}|^2 + |E_{\mathbf{M}+\mathbf{N}}|^2 + |E_{\mathbf{N}+\mathbf{L}}|^2) - 2] \quad (5)$$

where N is the number of atoms, assumed identical, in the unit cell. I_0 and I_1 are the modified Bessel functions. Then, provided that the number of phases is chosen sufficiently large, the minimal function $R(\varphi)$ has a constrained global minimum at the point at which all the phases are equal to their true values for some choice of origin and enantiomorph. Use of the minimal function with estimated values for triplets alone when anomalous-dispersion data are available is described elsewhere in this issue (Hauptman & Han, 1993).

There were two key steps in the transformation of the minimal function from a mathematical formula to a practical tool useful for crystal structure determination. The first of these steps was the recognition that the minimal function provides a useful basis for phase refinement when incorporated into an iterative procedure in which phase refinement is alternated with real space filtering techniques. The second step was the discovery that randomly positioned atomic models provide starting phases good enough for refinement to correct structures for a significant number of trials. The resulting six-part structure determination procedure is described below. The first two parts should be considered as preprocessing with the last four forming an iterative process which is repeated until a solution is achieved or a designated number of cycles have been performed (DeTitta, Hauptman, Miller, Pagels, Sabin, Thuman & Weeks, 1991). For present purposes, 'solution' does not necessarily imply a complete structure but rather a set of atomic positions which can be refined and extended by standard least-squares and Fourier techniques. Misplaced molecules are not solutions, and this method does not produce this type of false solution.

(1) Normalized structure-factor magnitudes ($|E|$) are generated by standard scaling methods from a Wilson plot, and the triplet and negative quartet

* Permanent address: Department of Computer Science, State University of New York at Buffalo, Buffalo, NY 14260, USA.

0907-4449/93/010179-03$06.00

invariants are generated which involve the largest corresponding $|E|$. The number of reflections is chosen such that the phase to atom ratio is in the traditional direct-methods range of 6–10 to 1 with fewer phases having been used for larger test structures to reduce computation time. Preliminary evidence indicates, however, that it may be beneficial to use the larger ratio. Successful applications have been made using triplets alone as well as combinations of the two types of invariants in which triplets and negative quartets have been given equal weights (*i.e.* $\sum A \simeq \sum |B|$). In all cases, the combined invariant-to-phase ratio has been at least 10 to 1.

(2) A trial structure or model comprised of randomly positioned atoms and their space-group-related symmetry mates is generated. A structure-factor calculation based on these atomic coordinates is then used to compute initial values for all the desired phases simultaneously. The starting coordinate sets are subject to the restrictions that no two atoms are closer than a specified distance (normally 1.2 Å) and that no atom is within bonding distance of more than four other atoms. Randomly positioned one-atom models have provided the initial phases for the results reported here because it has been found that the minimum initial mean phase error obtainable for any trial increases as the model size increases.

(3) The phases are refined by a parameter shift method while $R(\varphi)$, which measures the mean-square difference between estimated and calculated structure invariants, is minimized. In the examples reported here, the value of the first phase in the list was varied by five positive and five negative steps of 16°, $R(\varphi)$ was recalculated using each of these values and the value of the phase yielding the minimum $R(\varphi)$ was selected. This process was then repeated, for each of the other phases.

(4) Fourier summation is used to transform phase information into an electron-density map. Normalized structure-factor amplitudes, $|E|$, have been used at this stage (rather than F's) because phases are available for the largest E's but not for all of the largest F's. A grid size of ~ 0.33 Å has been utilized.

(5) Image enhancement has been accomplished by a discrete electron-density modification consisting of selecting a specified number of the largest peaks on the Fourier map. Simply choosing in each cycle a number of the largest peaks corresponding to the number of expected atoms has given satisfactory results. In trials which are moving towards a correct solution, the number of peaks matching true atomic positions begins to increase very slowly at first and then rapidly in the final few cycles.

(6) Inverse Fourier summation involving a structure-factor calculation in which the selected peaks are used as atoms is employed to generate new phase values.

Table 1. *Summary of results for 9α-methoxycortisol* (S28), *isoleucinomycin* (S84) *and the isoleucinomycin analog* (S127) *random trials*

Gramicidin A (S317) trials were pre-screened according to mean phase error. Computer times are for a DEC station 5000 model 200, but the complete calculations for S127 and S317 were performed on an Intel iPSC/860 hypercube. Atomic resolution data were used for all data sets.

	S28	S84	S127	S317
No. of independent non-hydrogen atoms, M	28	84	127	317
No. of phases, P (P/M)	280 (10)	600 (7.1)	900 (7.1)	2000 (6.3)
No. of triplets, T (T/P)	560 (2)	6000 (10)	9000 (10)	20000 (10)
No. of neg. quartets, Q (Q/P)	2800 (10)	45000 (75)	90000 (100)	186000 (93)
No. of trials	500	500	2000	346 of 240396
Time/cycle (s)	3.4	18.4	37.7	94.0
Cycles/trial	40	100	100	450
No. of solutions	79	16	4	3
% Success	16	3	0.2	≥0.001

Results

The minimal-function phasing procedure has been tested using the experimentally measured atomic resolution intensities for the known structures 9α-methoxycortisol (Weeks, Duax & Wolff, 1976), isoleucinomycin (Pletnev, Galitskii, Smith, Weeks & Duax, 1980), an isoleucinomycin analog (Pletnev, Ivanov, Langs, Strong & Duax, 1992), and an uncomplexed gramicidin A dimer (Langs, 1988). These structures, all in the space group $P2_12_12_1$, contain 28, 84, 127 and 317 non-hydrogen atoms respectively. The data for the smallest structure, a steroid, were used to test the programs. The other applications illustrate the results obtainable for peptides of increasing complexity including one of a size not routinely solvable by existing methods.

Summary information describing the numbers of phases and invariants used, success rates and required computing time is presented in Table 1. The results for the three smaller structures are based on random trials. Because of computer limitations, the gramicidin A experiment was not comprehensive, and the known structure was used to select trials. A 0.12 Å grid was used to generate 240 396 one-atom trial structures, and the mean phase error for each trial was computed. The 346 trials having an initial mean phase error $\leq 80°$ were run through the entire procedure and three solutions were obtained. The success rate of 0.001% reported in Table 1 for gramicidin A is based on the assumption that these are the only successful trials in the complete set, and

thus it represents the worst case scenario. An example of a contoured electron-density map for a gramicidin A solution is shown in Fig. 1.

Discussion

The results presented above demonstrate that the automated minimal-function procedure can be successfully applied to solve peptide structures containing as many as 300 non-hydrogen atoms. As illustrated in Fig. 1, the quality of the final electron-density maps is excellent. Many peaks are centered at atomic positions, and the main chains of the two gramicidin A monomers are readily recognized. In total 145 atoms in two chemically sensible fragments could be identified, and a subsequent structure-factor and Fourier calculation (using reflections having $F_{obs} > 2\sigma_F$ and $F_{calc}/F_{obs} > 0.3$) revealed the locations of an additional 117 atoms.

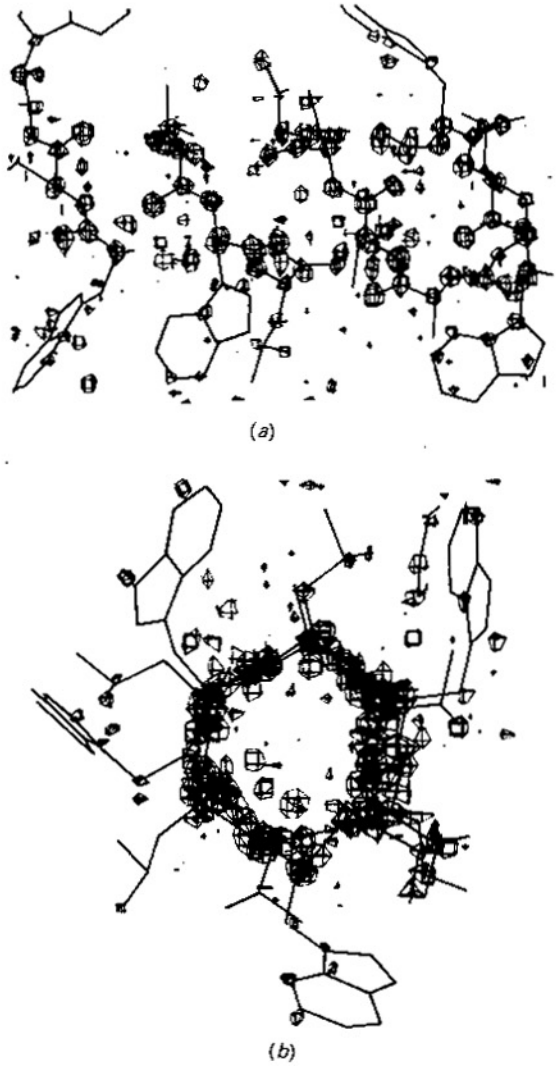

Fig. 1. (*a*) An 18 × 28 × 18 Å central section of a gramicidin A solution *E* map contoured at 6σ. (*b*) The same region as in (*a*) rotated 90° to look down the helical axis.

As the complexity of the structure increases, the percentage of trial structures converging to solutions decreases and the number of cycles necessary for successful convergence increases. The cycle time increases because larger structures require larger invariant sets and more grid points in the Fourier transform. Approximately one trial in 500 leads to a solution for the 127-atom isoleucinomycin analog. With the combined set of triplet and negative quartet invariants as defined in Table 1, 30 min of processing time was required for these trials using 32 nodes (8 megabytes/node) on an Intel iPSC/860 hypercube. The processing of sufficient gramicidin A trials to insure a solution might require months on a large currently available Thinking Machines Corporation CM-5.

The highest $R(\varphi)$ values for solutions are always less than the lowest $R(\varphi)$ values for non-solutions. However, neither $R(\varphi)$ itself nor any other quantity has yet been found which will identify, at an early stage, those trial structures which are most likely to become solutions. Consequently, present efforts are focused on finding ways to minimize the number of starting sets that need to be examined in order to insure a solution and to make each computational step as efficient as possible. Methods of fully exploiting the capabilities inherent in parallel architecture machines are also being investigated. It is hoped that these approaches will significantly reduce the computing time required and permit routine application of this method to many structures for which no routine phasing technique presently exists. General programs implementing this method on several computer systems are now being prepared.

This research was supported by NSF grants CHE-8822296 and IRT-9108288 and by NIH GM-46733. The authors also thank Dr R. Jones of the Thinking Machines Corporation and Dr R. Martino of NIH for computer time.

References

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.

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

Hauptman, H. A. & Han, F. (1993). *Acta Cryst.* D**49**, 3–8.

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

Pletnev, V. Z., Galitskii, N. M., Smith, G. D., Weeks, C. M. & Duax, W. L. (1980). *Biopolymers*, **19**, 1517–1534.

Pletnev, V. Z., Ivanov, V. T., Langs, D. A., Strong, P. & Duax, W. L. (1992). *Biopolymers*. In the press.

Weeks, C. M., Duax, W. L. & Wolff, M. E. (1976). *Acta Cryst.* B**32**, 261–263.