Acta Cryst. (1996). D**52**, 1092–1097

Approaches to Very Low Resolution Phasing of the Ribosome 50S Particle from *Thermus thermophilus* by the Few-Atoms-Models ands Molecular-Replacement Methods

A. G. URZHUMTSEV,† E. A. VERNOSLOVA† AND A. D. PODJARNY

UPR de Biologie Structurale, IGBMC, BP 163, 67404 Illkirch CEDEX, CU de Strasbourg, France. E-mail: sacha@igbmc.u-strasbg.fr

(*Received 30 October 1995; accepted 29 April 1996*)

Abstract

Estimates for the phases of the X-ray diffraction data from the 50S ribosomal particle of *Thermus thermophilus* has been made to an effective resolution around 80 Å using the few-atoms-modes *ab initio* technique [Lunin, Lunina, Petrova, Vernoslova, Urzhumtsev & Podjarny (1995). *Acta Cryst.* D**51**, 896–903]. This technique models the density with a small number of Gaussian spheres to generate a large number of possible phase sets and then uses clustering algorithms to identify the best ones. Independently, an envelope obtained from electron-micrograph image reconstruction [Yonath, Leonard & Wittmann (1987). *Science*, **236**, 813–816] was oriented and positioned using the molecular-replacement technique, specially adapted to the very low resolution case [Urzhumtsev & Podjarny (1995). *Acta Cryst.* D**51**, 888–895]. The two methods show similar packing arrangements.The electron density calculated by the few-atoms-models technique without any assumption on the number of molecules in asymmetric unit or on their shape shows recognizable features of the particle.

1. Introduction: phasing problem for the ribosome

Because of the biological importance of the ribosome a large effort has been made to obtain high-quality crystals, diffracting at best to 3.0 Å (von Böhlen *et al.*, 1991). The structure determination by X-ray diffraction methods demands the solution of the phase problem, which poses special challenges. Because of the large size of the ribosomal particles, the relative change in diffraction amplitudes resulting from single heavy-atom substitutions usually applied would be too small. Therefore, dense heavy-atom clusters are being employed. Even with these clusters, the application of the heavy-atom method poses difficulties; to validate the corresponding maps the knowledge of the molecular envelope should be very useful.

The determination of structures of complexes of similar size (the viruses) has been made possible by the use of the non-crystallographic symmetry averaging technique. The existing ribosome crystals do not have this symmetry. However, as the same ribosomal particles from different bacteria crystallize in different forms (three for the 50S case, Berkovitch-Yellin, Bennett & Yonath, 1992), this technique can be used to merge density maps from different crystal forms to improve the image quality. This requires the accurate knowledge of the corresponding envelopes. Therefore, a special effort is being developed to obtain these molecular envelopes by independent methods at very low resolution. In parallel, low-resolution image reconstruction from electron micrographs of ordered monolayers led to molecular envelopes.

This paper concerns the phasing results obtained at very low resolution for the 50S ribosome particle from *Th. thermophilus* (T50S) using the FAM (few-atoms-models, Lunin *et al.*, 1995) and the low-resolution molecular-replacement methods (Urzhumtsev & Podjarny, 1995*a*). This particle crystallizes in space group $P4_12_12$ (or $P4_32_12$), unit cell $a = b = 498$, $c = 198$ Å (Volkmann *et al.*, 1990). The X-ray diffraction data, practically complete for resolutions $d > 60$, were measured by a special experimental set up optimized for low-angle measurements (Berkovitch-Yellin *et al.*, 1994).

2. FAM phasing

2.1. *The FAM method*

The few-atoms-models method is an *ab initio* phasing technique for very low resolution diffraction data. The details of the method and two succesful applications are described by Lunin *et al.* (1995). In brief, it consists of the following main steps.

2.1.1. *Model generation.* A very large number (millions) of simple models are generated. Each of them consists of the same small number (usually less than ten) of large Gaussian spheres. The coordinates of these spheres are generated randomly and no conditions are imposed to prevent overlap.

2.1.2. *Model selection in reciprocal space.* A structure factor set at very low resolution is calculated from each model. If the correlation between the calculated and observed amplitudes, defined for example by,

$$\mathrm{Corr}(F^{\mathrm{obs}}, F^{\mathrm{calc}}) = \sum_s [F^{\mathrm{obs}}(s) \cdot F^{\mathrm{calc}}(s)] \div (\{\sum_s [F^{\mathrm{obs}}(s)]^2\} \cdot \sum_s [F^{\mathrm{calc}}(s)]^2)^{1/2}, \quad (1)$$

† Permanent address: Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region, 142292, Russia.

ISSN 0907-4449

is below a given threshold the model is rejected. The threshold is chosen such that the number of accepted models is adequate to sample the space; test calculations place this number around a few hundred. The uncentered correlation (1) is useful for maximizing the agreement between the strongest amplitudes at very low resolution.

2.1.3. *Merging the model phase sets.* The phase sets calculated from the selected models are then merged through a clustering procedure. This procedure is based on the definition of a distance in a multidimensional 'phase space'. All the phases of each set are considered as a single point in this phase space, where the dimensionality is the number of phases and the coordinates of this point are the values of the phases. The distance D_{jk} between points j and k is determined in terms of the corresponding densities ρ_j and ρ_k as,

$$D_{jk} = \kappa^* \{ \int [\rho_j(r) - \rho_k(r)]^2 \mathrm{d}^3 r \}^{1/2}, \qquad (2)$$

where $\kappa = \kappa_j = \int [\rho_j(r)]^2 \mathrm{d}^3 r \}^{-1/2}$ is the same for all models. Here ρ_j stands for an electron-density distribution calculated with the Fourier coefficients $\{F^{\mathrm{obs}}(s) \exp[i\varphi_j^{\mathrm{calc}}(s)]\}$, with the phases from the jth model. This distance (2) is related to the usual map correlation coefficient C_{jk} as,

$$C_{jk} = \kappa^2 * \{ \int [\rho_j(r) \rho_k(r)] \mathrm{d}^3 r \} = 1 - (D_{jk}^2/2). \qquad (3)$$

The process of clustering (Lunin, Urzhumtsev & Skovoroda, 1990) is then carried out as a function of a parameter δ. Initially, $\delta = 0$ and each point is an individual cluster. δ is then continuously increased, until it reaches the value of the smallest D_{jk}; then points j and k are merged into and replaced by a new point. The merging process, illustrated in Fig. 1(*a*), results in the arrangement of the selected phase sets in a 'cluster tree' (see Fig. 1*b*). In this tree, the Y axis represents the variable δ which is 0 at the bottom and increases as the correlation between maps from different cluster decreases; the value of δ corresponding to uncorrelated maps is $2^{1/2} \simeq 1.42$. The X axis represents the initial phase sets, with one point per phase set. As δ increases, the different points are merged, until finally they converge to a single one. The tree can then be read starting from the single unique cluster at the top and going down to the smaller ones; each cluster division represents a 'level'. Inside each cluster, the phase $\varphi(s)$ and its weight $w(s)$ can be obtained by averaging all the original phases from the starting models, as follows,

$$w(s) \cdot \exp[i\varphi(s)] = \sum_j \exp[i\varphi_j^{\mathrm{calc}}(s)]/N, \qquad (4)$$

where the index j goes over all the N points in the cluster.

Step (*c*) is applied as the points in phase space selected in (*b*) are generally observed to group in a small number of clusters. It is expected that one of the resulting phase sets corresponds to the correct solution, and that it can be identified by other criteria like molecular packing or symmetry considerations. Note that in the FAM solution the choice of enantiomer is arbitrary.

2.2. *FAM phasing for the T50S particle*

The FAM method was applied to experimental diffraction data of T50S crystals, space group $P4_12_12$ (or $P4_32_12$), $a = b = 498$, $c = 198$ Å (Volkmann *et al.*, 1990). The different parameters (number of atoms, number of trials, *etc.*) had been previously optimized in tests with simulated ribosome data (Lunin *et al.*, 1995). The observed amplitudes were used for $d > 60$ Å. The space group used in the calculations was $P4_12_12$; the results are also valid for $P4_32_12$ by changing z into $-z$. One million models of 5-Gaussian spheres ($B = 34000$ Å^2) were generated in step (*a*) of which 560 variants with amplitude correlation (1) higher than 86% were chosen in step (*b*).

The cluster tree showed clearly separated subsets (Fig. 2*a*). A path to the solution was established up to the fifth level through the clusters numbered 1–5 in Fig 2(*a*). The corresponding synthesis are shown in the right column of Fig. 2(*b*). The left-hand side of Fig. 2(*b*) shows the alternative synthesis at each level (except for level one, where the alternative cluster was too small). The criteria

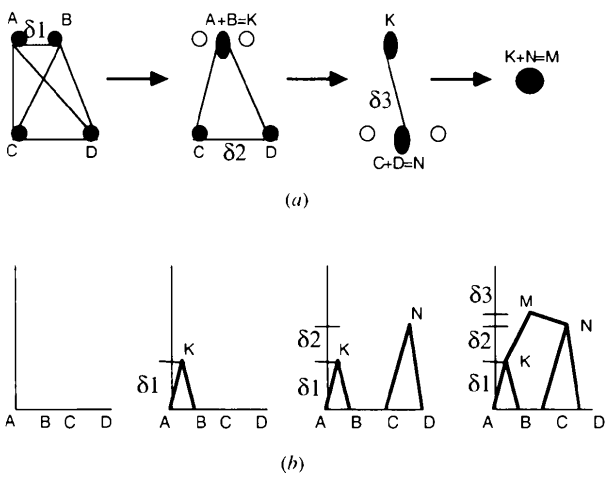
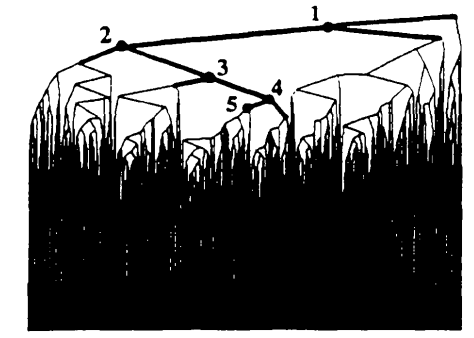
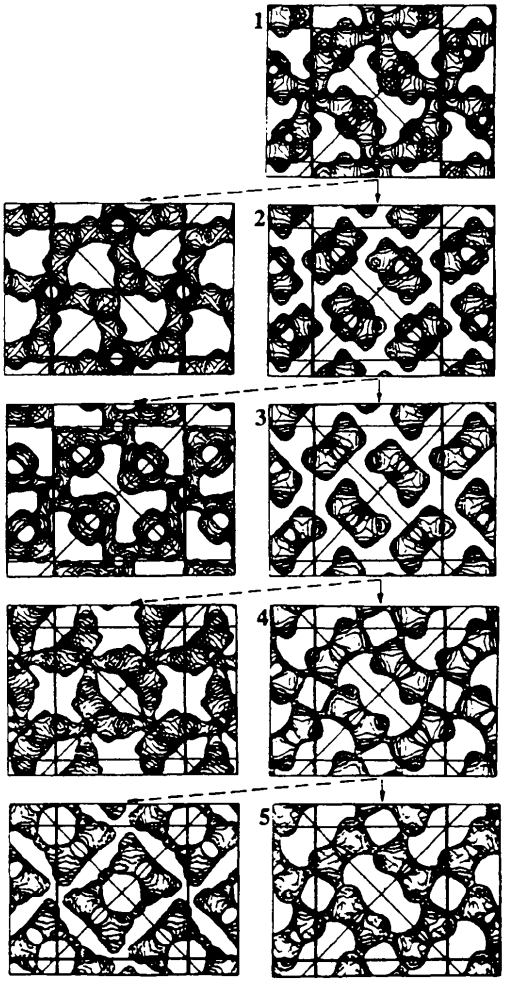


Fig. 1. Scheme of the clustering process. (*a*) The scheme of merging four points in solution space, *A*, *B*, *C* and *D*, as a function of the parameter δ, is presented. When $\delta = \delta 1$, the distance between *A* and *B*, these points are merged into a new cluster, *K*. When $\delta = \delta 2$, the distance between *C* and *D*, these points are merged into a new cluster, *N*. When $\delta = \delta 3$, the distance between *K* and *N*, these points are merged into a new cluster, *M*. (*b*) Graphical representation of this process in a 'tree', in which the *X* axis shows the solution number, ordered by proximity to avoid crossing of lines, and the *Y* axis shows the distance d. A new point, for example *K*, has as *Y* coordinate the distance $\delta 1$, and an *X* coordinate between *A* and *B*. To illustrate the merging, lines are drawn from the original points to the new one. Once full clusterization is achieved, the tree can be analyzed by levels. The zero level corresponds to the overall average (point *M*). The first level corresponds to the splitting of this point in its components (points *N* and *K*). The next levels are defined similarly.



(*a*)



(*b*)

Fig. 2. Cluster tree, the intermediate maps and the final map of T50S phased with FAM. The cluster tree is shown in (*a*). The bottom points ($\lambda = 0$) merge into larger clusters, the top point ($\lambda \simeq 1.2$) corresponds to the average of all models. The intermediate maps (*b*, *c*-axis projection) are calculated from cluster phases sets, starting from the top point and going down. At each step of decreasing λ, the more reliable map (shown in the right, and numbered following the points in the cluster tree) is chosen on the basis of less density on top of the crystallographic dyads (diagonals of the unit-cell face $c = 0$). The program *CAN* (Vernoslova & Lunin, 1993) was used for the map drawing.

for choosing the correct cluster at each level was that the density on crystallographic dyads should be weak for the correct synthesis. This criteria was chosen since the particles do not form dimers and should not have extensive interfaces. Note that this criteria is quite clear for levels 1–4; it becomes less so at level 5, since most of the models with density on top of the dyads were already ruled out. Therefore, the choice was no longer clear and the process was finished at this level.

Averaging phase sets inside the cluster according to (4) weighs down the higher resolution reflections because the phases are more dispersed; the mean weight decreases as a function of resolution. For example, a threshold in weight of 0.3 leads to an effective resolution around 110 Å. This limit could not be improved just by changing the parameters (*e.g.*, the number of pseudo-atoms). The strongest reflections (which are the ones of lowest resolution) dominated the correlation of the observed and the calculated amplitudes at step (*b*), leading in all cases to similar results.

To solve this problem and to increase the resolution of the image, the FAM method was developed further and re-run. To increase the sensitivity to weaker reflections, the uncentered amplitude correlations (1) were replaced by the more commonly used centered ones,

$$\mathrm{Corr}(F^{\mathrm{obs}},F^{\mathrm{calc}}) = \{\sum_s [F^{\mathrm{obs}}(s) - \langle F^{\mathrm{obs}}\rangle][F^{\mathrm{calc}}(s) - \langle F^{\mathrm{calc}}\rangle]\}$$
$$\div (\{\sum_s [F^{\mathrm{obs}}(s) - \langle F^{\mathrm{obs}}\rangle]^2\}$$
$$\times \{\sum_s [F^{\mathrm{calc}}(s) - \langle F^{\mathrm{calc}}\rangle]^2\})^{1/2}, \qquad (5)$$

calculated around the mean values $\langle F^{\mathrm{obs}}\rangle$ and $\langle F^{\mathrm{calc}}\rangle$ of $\{F^{\mathrm{obs}}(s)\}$ and $\{F^{\mathrm{calc}}(s)\}$, respectively. Three million ten-atoms models with $B = 10\,000$ Å^2 were generated (step *a*) inside a sphere of 110 Å radius centered in the middle of the envelope obtained previously thus covering it completely. In step (*b*) different filtering criteria were applied simultaneously in different resolution ranges, as follows:

(1) 60 Å $< d <$ 110 Å (69 reflections): centered amplitude correlation (5) larger than 0.45 (corresponding mean value in the original run was 0.23 for the selected models).

(2) 110 Å $< d$ (20 reflections): map correlation (3) larger than 0.7 with the map obtained previously (Fig. 3*a*).

These filters selected 950 variants and a cluster tree was built. The tree showed several major clusters, but only one corresponding map had no noise density on the crystallographic dyad ($x = y$, $z = 0$) and improved envelope definition (Fig. 3*b*). The effective resolution for this synthesis is about 80 Å (mean weight > 0.3), thus extending the phase information from the earlier limit of 110 Å. Note that the corresponding envelope (Fig. 3*c*) is not spherical and its peaks and valleys

resemble the shape of the ribosome particle as seen by electron microscopy (Yonath *et al.*, 1987). The results presented in Fig. 3 are shown in the space group $P4_32_12$ for easier comparison with the results of the molecular-replacement searches.

3. Molecular replacement with a particle envelope

3.1. *Molecular replacement at very low resolution*

Performing molecular replacement (Rossmann, 1990, and references therein) at very low resolution in the present case poses two particular problems.

(1) The very low resolution reflections have a very strong contribution from the disordered solvent.

(2) The available searching object is not a molecular model but an envelope, *e.g.*, from electron-microscopy reconstruction.

Urzhumtsev & Podjarny (1995*b*) investigated these problems using model and experimental data. They showed that at resolutions lower than approximately 20 Å the contribution of the disordered solvent to the observed structure factors becomes proportional to the contribution of the molecule, and that a simple envelope model can take the solvent contribution into account. Therefore, at this resolution range, molecular-replacement searches can be carried out with either atomic models or molecular envelopes. In this work a protocol for low-resolution molecular replacement was developed and tested (Urzhumtsev & Podjarny, 1995*a*). At this low resolution the signal peak can be both broad and displaced from the correct one. Therefore, the peaks of the rotation function are supplemented by their closest neighbours with a 15–20° scanning step. Also, all points of translation searches with a correlation value above a certain level (and not only the positions of the peaks) are kept for the final analysis by rigid-body refinement.

It should be noted that the problems outlined above are not faced during usual molecular-replacement ap-

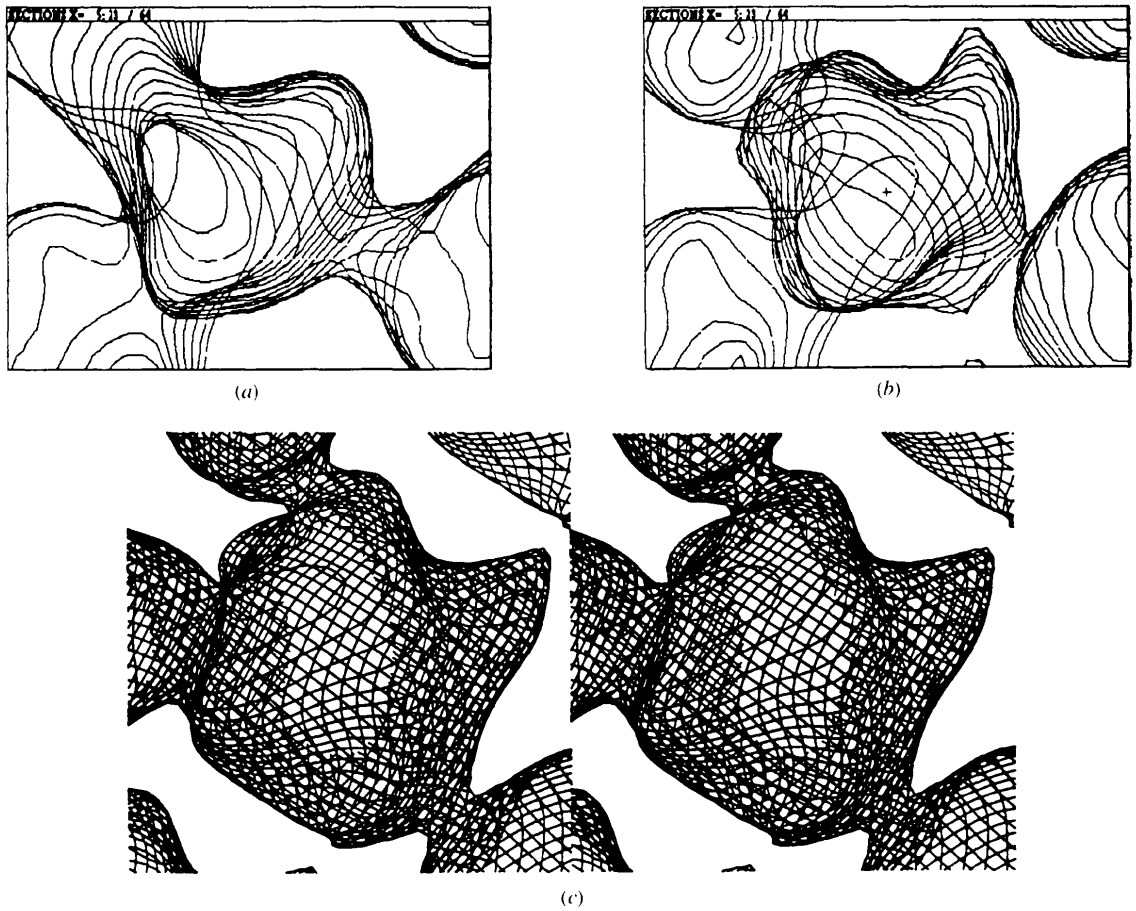


Fig. 3. FAM maps for the T50S particle (*a*) Electron-density map (*a*-axis projection) after the first FAM run showing a single particle at an effective resolution of 110 Å; (*b*) same view after the second FAM run, showing a single particle at an effective resolution of 80 Å. Note the increased details and improved particle separation; (*c*) stereoview (in the space group $P4_32_12$) of FAM density after the second run, showing the topological features of the T50S particle. The density level is chosen to include 30% of the unit-cell volume inside the envelope. The programs *CAN* (Vernoslova & Lunin, 1993) and *O* (Jones, Zou, Cowan & Kjeldgaard, 1991) were used for graphical presentation of the results.

Table 1. *Peaks of the low-resolution molecular replacement procedure (after the refinement step) with correlation value higher than 3σ (see text) for at least one of the resolution ranges studied*

The list is shown in the order of the corresponding rotation peak positions, ordered in decreasing height: peaks up to the 11th position (height 16.5) were studied. The intermolecular distance is measured between the centers of mass of neighbouring molecules. Larger values are preferred; a value lower than 100 Å implies strong overlap. Peaks *A*, *C*, *D* and *E* are close to each other through symmetry or enantiomer operations. Peak *C* is the best: it has the strongest value in both resolution ranges, an acceptable intermolecular distance, and the second highest value of the rotation function.

Peak name	Height (in σ) at 90–∞ Å	Height (in σ) at 90–150 Å	Space group	Intermolecular distance (Å)	Rotation peak Position	Rotation peak Height
A	3.0	2.5	$P4_12_12$	132	1	41.8
B	2.9	3.1	$P4_12_12$	76	1	41.8
C	3.7	3.3	$P4_32_12$	142	2	39.1
D	3.5	3.0	$P4_12_12$	120	5	37.0
E	3.1	2.9	$P4_12_12$	134	5	37.0
F	3.0	3.0	$P4_12_12$	132	6	35.7
G	2.9	3.0	$P4_12_12$	63	7	31.6

plications as the very low resolution data are seldom measured, however some of proposed measures can be possible useful for difficult cases at usual resolution.

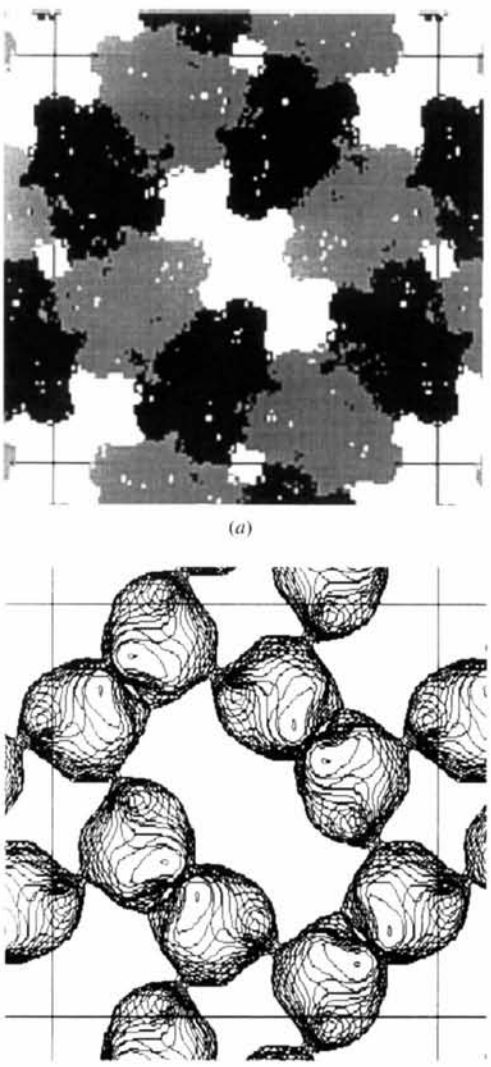
3.2. *Molecular replacement search using experimental ribosome diffraction data*

The low-resolution molecular-replacement protocol was applied to the T50S diffraction data (Volkmann *et al.*, 1990). The searching object was an image (envelope) of this particle (electron-microscopy envelope) as reconstructed at 28 Å from the electron-microscopy views of a tilt series of the crystalline arrays (Yonath *et al.*, 1987). The program *AMoRe* (Navaza, 1994) was used for the rotation function, the translation search was carried out with an amplitude correlation function developed independently, and the rigid-body refinement was carried out using *AMoRe* again. The result of the rigid-body refinement was confirmed by a complete six-dimensional search in a large 'box' around the indicated orientation and position, which also generated mean values and dispersions (σ). This work followed the protocols developed with simulated ribosome data (Urzhumtsev & Podjarny, 1995*a*).

In order to minimize the possible discrepancies between the electron-microscopy envelope and the crystallographic image, the molecular-replacement searches were carried out at the lowest possible resolution where the rotation function could yet show a signal. Such resolution was estimated as 90 Å. In contrast to FAM, in the molecular-replacement method the model fixes the enantiomer, and therefore both space groups need to be explored. The results of these searches are shown in Table 1; there are clear indications for the choice of peak *C* using the criteria of amplitude correlation and intermolecular distance. The resulting packing is very similar (Fig. 4) to the one obtained by the FAM method, where no assumptions were made either about the number of molecules or their shape.

4. Conclusions

The FAM method, starting on the amplitudes only, produced a clear image of the particle with identifiable topological features (Fig. 3*c*). Molecular-replacement searches indicated a preferred position and orientation for the electron-microscopy envelope (including the



(*a*)

(*b*)

Fig. 4. Packing of the T50S particle (*a*) *c*-axis projection of the symmetry related particle envelopes, placed by molecular replacement, showing their packing in a full unit cell; (*b*) corresponding projection of the FAM map after the second run. The figures show that the two approaches give a similar packing. Note the lower resolution of the current FAM map in comparison with the available electron-microscopy model.

choice of the space group). Several other methods (without heavy-atom derivatives) have been applied to the same data, and the comparison of their results and those presented in this paper is described in Volkmann *et al.* (1995). The results obtained by these various approaches are very similar. For example, the map correlation between the FAM-phased map and the one calculated with the likelihood aided positioning and shaping (LAPS, Volkmann, personal communication) approach is 98%. These techniques are currently being applied to other ribosome particles.

The data of T50S (X-ray diffraction amplitudes and electron-microscopy diffraction envelope) were provided by Dr A. Yonath and co-workers as a part of an ongoing collaboration; we thank them for making these data available. The FAM algorithms were programmed by Dr V. Lunin and co-workers, also as part of an ongoing collaboration; we thank them for making these programs available. The low-resolution molecular-replacement algorithms were developed in collaboration with J. Navaza. We also thank Drs Z. Berkovitch-Yellin, P. Dumas, D. Moras, B. Rees, M. Roth, F. Schülnzen and N. Volkmann for useful discussions and their continuing interest and encouragement, and Dr M. Zwick for his careful reading of the manuscript and useful suggestions. We thank Mr A. Litt, Mlle C. Mayer, Mr L. Moulinier and Dr R. Ripp for computer assistance. The collaboration with Dr V. Lunin is funded by the CNRS and the Russian Academy of Sciences. The stay of EAV in Strasbourg was funded by the Ministere de l'Enseignement Superieur et de la Recherche. This work is supported by the CNRS through the UPR 9004, by the Institut National de la Santé et de la Recherche Médicale and the Centre Hospitalier Universitaire Régional. The criticism of the referees was very useful for improving the presentation of the material.

References

Berkovitch-Yellin, Z., Bennett, W. S. & Yonath, A. (1992). *Crit. Rev. Biochem. Mol. Biol.* **27**, 403–404.

Berkovitch-Yellin, Z., Hansen, H. A. S., Weinstein, S., Eisenstein, M., von Böhlen, K., Agmon, I., Evers, U., Thygesen, J., Volkmann, N., Bartels, H., Schlnzen, F., Zaytzev-Bashan, A., Saron, R., Levine, I., Dribin, A., Kriger, G., Bennett, W. S., Franceschi, F. & Yonath, A. (1994). *Synchrotron Radiation in Biosciences*, edited by B. Goodhead, J. H. Helliwell, H. E. Huxley & B. Chance, pp. 61–69. Oxford: Clarendon Press

von Böhlen, K., Makowski, I., Hansen, H. A. S., Bartels, H., Berkovitch-Yellin, Z., Zaytzev-Bashan, A., Meyer, S., Paulke, C., Franceschi, F. & Yonath, A. (1991). *J. Mol. Biol.* **222**, 11–15

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

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

Lunin, V. Yu., Urzhumtsev, A. G. & Skovoroda, T. P. (1990). *Acta Cryst.* A**46**, 540–544.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.

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

Urzhumtsev, A. G. & Podjarny, A. D. (1995*b*). *Jnt CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **32**, 12–16.

Vernoslova, E. A. & Lunin, V. Yu. (1993). *J. Appl. Cryst.* **26**, 291–294.

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.

Volkmann, N., Schlünzen, N., Urzhumtsev, A. G., Vernoslova, E. A., Podjarny, A. D., Roth, M., Pebay-Peyroula, E., Berkovitch-Yellin, Z., Zaytsev-Bashan, A. & Yonath, A. (1995). *Jnt CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **32**, 23–31.

Yonath, A., Leonard, K. R. & Wittmann, H. G. (1987). *Science*, **236**, 813–816