Acta Cryst. (1996). D**52**, 909–915

Extraction of Molecular Edges on the Average Difference Map by a Monte-Carlo Method

QINGPING XU, LIWEN NIU,* MAIKUN TENG AND WEIMIN GONG

Young Scientist Laboratory of Structural Biology, and Department of Biology, University of Science and Technology of China, Hefei, Anhui 230027, People's Republic of China. E-mail: niu@xtal.bio.ustc.edu.cn

(*Received 14 June 1995; accepted 8 January 1996*)

Abstract

A method giving a low-resolution image of the molecules in the unit cell has been described, which was based only on the observed structure factors. An operator, called the average difference operator (ADO), was introduced in reciprocal space to flatten the electron densities everywhere but the regions on either side of the molecular envelope in real space. The observed structure factors were first modified by ADO, then a Monte-Carlo condensing protocol [Subbiah (1991). *Science*, **252**, 128–133; (1993). *Acta Cryst.* D**49**, 108–119] was employed to stimulate the modified electron-density map at low resolution. It was found that molecular edges could be extracted, especially when there was relatively large solvent content in the unit cell.

1. Introduction

The diffraction phases are obtained mainly *via* the multiple isomorphism replacement (MIR) method in structure determination of biological molecules by X-ray crystallography. As the number of known structures increases, the molecular-replacement method (MR, Rossmann & Blow, 1962) has been playing an increasingly important role in elucidating the structures of homogenous proteins. If threefold or greater non-crystallographic symmetry is present, the phases can be extended to medium or high resolution provided that a crude image of the molecule (molecular envelope) in the unit cell is known (Rossmann, 1990). Although much work has been devoted to expand the success of direct methods into the macromolecular X-ray crystallography (*e.g.*, Sayre, 1976; Woolfson & Yao, 1990), the invalidation of basic assumptions (such as atomicity), the limited quality (accuracy, completeness and resolution) and the bulk of the data make the work still challenging in both theory and practice.

It has been found that biological macromolecules incorporate a high degree of structural hierarchy. Well marked troughs were found in resolution ranges which separate scales of structural elements corresponding to successive levels of structure (Bricogne, 1984). Therefore, it seems reasonable to break the phasing procedure into several steps according to resolution. One can start from very low resolution and then increase the resolution step by step. At each resolution range, some physically meaningful properties of the electron density, in addition to positivity, should be employed to impose more restrictions on the phases and to obtain phases *ab initio* or to refine and extend the phases. Cannillo, Oberti & Ungaretti (1983) successfully improved and extended phases by proposing a binary modification of the electron-density map. Podjarny, Bhat & Zwick (1987) used Guassian spheres to simulate the packing in the unit cell. Subbiah (1991, 1993; David & Subbiah, 1994) proposed a simple and elegant pseudo-atom method to simulate the molecular packing and obtained encouraging results. At medium and high resolution, a pseudo-atom or grid-atom concept combined with the physical requirements, such as connectivity, has also proved to be powerful. Ramamchandran (1990) used grid atoms to simulate the crystal structure. Greer (1985) and Wilson & Agard (1993) used pseudo-atoms and connectivity to refine and extend the phases at 3.5 Å resolution or above.

What we are concerned with at ultra-low/low resolution is the molecular envelope or the packing of molecules in the unit cell. Since the number of reflections is small, the implementation of ideas is easy in both real and reciprocal space. Encouraged by the work of Subbiah (1991, 1993; David & Subbiah, 1994), we decided to work with a low-resolution model to simulate the crystal packing in real space.

2. Method

2.1. *Average difference operator (ADO)*

It is known that finding a local average or weighted local average of electron density in real space corresponds to an operator in reciprocal space (Leslie, 1987). By defining the locally averaged electron density $\langle\rho(r)\rangle$ within a sphere of radius $\langle R\rangle$ as,

$$\langle\rho(\mathbf{r})\rangle = \int_{R^3} a(|\mathbf{r}-\mathbf{u}|)\rho(\mathbf{u})\mathrm{d}V_u, \qquad (1)$$

we have,

$$\langle\rho(\mathbf{r})\rangle = \frac{1}{V}\sum_{\mathbf{h}} F(\mathbf{h})\tau(2\pi s\langle R\rangle)\exp(-2\pi i\mathbf{h}\cdot\mathbf{r}), \qquad (2)$$

where $s = 2\sin(\theta)/\lambda$. $\tau(2\pi s\langle R\rangle)$ is called the shape function. It has different forms corresponding to different

weight functions $a(t)$. For a Gaussian weight function the only one we consider in our paper is,

$$a(t) = [3/(2\pi\langle R\rangle^2)]^{3/2} \exp -[3t^2/(2\langle R\rangle^2)], \quad (3)$$

and the shape function $\tau(2\pi s\langle R\rangle)$ is (Urzhumtsev, Lunin & Luzyanina, 1989),

$$\tau(2\pi s\langle R\rangle) = \exp[-(2\pi s\langle R\rangle)^2/6]. \quad (4)$$

$\tau(2\pi s\langle R\rangle)$ is a function of $s[= 2\sin(\theta)/\lambda]$ and the average radius $\langle R\rangle$ (Fig. 1). It can be seen that $\tau(2\pi s\langle R\rangle)$ decreases rapidly as $2\sin(\theta)/\lambda$ increase if $\langle R\rangle$ remains constant. Therefore, low-order reflections are most important in calculating the locally averaged electron-density map. This is expected because taking a local average or weighted local average smears out high-resolution details and leaves only low-resolution features.

Given a structure, one can consider the so-called average difference map (ADM) which can be obtained from the difference of two electron-density maps locally averaged with two different radii $\langle R_1\rangle$ and $\langle R_2\rangle$, respectively. In the ADM, the electron densities of solvent region are zero because they are set to be constant in the averaging process. The density variations will become much smaller in the molecular region. The densities near the molecular boundaries, although dampened, still remain significant compared with those in the solvent and molecular regions. Therefore, the ADM will show larger fluctuations in the boundary regions (see Fig. 2).

The difference of the densities under the two average radii $\langle R_1\rangle$ and $\langle R_2\rangle$ corresponds in reciprocal space to a difference in the shape functions by which the observed amplitudes are multiplied. From (2) and (4), we have

$$\langle\rho_1(\mathbf{r})\rangle - \langle\rho_2(\mathbf{r})\rangle = \frac{1}{V}\sum_{\mathbf{h}} F(\mathbf{h})\Delta\tau(s, \langle R_1\rangle, \langle R_2\rangle) \times \exp(-2\pi i\mathbf{h}\cdot\mathbf{r}), \quad (5)$$

where,

$$\Delta\tau(s, \langle R_1\rangle, \langle R_2\rangle) = \tau(2\pi s\langle R_1\rangle) - \tau(2\pi s\langle R_2\rangle). \quad (6)$$

The function $\Delta\tau(s, \langle R_1\rangle, \langle R_2\rangle)$, called the average difference operator (ADO), is plotted against $2\sin(\theta)/\lambda$ in Fig. 1. It can be seen easily that the ADO truncates not only the high-resolution amplitudes but also dampens the ultra-low resolution amplitudes. Only the data within a specific resolution range are relatively strengthened. In other words, it has the effect of a window function which selects the reflections within a narrow resolution range. This characteristic of ADO is useful at low resolution for two primary reasons. First, reflections of ultra-low resolution are often practically unavailable from most conventional experiments. Second, the high-resolution data are numerous and unnecessary for low-resolution phasing. The peak of $\Delta\tau(s, \langle R_1\rangle, \langle R_2\rangle)$ can be easily found to be $s_{\max}\{= [2\sin(\theta)]/\lambda\} = [3\ln(\langle R_2\rangle/\langle R_1\rangle)/\pi^2(\langle R_2\rangle^2 - \langle R_2\rangle^2)]^{1/2}$. It should be noted that the analysis above holds well only for a relatively large proportion of the solvent, restricted to low resolution, and is not valid in a densely packed unit cell. In this case the ADM will show nearly the same fluctuations as the whole space.

2.2. *Extraction of molecular edges*

Suppose that the ADM is a binary-style map in which all the electron densities take only two values besides zero, ρ_{pro} and ρ_{sol}, then the following relationships hold,

$$F_{\text{obs}}(\mathbf{h}) = F_{\text{pro}}(\mathbf{h}) + F_{\text{sol}}(\mathbf{h}), \quad (7)$$

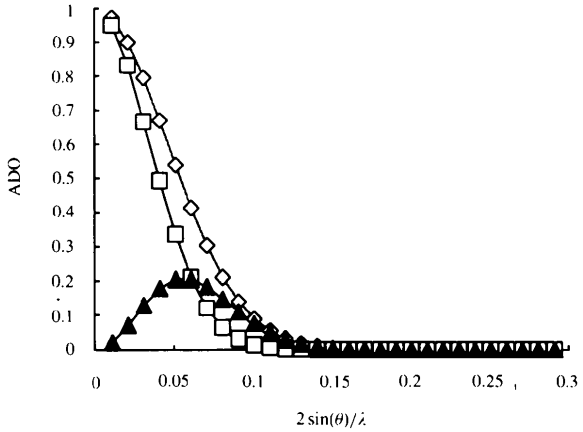

Fig. 1. The plot of the shape function $\tau(2\pi s\langle R\rangle)$ and average difference operator (ADO) $\Delta\tau(s, \langle R_1\rangle, \langle R_2\rangle)$ against s $[= 2\sin(\theta)/\lambda]$. ◇ represents the shape function under an average radius of 6 Å, □ shows the shape function under an average radius of 8 A. The ADO function calculated from the two shape functions is denoted by ▲.

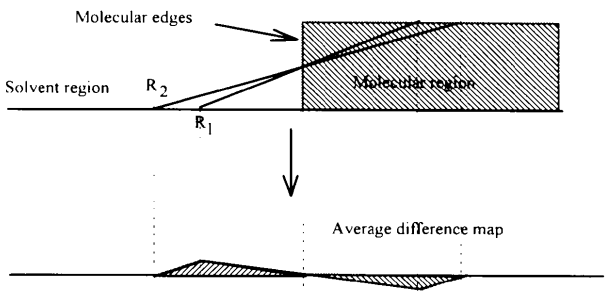

Fig. 2. A simple illustration of the average difference map. The electron densities in the molecular region and the solvent region are assumed to have two constant values (ρ_{pro} and ρ_{sol}, $\rho_{\text{pro}} > \rho_{\text{sol}}$), respectively. After locally averaging within two different radii $\langle R_1\rangle$ and $\langle R_2\rangle$, the difference of the two locally averaged structures is shown in the shaded region. Note that the difference map has a different sign and a relatively large value at either sides of the molecular edges.

$$F_{\rm sol}(\mathbf{h}) = \rho_{\rm sol} \int_{V_{\rm sol}} \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\mathrm{d}\mathbf{r}, \quad (8)$$

$$F_{\rm pro}(\mathbf{h}) = \rho_{\rm pro} \int_{V_{\rm cell}-V_{\rm sol}} \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\mathrm{d}\mathbf{r}$$
$$= \rho_{\rm pro} V\delta_{\mathbf{h},0} - \rho_{\rm pro} \int_{V_{\rm sol}} \exp(2\pi i \mathbf{h}\mathbf{r})\mathrm{d}\mathbf{r}. \quad (9)$$

From these equations, we have,

$$|F_{\rm obs}(\mathbf{h})| \propto \left| \int_{U} \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\mathrm{d}\mathbf{r} \right|, \quad (10)$$

where U is the solvent region or the molecular region. (10) can be solved approximately by the method of Subbiah (1991, 1993): a small number of hard-sphere point scatterers with uniform scattering factor were placed in the unit cell by a Monte-Carlo condensing protocol. Three constraints were enforced, two in real space and one in reciprocal space. These scatterers are placed in the unit cell so that they should comply with the known modified amplitudes. At the same time, the distance of any two point scatterers should not fall below a minimum value (*e.g.* 3 Å). The distance constraint has the effect of preventing over sampling. Another constraint in real space is the compactness criterion (Comp) defined by Subbiah (1993). To make this model agree with the known constraints (amplitudes, colliding distance and compactness), we can maximize the correlation coefficient (CC) using the amplitudes, colliding distance and compactness criteria as constraints. The CC is defined by,

$$\mathrm{CC} = \left\{ \sum_{\mathbf{h}} [|F_c(\mathbf{h})| - \langle |F_c(\mathbf{h})| \rangle][|F_o(\mathbf{h})| - \langle |F_o(\mathbf{h})| \rangle] \right\} \Big/$$
$$\times \left\{ \sum_{\mathbf{h}} [|F_c(\mathbf{h})| - \langle |F_c(\mathbf{h})| \rangle]^2 \right.$$
$$\left. \times \sum_{\mathbf{h}} [|F_o(\mathbf{h})| - \langle |F_o(\mathbf{h})| \rangle]^2 \right\}^{1/2} \quad (11)$$

The implementation is as follows.

2.2.1. *Step 1.* *Na* point scatterers per asymmetric unit are generated randomly in the unit cell. *Na* is chosen to be large enough to sample the fluctuations in the map. The required number of point scatterers is greatly decreased because the density in the main solvent and molecular region is near zero. A good estimation for *Na* is to be near $0.67N_{C\alpha}$ in which $N_{C\alpha}$ is the number of Cα atoms per asymmetric unit (Subbiah, 1991). We dampen this value by 20–40% because of data incompleteness. In fact, the results are insensitive to the *Na* value.

2.2.2. *Step 2.* Reflections are selected within a specific resolution range. It is required that the overdeterminacy condition should be satisfied. It is recommended that the low-resolution data (the more complete the better) should be used. High-resolution reflections are to be omitted in order to speed up the calculations. In general, the ratio of unique reflections to the number of independent point scatterers should be in the range 2–4. Two average radii $\langle R_1 \rangle$ and $\langle R_2 \rangle$ are then defined, and the amplitudes are multiplied by $\Delta\tau(s, \langle R_1 \rangle, \langle R_2 \rangle)$. The average radii $\langle R_1 \rangle$ and $\langle R_2 \rangle$ should be carefully chosen so that the density fluctuations near the boundaries of the solvent (or the molecular regions) should be sufficiently large. In general, we choose $\langle R_1 \rangle$ to be near 1–2 Å below the upper resolution limit and $\langle R_2 \rangle$ around 2–4 Å above the upper resolution limit.

Table 1. *Condensing-protocol parameters for the test examples*

$N_{\rm ref}$ is the reflections selected within the given resolution range. *Na* is the number of point scatters per asymmetric unit. $\langle R_1 \rangle$ and $\langle R_2 \rangle$ are the two average radii. μ_i, μ_f are the initial and final search steps, respectively.

	AaHIII	AaHI	Dau-d(CG)$_3$
Resolution range (Å)	∞–8	∞–9	∞–4
$N_{\rm ref}$	264	183	209
Completeness (%)	58	67	62
Na	100	80	40
$\langle R_1 \rangle$, $\langle R_2 \rangle$ (Å)	7, 9	7, 12	4, 8
μ_i, μ_f (Å)	25, 5	35, 8	20, 5
Colliding distance (Å)	3.0	3.0	3.0

2.2.3. *Step 3.* The maximization procedure has been accomplished by the Monte-Carlo random-walk method (Subbiah, 1991, 1993). The basis of the process is that scatterers are moved individually by steps which are first very large. As the process continues, the search steps are reduced in length and the search is then repeated at the decreased step. Once the step size is near the upper resolution limit, the whole search procedure is then terminated. The point scatterers should be sampled sufficiently until the target function CC does not increase for most of the moves at each step (as we will see in our examples, having fewer searches at each step is also valid). At the end of the maximization procedure all point scatterers will fall into a stable configuration, which should represent the molecular region or the solvent region. The molecular packing or molecular envelope can be readily extracted then. The algorithm similar to that in direct methods can be employed to solve the origin and/or enantiomer ambiguities if necessary, *i.e.*, assigning the phases of several strong reflections first, then generating the random starting map that complies with these initial phases, and allowing only the movements that cause no defined phase changes in the condensing protocol.

3. Test examples

The first examples is a hemorrhagin protein (AaHIII) (Gong, Zhu, Niu & Teng, 1996*a*) from snake *Agkistrodon acutus* venom, which crystallizes in $P2_12_12_1$ ($a = 95.3$, $b = 49.9$, $c = 46.8$ Å) with one molecule (195 residues) per asymmetric unit. Its structure is solved using by the molecular-replacement method to 2.7 Å (Weimin Gong *et al.*, in preparation) using the model of

Adamalysin II kindly provided to us by Professor Bode (Gomis-Rüth *et al.*, 1994). The estimated solvent content is around 50%. The parameters used for the condensing protocol are listed in Table 1. Because the compactness criterion tends to compress all point scatterers into a cluster, we performed fewer searches under each step first. At each search step, only one or two consecutive condensed macrocycles (instead of 20–40, with a

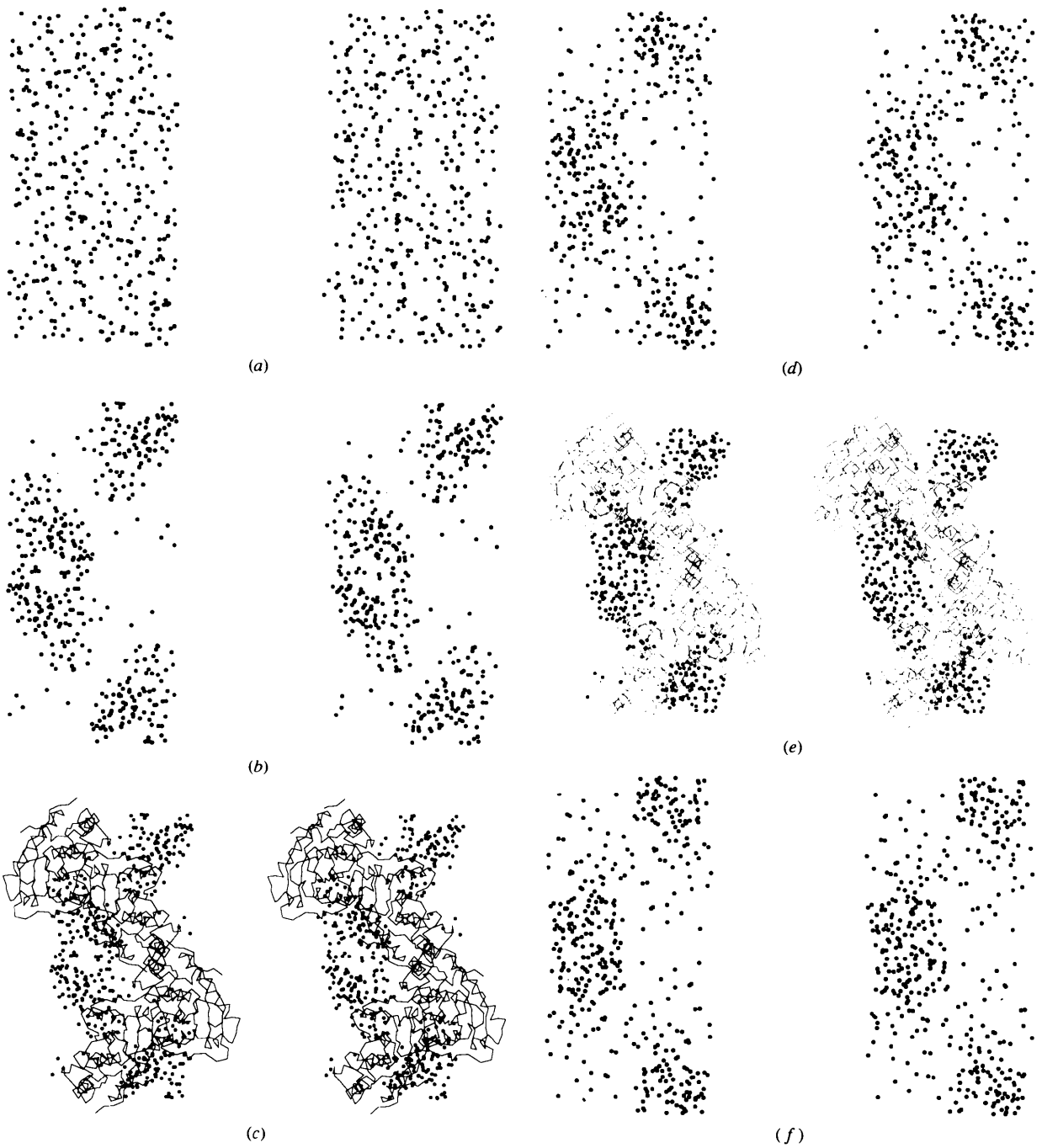

Fig. 3. Results of AaHIII test case shown in stereo pairs. (*a*) The random start map; (*b*) final configuration of the condensing protocol with ADO (a search step consists only one condensed macrocycle); (*c*) (*b*) with a C_α model and its symmetric mates superimposed on it; (*d*) final configuration of the condensing protocol without ADO, other parameters are the same as (*b*); (*e*) final configuration of the condensing protocol with ADO. Some control parameters have been changed ($\langle R_1 \rangle = 7$ Å, $\langle R_2 \rangle = 12$ Å, a search step consists of 20 condensed macrocycles); (*f*) final configuration of the condensing protocol without ADO (a search step consists of 20 condensed macrocycles), compared with (*e*).

condensed macrocycle being defined by Subbiah, 1991) terminated the search under this step. The CC value increased from −0.05 to 0.71 and compactness (Comp) decreased from 108 780 to 88 730. The results are shown in Figs. 3(*b*) and 3(*c*). It was found that most of the point scatterers converged to the solvent region, especially the edges. Another run with average radii $\langle R_1 \rangle = 7$, $\langle R_2 \rangle = 12$ Å and sufficient samples at each search step (20 condensed macrocycles) gave similar results (Fig. 3*e*) with the numeric results: initial CC = −0.04, final CC = 0.83; initial Comp = 108 780, final Comp = 80 148. The run with the same parameters but without ADO gave a poorer indication of the solvent region (Fig. 3*f*). Corresponding numeric results were: initial CC = 0.04; final CC = 0.78, initial Comp = 108 780; final Comp = 92 261.

AaHI (Gong *et al.*, 1996*b*; Gong, Teng & Niu, 1996) a homogenous molecule of AaHIII, was used as the second example. It crystallizes in $P4_32_12$ ($a = 63.5$, $b = 63.5$, $c = 95.5$ Å). Its solvent content is near 40%. The CC for the initial random configuration was −0.05, the final CC = 0.82. The compactness also decreased from 69 199 to 49 505. Table 1 lists the parameters used for the condensing protocol. The results (Figs. 4*a* and 4*b*) are better than those of Subbiah's original method using the same parameters.

The third example is a DNA–drug complex, daunomycin and d(CGCGCG) [Dau–d(CG)$_3$] (Teng, unpublished work) which crystallizes in space group $P4_32_12$ with cell dimensions $a = 27.943$, $b = 27.943$, $c = 52.243$ Å. The final R factor for the refined structure was 0.20. The solvent content is 60%. The condensing protocol control parameters are given in Table 1. The numeric results of the condensing protocol were: CC value for the initial start map = −0.0; final CC = 0.76; initial Comp, compactness of random start = 8388; final Comp = 6325. As can be seen in Figs. 5(*a*) and 5(*b*), most of the point scatterers converged to the solvent region. Another run with the same parameters except with the upper resolution limit changed to 6 Å, in which only 72 reflections were selected (73% completeness)

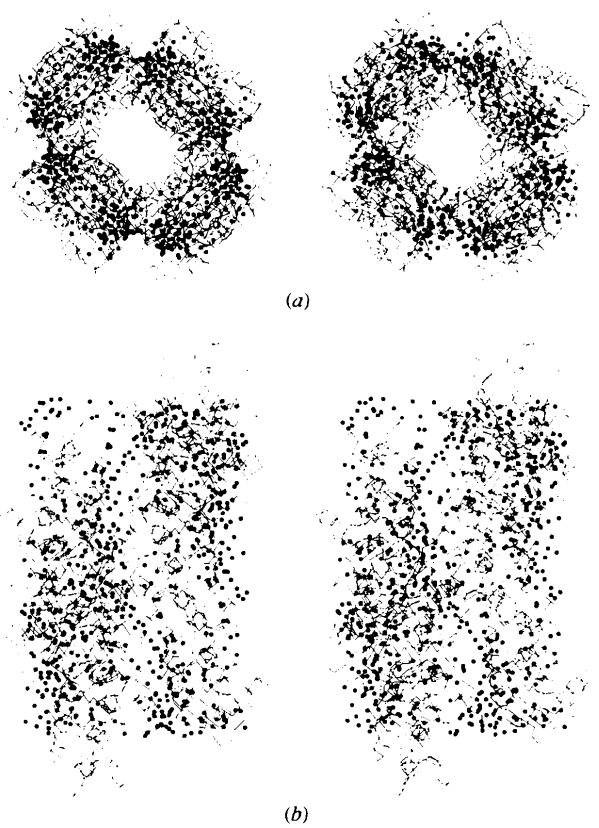

Fig. 4. The distribution of point scatterers in the AaHI unit cell with a C_α model of AaHI and all its symmetric mates superimposed on it at the end of the Monte-Carlo random-walk maximization protocol. Most of the scatterers converged into a compact region, which is the molecular region. (*a*) The view is along z axis; (*b*) the view is along x axis.

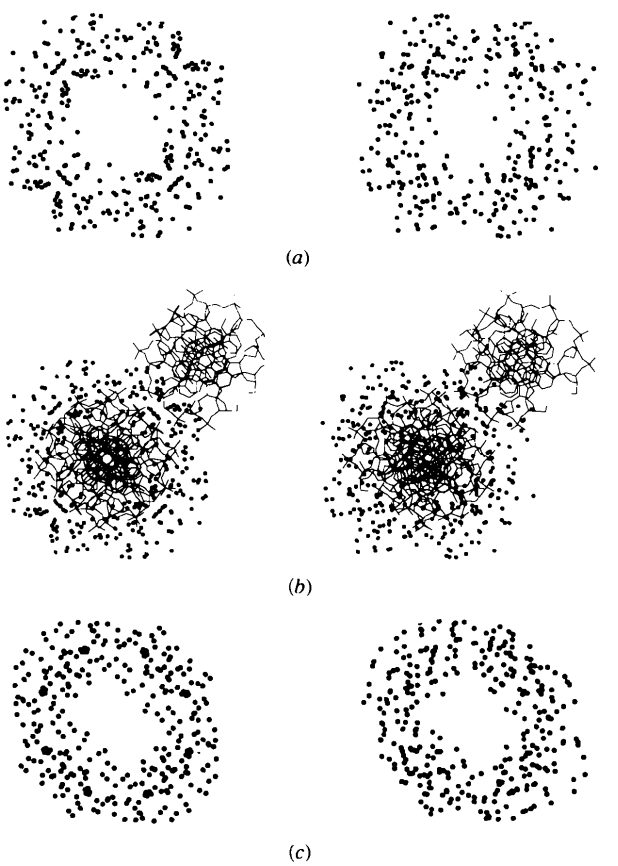

Fig. 5. The final configuration of the condensing protocol of the test Dau–d(CG)$_3$, shown as a stereo pair. Most of the point scatterers fill the solvent region. (*a*) The case of 4 Å resolution: the distribution of all point scatterers in the unit cell. (*b*) The case of 4 Å resolution: the distribution of point scatterers in the unit cell and three symmetry-related Dau–d(CG)$_3$ molecules superimposed; (*c*) the case of 6 Å resolution, in this case the condensing protocol without ADO converged to a very poor map using the same parameters.

also gave the similar results. The numeric results were: initial CC = 0.09, final CC = 0.915; initial Comp = 8388, final Comp = 7046. The distribution of point scatterers is shown in Fig. 5(*c*). Running the condensing protocol with the same controlling parameters but without ADO gave much more sparse distribution of point scatterers in the unit cell.

Similar results were obtained with different crystals with different solvent contents, space groups and cell sizes. Our tests suggested that the introduction of ADO into Subbiah's condensing protocol makes Subbiah's method more stable with respect to choosing parameters and a more rapid convergence to correct solutions. In general, the condensing protocol with ADO implemented gave higher correlation of structure factors and more a compact distribution of point scatterers than that without ADO, if running with the same parameters. The edges extracted are somewhat superior to that obtained by Subbiah's method. Also, our method performs especially well if the solvent contents are relatively large. The incompleteness of data at ultra-low resolution is not very crucial, but the more complete the reflections relatively strengthened by ADO are the better. This requirement presents no further problems in practice. Furthermore, the whole procedure is stable with respect to choosing parameters within the defined ranges.

All our tests used experimental data without particular attention to low-order reflections. The computing time used was moderate, it ranged from a few minutes to several hours on PC386 or PC486 computers.

4. Discussion

Unfortunately, we cannot tell whether the resultant map is of the molecular region or the solvent region because the amplitudes are used in our calculations only. This problem has been solved by Subbiah (1993) by proposing a sign-fixing method. Also, the maximum-entropy principle may be helpful in selecting the correct phase set [the two phase sets differentiate by 180° as seen in (8) and (9)], *i.e.* the phase set with the larger entropy is preferred as the true phase set (Sjölin, Prince, Svensson & Gilliland, 1991). All these methods need lowest order reflections.

It has often been suggested that this condensing protocol is related to simulated-annealing techniques (for a review of SA and its applications, see *e.g.*, Kirkpatrick, Gelatt & Vecchi, 1983; Brünger, 1991). This may not be the case. The essence of Subbiah's method lies in judicious choice of parameters in order to solve (10) while the unknowns are the integrated region U(solvent region or molecular region). The left-hand side of (10) can be evaluated by two methods: the random-sample method or the regular-grid sample method. Subbiah used the random-sample method to access the integrated region instead of using the grid method purely because the grid method is more difficult to implement. Another advantage of Subbiah's method is the use of correlation coefficient (CC) as target function. First, CC is not affected by a scaling problem. Second and more importantly, CC is more indicative than other indicators such as R factor when the images are crudely similar. We think that other properly implemented optimization techniques which overcome these problems should also be able to give the same results.

In the above, the ADM was simulated in real space by a simple model with a Monte-Carlo method. However, it should be noted that it has the potential of further development in reciprocal space. The contribution of strong reflections of the lowest order is dampened and series truncation errors are also decreased because of the windows effect of ADO. These would make low-resolution implementation of ADO in reciprocal space an attractive prospect. If we add a constant to the whole map to ensure that all the values of electron density in the unit cell are greater than zero, maximum-entropy methods (Bricogne, 1984; Harrison, 1987; Sjölin *et al.*, 1991) can be used readily to refine or extend the phases. Further, this method may be combined with the method suggested by Rees (1990), which depends only on the envelope.

We are greatly indebted to Dr Subbiah for his kind suggestions and encouragement to continue this work. Grants to LN were provided by Chinese Academy of Sciences and State Education Commission of China.

References

Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–450.
Brünger, A. T. (1991). *Annu. Rev. Phys. Chem.* **42**, 197–223.
Cannillo, E., Oberti, R. & Ungaretti, L. (1983). *Acta Cryst.* A**39**, 68–74.
David, P. & Subbiah, S. (1994). *Acta Cryst.* D**50**, 132–138.
Gomis-Rüth, F. X., Kress, L. F., Kellermann, J., Mayr, I., Lee, X., Huber, R. & Bode, W. (1994). *J. Mol. Biol.* **239**, 513–544.
Gong, W., Teng, M. & Niu, L. (1996). *Chin. Sci. Bull.* In the press.
Gong, W., Zhu, Z., Niu, L. & Teng, M. (1996*a*). *Chin. Sci. Bull.* In the press.
Gong, W., Zhu, Z., Niu, L. & Teng, M. (1996*b*). *Acta Cryst.* D**52**, 201–202.
Greer, J. (1985). *Methods Enzymol.* **115**, 206–224.
Harrison, R. W. (1987). *Acta Cryst.* A**43**, 428–430.
Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P. (1983). *Science*, **220**, 671–680.
Leslie, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.
Podjarny, A. D., Bhat, T. N. & Zwick, M. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **6**, 351–373.
Ramachandran, G. N. (1990). *Acta Cryst.* A**46**, 359–365.
Rees, D. (1990). *Acta Cryst.* A**46**, 915–922.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Sayre, D. (1976). *Crystallographic Computing Techniques*, pp. 322–327. Copenhagen: Munksgaard.
Sjölin, L., Prince, E., Svensson, L. A. & Gilliland, G. L. (1991). *Acta Cryst.* A**47**, 216–223.
Subbiah, S. (1991). *Science*, **252**, 128–133.
Subbiah, S. (1993). *Acta Cryst.* D**49**, 108–119.
Urzhumtsev, A. G., Lunin, V. Y. & Luzyanina, T. B. (1989). *Acta Cryst.* A**45**, 34–39.
Wilson, C. & Agard, D. A. (1993). *Acta Cryst.* A**49**, 97–104.
Woolfson, M. M. & Yao, J. X. (1990). *Acta Cryst.* A**46**, 409–413.