Acta Cryst. (1995). D**51**, 695–702

Fast *Ab Initio* Calculation of Solvent Envelopes for Protein Structures

By Gillian W. Harris

Department of Protein Engineering, Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 6BZ, England

(*Received* 9 *March* 1993; *accepted* 22 *December* 1994)

Abstract

A fast and simple method has been developed for the *ab initio* calculation of low-resolution solvent envelopes for macromolecular structures. In essence, a sphere of point scatterers is moved through the asymmetric unit cell in a part random, part systematic search for the configuration which corresponds to the lowest R value. The spheres correspond to the solvent regions in the cell. The program has been shown to work successfully for a number of test structures in a variety of space groups. No prior knowledge of the structures is needed, and c.p.u. requirements are extremely modest.

Introduction

The phase problem is the central stumbling block to rapid progress in X-ray macromolecular crystallography. For many years it was not thought possible to go directly from a diffraction pattern to a determination of the crystal structure, because the X-ray diffraction pattern of a crystal gives only the amplitudes of the Fourier transformation of the crystal contents, the phase angles are lost. This is the so-called 'phase problem'. For small molecules (less than 200 atoms), the problem has been successfully overcome by the mathematical direct methods of Karle & Hauptman (Karle & Hauptman, 1950).

The solution of a novel protein structure of no known homology to existing structures presents a formidable challenge. Molecular-replacement techniques can only be used where there is a high homology to a structure already known (Crowther & Blow, 1967). Established methods of isomorphous replacement are work and time intensive, requiring the search for and study of heavy-metal-bound isomorphous derivatives (Leslie, 1991). An *ab initio* solution to the phase problem is needed for macromolecules, for which only the native macromolecule needs to be crystallized. Attempts to extend direct methods to small macromolecules, *e.g.* polypeptides, have relied on having good quality data to at least 1.1 Å resolution (Gilmore, 1992), which is not usually the case for proteins. The problem is getting an initial set of low-resolution phases for a native protein data set.

Protein structure consists in essence of a large well defined protein envelope surrounded by generally disordered solvent. Knowledge of the molecular envelope, combined with established maximum-entropy techniques is sufficient for phase improvement and extension (Bricogne, 1984, 1993; Sato, 1992; Sjölin & Svensson, 1993). Density averaging can also be used where there is more than one copy of the protein unit in the asymmetric unit cell (Bricogne, 1976; Tsao, Chapman & Rossmann, 1992). Such methods are relatively well established. However, all these methods depend on an initial set of low-resolution phases. Recently Lunin (1993) has applied the method of electron-density histogram matching, routinely used in phase refinement and extension (Zhang, 1993), to *ab initio* phase calculation.

In the early days of protein crystallography, Kraut (1958) predicted the molecular envelope of chymotrypsinogen years before the structure was known. Kraut used as a model a sphere with a radius of 19.3 Å, corresponding to the expected solvent fraction. The sphere was used to systematically search the unit cell and structure factors were calculated and compared with the observed amplitudes (28 reflections to 11 Å). Teeter & Hendrickson (1979) applied Kraut's method to crambin, successfully predicting the solvent void using a sphere of radius 12.5 Å and 19 reflections to 11.6 Å. Later this prediction was verified when Hendrickson & Teeter (1981) solved the structure of crambin at 1.5 Å resolution by the method of anomalous scattering of sulfur.

Recently, Lunin (Lunin *et al.*, 1994) has developed a method in which the electron density at very low resolution (50 Å) is approximated by a few atoms model. Correct solutions are selected on the basis of cluster analysis and packing considerations. This method has been successfully applied to $tRNA^{Asp}$-synthetase, although the result cannot be considered a completely independent *ab inito* solution as choices were made on the basis of knowledge of the correct model.

A bold attempt at direct phase determination for protein structures was made by Subbiah (1991). Hard point scatterers are moved around randomly until they converge in a cluster which defines the molecular envelope of the protein. This method employs direct minimization of the difference between the observed and calculated diffraction pattern by continually making slight random adjustments to the initial random structural model. What is usually obtained is the negative or

ISSN 0907-4449

Babinet image, corresponding to the solvent void. The solvent void is more nearly featureless at the low resolutions used (10 or 15 Å) than the protein regions. Subbiah (1993) has further developed a procedure to establish the 'sign' of the envelope, *i.e.* whether the solvent or protein region is found. This method has been demonstrated to work for a number of test structures in a variety of space groups, with calculated structure factors (David & Subbiah, 1994).

This paper presents a simple and fast method that has been developed for the *ab initio* calculation of low-resolution envelopes for macromolecular structures. No prior knowledge of the stucture is assumed. A rough estimate of the number of residues or CA (C_α) atoms and the solvent fraction (50% can be used unless there is reason to suspect an unusually high or low solvent content) is needed to choose the initial input parameters, as will be discussed.

In essence, a point is moved randomly in the asymmetric unit cell, and for each position of this origin point, a sphere of points is set up centred on this origin, and structure factors for the sphere of point scatterers thus obtained are calculated, together with the conventional crystallographic R value. This procedure is repeated to minimize R. Various methods for ensuring that all of the asymmetric unit cell is traversed in the course of this search are discussed below.

Methodology

The initial origin point is generated randomly. The sphere is set up as a cubic grid centred on this origin

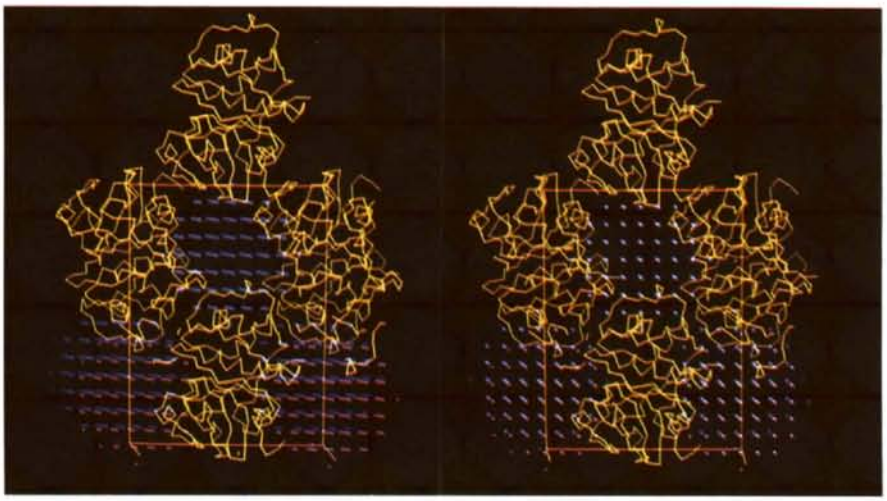

Fig. 1. Stereoview of monoclinic papain (1PPN) packing with solvent spheres from test run.

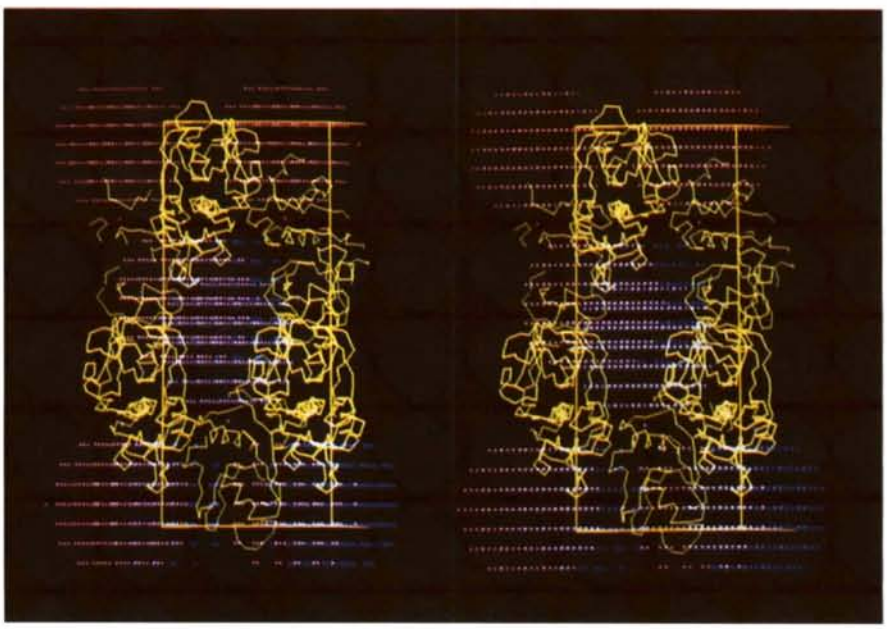

Fig. 2. Stereoview of orthorhombic papain (9PAP) packing with solvent spheres from test run.

point and the points that lie outside a sphere of radius r are discarded, giving a 'spherically rounded cubic grid'. The number of grid layers and the spacing between the layers are input to the program and have been chosen by pre-selector programs based on the number of CA atoms and the solvent fraction. In addition, the program reads as input the cell dimensions, asymmetric unit as a fraction of the unit-cell lengths, and the space-group symmetry

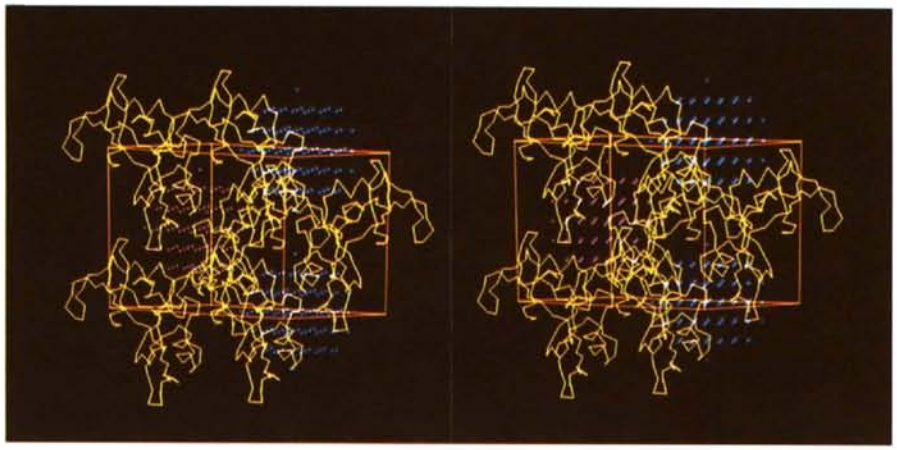

Fig. 3. Stereoview of monoclinic bovine pancreatic ribonuclease A (3RN3) packing with solvent spheres from the test run.

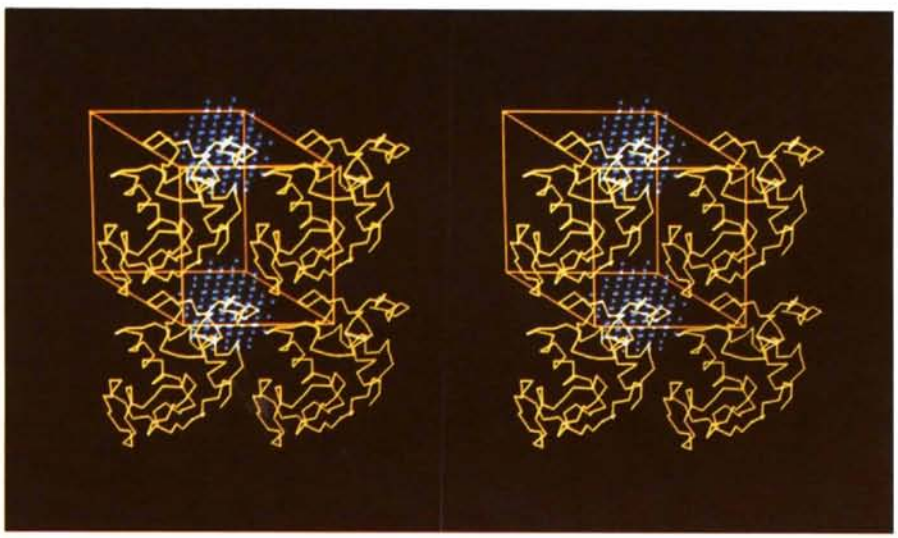

Fig. 4. Stereoview of triclinic hen egg-white lysozyme (2LZT) packing with solvent spheres from test run.

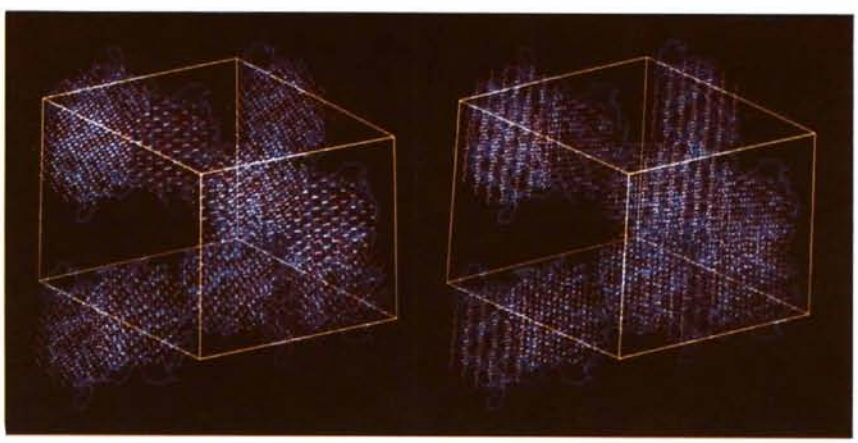

Fig. 5. Stereoview of trigonal papaya protease ω (1PPD) showing the packing of the spheres from a test run superimposed on the molecules.

operators, as given in *International Tables for Crystallography*, Vol. A (1983).

According to Babinet's principle, the scattering due to the solvent has the same amplitude, but opposite phase, as the scattering that would be produced by the solvent if it filled the molecular volume of the protein (Langridge *et al.*, 1960). The number of points in the sphere is taken as roughly the same as n.c.a., the number of CA atoms, and this fixes the number of layers in the grid which is constructed outward from the origin point. The density, *i.e.* spacing between grid layers, depends on the radius of the sphere. The radius r of the sphere can be estimated from the volume of the sphere $V_s = \text{sf} \times V_r/(\text{nmols} \times Z \times 100)$, where sf is the solvent fraction by volume in the protein crystal, V_r is the real volume of the unit cell (in Å^3), Z is the number of asymmetric units or symmetry equivalents in the unit cell, and nmols is the number of protein molecules or units per asymmetric unit cell. An estimate of the solvent fraction (sf) can be obtained from Matthews (1968). The spacing between grid layers is given by the radius r/the number of layers in the grid.

This gives initial values for input parameters. A range of values for the radius of the sphere r, and the number of layers in the grid, may be tested; the correct set of parameters can be distinguished by two criteria: the ability to produce solutions with the lowest R values, and to consistently produce the same solution, *i.e.* converge. The number of points in the sphere, *i.e.* the density of point scatterers, is less critical than the appropriate choice of radius r, as the volume of the sphere should reflect the volume of the solvent void.

The initial origin point is chosen randomly and this initial seed is not stored, so that it is not possible in practice to duplicate two runs, even using identical input parameters. Every run is thus unique. In this way a random element is built into the program. The program does not employ Monte Carlo or Metropolis techniques of selection (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953; Kirkpatrick, Gelatt & Vecchi, 1983) in that only favourable moves are accepted, bad moves are not accepted probabilistically. The decision to select one ensemble over another is based solely on the lowest R value.

At the start of the program, the reflection file (h,k,l,F_o) is read and those reflections in the resolution range specified are written out to a separate file which is then used for all further structure-factor calculations to speed up the process. The structure-factor calculating routine is adapted from that of the macromolecular-refinement program *RESTRAIN* (Driessen *et al.*, 1989). A crude scaling procedure is used where the scale of F_o to F_c is estimated as the average of the sum of the ratio of F_c/F_o, *viz.*,

$$G = \sum(F_c/F_o)/n,$$

where n is the number of reflections.

The R value is given by,

$$R = \sum \big||F_o| - G|F_c|\big| / \sum |F_o|,$$

where F_o and F_c are the observed and calculated structure factors, respectively, and G is the scale factor.

For each position of the origin point tried, the sphere of point scatterers is set up and the structure factors and crystallographic R value are calculated for this ensemble of point scatterers (scattering factors for C atoms are used). If the R value is lower than the last seen best or lowest R value which is stored, this origin point is accepted as representing the current 'best sphere' and the stored best R value updated accordingly. This 'best origin' point is then used as the starting point for methods improving the best R value, as discussed below.

Two methods are used to direct the search. These are aimed at speeding up and increasing the efficiency of the search through the asymmetric unit cell and avoiding wasteful searching of unfavourable regions of the cell. To achieve the latter, 'exclusion zones' are set up. The origin point of a spherical configuration which has the best (lowest) R value to date is stored. Whenever a spherical configuration is found with a higher R, the origin point that generated this configuration is added to an exclusion zone list. This list of 'bad origins' is consulted every time a potential move of the origin point is made, either randomly or by the method of improvement described below. If the new origin point falls within a sphere of radius ra (default 2.0 Å) of any of these points in the exclusion zone list, the new point is discarded and another new point is selected for trial. This saves unproductive structure-factor calculation.

The method of improving the R value works on the concept of minimizing the R value in the region of a configuration with a low R value, and also goes beyond this to rapidly transverse the cell in all six directions $(\pm x, y, z)$ starting from the origin point, searching for a configuration with an even lower R value. Taking each direction in turn, 'improving' steps of size s (default 1.0 Å) are taken from the current best origin point, progressively updating the stored best origin when configurations with lower R values are found. By this process, a three-dimensional zig-zag walk through the cell is essentially taken along a route of successively lower R value. When this iterative process is exhausted, and no further improvement can be found by taking steps in any of the six directions from the current best origin point, a new point is generated randomly and the whole test process repeated for this try.

A run terminates when the number of tries (randomly generated new points, not counting improved configurations of a try) since the latest best R value found is greater than an input limit (default 1000). The coordinates of the 'best sphere', *i.e.* configuration with the lowest R value, are written to a file at the end of each run. The best sphere is usually found to be a configuration in a relatively early try, thus demonstrating the efficacy of

the method of improving the initial configuration of a random try. However, sometimes the lowest R value configuration is found only after a large number of attempts, and the process should always be continued for a reasonably large number of attempts to ensure optimal results.

Test cases

The following structures were used as test cases: monoclinic papain (Pickersgill, Harris & Garman, 1992); orthorhombic papain (Kamphuis, Kalk, Swarte & Drenth, 1984); trigonal papaya protease ω or ppo (Pickersgill, Rizkallah, Harris & Goodenough, 1991); bovine pancreatic ribonuclease A or rnaseA (Howlin, Moss & Harris, 1989) and triclinic hen egg-white lysozyme or HEWL (Ramanadham, Sieker & Jensen, 1990). Crystal details for these structures are given in Table 1, together with Protein Data Bank (PDB) codes (Bernstein *et al.*, 1977). Low-resolution (10–100 Å) structure factors were generated from the coordinates using the *CCP*4 program *GENSFC* written by E. Dodson (Collaborative Computational Project, Number 4, 1994). All runs used resolution limits of 15–100 Å. Tables 2 and 3 show the input parameters and corresponding results for typical runs. The results were displayed (on an Evans & Sutherland PS390) using the graphics program *FRODO* (Jones, 1978). When the protein molecules (CA traces) and 'best spheres' were packed in the crystal cell, the spheres were found to fill the solvent voids between the protein molecules, forming solvent channels (see Figs. 1, 2, 3 and 4).

As a measure of the success of the results, an envelope was drawn around the molecules and each point scatterer in the 'best spheres' tested against this. The fraction of point scatterers within the molecular envelope are given in Table 4. The majority of points do not overlap with the molecules.

The program was run on a Vax 3600. C.p.u. times are given in Table 3. These vary from 3 min for triclinic lysozyme, with low symmetry and a small unit cell, to typically one to two hours for higher symmetry cases. The c.p.u. time increases with the number of reflections and the number of points in the sphere. The modest computational requirements (access to a supercomputer is not necessary) and the speed with which the program runs allow for multiple runs and graphical analysis of the results within a matter of hours. Results of runs which converge to similar solutions can be combined by taking only those points common to all solutions.

For one test case, ppo, the 'best spheres' packed to give the protein envelope, rather than the solvent void, with the spheres coinciding with the protein molecules (see Figs. 5 and 6). In this case, the molecules pack in a threefold arrangement around the large solvent channel down z. Subbiah's (1993) method could be used to resolve the Babinet ambiguity.

Table 1. *Crystal details of protein structures used as test cases*

PDB code	Protein	Space group	Z	No. of CA atoms	sf*
1PPN	Papain	$P2_1$	2	212	0.40
9PAP	Papain	$P2_12_12_1$	4	212	0.47
1PPO	Ppo	$P3_112$	6	216	0.48
3RN3	RnaseA	$P2_1$	2	124	0.39
2LZT	HEWL	$P1$	1	129	0.27
1GAL	Glucose oxidase	$P3_121$	6	583	0.37

* Solvent fraction estimated by the method of Matthews (1968).

Table 2. *Input parameters for test runs: details of sphere*

Test case	No. of points	No. of grid layers	Distance between layers (Å)	Radius of sphere (Å)
1PPN	256	4	4.4	17.6
9PAP	256	4	5.0	20.0
1PPO	256	4	4.75	19.0
3RN3	107	3	4.67	14.01
2LZT	106	3	3.33	9.99
1GAL	912	6	4.0	24.0

Table 3. *Results for typical test runs*

Test case	R value	No. of reflections*	C.p.u.†	Try (configuration)
1PPN	0.4910	36	1 h 42 min	608 (52)
9PAP	0.5808	55	1 h 33 min	210 (22)
1PPO	0.4820	227	7 h 05 min	60 (9)
3RN3	0.5222	24	7 min	59 (19)
2LZT	0.6381	16	3 min	23 (16)
1GAL	0.5518	105	30 h	10 (36)

* Resolution range 15–100 Å.

† C.p.u. (central processing unit) time in hours (h) and minutes (min) on a VAX 3600.

Table 4. *Fraction of point scatterers within molecular envelope*

Test case	Fraction
1PPN	0.019
9PAP	0.093
3RN3	0.073
2LZT	0.019
1GAL	0.039

A real test case, an unknown structure

The program was run for glucose oxidase, using data collected on a Xentronics area detector. Data in the resolution range 15–100 Å was used. At the time the structure of glucose oxidase was not known, though the structure has since been published (Hecht, Kalisz, Hendle, Schmid & Schomburg, 1993). The crystal data, input parameters and result are given in Tables 1, 2 and 3, respectively, and the graphical results are shown in Fig. 7. The packing of the 'best spheres' corresponds quite well to the solvent channels, with less than 4% of point scatterers within the molecular envelope (Table 4). This demonstrates that the program can successfully be used to predict the solvent envelope for unknown

structures and this could be exploited in structure determination, in conjunction with existing methods such as isomorphous replacement, and ultimately, possibly in *ab initio* structure determination.

Discussion

The method has proved able to calculate *ab initio* solvent envelopes for protein structures. This is the first step towards *ab initio* solution of protein structures.

As has been pointed out by Yeats & Zhang (1993), ambiguities arise in *ab initio* phasing methods that rely on the comparison of structure factors calculated from models that give rise to structure factors which differ from each other in phase but not in amplitude. There are three sources of phase ambiguity, namely, twofold Babinet ambiguity (discussed above), twofold enantiomer ambiguity, and origin ambiguity, depending on the space group. For example, in the test cases, in $P2_1$, the origin is not fixed in y; in $P2_12_12_1$, there is a choice of eight origins; in $P1$, the choice of origin is arbitrary.

For monoclinic papain, the test run was repeated 20 times, with the same input parameters, and the solutions compared using a graphics program. Of the 20 runs, 11 gave the correct solvent envelope (shown in Fig. 1); three gave the Babinet image, *i.e.* the protein molecules, and six gave a solution shifted only in y from the best solvent image, *i.e.* these correspond to alternate choices of origin.

A similar study was carried out for orthorhombic papain. Of the 20 runs, nine gave the correct solvent envelope (shown in Fig. 2), four gave the Babinet image, while the remaining seven gave one of two alternate origin choices for the correct solvent image.

Alternative solutions (the minority) were not incorrect; rather these represented alternative solutions due to origin ambiguity and hence, in an unknown case, would be equally valid, since the choice of origin is made when the phase problem is solved.

The R values for these solutions differ by about 0.01; *e.g.* for monoclinic papain, the best R value for the correct solvent void was 0.488, the R value for the Babinet image was 0.493, and the R value for y-shifted origin solutions ranged from 0.489 to 0.502. The R value criterion cannot distinguish between these alternate (but equally correct) solutions. However, the last incorrect solution found in the course of any one run typically

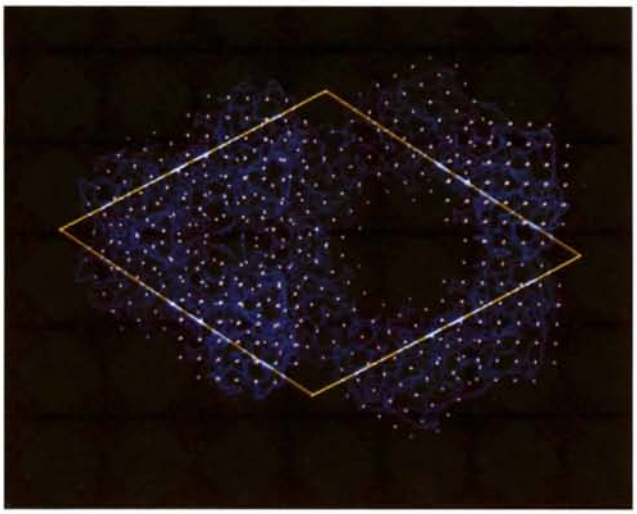

Fig. 6. Projection down z of trigonal papaya protease ω (1PPD) showing the packing of the spheres and the corresponding molecules about the threefold axis.

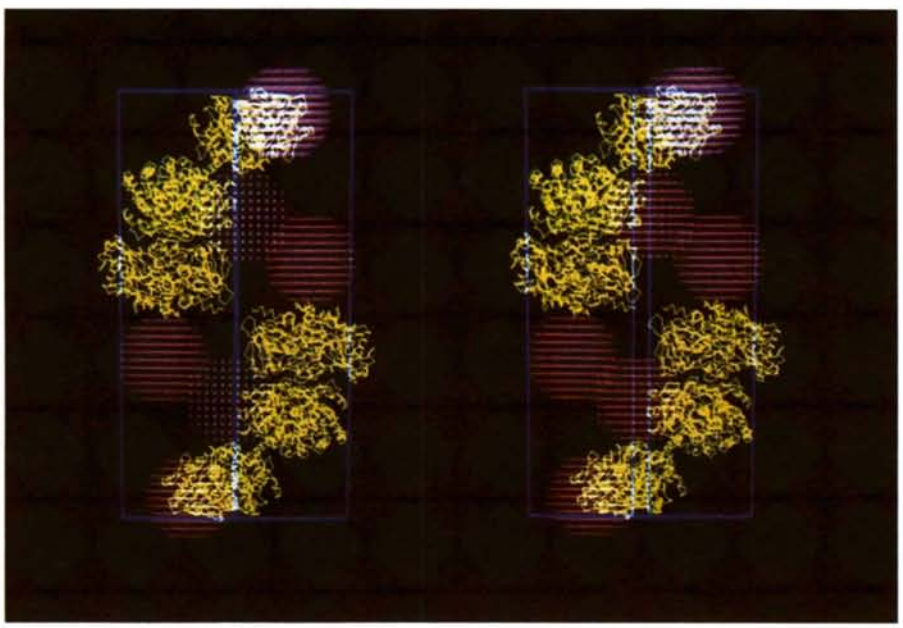

Fig. 7. Stereoview of trigonal glucose oxidase (1GAL) packing with solvent spheres from test run.

corresponded to an R value of around 0.56–0.60. Thus, the tolerance in R value is of the order of 0.1 in order to correctly place the origin of the sphere.

Improvements and extensions to the basic method are envisaged. The calculation could be speeded up by using a fast Fourier transform method (Ten Eyck, 1977), or direct simple trigonometric methods. This would allow an exhaustive search of the unit cell without the price of high computational times.

The method in its present form is limited by the spherical nature of the search model to compact molecules which approximate a sphere at low resolution. Non-compact molecules, *i.e.* rod-like or dumbbell shaped, *e.g.* tropomyosin (2TMA) and calmodulin (3CLN), would require correspondingly odd-shaped search models. Even in less extreme cases, where the protein molecules or solvent void is a poor approximation to a sphere, the method will not work well. Such cases would require the use of elliptical search models and the development of a method to optimize the shape of these for an unknown structure.

The method worked well for six of the eight cases studied. The other two cases, trigonal ($P3_121$) porcine phospholipase A2 or pla2 (1P2P; Dijkstra, Renetseder, Kalk, Hol & Drenth, 1983) and tetragonal ($P4_32_12$) hen egg-white lysozyme or HEWL (2LYM; Kundrot & Richards, 1987) provide valuable information on the limitations of the current method.

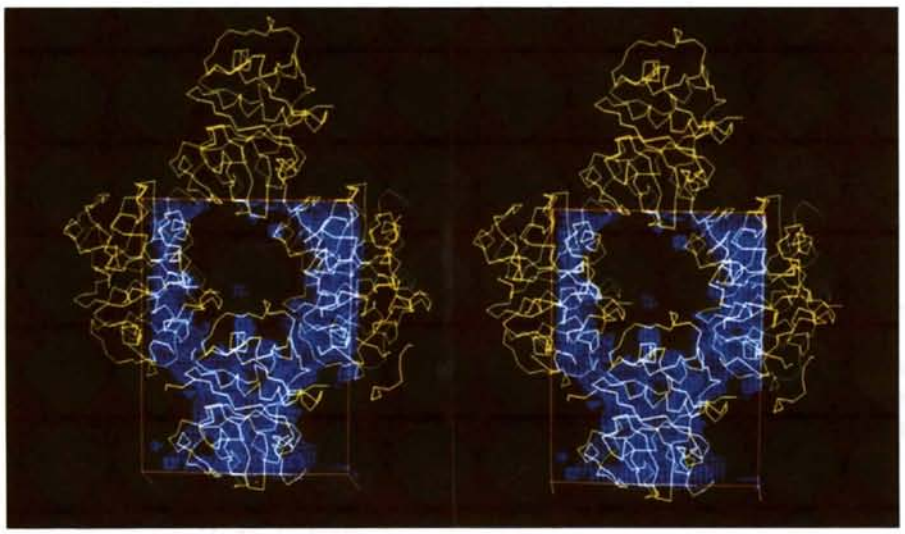

Fig. 8. Stereoview of electron-density envelope for monoclinic papain (1PPN), superimposed on the molecules in the unit cell (orientation as in Fig. 1).

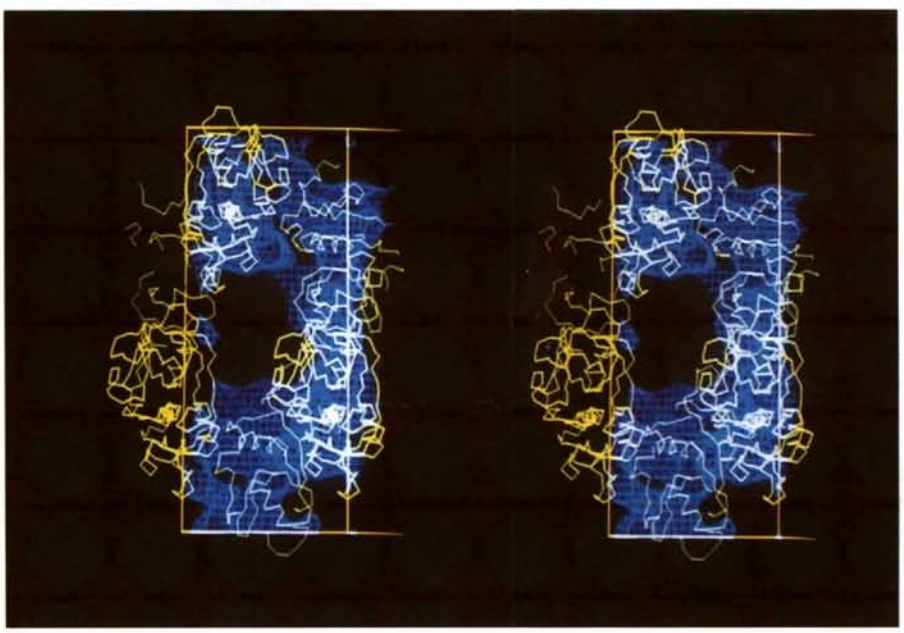

Fig. 9. Stereoview of electron-density envelope for orthorhombic papain (9PAP), superimposed on the molecules in the unit cell (orientation as in Fig. 2).

In the case of pla2, the elongated rather than spherical shape of the molecule and the high solvent content (62%) means that the spheres cannot adequately fill the large solvent channels between the molecules without considerable overlap with the molecules. The use of smaller spheres can fill the solvent channels but cannot also fill the spaces between molecules.

For tetragonal HEWL, the packing of the eight molecules in the thin unit cell ($a = b = 79.17$, $c = 37.96$ Å) means that the solvent voids are tablet-shaped rather than spherical and a good solvent envelope cannot be obtained using the present method. The asymmetric unit cell of $z/8 = 4.75$ Å means that the search model must be correspondingly thin in this direction if the method is to be sufficiently sensitive to accurately locate the solvent.

Where there is no *a priori* knowledge of the shape of the molecule or the solvent voids, warning signs that a more sophisticated search model may be needed are a high solvent fraction (greater than 0.5) or a unit cell/asymmetric unit cell that is very thin in one direction. Selecting a suitable shape for a search model in an unknown case poses a challenge.

Currently the method only deals with cases where there is one protein molecule or unit per asymmetric unit, but there is no reason in principle why the method cannot be extended to cases with two (or more) molecules in the asymmetric unit. This will be the subject of a future study.

In a further development, low-resolution electron-density envelopes have been obtained from the solvent spheres. Point atoms are packed into the region in the asymmetric unit cell not occupied by the solvent spheres and from this crude model, structure factors and phases are calculated at 10 Å. An electron-density map is calculated (*CCP*4 program *FFT* written by Lynn F. Ten Eyck) and displayed on the graphics program. Examples of these 10 Å electron-density envelopes are shown for monoclinic papain (Fig. 8) and for orthorhombic papain (Fig. 9), superimposed on the protein molecules in the unit cell. As can be seen, the electron-density envelopes are a good approximation to the expected molecular envelopes.

The phasing potential of this method is being investigated for known as well as new protein structures.

I would like to thank Sean Jones for encouragement, Dr Jim Warwicker for casting a critical eye over some of the results on the graphics and Dr Richard Pickersgill for assistance with the stereoviews. I am grateful to Dr Harry Powell for allowing me to use his glucose oxidose data to test my program. Further, I would like to thank one of the referees for drawing my attention to the early papers by Kraut, Teeter & Hendrickson, as well as for valuable comments.

References

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.
BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.
BRICOGNE, G. (1984). *Acta Cryst.* A**40**, 410–445.
BRICOGNE, G. (1993). *Acta Cryst.* D**49**, 37–60.
COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4 (1994). *Acta Cryst.* D**50**, 760–763.
CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.
DAVID, P. R. & SUBBIAH, S. (1994). *Acta Cryst.* D**50**, 132–138.
DIJKSTRA, B. W., RENETSEDER, R., KALK, K. H., HOL, W. G. J. & DRENTH, J. (1983). *J. Mol. Biol.* **168**, 163–179.
DRIESSEN, H., HANEEF, M. I. J., HARRIS, G. W., HOWLIN, B., KHAN, G. & MOSS, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.
GILMORE, C. J. (1992). *Curr. Op. Struct. Biol.* **2**, 806–810.
HECHT, H. J., KALISZ, H. M., HENDLE, J., SCHMID, R. D. & SCHOMBURG, D. (1993). *J. Mol. Biol.* **229**, 153–172.
HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **209**, 107–113.
HOWLIN, B., MOSS, D. S. & HARRIS, G. W. (1989). *Acta Cryst.* A**45**, 851–861.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
KAMPHUIS, I. G., KALK, K. H., SWARTE, M. B. A. & DRENTH, J. (1984). *J. Mol. Biol.* **179**, 233–256.
KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181–187.
KIRKPATRICK, S., GELATT, C. D. & VECCHI, M. P. (1983). *Science*, **220**, 671–680.
KRAUT, J. (1958). *Biochim. Biophys. Acta*, **30**, 265–270.
KUNDROT, C. E. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 157–170.
LANGRIDGE, R., MARVIN, D. A., SEEDS, W. E., WILSON, H. R., HOOPER, C. W., WILKINS, M. H. F. & HAMILTON, L. D. (1960). *J. Mol. Biol.* **2**, 38–64.
LESLIE, A. G. W. (1991). *Isomorphous Replacement and Anomalous Scattering: Proceedings of the CCP4 Study Weekend, January* 1991, pp. 9–22. Warrington, England: SERC Daresbury Laboratory.
LUNIN, V. Y. (1993). *Acta Cryst.* D**49**, 90–99.
LUNIN, V. Y., LUNINA, N. L., PETROVA, T. E., VERNOSLOVA, E. A., URZHUMTSEV, A. G. & PODJARNY, A. D. (1994). *Daresbury Laboratory Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, Vol. 30, pp. 37–44.
MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953). *J. Chem. Phys.* **21**, 1087–1092.
PICKERSGILL, R. W., HARRIS, G. W. & GARMAN, E. (1992). *Acta Cryst.* B**48**, 59–67.
PICKERSGILL, R. W., RIZKALLAH, P., HARRIS, G. W. & GOODENOUGH, P. W. (1991). *Acta Cryst.* B**47**, 766–771.
RAMANADHAM, M., SIEKER, L. C. & JENSEN, L. H. (1990). *Acta Cryst.* B**46**, 63–69.
SATO, T. (1992). *Acta Cryst.* A**48**, 842–850.
SJÖLIN, L. & SVENSSON, L. A. (1993). *Acta Cryst.* D**49**, 66–74.
SUBBIAH, S. (1991). *Science*, **252**, 128–133.
SUBBIAH, S. (1993). *Acta Cryst.* D**49**, 108–119.
TEETER, M. M. & HENDRICKSON, W. A. (1979). *J. Mol. Biol.* **127**, 219–223.
TEN EYCK, L. F. (1977). *Acta Cryst.* A**33**, 486–492.
TSAO, J., CHAPMAN, M. S. & ROSSMANN, M. G. (1992). *Acta Cryst.* A**48**, 293–301.
YEATS, T. O. & ZHANG, K. Y. J. (1993). *Science*, **259**, 1771–177?.
ZHANG, K. Y. J. (1993). *Acta Cryst.* D**49**, 213–222.