Acta Cryst. (1999). A**55**, 739–745

A likelihood-based search for the macromolecular position in the crystalline unit cell

T. E. Petrova,[a] V. Y. Lunin[a] and A. D Podjarny[b]*

[a]*Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region 142292, Russia, and* [b]*UPR de Biologie Structurale, IGBMC, BP 163, 67404 Illkirch CEDEX CU de Strasbourg, France. E-mail: podjarny@igbmc.u-strasbg.fr*

(*Received 20 November 1998; accepted 28 January 1999*)

Abstract

The goal of this work is to analyse whether the generalized likelihood criterion can be used to find the best spherical envelope of a macromolecule in a unit cell. A family of spherical envelopes is ranged in accordance with their likelihood values calculated by means of a Monte Carlo-type computer procedure. Two kinds of envelope families were tested. The first one was composed of spherical envelopes of fixed radius but different positions in the unit cell. In the second case, the sphere radii were linked to their centre position so that the total volume occupied by all symmetry-related spheres was roughly equal to the total volume occupied by the real molecule. The experiments showed that when using the first type of envelope the level of the signal for the right solution is higher than the one obtained with the straightforward *R*-factor-based single-Gaussian-atom search, but spurious maxima (usually placed on the symmetry axes) may still exist. The use of the second type of envelope family reduces the level of the spurious maxima.

1. Introduction

During the past few years, the problem of phasing macromolecules at low resolution has been a subject of considerable interest. This is explained by the fact that traditional methods of structure determination do not always succeed in the case of large macromolecular complexes. The molecular replacement (MR) method can be applied only if the model of a highly homologous structure is available, while the multiple isomorphous replacement (MIR) techniques depend on obtaining isomorphous derivatives, which is often problematic. As an alternative, *ab initio* low-resolution phasing can be attempted as the first stage in the structure determination of macromolecules. The information thus obtained can be used for phase extension or as a help during the application of MIR or MR.

At very low resolution, the unit cell can be roughly divided into two regions, one of which is occupied by the molecule and the other by solvent. The boundary between these two regions is the envelope. One of the possible approaches to the low-resolution phasing of proteins involves obtaining the envelope, calculating the corresponding low-resolution phases and using them as a starting set for phase extension and refinement. These low-resolution phases can be calculated *e.g.* from the models consisting of a small number of large Gaussian scatterers (Podjarny *et al.*, 1987; Lunin *et al.*, 1995) or even from a simple spherical envelope (Andersson & Hovmöller, 1996).

The first step in obtaining a fixed-form envelope is to determine the centre position. Usually, two approaches have been used for this determination: the MR method and the single-sphere search. Using the very low resolution reflections, MR can be used even if the sequence similarity is low or there is only a very low resolution electron-microscopy image (Urzhumtsev & Podjarny, 1995). In the course of the single-sphere search, the correct position of the centre can be identified either by *R* factors or by correlation coefficients of the observed and calculated magnitudes. This search should work if the low-resolution envelope can be approximated by a single sphere. This is true for proteins of globular form and for the cases when nonspherical proteins are packed in such a way that the solvent regions are approximately spherical. The single-sphere search has been used for determining the centre in several cases (Kraut, 1958; Teeter & Hendrickson, 1979; Harris, 1995; Andersson & Hovmöller, 1997). Unfortunately, both these and our calculations show that this is prone to spurious solutions. Even when it is sometimes possible to calculate correctly either the centre of the molecule or the solvent region, the method is not fully reliable and often the results cannot be interpreted.

The goal of this work is to analyse whether the generalized likelihood (GL) approach (Lunin *et al.*, 1998) can be used for increasing the single-sphere-search reliability. We consider envelopes of spherical shape as the simplest case, range them in accordance with the likelihood and take the centre of the region possessing of the maximal GL value to be the centre of the molecule. To apply the GL criterion, a great number of pseudoatomic models are generated randomly for every trial envelope and the frequency of appearance of models with a magnitude correlation greater than some

fixed level is evaluated. As was shown previously (Lunin *et al.*, 1998), this criterion can be considered as an analogue of the statistical maximal-likelihood principle.

All tests were performed with experimental rather than calculated sets of low-resolution data.

2. Methodology

2.1. *The generalized likelihood*

In the framework of the statistical approach, a given structure is considered as one of the possible trial sets. In these trials, N atoms are placed independently in the asymmetric part of the unit cell following randomly some prior probability density function $q(\mathbf{r})$. The corresponding structure factors can then be calculated and become random variables. The problem is to determine the prior $q(\mathbf{r})$ that produces the maximum agreement between observed and calculated data. Bricogne (1988) suggested in this situation choosing as the optimal prior the one that satisfies the statistical principle of maximum likelihood (Cox & Hinkley, 1974). This principle was also used for the choice of the most appropriate structure-factor distribution from the class of Gaussian distributions (Lunin & Urzhumtsev, 1984; Read, 1986; Lunin & Skovoroda, 1995; Urzhumtsev, Skovoroda & Lunin, 1996) when estimating errors in phases calculated from preliminary models.

For every prior $q(\mathbf{r})$, the value of the likelihood $L(q(\mathbf{r}))$ may be defined as the probability of getting the observed set $\{F^o_{\mathbf{h}}\}$ of magnitudes of the structure factors provided the atoms are randomly generated according to this prior:

$$L(q(\mathbf{r})) = P\{F_{\mathbf{h}} = F^o_{\mathbf{h}}, \quad \text{for all } \mathbf{h}\}. \tag{1}$$

The simplest possible envelope is a sphere with radius R. We call it the envelope-generating sphere. In the unit cell, the full envelope is a boundary of the region that is the union of all spheres related by crystallographic symmetries with the envelope-generating sphere. Below, we will call this full envelope the spherical envelope. It is defined by the centre and radius of the envelope-generating sphere. We assume that the prior probability distribution is equal to a positive constant inside the envelope and is equal to zero outside it. We can scan the unit cell and consider many different possible priors of this kind. One can expect that the prior that is chosen on the basis of the maximum-likelihood principle would correspond to the 'best' spherical envelope, and the centre of this region would be the centre of the molecule. Therefore, we maximize $L(q(\mathbf{r}))$, where

$$q(\mathbf{r}) = \begin{cases} 1/V & \text{inside the envelope} \\ 0 & \text{outside the envelope,} \end{cases} \tag{2}$$

provided the envelope is spherical and is defined by the centre coordinates and radius and V is the envelope volume.

The calculation of the likelihood function (1) is quite complex. The procedure involves the derivation of the joint probability distribution function for the set of structure factors and the integration of this function over the phases provided that the magnitudes are equal to the observed ones. Since it is still impossible to obtain the exact expression for the joint probability function of the magnitudes and phases, asymptotic expansions are employed. The Edgeworth series (Klug, 1958) is the most widely used asymptotic expansion but it yields an accurate representation only if the deviations of the unitary structure factors from their mean values are small. The case of large deviations was considered by Bricogne (1984); his estimation based on the saddle-point approximation is more accurate but the final formula includes an implicit function, obtained from a large system of nonlinear equations, which is difficult to solve. Integration with respect to phases also induces serious mathematical difficulties, which require numerous simplifications. One of such simplifications is the so-called 'diagonal approximation' (Bricogne & Gilmore, 1990) in which nondiagonal elements of the covariance matrix are replaced by zeros as if the structure factors were independent. This procedure may result in the loss of a considerable part of the phase information.

In this work, we used a simpler approach, which proved to be quite effective (Lunin *et al.*, 1998). It employs the generalized likelihood as a criterion of choice instead of the usual likelihood function. The GL is an analogue of the likelihood function and is defined as the probability of obtaining a set of magnitudes that are equal or close to the observed magnitudes:

$$L_\omega(q(\mathbf{r})) = P\{C(\{F_{\mathbf{h}}\}\{F^o_{\mathbf{h}}\}) \geq \omega\}, \tag{3}$$

where C is a measure of closeness of two sets of structure-factor magnitudes and ω is a chosen cut-off level. Obviously, the definition of the generalized likelihood depends on the choice of the measure of closeness C and cut-off level ω, which are the parameters of the method. In our tests, C was the coefficient of the correlation of the magnitudes, which was calculated by the following formula:

$$\mathrm{CF} = \frac{\sum_{\mathbf{h}}(F_{\mathbf{h}} - \langle F\rangle)(F^o_{\mathbf{h}} - \langle F^o\rangle)}{\left[\sum_{\mathbf{h}}(F_{\mathbf{h}} - \langle F\rangle)^2 \sum_{\mathbf{h}}(F^o_{\mathbf{h}} - \langle F^o\rangle)^2\right]^{1/2}}, \tag{4}$$

where $\langle F\rangle$ is a magnitude averaged over the considered set of reflections. The GL may be estimated with a Monte Carlo-type computer procedure. A large number of sets, consisting of N_{glob} pseudoatoms each, are placed into the unit cell randomly with a given prior probability function. (If a prior is a function of type (2), this can be performed easily, by generating atoms only inside the envelope being considered). For every generated model, a set of magnitudes $\{F^c_{\mathbf{h}}\}$ and the coefficient (4) of their correlation with the observed magnitudes are calcu-

lated. The probability (3) is estimated as the ratio of the number of generated models resulting in C values higher than ω to the total number of the generated models:

$$L_\omega \approx \frac{\text{number of models with } C \geq \omega}{\text{total number of generated models}}. \quad (5)$$

When calculating structure factors from the model, we suppose that the pseudoatoms have an electron-density Gaussian distribution

$$\rho(r) = K_{\text{glob}}(4\pi/B_{\text{glob}})^{3/2} \exp(-4\pi^2 r^2/B_{\text{glob}}) \quad (6)$$

with the same K_{glob} and B_{glob} values. It must be emphasized that, in the statistical model, when working at a low resolution we use not the real number of usual atoms but a relatively small number N_{glob} of artificially huge pseudoatoms ('globs'). So the parameter B_{glob} in (6) is not the usual temperature factor but a parameter defining the size of the glob, which may have very large values in comparison with the usual temperature factors. The value B_{glob} in (6) is chosen in the course of special computer tests (§2.5). It must be noted that the magnitude correlation (4) and hence GL value (5) depend on the B_{glob} value but do not depend on the K_{glob} value. Consequently, the problem of choice of K_{glob} does not exist.

Given only the magnitudes of the structure factors, it is impossible to distinguish solutions that are related by possible shifts of the origin and enantiomorph transform. Thus, we cannot decide whether the atoms are placed into the sphere with a centre $\mathbf{r}$ or into the sphere with a centre $\mathbf{r} + \mathbf{t}$ if $\mathbf{t}$ is any permitted shift of the origin. In fact, we calculate the generalized likelihood based on the following statistical hypothesis: the atoms are placed either into a sphere with the centre at a given position or into any other sphere that is related to this sphere by any permitted origin–enantiomorph transform.

One can define some grid in the unit cell, consider every point of this grid as the centre of the envelope generating sphere, calculate the generalized likelihood (5) for every full spherical envelope and build an L_ω map in the unit cell. Below, we will call the maxima of this map the peaks.

2.2. *Likelihood-based search for the molecule position*

We expect that the ranking of the envelopes in accordance with their GL values allows one to find the envelope that is centred in the position mostly close to the centre of the molecule. In the unit cell, the full envelope comprises not a single sphere but the union of all symmetrically related spheres. Since the spheres can overlap, two approaches can be used to define the class of feasible regions:

(i) the radius of the sphere is fixed; while scanning the unit cell, we compare the regions with different volumes as the spheres may have different overlapped parts depending on centre positions;

(ii) the volume of the full envelope is fixed; for every position of the centre of the envelope-generating sphere, its radius is adapted so that the volume of the union of all symmetrically related spheres is equal to this fixed value.

Computer tests were performed for both classes of regions.

2.3. *One-Gaussian search for the molecule position*

The results of the maximum-likelihood-based search were compared with the results of a one-Gaussian-atom search. The model consisting of one Gaussian atom (6) is placed at every point of some grid in the unit cell. For this model, a set of magnitudes $\{F_{\mathbf{h}}^c\}$ and the coefficient of correlation CF (4) are calculated. On this grid, a CF map can be built. It is supposed that the correct position of the model will coincide with the position of the main peak in the CF map.

2.4. *Control functions*

Two functions were used to judge the success of the search when the atomic coordinates for the structure being studied were known. The first function was the mean value of the coefficient of phase correlation (Lunin & Woolfson, 1993):

$$C_\varphi = \sum_{\mathbf{h}} (F_{\mathbf{h}}^o)^2 \cos(\varphi_{\mathbf{h}}^{\text{true}} - \varphi_{\mathbf{h}}) \Big/ \sum_{\mathbf{h}} (F_{\mathbf{h}}^o)^2, \quad (7)$$

where $\{\varphi_{\mathbf{h}}^{\text{true}}\}$ is the set of the true phases calculated from the known model and $\{\varphi_{\mathbf{h}}\}$ is a phase set calculated for the pseudoatom model. Each time before calculating (7), the phase alignment was produced (Lunin & Lunina, 1996). This procedure is necessary because the experimental magnitudes do not fix the origin and the enantiomorph unambiguously. The mean phase correlation $\langle C_\varphi \rangle$ was defined for a trial envelope as a value (7) averaged over all pseudoatom models generated.

The second control function was the trapping function defined as the ratio

$$T = \frac{\text{number of model atoms inside the envelope}}{\text{total number of atoms in the model}}. \quad (8)$$

2.5. *The strategy of the tests*

The general strategy for every data set involved the following steps:

(i) a grid for the considered centres of envelope-generating spheres was introduced and the number of globs N_{glob} was specified;

(ii) the optimal B_{glob} value was chosen; a great number of pseudoatoms models were placed using the uniform prior distribution of atomic coordinates and mean values of the magnitude correlation (4) were calculated with different B_{glob} values; the value resulting

in the highest mean magnitude correlation was suggested to be the optimal;

(iii) for every grid node, the full sphere-based envelope was constructed and the value of the generalized likelihood L_ω was calculated;

(iv) the CF map was calculated for the one-Gaussian-atom search;

(v) when the atomic models were known, the values of control functions $\langle C_\varphi \rangle$ and T were calculated for every tested envelope.

The calculations were performed for different values of the resolution zone, sphere radii or the full envelope volume, number $N_{\rm glob}$ of atoms and cut-off level ω.

3. Tests and results

3.1. *Data sets*

Six sets of experimental data were used in tests:

(*a*) 50 Å neutron diffraction data for the $tRNA^{Asp}$–Asp RS complex (Moras *et al.*, 1983); $I432$, $a = b = c =$ 354 Å; the structure was previously solved by the molecular replacement method (Urzhumtsev *et al.*, 1994);

(*b*) X-ray diffraction data of the ribosomal particle 50S from *Thermus thermophilus* (Volkmann *et al.*, 1990); $P4_32_12$, $a = b = 495$, $c = 196$ Å; the position of the particle in the unit cell was previously found independently by the few-atoms model (FAM) (Lunin *et al.*, 1995) and molecular replacement methods (Urzhumtsev, Vernoslova & Podjarny, 1996).

(*c*) X-ray diffraction data for four proteins crystallized in the space group $P2_12_12_1$: RNAase Sa (Sevcik *et al.*, 1991), $a = 64.90$, $b = 78.32$, $c = 38.79$ Å; protein G (kindly supplied by E. Dodson), $a = 34.9$, $b = 40.3$, $c = 42.2$ Å; γ-crystallin IIIb (Chirgadze *et al.*, 1991), $a = 58.7$, $b = 69.5$, $c = 116.9$ Å; and ribosomal elongation factor G (Aevarsson *et al.*, 1994), $a = 75.9$, $b = 105.6$, $c = 115.9$ Å.

3.2. *The complex $tRNA^{Asp}$–Asp RS*

In the first series of calculations, the envelopes were defined as the union of fixed-size spheres related by crystallographic symmetries. The likelihood maps calculated at the resolutions 40 and 60 Å for different choices of the spheres' radius and ω levels revealed a similar picture. The map showed a few peaks, one corresponding to the right solution and the others being strong false peaks, which are situated on the symmetry axes. It is necessary to point out that the maximum corresponding to the correct solution was only the fifth one in the list of the strongest peaks. Similar results were obtained for the one-Gaussian-atom search. The results obtained at a resolution of 40 Å are presented in Table 1.

Table 1. *The results of the likelihood-based search with fixed sphere radius and one-Gaussian-atom search for the position of the Asp RS complex in the unit cell*

40 Å resolution neutron diffraction data were used The generalized likelihood value L_ω [defined by equation (6)], the trapping function T [defined by equation (8)] and the mean value $\langle C_\varphi \rangle$ of the map correlation coefficient (7) were calculated for envelopes composed by the symmetry-related spheres of fixed radii 30 Å. 100-atom models with $B_{\rm glob} = 50\,000$ and cut-off level $\omega = 0.60$ were used when calculating L_ω values The coefficient of correlation of the magnitudes CF was calculated as in equation (4) with $B = 50\,000$. The peaks $p1$–$p5$ are the five highest peaks found by the single-Gaussian search; note that they also correspond to the five highest values of the likelihood function.

Peaks	Likelihood L_ω	Trapping function T	Mean map correlation $\langle C_\varphi \rangle$	One-Gaussian-atom magnitude correlation CF
$p1$	0.48	0.46	0.76	0.63
$p2$	1.0	0.0	−0.15	0.70
$p3$	1.0	0.05	−0.03	0.71
$p4$	1.0	0.03	−0.11	0.67
$p5$	1.0	0.02	0.18	0.66

The situation changes if instead of scanning the unit cell with spheres of constant radius we recalculate the radius at every position so that the summed volume of all symmetrically related spheres has a fixed value. The tests showed that the value of this volume is a crucial parameter. If the value of this volume is much less than the real volume of the molecule, the likelihood map follows the one-Gaussian search map and false peaks of the likelihood function are higher than the peak corresponding to the correct solution. If this value is roughly equal to the real volume of the molecule, the peak corresponding to the correct solution becomes the highest. This peak continues to be the highest for a somewhat increased value of the volume of the envelope. However, a tendency for levelling of the peaks appears. Figs. 1(*a*)–(*f*) show the histograms of the number of models with a given CF value (Figs. 1*a*, *c*, *e*) and the dependence of the likelihood function on the cut-off level ω (Figs. 1*b*, *d*, *f*) for those positions of the centre of the sphere that correspond to the correct solution and two spurious peaks. These dependencies are shown for the different values of the volume of the envelope. The values of control functions for the five highest peaks of the likelihood map are listed in Table 2.

3.3. *Ribosomal particle T50S*

Similar calculations were performed using 60 Å resolution experimental data for the ribosomal particle T50S (Volkmann *et al.*, 1990). In the cases of both the fixed radii of the spheres composing the envelopes and the fixed total volume of the envelope, the position of the main maximum on the likelihood map coincided well with the particle centre obtained by the FAM and the molecular replacement methods (Urzhumtsev, Verno-

slova & Podjarny, 1996). The one-Gaussian search resulted in this case in the same position. Nevertheless, the procedure based on the likelihood maximization gave a higher signal contrast in a wide range of cut-off levels ω (Table 3). It is worth noting that generation of globs with the uniform prior resulted in nearly the same

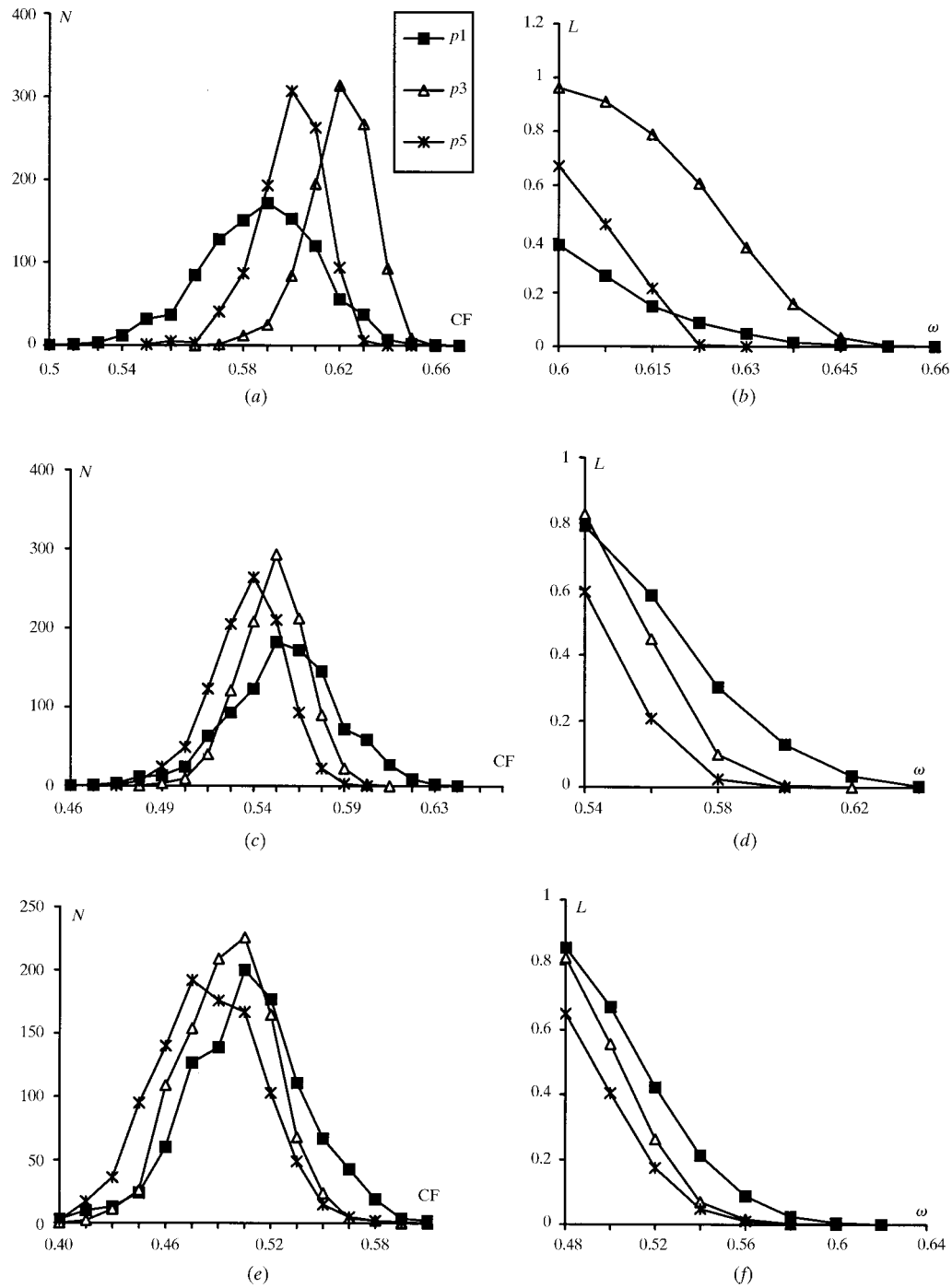

Fig. 1. Generalized likelihood-based testing of envelopes centred at the points corresponding to peaks $p1, p3, p5$ shown in Tables 1 and 2. The ratio of the volume of the envelope to the unit-cell volume was fixed and equal to: (*a*) and (*b*) 15%; (*c*) and (*d*) 30%; (*e*) and (*f*) 60%. 40 Å resolution neutron diffraction data for the AspRS complex were used. The plots are: (*a*), (*c*), (*e*) distribution of the values of the coefficient of correlation (4) for 100-atom models (B_{glob} = 50 000) generated into the trial envelopes; (*b*), (*d*), (*f*) dependence of the generalized likelihood value L on the cut-off level ω. The corresponding values of the control criteria are presented in Table 2.

Table 2. *The results of the likelihood-based search with the fixed volume of the envelope for the position of the Asp RS complex in the unit cell*

The values of the generalized likelihood value L_ω [defined by equation (6)], the trapping function T [defined by equation (8)] and the mean value $\langle C_\varphi \rangle$ of the map correlation coefficient (7) are listed for the first five peaks using three different conditions of V/V_{cell} and ω cut. 40 Å resolution neutron diffraction data were used; L_ω, T and $\langle C_\varphi \rangle$ were calculated for envelopes composed of the symmetry-related spheres. The sphere radii were adapted to fix the ratio of the envelope volume V to the unit-cell volume V_{cell}. 100-atom models with B_{glob} = 50 000 and different cut-off levels ω were used when calculating L_ω values. The peaks $p1$–$p5$ are the same peaks presented in Table 1. The corresponding values of the likelihood function at different cut-off levels are presented in Figs. 1(*b*), (*d*), (*f*) for peaks 1, 3 and 5.

	V/V_{cell} = 0.15, ω cut = 0.60			V/V_{cell} = 0.3, ω cut = 0.58			V/V_{cell} = 0.6, ω cut = 0.52		
Peaks	L_ω	T	$\langle C_\varphi \rangle$	L_ω	T	$\langle C_\varphi \rangle$	L_ω	T	$\langle C_\varphi \rangle$
$p1$	0.38	0.77	0.66	0.30	0.54	0.72	0.42	0.96	0.45
$p2$	0.00	0.16	−0.25	0.00	0.00	−0.17	0.13	0.82	0.33
$p3$	0.96	0.33	0.15	0.10	0.18	−0.15	0.27	0.60	0.10
$p4$	0.00	0.44	0.22	0.00	0.15	−0.01	0.08	0.89	0.40
$p5$	0.65	0.36	0.36	0.02	0.15	0.30	0.18	0.71	0.30

Table 3. *The contrast of main peaks in one-Gaussian and likelihood-based searches for the position of the T50S particle*

$H_{CF} = (CF^{max} - \langle CF \rangle)/\sigma(CF)$, $H_{L_\omega} = (L_\omega^{max} - \langle L_\omega \rangle)/\sigma(L_\omega)$; CF^{max}, $\langle CF \rangle$ and $\sigma(CF)$ are the maximum, mean and r.m.s.d. values of the coefficient of correlation of calculated and observed magnitudes obtained in the one-Gaussian search; L_ω^{max}, $\langle L_\omega^{max} \rangle$ and $\sigma(L_\omega)$ are the similar values for the likelihood-based search The results were obtained with d = 60 Å, R = 0 Å, B_{glob} = = 90 000.

H_{CF}	H_{L_ω}		
	ω = 0.74	ω = 0.78	ω = 0.82
2.15	4.75	7.08	13.06

mean and root-mean square deviation (r.m.s.d.) values (0.52 and 0.12) of the coefficient of correlation (4) as in the case of the one-Gaussian search (0.53 and 0.12, respectively).

3.4. *Protein G*

A similar increase in the contrast of the signal for the right solution was observed in the case of protein G (Table 4). The position of the molecular centre obtained by the maximum-likelihood procedure agrees well with the positions of the maxima of both control functions. The one-Gaussian-atom search results in the same position. However, in the case of the likelihood-based search, the contrast for the correct solution is much higher than the contrast for the next false solution, as compared with the one-Gaussian-atom search (Table 4).

3.5. *RNAase Sa, γ-crystallin IIIb and elongation factor G*

The method has not led to the correct solution for RNAase, γ-crystallin IIIb and G-factor. Possible reasons are the presence of two molecules in the independent part of the unit cell in crystals of RNAase and γ-crystallin and the essentially nonspherical form of the low-resolution envelope for G-factor. In the RNAase model, two identical molecules are very densely packed in the asymmetric unit, which makes it difficult to identify the correct answer. In this case, no correlation between the maximum of the likelihood map and the position of the model in the cell is observed. The low-resolution envelope for γ-crystallin resembles an ellipsoid rather than a sphere. However, a visual comparison of the model with the likelihood maps shows that the positions of the maxima of the likelihood at different cut-off levels ω are not far from the centre of one of the molecules. The low-resolution envelope for the G-factor is very nonspherical and the maxima of the maps of likelihood and trapping function do not coincide. However, the visual comparison of the likelihood-function map and the model shows that the maximum of the likelihood-function map is situated close to the region of the highest atomic density. The interpretation of the results for RNAase, γ-crystallin and G-factor is complicated by the fact that all of them belong to the space group $P2_12_12_1$, in which there are eight possible shifts of the origin and the additional enantiomorph ambiguity.

4. Concluding remarks

The present method made it possible to obtain the true positions of the centres for three structures: synthetase complex, T50S ribosome particle and protein G. In the cases of T50S and protein G, the method enabled a higher contrast signal to be obtained than in the one-Gaussian search. In the case of the synthetase complex, the one-Gaussian search led to spurious solutions on symmetry axes, while the generalized likelihood approach eliminated them when being accomplished by calculating the radius of the sphere for every trial position of the centre so that the summed volume of all symmetrically related spheres was roughly equal to the volume of the molecule. However, the method failed to provide the correct solution for three proteins in the space group $P2_12_12_1$. Possible reasons are the large number of possible shifts of the origin and the enantiomorph ambiguity in this space group, and nonglob-

Table 4. *The contrast of the two highest peaks in one-Gaussian and likelihood-based searches for the position of the protein G molecule*

The H_{CF} and H_{L_ω} values are defined as in Table 3. The results were obtained with d = 14 Å, B_{glob} = 1000. The value of the volume of the summed envelope was fixed and equal to 30% of the unit-cell volume.

	H_{CF}	H_{L_ω}		
		$\omega = 0.56$	$\omega = 0.58$	$\omega = 0.60$
Main peak	1.88	6.54	7.57	9.10
Highest false peak	1.68	3.06	2.44	2.05

ular shapes of these proteins. The procedure of structure solution using the molecular envelopes of more complicated forms is under development. In the cases of RNAase, γ-crystallin and G-factor, it seems reasonable to try an ellipsoid shape or to use several spheres instead of a single one to approximate the molecule. Nevertheless, in this case, the search for the maximum of the likelihood function must be performed in a larger parameter space and it will be much more time consuming.

The authors thank A. G. Urzhumtsev and T. P. Skovoroda for useful discussions and N. L. Lunina for providing the programs. This work was supported by the Centre National de la Recherche Scientifique (CNRS) through the UPR 9004, through grant No. 97-04-48319 of the RFBR, by the RAS and the CNRS through a joint RAS–CNRS collaboration, by the Institut National de la Santé et de la Recherche Médicale (INSERM) and by the Centre Hospitalier Universitaire Régional.

References

Aevarsson, A., Brazhnikov, E., Garber, M., Zhelnotsova, J., Chirgadze, Yu., Al-Karaoabhi, S., Svensson, L. A. & Liljas, A. (1994). *EMBO J.* **13**, 3669–3677.

Andersson, K. M. & Hovmöller, S. (1996). *Acta Cryst.* D**52**, 1174–1180.

Andersson, K. M. & Hovmöller, S. (1997). *Seventeenth European Crystallographic Meeting, Lisboa, Portugal, 24–28 August 1997*. Abstracts, p. 54.

Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.

Bricogne, G. (1988). *Acta Cryst.* A**44**, 517–545.

Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* A**46**, 284–297.

Chirgadze, Yu. N., Brazhnikov, E. V., Nikonov, S. V., Fomenkova, N. P., Garber, M. B., Urzhumtsev, A. G., Lunin, V. Yu., Chirgadze, N. Yu. & Nekrasov, Yu. V. (1991). *Dokl. Acad. Nauk SSSR*, **320**, 488–491.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics.* London: Imperial College.

Harris, G. W. (1995). *Acta Cryst.* D**51**, 695–702.

Klug, A. (1958). *Acta Cryst.* **11**, 515–543.

Kraut, J. (1958). *Biochim. Biophys. Acta*, **30**, 265.

Lunin, V. Y. & Lunina, N. L. (1996). *Acta Cryst.* A**52**, 365–368.

Lunin, V. Y., Lunina, N. L., Petrova, T. E., Urzhumtsev, A. G. & Podjarny, A. D. (1998). *Acta Cryst.* D**54**, 726–734.

Lunin, V. Y., Lunina, N. L., Petrova, T. E., Vernoslova, E. A., Urzhumtsev, A.G. & Podjarny, A. D. (1995). *Acta Cryst.* D**51**, 896–903.

Lunin, V. Y. & Skovoroda, T. P. (1995). *Acta Cryst.* A**51**, 880–887.

Lunin, V. Y. & Urzhumtsev, A. G. (1984). *Acta Cryst.* A**40**, 269.

Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.

Moras, D., Lorber, B., Romby, P., Ebel, J.-P., Giegé, R., Lewitt-Bentley, A. & Roth, M. (1983). *J. Biomol. Struct. Dynam.* **1**, 209–223.

Podjarny, A. D., Rees, B., Thierry, J.-C., Cavarelly, J., Jesior, J. C., Roth, M., Lewitt-Bentley, A., Kahn, R., Lorber, B., Ebel, J.-P., Giege, R. & Moras, D. (1987). *J. Biomol. Struct. Dynam.* **5**, 187–198.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Sevcik, J., Dodson, E. & Dodson, G. G. (1991). *Acta Cryst.* B**47**, 240–253.

Teeter, M. M. & Hendrickson, W. A. (1979). *J. Mol. Biol.* **127**, 219.

Urzhumtsev, A. G. & Podjarny, A. D. (1995). *Acta Cryst.* D**51**, 888–895.

Urzhumtsev, A. G., Podjarny, A. D. & Navaza, J. (1994). *Jnt CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **30**, 29–36.

Urzhumtsev, A. G., Skovoroda, T. P. & Lunin, V. Y. (1996). *J. Appl. Cryst.* **29**, 741–744.

Urzhumtsev, A. G., Vernoslova, E. A. & Podjarny, A. D. (1996). *Acta Cryst.* D**52**, 1092–1097.

Volkmann, N., Hottenträger, S., Hansen, H. A. S., Zaytsev-Bashan, A., Sharon, R., Yonath, A. & Wittmann, H. G. (1990). *J. Mol. Biol.* **216**, 239–241.