On the solution of the molecular replacement problem at very low resolution

By A.G.Urzhumtsev*, A.D.Podjarny,

UPR de Biologie Structurale, 15, rue Descartes, 67084 Strasbourg, France

and J.Navaza

Laboratoire de Physique, Faculté de Pharmacie, 92290 Chatenay Malabry, France

1. Introduction.

The molecular replacement method (MR) has become a widespread and powerful technique for solving the phase problem when a structure closely related to the one under study (or a fragment of it) is available. This method is generally applied to diffraction data of 6 Å or higher resolution. A similar problem appears when an image of the structure under study is known at a much lower resolution, e.g., when an electron microscopy image of 30 to 50 Å is known, or when the search structure is not accurate enough and can provide only the molecular envelope.

The correct placement of such an image in the crystallographic unit cell is useful in several ways. For the case of a totally unknown structure, it provides a starting phase set for phase extension procedures, as well as the information (e.g., the molecular envelope, placement of NCS related units) necessary for some of these procedures. For the case when the model of an homologous structure is known, but it is significantly different, the placement of a low resolution envelope might provide the right solution when standard molecular replacement has failed.

The cubic form of the $tRNA^{Asp}$-synthetase complex provides an ideal case where the possibility of correctly placing an envelope can be tested, since:

a) a high resolution structure of the same complex is available (Ruff et al., 1991), and therefore a search can be done with either model or envelope at different resolutions;
b) due to the high solvent content (Lorber et al., 1983) the envelope is well defined inside the unit cell;
c) neutron data is available with excellent very low resolution completeness (Moras et al., 1983).

The work described below shows that indeed, it is possible to use very low resolution data to position an envelope. However, in order to do so current molecular replacement techniques need to be improved, since their direct application may fail to solve the problem. These improvements are described below.

* Permanent position : Institute of Mathematical Problem of Biology, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

2. Molecular replacement strategy.

The methods herein developed are based on the MR package AMoRe (Navaza, 1994). The AMoRe package consists of the same steps of rotation function, translation function and rigid body refinement that conventional MR methods. Its main advantages are :

- it samples by using fast algorithms a much larger portion of six-dimensional space than conventional MR methods;
- there is a high degree of automation; for example, when many molecules need to be found within the asymmetric unit, the information coming from already located models is automatically incorporated into the procedure;
- several search criteria are used simultaneously;
- to improve speed, the main programs of the package do not use atomic coordinates but Fourier coefficients;
- the search models may be approximate electron densities such as molecular envelopes or images coming from electron microscopy.

The domain of potential orientations is sampled by the fast rotation function, which has several improvements over currently used ones, e.g. MERLOT (Fitzgerald, 1988) or XPLOR (Brünger, Kuriyan & Karplus, 1987). In particular, it is casted in the form of a correlation coefficient between truncated Patterson functions and the angular resolution of peaks is enhanced by skipping low angular resolution terms; furthermore, there is no limitation on the value of the integration radius, for a given data resolution.

A large number of retained orientations are then used to compute translation functions. First a selection of peaks is made by computing the centred overlap of observed and calculated Patterson functions:

$$\text{Overlap}= \Sigma_h [I_{obs}(h)-\langle I_{obs}\rangle] \times [I_{calc}(h; \mathbf{R,t})-\langle I_{calc}\rangle(\mathbf{R,t})] \qquad (1)$$

which is computed from the amplitudes, both observed, Iobs(h), and calculated, Icalc(h; **R,t**), where **R** represents the three rotation parameters and **t** represents the three translation parameters. Another option of the package is the full-symmetry phased-translation function (Colman et al., 1976; Bentley & Houdusse, 1992), which involves essentially amplitudes instead of intensities.

The output is however the full correlation function,

$$\text{Corr}(F)= \frac{\Sigma_h [F_{obs}(h)-\langle F_{obs}\rangle] \times [F_{calc}(h; \mathbf{R,t})-\langle F_{calc}\rangle(\mathbf{R,t})]}{\{\Sigma_h [F_{obs}(h)-\langle F_{obs}\rangle]^2\}^{1/2} \times \{\Sigma_h[F_{calc}(h; \mathbf{R,t})-\langle F_{calc}\rangle(\mathbf{R,t})]^2\}^{1/2}} \qquad (2)$$

defined in terms of amplitudes, $F_{obs}(h)$ and $F_{calc}(h; \mathbf{R,t})$, and calculated for the top peaks of the overlap function. The final list, sorted in descending order of correlation, includes also the R-factor.

It has been observed that in general high overlaps correspond to high correlations, so that only a rather small number of peaks are retained for each orientation. However this is not a general rule, and sometimes the greatest correlation may correspond to a peak which has a very low rank in overlap. This is particularly true for the present studies (see section 3.3).

Eventually the **R, t** parameters are refined by a fast rigid body refinement procedure.

In the case where many molecules or fragments have to be located within the asymmetric unit, the contribution of the best n-body solution is used to compute (n+1)-body translation functions.

AMoRe has proven very effective in solving the MR problem at high and medium resolution ranges, and therefore its parameters are tuned accordingly. Since the nature of the signal varies when lowering the resolution range, the nature of the function used to detect it will

vary also. The purpose of this paper is to develop the necessary changes to make AMoRe work at very low resolution ranges.

3. Low resolution MR. Tests with an atomic model.

3.1) Initial check of standard AMoRe protocol.

To develop the low resolution protocols, the neutron data from the cubic form of the $tRNA^{Asp}$-synthetase complex was used (Moras et al., 1983). This structure was solved by AMoRe using the standard protocol, a high resolution model from a different crystal form (Ruff et al., 1991) and X-ray diffraction data at 8 Å resolution (see appendix 1). The solution showed very clearly one single dimer in the asymmetric unit, placed in a general position.

During the initial stages of the solution of this structure (Podjarny et al., 1987), neutron data had been collected in the resolution range ∞-20 Å for three different D_2O/H_2O contrasts, corresponding to the diffraction of the tRNA, the synthetase and the full complex (Moras et al., 1983), respectively. This data is very complete to 26 Å; moreover, the number of reflections (31) at 50 Å resolution is enough to have a significant observations/parameters ratio in a six-dimensional search

To test the power of the standard protocol of AMoRe at low resolution, several searches with the complex dimer were conducted using the neutron diffraction data from the full complex from 20 Å to infinity (although the high resolution end is not complete). These searches gave no solution when using the centred-overlap translation function, even when all positive peaks were considered. Although for the best positions the value of Corr(F) after refinement was high (max = 81%), the R-factor was systematically too high (min = 47%) and no position was close to the correct one. Moreover, the peaks of the translation function correspond to positions where the dimer is placed on a rotation axis, and even on the origin of the cubic space group. The reason for this failure is not the quality of the neutron data, since a comparison of the experimental neutron diffraction data and the model data from correct solution gives a Corr(F) of 92% and an R-factor of 28.3%. This check shows that very low resolution data can be accurately calculated from a model, which is therefore a proper search object. Therefore, the problem is that in this case the use of the centred Patterson overlap as the translation function missed the global maxima of the correlation function between F_{obs} and $F_{calc}(h; \mathbf{R,t})$, which certainly would be present in a full six-dimensional search.

It should be noted that the standard protocol did give the correct solution when using the full-symmetry phased-translation function, but the peak for the correct position (which has the largest full correlation factor, see formula (2)) appeared beyond the 50th position of this phased-translation function. Therefore, this procedure could also miss the right solution if only the top peaks of the translation function were investigated, as is usually the case .

3.2) Tests with calculated diffraction data.

In order to find the right protocol for searches at low resolution, diffraction data calculated from the atomic model without solvent modelisation was used at various resolution ranges. The use of model data is needed in the first stages to avoid the effect of the error in the amplitudes or of their lack of completeness.

Since the overall resolution range is considerably wider than the one available from the experimental data, several intermediate resolution ranges could be tested (d_{min} varied from 8 to 30 Å, d_{max} varied from 15 Å to infinity). The original AMoRe protocol works for almost all cases. For example, in the 15-30 Å resolution range, the rotation error is 6° for the first peak (height = 18.6) and 25° for the second peak (height = 14.6), and after translation searches and refinement the correlation in F is 96% for the first peak, which is correct, and 48% for the second peak, which is wrong.

However, there are two exceptions:

a) When the higher resolution limit is lower than 30 Å, the rotation function has a large error (about 20°), which causes overall failure. This poses a major problem for the use of envelopes as a search object.

b) When the "inner core" of reflections (d > 50 Å) are included, the translation function based on Patterson overlap fails even for cases were the rotation function has worked nicely (5° error). This arises from the fact that the Patterson overlap is not properly normalised and is dominated by a few ultra-low resolution very strong terms.

Unfortunately, problem (b) precludes the use of the "central reciprocal space zone" terms to diminish the sensitivity of the translation searches to errors in the rotation parameters, which causes problem (a)

The solution of these problems could not be done by simple adjustments of the existing protocol but needed the development of a new one, based on correlation functions instead of Patterson overlap.

3.3) Development of a new protocol. Model data.

To solve the problem highlighted above, search programs (Urzhumtsev & Podjarny, 1994) which use directly as the target value the correlation of amplitudes (F or F^2, see formula (2)) were used instead of the Patterson overlap. The difference of these two functions is specially evident when the values of Fcalc are strongly dependent of the rotation and translation parameters. This is the case for very low resolution reflections, where the value of Fcalc will increase considerably when the molecules overlap with their symmetry related images. Since the Patterson overlap has Fcalc only in the numerator (1), its value will increase artificially when molecules overlap. For the correlation, where Fcalc appears both in the numerator and the denominator (2), this effect is cancelled.

This strategy has solved the cases which had previously failed under condition that the translation searches were done at the resolution of 50Å-∞. All cases confirm the previous observation that the translation searches using the lowest resolution ranges and based on correlation functions show the signal clearly and correct large rotation errors, if necessary. For example, for the resolution range 50Å-∞, this procedure corrected an initial translation error of 20° (second peak of the rotation function) to end after refinement with a correlation of 99%, a rotation error of 2° and a translation error of 1 Å. The correlation in F worked slightly better than the correlation in F^2.

Further tests showed that the rotation error limit for correct translation searches lies around 20°. However, the rotation function error can be as large as 30°, particularly when the data used goes only to 40 or 50 Å. To assure that the error in the rotation parameters is small enough to fall within the 20° limit, a scanning with a step of ±20° around the rotation peaks was introduced. The peak with its neighbourhood of ±20° is called the "expanded rotation peak".

This protocol was also applied to a model case at 50 Å resolution, where the Patterson radius was displaced on purpose from the optimal value. This increased the rotation error from 20° to 27°, but even this larger value can be corrected by using the "expanded rotation peak" as input to the translation function and the refining with the "inner core" reflections.

3.4) Tests with experimental neutron diffraction data.

In order to test the protocol using real diffraction data, the same procedure was applied for the experimental neutron data at very low resolution. It converged to the correct solution in all cases. Note that, as before, the limitation of data at the high resolution end causes a large error (27°) in the rotation function, but the translation search and rigid body refinement give the correct solution and correct the rotation error. For the final solution, the Corr(F) was 97%, the rotation error was 2° and the translation error was 1Å.

In the case of the effect of changing some of the parameters (e.g., Patterson radius), the rotation function finds a position with larger error (28°), enough to cause failure of the translation function. Therefore the expansion of rotation peaks is necessary, and it finds the right solution.

3.5) Summary of searches with an atomic model.

The low resolution searches described above show that different resolution ranges should be used for different types of searches. For rotation searches the highest available resolution is preferable, and very low resolution terms can be excluded. Translation searches should be done using correlation searches against very low resolution data. The neighbourhood of rotation peaks should be explored. Rigid body refinements should be done against all available data. However, using lowest resolution data may be important for correcting large errors, even when it may also lower the signal/noise contrast (for example, the ratio of the correlation for the correct and the first false peaks are 99%/87% at the resolution 30Å-∞ against 98%/63% for 30-40Å).

4. Tests with envelopes.

The object available for conducting low resolution searches might come in the form of an envelope (and not of a detailed atomic model), for example from electron microscopy or from very approximate models. Therefore it is important to check whether the protocol described above can be used to correctly place a molecular envelope inside the unit cell.

Preliminary tests on structures solved at high resolution (Urzhumtsev & Podjarny, 1993) have shown that at a resolution lower than 15 Å there is good correspondence between the observed diffraction amplitudes and the ones calculated from envelopes. This limit is necessary since at higher resolution (10 to 15 Å) the solvent region cannot be considered flat, and therefore the diffraction amplitudes have a significant solvent contribution.

As before, the studies with envelopes were conducted using the neutron data from the cubic form of the $tRNA^{Asp}$-synthetase complex.

4.1) Exact flat envelope, 50Å neutron data.

To obtain an exact envelope, each atom of the model was surrounded by a ball of 2.5Å radius. For the purpose of using AMoRe, the resulting points were placed in a box 6 times larger than the model in each linear direction, where it occupied 0.0012 of the volume, and assigned a density value of 1; the rest of the points are set to 0. This density was then used to calculate structure factors**.** The R-factor at 50Å between modules from the atomic model and from this flat envelope was equal to 4%. This shows that at this low resolution range, the transform of an atomic model is virtually the same as that of the corresponding exact envelope. The protocol optimised using the searches with a model was used, and the results were, as expected, virtually the same as those obtained using an atomic model.

4.2) Density-based envelope, 50Å neutron data.

To obtain a more realistic envelope, the model structure factors were used to calculate a density distribution at 50Å, which was placed in the same box as before. The envelope is defined

by all the points above a cutoff level. In this synthesis, the envelope can be expanded to 0.0035 of the volume by lowering the cutoff level before it includes noise peaks.

Two possibilities were considered: a flat envelope or a modulated one, where the density values inside the molecular envelope are kept. For both cases, two different cutoff levels were considered, corresponding to the exact molecular volume and to three times the molecular volume. It was found that for the flat envelope the exact molecular volume fits the data better (R-factor is 15% against 39%), while for the modulated envelope the larger molecular volume is better (31% against 12%).

Searches were done using both the flat and the modulated envelopes at their optimal volumes. The solution was clearly found for both cases. Again, the orientation errors are large (17° and 18°, respectively) but they are within the rotation peak expansion.

4.3) Modulated envelope; 20Å neutron data.

These tests were repeated at higher resolution. The model structure factors were used to calculate a density distribution at 20Å. In this synthesis, the envelope can be expanded to 0.0020 of the volume by lowering the cutoff level before it includes noise peaks. Like in the 50 Å case, for the flat envelope the exact molecular volume fits the data better (R-factor is 14% against 37%), while for the modulated envelope the larger molecular volume is better (29% against 11%). Searches were conducted using the modulated envelope only, and the correct solution is found. It should be noted that the signal contrast here is higher than in the 50Å-resolution searches (the ratio of the correlation of the correct peak to the first false one is 84%/81% at 50Å and 96%/84% at 20Å).

4.4) Protocol for low resolution searches with envelopes.

>From this work the following rules can be derived:

1) To get a search model, a synthesis should be calculated, then cut at the "noise level" (1-3 times larger than the atomic model volume);

2) The same procedure of MR which was used for atomic models at very low resolution can be applied also for envelopes.

3) Both modulated and flat envelopes work; however, modulated envelopes are less dependent on an exact definition of the cutoff level.

5. CONCLUSIONS.

This study highlights the importance of the use of correlation functions directly as the search criterion. It also shows that, in the very low resolution case, the optimum resolution ranges are not the same for the rotation function, the translation function and the refinement steps. The rotation function works best using higher resolution terms; but, for experimental data the errors in rotational parameters are large. The translation function and refinement based on the lowest resolution data only are capable of correcting this error. In some cases, the rotation error is too large and the rotation peak needs to be expanded to fall within the radius of convergence of the translation and refinement procedures. The low resolution case is therefore a particular one, and needs the change of the standard protocol according to these rules. With these changes, the AMoRe package can be used successfully to solve the molecular replacement problem at very low resolution.

The experience gained in the atomic model searches has been applied to searches with molecular envelopes. These searches are very important since they open a whole new field, that

of using experimentally determined envelopes as the search model in the MR procedure. We have been able to solve this problem for the case of the cubic tRNA^{Asp}-synthetase complex. The behaviour of the envelope searches is similar to that of searches using the atomic models. Thus, this work has succeeded in finding a molecular replacement protocol that can use envelopes as the search object at very low resolution.

These experience is currently being applied to the case of the ribosome, both for simulated ribosome crystals and experimental X-ray diffraction data. The results are so far encouraging, and will be published elsewhere.

The experimental data of the tRNA^{Asp}-synthetase complex used in this work were collected by Drs. J.Cavarelli, J.-C.Thierry and D.Moras (X-ray data) and by Drs. M.Roth, A.Lewitt-Bentley and D.Moras (contrast variation neutron data). We thank them for making these data available to us. We also thank Dr. B.Rees for useful discussions and making available all his previous work on the different aspects of the crystallography of the cubic form of the tRNA^{Asp}-synthetase complex and Dr. P.Dumas for his continuing interest and encouragement on the current work. This work is supported by the CNRS through the budget of the UPR 9004, and by EMBL through a fellowship to A.G.U. .

Appendix I

Solution of the cubic form of the tRNA^{Asp}-synthetase - tRNA^{Asp} complex using X-ray data.

The cubic form of the tRNA^{Asp}-synthetase - tRNA^{Asp} complex crystallises in space group I432 (a=354 Å), with one dimer of the complex per asymmetric unit (Giegé et al., 1980; Lorber et al., 1983). It diffracts to a resolution of 8Å. X-ray data from native crystals and one Hg derivative were collected. Chemically, this derivative should bind the six reactive cysteines in the dimer. A first attempt was done to solve the phase problem ab-initio using translation searches with a single huge sphere followed by a phase extension procedure with multiple sphere envelopes and tRNA backbones (Podjarny et al., 1987). The single sphere search was successful, as well as the first steps of phase extension with few spheres (see accompanying paper by Lunin et al.), but the consequent phase extension procedure with many more spheres and the tRNA backbones diverged to a wrong envelope. The problem was finally solved with AMoRe using the 8-15Å resolution X-ray data and a high resolution model (3Å) of the complex dimer obtained for an orthorhombic crystal form (Ruff et al., 1991), which crystallised in essentially the same conditions and the cubic form (Ruff et al., 1988). The right solution appears as the first one in rotation function (the first peak has a height of 11.7 and a rotation error of 2° while for the second peak these numbers are 8.4 and 43°, respectively), in translation function (F-correlation for the right position is 51%, while it is equal to 21% for the next translation peak with correct orientation and 24% for the first peak when searched with a wrong orientation; corresponding R-factors are 35%, 47% and 48%, respectively), and specially after refinement: F-correlation and R-factor for the correct solution are equal to 61% and 31%, respectively while for the next peak they are 24% and 45%.

The solution was checked by the following criteria:

- the packing showed extensive contacts at the right distance, both tRNA-tRNA and tRNA-synthetase, even when no packing criteria were used during the solution with AMoRe;

- all packing interactions observed in the cubic form are essentially the same than those of the high resolution orthorhombic form, which has some additional ones (synthetase-synthetase) that cause the difference in space group;

- the dimer dyad has an orientation agreeing closely with that shown by the self rotation function (Podjarny et al., 1987);

- an 8Å-resolution difference synthesis with coefficients ($|F_{der}-F_{nat}|$, ϕ_{model}) showed clearly that the strongest peaks of the map, corresponding to the mercury positions, mark all the Sγ from the cysteines. The 6 Sγ's are marked, within a distance of 3 to 5 Å, by the first 5 significant peaks (with the relative height from 1000 to 681) and the 11th peak (the height is 394); as a comparison, the height of the 14th peak is 350. This is a completely independent and very reliable test of the solution.

References

Bentley,G.A. & Houdusse,A. (1992) *Acta Cryst.,* **A48**, 312-322

Brünger,A.T., Kuriyan,J. & Karplus,M. (1987) *Science,* **235**, 458-460

Colman,P.M., Fehlhammer,H. & Bartels,K. (1974) In *Crystallographic Computing Techniques,* edited by F.R.Ahmed, K.Huml & B.Sedlacek, 248-258. Copenhagen: Munksgaard

Fitzgerald,P.M.D. (1988) *J.Appl.Cryst.*, **21**, 273-278

Giegé,R., Lorber,B., Ebel,J.-P., Thierry,J.-C. & Moras, D. (1980) *Comp.Rend.Acad.Sci. (Paris), série D,* **291**, 393-396

Lorber,B., Giegé,R., Ebel,J.-P., Berthet,C., Thierry,J.-C., Moras,D. (1983) *J.Biol.Chem.*, **258**, 8429-8435

Moras,D., Lorber,B., Romby,P., Ebel,J.-P., Giegé,R., Lewit-Bentley,A., Roth,M. (1983) *J.Biomol.Structure & Dynamics,* **1**, 209-223

Navaza,J. (1994) *Acta Cryst.*, **A50**, 157-163

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

Ruff,M., Cavarelli,J., Mikol,V., Lorber,B., Mitschler,A., Giegé,R., Thierry,J.-C. & Moras,D. (1988) *J.Mol.Biol.*, **201**, 235-236

Ruff,M., Krishnaswamy,S., Boeglin,M., Poterszman,A., Mitschler,A., Podjarny,A., Rees,B., Thierry,J.-C. & Moras,D. (1991) *Science*, **252**, 1682-1689

Urzhumtsev,A.G. & Podjarny,A.D. (1993) *Réunion Francophone de Cristallographie de Macromolecules, Ste-Odile, 2-5 Juin 1993*, A1

Urzhumtsev,A.G. & Podjarny,A.D. (1994) *J.Appl.Cryst.*, **27**, 122-124