electronic reprint

Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

On the use of low-resolution data for translation search in molecular replacement

Andrei Fokine and Alexandre Urzhumtsev

Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 16 June 2001
Accepted 3 October 2001

On the use of low-resolution data for translation search in molecular replacement

Andrei Fokine and Alexandre Urzhumtsev*

LCM3B, UMR 7036 CNRS, Faculté des Sciences, Université Henry Poincaré, Nancy I, 54506 Vandoeuvre-lés-Nancy, France. Correspondence e-mail: sacha@lcm3b.uhp-nancy.fr

Low-resolution reflections (approximately 15 Å and lower) are very useful for the translation search in molecular replacement because they are less sensitive to model errors compared with the traditionally used reflections of resolution 4–10 Å. At low resolution, however, the contribution from the bulk solvent is quite significant and corresponding structure factors calculated from a macromolecular model cannot be compared with experimental values if this contribution is neglected. The proposed method provides a way of fast translation searches where low-resolution reflections are taken into account. Test calculations using several experimental data sets show a dramatic improvement in the signal after the bulk-solvent correction and low-resolution reflections were included in the calculation; this improvement allowed unambiguous identification of the solution.

1. Introduction

Molecular replacement (MR hereinafter; Rossmann, 1972) is one of the key methods for macromolecular structure determination which becomes even more important with the development of structural genomics projects. Unfortunately, the quality of MR models is sometimes quite poor and the solution cannot be found because it does not correspond to the optimum of usual search functions (for example, to the maximum of the correlation of experimental magnitudes of structure factors with corresponding values calculated from the model placed differently in the unit cell). As a consequence, even the simultaneous rotation and translation search does not help; on the contrary, a separate rotation and translation analysis can be useful. It has been shown recently (Urzhumtseva & Urzhumtsev, 2001) how the rotation problem can be solved even with poor models using multiple rotation functions. Another approach which facilitates the solution of the rotation problem has been reported by Read (2001). In the following, only the translation problem is addressed, supposing that the correct model orientation has already been found.

Working with low-resolution models, Urzhumtsev & Podjarny (1995*b*) showed that the sets of structure factors suitable for an easy resolution of the rotation and translation problems are different. In particular, it has been shown that the use of the lowest-resolution reflections, which are less sensitive to model imperfections (including the errors in orientation), helps greatly in the translation search. Standard MR protocols, however, do not use reflections of resolution lower than 10–15 Å because they are strongly influenced by the bulk solvent. A soft low-resolution cut-off by exponential weighting of these reflections in order to reduce the influence of low-resolution data has been suggested by Vagin *et al.* (1998). Some attempts to use and not to suppress low-resolution data have been announced by Glykos & Kokkinidis (2000), who implemented the exponential bulk-solvent correction (Moews & Kretsinger, 1975; Tronrud, 1997).

The tests described below show that indeed the inclusion of low-resolution reflections with the bulk-solvent correction following a more advanced method by Jiang & Brünger (1994) makes the translation search much more efficient. A fast way for such a search is suggested.

2. Methods and test calculations

It may be noted that the basic assumption of the exponential solvent model, namely the proportionality of the molecular and solvent structure factors, is not usually fulfilled at a resolution of about 10–20 Å (Urzhumtsev & Podjarny, 1995*a*) and that the flat mask approach (Jiang & Brünger, 1994) is of a superior quality (Kostrewa, 1997).

The principal steps of this latter procedure are as follows:

(*a*) the molecular envelope (a binary function) is determined from the atomic model;

(*b*) structure factors $\mathbf{F}_{\rm env}$ are calculated as the Fourier coefficients of this function;

(*c*) scale parameters $k_{\rm sol}$ and $B_{\rm sol}$ are estimated by minimizing the residue

$$\sum_{\mathbf{s}} \left\{ \left| \mathbf{F}_{\rm model}(\mathbf{s}) - k_{\rm sol} \exp(-B_{\rm sol}|\mathbf{s}|^2/4)\mathbf{F}_{\rm env}(\mathbf{s}) \right| - F_{\rm obs}(\mathbf{s}) \right\}^2;$$

(*d*) the complex numbers $\mathbf{F}_{\rm solv} = -k_{\rm sol} \exp(-B_{\rm sol}|\mathbf{s}|^2/4)\mathbf{F}_{\rm env}(\mathbf{s})$ calculated with the optimal parameter values are taken as the solvent structure factors.

Here, $\mathbf{F}_{\rm model}(\mathbf{s})$ is the structure factor calculated from an available atomic model and $F_{\rm obs}(\mathbf{s})$ is an experimental structure-factor magnitude for the reflection $\mathbf{s}$; for simplification, all values are supposed to be known on the absolute scale.

For the translation search, such a solvent correction can eventually be performed at every position of the search model (while the obtained $k_{\rm sol}$ and $B_{\rm sol}$ can be completely unreasonable for incorrect positions). However, such a method of solvent correction cannot be included in fast translation algorithms (Navaza, 1994; Navaza & Vernoslova, 1995), making its practical application inefficient. In

Table 1
Test structures: summary information.

Protein name (reference)	Crystal structure: PDB ID / NMR ID	Space group / Unit-cell parameters a, b, c (Å)	Percentage of solvent in unit cell
Human interleukin-4 (Müller *et al.*, 1995)	1hik / 1bcn	$P4_12_12$ / 92.1, 92.1, 46.4	63
P53 tetramerization domain (Mittl *et al.*, 1998)	1aie / 1pet	$P422$ / 45.5, 45.5, 32.2	53
Corn Hageman factor inhibitor (Behnke *et al.*, 1998)	1bea / 1bip	$P4_22_12$ / 57.12, 57.12, 80.24	49

order to realize a fast search, we suggest a technique based on the following observations:

(*a*) Our statistical analysis of the structures deposited in the Protein Data Bank (PDB) (Bernstein *et al.*, 1977) shows that a large majority of the models have a k_{sol} value between 0.30 and 0.40 e Å^{-3} and a B_{sol} value between 30 and 70 Å^2, and that some existing extreme values are rather unjustified (Fokine & Urzhumtsev, 2002); the much smaller dispersion of these values in comparison with the dispersion of the parameters for the exponential model (Glykos & Kokkinidis, 2000) is due to a more physical meaning of the parameters of the flat mask model; the use of the mean values of k_{sol} = 0.35 e Å^{-3} and B_{sol} = 50 Å^2 instead of the optimal parameters does not influence the value of the search function greatly.

(*b*) For the positions in the unit cell where the search model does not overlap with its symmetrically related images, the mask of the region occupied by all molecules can be calculated as a junction of masks of individual molecules related by symmetries; as a consequence, the structure factors of such a total molecular envelope can be rapidly recalculated from the structure factors of the envelope of a single model; if the standard mean values of the parameters k_{sol} and B_{sol} are used for all model positions, this allows one to calculate the solvent correction for an isolated model, add it to the model structure factors of this model and perform a search with the known FFT-based procedures (see, for example, Navaza & Vernoslova, 1995) using bulk-solvent-corrected structure factors instead of those calculated directly from the atomic model.

(*c*) Such an approach does not provide an adequate bulk-solvent correction for the positions where the models overlap, and spurious peaks in the translation function are eventually possible at such points; these spurious peaks, however, will be eliminated by the packing criterion and will not appear in the final list.

In order to construct a good molecular envelope, an atomic model is supposed to be more or less complete. Such a situation is usual when NMR models are used as templates [for a review of the MR searches with NMR models, see Chen *et al.* (2000); cases reported in this work as most difficult were chosen for our tests below (Table 1)].

All test calculations were performed using experimental data, and the orientation of the search models was supposed to be known (it may be noted that typical errors in model orientation practically did not influence the searches when low-resolution reflections were included; not shown in this paper). All translation searches were made using *CNS* (Brünger *et al.*, 1998) using the fast translation function (Navaza & Vernoslova, 1995). The translation search parameters were taken without any optimization; complete NMR models were taken as they are in the PDB; the *B* factors for all atoms of the search models were assigned to be equal to 20 Å^2 which is far from the optimal scheme (Chen *et al.*, 2000). In each test, a single NMR model was used for the translation.

3. Results and discussions

Fig. 1 shows the results of the translation searches performed with and without low-resolution data, with and without the bulk-solvent correction, using three experimental data sets. Each pair of diagrams shows the results of the translation search at a given resolution shell:

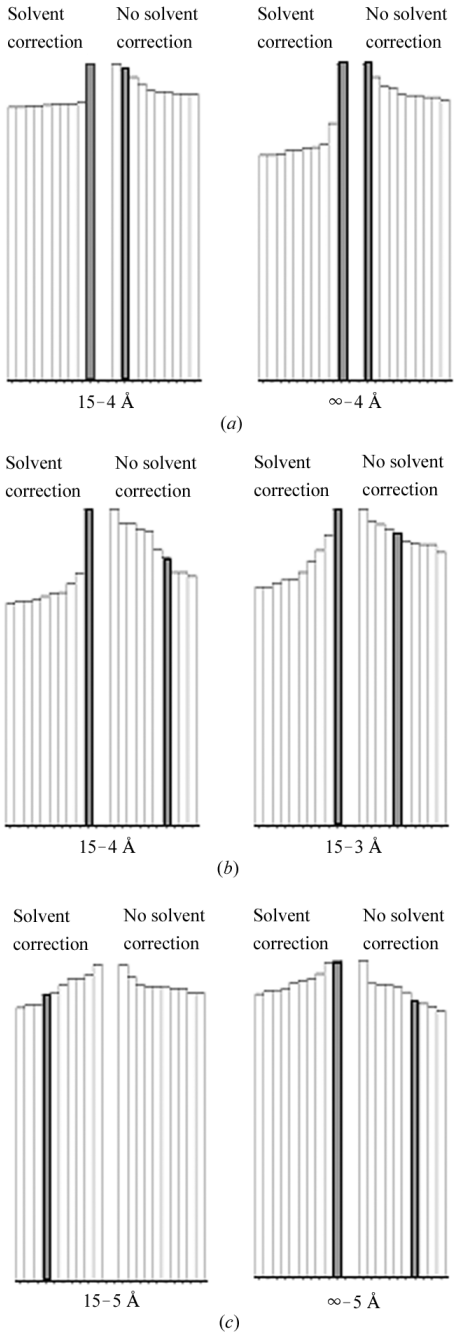

Figure 1
Relative heights of peaks of the translation function. Each bar represents a peak of the translation function and has a height as a percentage of the value of the first (highest) peak obtained at the same translation search. For each resolution range, ten highest peaks corresponding to searches without solvent correction are shown on the right and ten highest peaks corresponding to searches with such a correction are shown on the left. Peaks are positioned from the centre of the histogram to the outside in decreasing order. Peaks corresponding to the correct solution are shown in grey. The figures represent the results of the searches for (*a*) human interleukin-4 (Müller *et al.*, 1995), (*b*) p53 tetramerization domain (Mittl *et al.*, 1998) and (*c*) Corn Hageman factor inhibitor (Behnke *et al.*, 1998).

the right-hand diagrams show the peaks obtained in the translation search without bulk-solvent correction; the left-hand diagrams show the peaks obtained under the same conditions when the bulk-solvent correction was taken into account, as suggested above. The height of each peak is shown as a percentage of the height of the first peak of the corresponding search, and the correct solution is indicated in grey.

For human interleukin-4 (Müller *et al.*, 1995), the translation search performed at the standard resolution of 4–15 Å without solvent correction gave the solution as the second peak (Fig. 1*a*, left). When all available reflections with resolution lower than 15 Å were also included in the calculations, this brought the correct peak to the first position without significant contrast (Fig. 1*a*, right). The following bulk-solvent correction increased the contrast of the signal drastically; the best results were obtained when low-resolution data were included.

In the case of p53 tetramerization domain (Mittl *et al.*, 1998), no experimental data below 15 Å are available. Without solvent correction, the translation search at the standard 4–15 Å resolution gave the correct solution hidden in noise (Fig. 1*b*, left) and the search at 3–15 Å resolution gave it slightly higher in the list (Fig. 1*b*, right). With the solvent correction, the peak for the solution became the first with the best contrast at 4 Å when higher-resolution data were suppressed [see Urzhumtsev & Podjarny (1995*b*) for similar observations when searches are performed with molecular envelopes].

Corn Hageman factor inhibitor (Behnke *et al.*, 1998) was reported as the worst case among all NMR-based searches (Chen *et al.*, 2000). The orientation of the search model can be found very surely and precisely by a multiple rotation function (Urzhumtsev & Urzhumtseva, 2002). Without the bulk-solvent correction, the solution did not appear among the ten highest peaks, neither at 4–15 Å nor at 5–15 Å resolution (Fig. 1*c*, left), and appeared only as the seventh peak when all available magnitudes with resolution lower than 5 Å were used (Fig. 1*c*, right). At the same time, with the bulk-solvent correction, this peak became the first for a resolution lower than 5 Å while the contrast was not as high as for the two previous cases. Better contrast can be obtained by variation of the MR parameters, which is the subject of further analysis.

More complete results are out of the scope of this paper.

4. Conclusions

Low-resolution reflections are very useful for the translation search in molecular replacement. Because the solvent contribution to them is very significant, a corresponding correction is necessary in order to use these reflections properly. The flat model for the bulk-solvent correction, the best model currently available, can be easily included in the translation search. The use of low-resolution reflections, especially with the bulk-solvent correction, solves a number of translation problems which are difficult to solve by standard procedures.

The authors thank V. Lunin for useful discussions on the project and the manuscript preparation, C. Lecomte for his support of the work, L. Torlay for computer assistance and D. Teller for making the Corn Hageman factor inhibitor experimental data available. The work was performed in the frame of the 'CP Etat-Lorraine'.

References

Behnke, C. A., Yee, V. C., Le Trong, I., Pedersen, L. C., Stenkamp, R. E., Kim, S.-S., Reeck, G. R. & Teller, D. C. (1998). *Biochemistry*, **37**, 15277–15288.
Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Brünger, A., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chen, Y. W., Dodson, E. J. & Kleywegt, G. J. (2000). *Structure*, **8**, 213–220.
Fokine, A. & Urzhumtsev, A. (2002). *Acta Cryst.* D. Submitted.
Glykos, N. M. & Kokkinidis, M. (2000). *Acta Cryst.* D**56**, 1070–1072.
Jiang, J.-S. & Brünger, A. (1994). *J. Mol. Biol.* **91**, 201–228.
Kostrewa, D. (1997). *CCP4 Newslett. Protein Crystallogr.* **34**, 9–22.
Mittl, P., Chène, P. & Grütter, M. G. (1998). *Acta Cryst.* D**54**, 86–89.
Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.
Müller, T., Oehlenschläger, F. & Buehner, M. (1995). *J. Mol. Biol.* **247**, 360–372.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J. & Vernoslova, E. (1995). *Acta Cryst.* A**51**, 445–449.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Rossmann, M. G. (1972). *The Molecular Replacement Method.* New York/London/Paris: Gordon and Breach.
Tronrud, D. E. (1997). *Methods Enzymol.* **277**, 306–319.
Urzhumtsev, A. G. & Podjarny, A. D. (1995*a*). *CCP4 Newslett. Protein Crystallogr.* **31**, 12–16.
Urzhumtsev, A. G. & Podjarny, A. D. (1995*b*). *Acta Cryst.* D**51**, 888–895.
Urzhumtsev, A. G. & Urzhumtseva, L. (2002). In preparation.
Urzhumtseva, L. & Urzhumtsev, A. G. (2001). *CCP4 Newslett. Protein Crystallogr.* **39**, 79–85.
Vagin, A. A., Murshudov, G. N. & Strokopytov, B. V. (1998). *J. Appl. Cryst.* **31**, 98–102.