Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Combining experimental data for structure determination of flexible multimeric macromolecules by molecular replacement

Stefano Trapani,[a]* Chantal Abergel,[b] Irina Gutsche,[a] Cristina Horcajada,[c] Ignacio Fita[c] and Jorge Navaza[a]

[a]Laboratoire de Virologie Moléculaire et Structurale, CNRS UMR 2472 LVMS, 1 Avenue de la Terrasse, Bâtiment 14B, 91198 Gif-sur-Yvette, France, [b]Information Génomique et Structurale, CNRS UPR 2589 IBSM, Parc Scientifique de Luminy, 163 Avenue de Luminy, 13288 Marseille CEDEX 20, France, and [c]Institut de Biologia Molecular de Barcelona, CSIC, Parc Científic de Barcelona, c/Josep Samitier 1-5, 08028 Barcelona, Spain

Correspondence e-mail: trapani@vms.cnrs-gif.fr

Received 15 November 2005
Accepted 14 February 2006

A major effort has been made by the structural biology community to develop user-friendly software for the use of biologists. However, structural projects become more and more challenging and their solution often relies on a combination of information from various sources. Here, it is described how X-ray data, normal-mode analysis (NMA) and electron-microscopy (EM) data can be successfully combined in order to obtain a molecular-replacement (MR) solution for crystal structures containing multimeric molecules. NMA is used to simulate computationally the inherent internal flexibility of the monomer and thus enhance, together with the crystal noncrystallographic symmetry (NCS), the MR capabilities. NCS is also used to obtain a reliable EM reconstruction, which is then employed as a filter to construct oligomers starting from monomers. The feasibility of the direct use of EM reconstructions as a template for MR when the X-ray and EM data resolutions overlap is also discussed.

1. Introduction

With the post-genomic era and the start of structural genomics programs, the number of structures available in the PDB has increased tremendously over the past few years. This stream of data now makes it possible to identify a representative for most structural families. Consequently, for each new crystallographic project, molecular replacement (MR) is first considered to solve its structure and the correct template is sought, usually relying on sequence identity to rank the models to be used. In some cases, the link with a structure of the PDB is obvious, while in some other cases the homology can be remote and will only be identified through more sophisticated homology-search programs such as *FUGUE* (Shi *et al.*, 2001) and, even if the structural homology is certain, such cases are likely to fail when standard MR procedures are used. This can even worsen when the structural family corresponds to multidomain proteins that are able to assume various conformations (relative domain orientations) upon cofactor or ligand binding or more generally depending on the crystallization conditions.

In these tricky cases, exhaustive MR protocols should be worked out with strategies suited to the origin of the difficulty. In the case of flexible multidomain proteins, the use of normal-mode analysis (NMA) to generate a series of models with various domain orientations has been shown to be a very promising approach (Suhre & Sanejouand, 2004*b*). Furthermore, the exploitation of noncrystallographic symmetry (NCS) in the MR protocol can also be a successful strategy when dealing with oligomeric structures (Tong, 2001). In this context, the availability of accurate oligomeric models with their subunits/domains correctly oriented to replace mono-

Table 1
PaGS self-rotation function top peaks.

Resolution range, 15–3.58 Å; integration radius, 39.80 Å; minimal spherical harmonic component, $l = 6$; angular step, 1.0°.

	Rotation angle (°)	Rotation-axis direction			Patterson correlation (%)
		Cos x	Cos y	Cos z	
1	0.00	0.00000	0.00000	1.00000	100.0
2†	120.23	−0.36600	−0.46094	0.80844	25.6
3†	120.23	0.36601	−0.46094	−0.80844	25.5

† Peaks 2 and 3 are related by the crystal symmetry.

meric models is obviously an advantage during an MR search. Interestingly, other experimental techniques, such as electron microscopy (EM) and small-angle X-ray scattering (SAXS), can provide valuable information on the relative positions and orientations of each molecule (or domain) in the oligomer.

The worst scenario will occur when crystals contain oligomers of flexible structures; indeed, standard MR will often fail. In the present work, we used a non-trivial MR case of this kind to verify that the MR technique can still be used and lead to the correct solution when supported by an adequate strategy combining all available information from experimental data. This case study was performed on the *Pyrococcus abyssi* glycogen synthase (PaGS) crystal structure (Horcajada *et al.*, 2006). PaGS is a homotrimeric protein that forms monoclinic crystals belonging to the *C*2 space group, with one homotrimer with approximate threefold NCS in the asymmetric unit. After several unsuccessful MR trials based on the monomeric structure of *Agrobacterium tumefaciens* glycogen synthase (AtGS; Buschiazzo *et al.*, 2004), the PaGS structure was eventually solved by crystallizing fragments of the protein.

We report here the crystal structure determination of PaGS by MR using the AtGS structure as a first template to generate models verifying experimental results obtained from various sources. A general strategy was developed to make use of information from NCS, NMA and EM reconstructions. Three protocols have been designed. The first protocol consists of imposing NCS orientational constraints during a standard MR procedure. The second employs the NCS locked-translation approach (Tong, 2001) to build a trimeric model, which is then used in the translational search. The third and last exploits low-resolution EM reconstructions to build the trimeric MR search model.

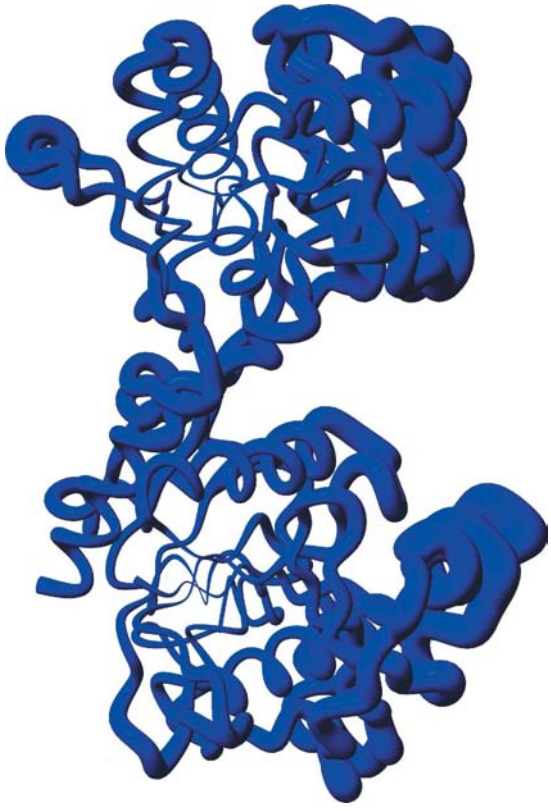

Figure 1
Tube representation of the AtGS model perturbed along normal mode No. 7. The radius of the tube varies according to the C^{α}-position spread in the perturbed models. The figure was generated using *VMD* (Humphrey *et al.*, 1996).

2. The PaGS case

The first glycogen synthase to be solved was that from *A. tumefaciens* (AtGS) and revealed a tertiary structure made up of two globular Rossmann domains separated by a large cleft constituting the catalytic centre (Buschiazzo *et al.*, 2004). This cleft has also been reported as being responsible for the overall flexibility of the molecule. Recently, Horcajada *et al.* (2006) suggested that the two monomers found in the asymmetric unit of the AtGS crystals may arise from a natural dimeric arrangement of the protein. The *P. abyssi* glycogen synthase (PaGS) belongs to the same family and was characterized as a trimeric enzyme.

PaGS sequence shares less than 26% identity over 490 residues with AtGS. The structural comparison of the AtGS structure (PDB code 1rzu) with the newly solved PaGS structure (PDB code 2bis) reveals that 376 amino acids of the 437 amino-acid PaGS sequence are equivalent in the two structures (equivalence based on a 3.5 Å $C^{\alpha}-C^{\alpha}$ distance cutoff after independent superposition of the N- and C-terminal domains). The rigid superimposition of the monomers of the two structures resulted in a total r.m.s.d. of 2.33 Å for the 1504 main-chain atoms involved. The successive rigid superimposition of the two AtGS domains individually onto the PaGS structure involved an inter-domain rotation of 12.2°, with a final r.m.s.d. of 1.63 Å.

2.1. Detecting NCS

The PaGS self-rotation function was computed using the *AMoRe* software (Navaza, 1994). It shows a clear threefold NCS axis approximately 63° away from the crystallographic twofold axis. In Table 1, we list the top peaks of the self-rotation function calculated at 3.5 Å resolution with an integration radius corresponding to the AtGS model radius. Calculations with different high-resolution limits (3.5–7.0 Å) and integration radii (26.3–43.9 Å) gave similar results with

very small variations of the NCS axis, the largest variation corresponding to an angular distance of 1.5°.

In the PaGS crystal structure, the three N-terminal domains of the trimer, rigidly bound around the centre of the oligomer, are related by an accurate threefold axis. On the other hand, owing to variations of the aperture of the interdomain clefts, the outer C-terminal domains present clear deviations from this threefold symmetry. As a consequence, the observed NCS is a result of the combination of an accurate threefold axis relating the N-terminal domains with an imperfect threefold symmetry relating the C-terminal domains. The peaks of the self-rotation function correspond to the threefold NCS axis relating the N-terminal domains. Indeed, upon rotation of the refined trimer around the self-rotation axis (120.23°), the r.m.s.d. between the N-terminal domains for each of the three superimposed monomers is 0.37, 0.5 and 0.53 Å. However, the same rotation results in larger deviations of the superimposed C-terminal domains (r.m.s.d. 1.49, 1.92 and 2.29 Å).

2.2. Models for MR

We chose the AtGS monomer structure to generate the models used for the determination of the PaGS structure by MR. The *ElNemo* software, based on the elastic network model (Suhre & Sanejouand, 2004*a*,*b*), was used to obtain 11 perturbed models of AtGS corresponding to different amplitudes of the calculated lowest frequency normal mode (normal mode No. 7 in the *ElNemo* numbering scheme). The calculated normal mode (Fig. 1) corresponds to a quasi-rigid relative rotation of one domain with respect to the other, with a maximum amplitude of 13.2° along both the 'closing' and 'opening' directions. It is worth noticing that while the maximum main-chain r.m.s.d. between the original and the perturbed AtGS models was 1.66 Å after rigid-body superposition of the complete monomers, it became as high as 3.39 Å when the superimposition was performed using solely the N-terminal domains (N-terminal, 0.18 Å; C-terminal, 5.09 Å). NMA could thus reproduce the variations of the relative orientations of the N- and C-terminal domains observed between monomers of the PaGS structure.

The best MR results were obtained for amplitude 9 of normal mode 7. The r.m.s.d. between this perturbed model and the PaGS structure is 1.84 Å, compared with 2.33 Å for the unperturbed model. In order to better describe the discrepancies between the various models, we substituted the r.m.s.d. criterion by the correlation between the electron densities of the whole molecules. In the 15–7 Å resolution shell, the correlation goes from 42% for the unperturbed model to 49% for the perturbed one after optimal superimposition of their electron densities onto the PaGS structure. These values drop to 29 and 36%, respectively, for the 15–4 Å resolution range.

2.3. Negative-stain EM and image analysis

Given the size of the particle (∼140 kDa), which is well below the threshold of applicability of cryo-EM, a low-resolution map of the trimeric PaGS was obtained using negative-stain EM. For the preparation of negatively stained grids, the sample was diluted to 0.05 mg ml^{-1} in 20 m*M* Tris–HCl, 50 m*M* NaCl buffer pH 8, applied to the clean side of the carbon on a carbon–mica interface and stained with 1%(*w*/*v*) sodium silicotungstate pH 7. The grids were observed under low-dose conditions with a Philips CM 12 transmission electron microscope with an LaB6 filament at 120 kV. Image processing was carried out on a Linux workstation using the *EM* (Hegerl, 1996; Hegerl & Altbauer, 1982), *EMAN* (Ludtke *et al.*, 1999), *PFT*2 (Baker & Cheng, 1996; Belmap *et al.*, 2004) and *Bsoft* (Heymann, 2001) software packages.

An example of the micrographs obtained is shown in Fig. 2. Most of the protein particles were absorbed on the grid in a top orientation or in a slightly tilted orientation. Therefore, in order to avoid the under-representation of side views of the particles in the final set of extracted views, the micrograph views were carefully screened visually. A total of 1956 particle images were included in the final reconstruction, 4% of which represented side-view orientations. Based on the crystal NCS, a threefold symmetry was imposed during the reconstruction process. The resolution of the reconstruction, determined *via* Fourier shell correlation, was about 25 Å according to the 0.5 threshold. The handedness was not assessed.

3. Methodology

The standard MR protocol to solve a crystal structure containing several independent molecules consists of building up the complete structure through successive steps. Firstly, the position of the molecule giving the highest score in the search process is assumed to be correct. Next, the position of the second molecule is screened using score functions that take

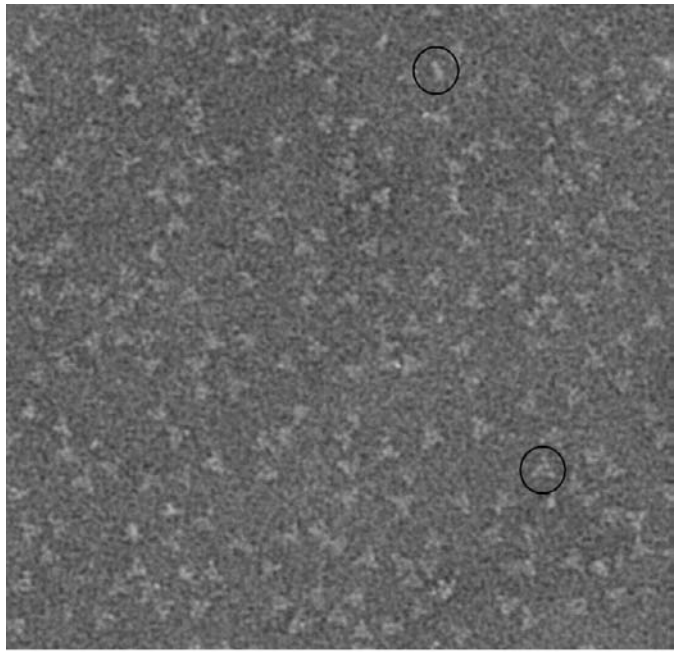

Figure 2
EM negative-stain image of a PaGS sample.

into account the contribution of the already positioned molecule (Navaza, 1994). In general, we call the known part the 'fixed molecules' and that whose position is to be determined the 'moving molecule'. This is repeated until all molecules are placed. In an ideal case with accurate models, the correct positional parameters are expected to appear as the top peaks of the rotation and translation functions. In real cases, the solution may appear at lower positions (or not appear at all) and/or with a lower contrast with false peaks. Various putative fixed-molecule positions are tested and several orientations for the moving molecules are included in the translational search. Also, different values of the parameters used in the calculations (resolution range, integration radius, series-expansion limits *etc.*) are examined.

When several copies of the same molecule are present in the asymmetric unit, each molecule can in principle be superimposed on another of the same type by a rigid-body movement, although the structural correspondence between the two molecules may not be exact owing to the different crystalline environments. This movement is not an element of the crystal symmetry space group; it defines an NCS operation (see Rossmann, 1990; Blow, 2001 and references therein). The rotational component of the NCS operations can be detected by analysis of the self-rotation function, while no straightforward method exists for the determination of the NCS translational components. An exception occurs when there is a pure translational NCS, which should result in very strong peaks in the Patterson map.

The knowledge of the NCS operations can be exploited to help the MR search when the standard procedures fail.

(i) If pure translational NCS is detected, a larger composite model formed by several copies of the search model separated by the NCS translation vectors can be used for the calculation of the translation function (Navaza *et al.*, 1998).

(ii) If the NCS rotational components are determined, these can be included in the rotation search in order to enhance the signal-to-noise ratio of the rotation function. Also, the translation-function steps in the standard MR procedure can be limited to only orientations of the moving molecules related to the fixed molecules by the NCS.

(iii) The NCS operations may form an approximate symmetry point group; often this can be reliably postulated if the equivalent molecules are known to exist as an oligomeric assembly in solution. In this case, the oligomeric assembly can be constructed by using the so called 'locked' search functions (Tong, 2001), which allow one to position, under the NCS restraints, several molecules at once around the centre of the assembly by varying the parameters of one molecule only. The oligomeric assembly is subsequently translated inside the unit cell. This procedure has the advantage of reducing the number of parameters to be determined and permits the use of larger models during the translation step. On the other hand, no crystal symmetry (other than the unit-cell translations) can be taken into account during the construction of the oligomeric assembly, which implies using data extended to *P*1.

The most crucial point in an MR problem is always the model quality, *i.e.* the structural similarity between the model and the unknown molecule. A model with an r.m.s.d. less than 2.5 Å from the unknown molecule is generally necessary to obtain an MR solution. According to Chothia & Lesk (1986), this may correspond to a remote homology between the model and the unknown molecule. However, accumulated experience shows that at least 30% identity is generally required to solve the unknown structure and even more when there is more than one independent molecule.

A low-quality search model may also be produced when the molecules correspond to flexible multidomain proteins which can assume various conformations (relative domain orientations) upon cofactor or ligand binding or more generally depending on the crystallization conditions. In such cases, to overcome the potential structural differences existing between the model and the unknown molecules, one can decide to split the search model into domains and use the latter as search models. This obviously has the disadvantage of increasing the number of search parameters and reducing the peak contrast during the first MR steps. As an alternative approach, NMA can be used to anticipate the most likely conformational changes of a given model. This has been demonstrated to be a powerful tool (Suhre & Samejouand, 2004*b*) and the screening for MR solutions with templates perturbed in the direction of one or two low-energy normal modes may allow one to find a valid MR solution when the method using the original template fails.

3.1. Exploitation of orientational NCS

The first MR strategy used was a standard MR procedure in which the orientations of the monomer to be translated were chosen to be consistent, according to the NCS, with the orientations of the already positioned molecules. The following MR protocol was used.

(i) Cross-rotation and locked cross-rotation function calculation.

(ii) Translation of the first monomer using orientations coming from the top peaks of the rotation functions (sorted by structure-factor amplitude correlation coefficient).

(iii) Rigid body least-squares fit of the translation-function results.

(iv) For each refined position of the translation function, generation of a copy of the monomer by applying a rotation verifying the NCS. Two opposite rotations of 120.23° for each one of the two self-rotation axes (Table 1) were used to generate the second monomer and perform the translational search.

(v) Rigid-body least-squares fit of the two-body translation function results. Both monomers were fitted independently.

(vi) Along the same principle, a third copy of the already positioned monomers satisfying the NCS was generated to perform the translation search for the third monomer.

(vii) Rigid-body least-squares fit of the three-body translation function results. All monomers were fitted independently.

(viii) Splitting of the monomers into two domains and rigid-body least-squares fit of the six resulting domains.

Calculations were carried out using the *AMoRe* package (Navaza, 2001). A software utility was written to read the *AMoRe* fitting output and produce an input to the translation-function program *TRAING*, taking into account a given set of NCS rotations to generate additional moving bodies. Another utility program was used to compute the positional variables of the various subdomains from the positional variables of the whole molecules.

3.2. Building the trimer using NCS

The second MR strategy consisted of the prior construction of the trimeric model based on the NCS, following the approach described by Tong (2001). The following protocol was used.

(i) Cross-rotation and locked cross-rotation functions.

(ii) Locked translation function; for each given orientation taken from the cross-rotation function, one monomer was translated on a plane perpendicular to the NCS axis, which was arbitrarily placed at the origin of the reference frame. For each position of the monomer on the translation plane, the trimer was generated by application of the NCS rotations and the corresponding structure factors, calculated in a *P*1 unit cell, were compared with the experimental data extended to *P*1. Monomer translation and trimer generation were carried out directly in reciprocal space using the fast structure-factor generation and retrieval algorithm of the *AMoRe* suite. The monomer translation was limited to a region defined by a minimum and maximum distances from the rotation axis, taking into account the molecule size. Both threefold NCS directions (Table 1) were considered.

(iii) Rigid-body least-squares fit of the three monomers in the *P*1 cell.

(iv) Translation of the trimeric models in the *C*2 cell.

(v) Rigid-body least-squares fit of the three monomers in the *C*2 cell.

(vi) Splitting of the monomers into two domains and rigid-body least-squares fitting of the six resulting domains.

In comparison with the protocol described in the previous section, this strategy has the disadvantage of neglecting the crystal symmetry during steps (ii) and (iii); nevertheless, it avoids the generation of a large number of false solutions that do not satisfy the expected local point-group symmetry of the trimeric assembly.

Calculations were carried out using the *AMoRe* package. Several programs were written to permit the locked translation calculations. In particular, the *NCS_CORREL* program calculates the locked translation function given a set of

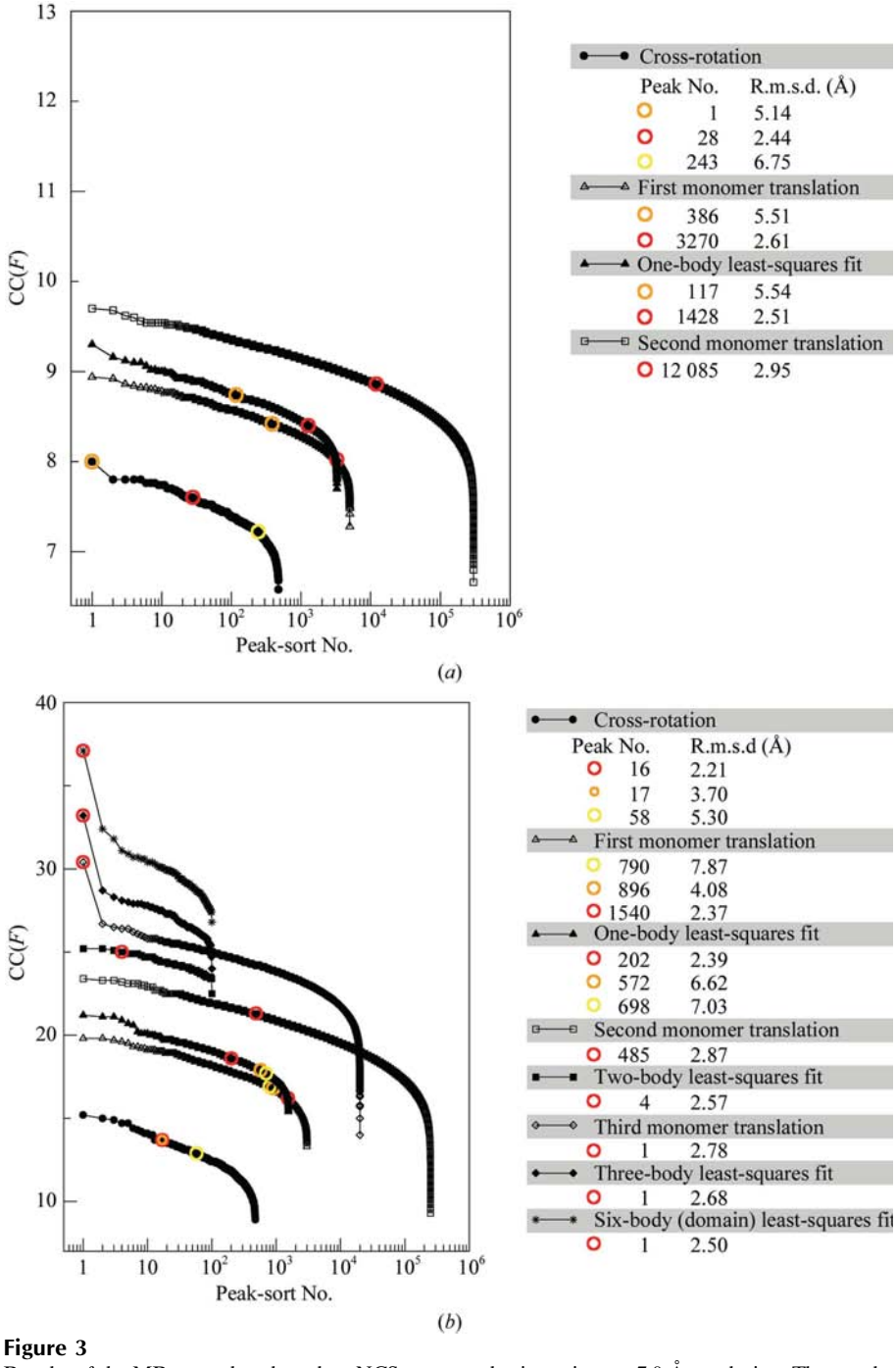

Figure 3
Results of the MR procedure based on NCS-generated orientations at 7.0 Å resolution. The graphs show the search-function results sorted by reflection-amplitude correlation [CC(*F*)]. The positions closest to the correct solution are highlighted by circles in the graphs. (*a*) AtGs model. (*b*) NMA model corresponding to amplitude 9 of normal mode 7. The r.m.s.d.s were calculated on a subset of residues (main-chain atoms only) considered to be structurally equivalent in the PaGS and AtGS refined structures.

monomer orientations and the NCS rotations. This program performs a point-by-point calculation of the correlation coefficient in terms of amplitudes.

3.3. Building the trimer combining NCS with EM data

Finally, it is possible to take advantage of the oligomerization state of a molecule both in solution and in the crystal by combining EM data with the information from X-ray data, *i.e.* the existence of NCS. This symmetry is usually difficult to recover from the EM micrographs. However, X-ray data can provide the point-group symmetry to be imposed to the reconstruction of the isolated trimeric molecule. The resulting map can then be used directly to search for an MR solution. This strategy has been successfully employed in the case of viruses and other very large molecular assemblies (Dodson, 2001). However, for medium-size assemblies this approach can rarely be used, either because the EM reconstructions are hard to produce or because the resolutions of the EM and X-ray data do not match.

Alternatively, EM maps of oligomeric proteins can be used to fit atomic models into them and to extract the orientations and relative positions of each molecule in the oligomer. This allows either the construction of oligomeric models that possess higher resolution than the EM data or the filtering of putative MR solutions by comparison with the EM-fitted model.

4. Results

4.1. Taking orientational NCS into account

When the AtGS model was employed, this strategy was not successful. Fig. 3(*a*) gives a graphical overview of the results obtained using data to 7.0 Å resolution and cross-rotation orientations. Similar results were obtained using data to 5.5 Å resolution (data not shown). Analysis of the translation function for the first monomer reveals that the peak corresponding to the best position (when compared with the solved PaGS structure) is found only after a huge number of better ranking peaks (3270 and 1428 peaks before and after rigid-body refinement, respectively). This makes the second monomer search rather time-consuming, as thousands of one-body putative solutions must be included in the successive search. When the second monomer is translated, just one of the two possible orientations could be located close to the correct position. Its rank (12 085; Fig. 3*a*) in the sorted peak list renders a search for the third monomer position computationally prohibitive. Furthermore, and surprisingly, all attempts we made to translate the third monomer using just the best positions found for the first two monomers did not lead to the correct solution.

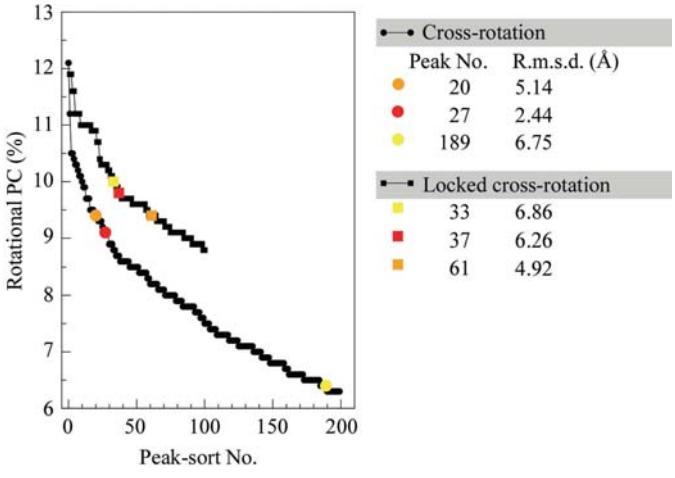

Figure 4
Comparison of the cross-rotation and locked cross-rotation function peaks for the AtGS model at 7.0 Å resolution. The graph shows the rotation-function peaks sorted by Patterson overlap correlation. The positions closest to the correct solution are highlighted by circles in the graphs. Their r.m.s.d.s from the correct orientations are reported in Table 1.

Interestingly, the same strategy led to an MR solution when normal-mode models were used as template. Fig. 3(*b*) shows the results for amplitude 9 of normal mode 7. If a sufficiently large number of fixed-body positions are included in the search, the correct solution appears with an appreciable contrast after translation of the third monomer. Similarly, calculations at 5.5 Å resolution (data not shown) using a large number of fixed-body positions led to a solution using this normal-mode model, though with no contrast until the very last step (rigid-body refinement of the six independent domains).

4.2. Building the trimer using NCS

The locked translation strategy led to detectable MR solutions with both the original AtGS and the normal-mode models, provided that cross-rotation orientations were used. The MR protocol systematically failed when locked cross-rotation orientations were used. The reason is that the NCS axis only correctly describes the N-terminal domains (see Table 2 in Horcajada *et al.*, 2006). In Fig. 4, we report the smallest deviations from the correct orientations found among the first peaks of the cross- and locked cross-rotation functions of the AtGS model at 7.0 Å resolution. While the normal cross-rotation function provides one sufficiently accurate orientation (r.m.s.d. 2.44 Å), all locked cross-orientations are far too inaccurate (r.m.s.d. > 4.9 Å). It is generally expected that averaging over the NCS operations should enhance the rotation-function signal-to-noise ratio. However, as shown in the AtGS case, if the model and/or the imposed NCS transformations are not sufficiently accurate, averaging may just hide good orientations. Hence, to avoid this, the NCS parameters can be refined using the program *COMPANG* (Urzhumtseva & Urzhumtsev, 2002) to obtain an accurate set of rotation angles consistent with NCS.

Fig. 5(*a*) shows the results obtained using the AtGS model and data to 7.0 Å resolution. The three independent orientations of the monomer are found at positions 1, 28 and 243 in the sorted list of the cross-rotation function peaks. However, just one of them (highlighted in red) seems to have a sufficiently low discrepancy (expressed as r.m.s.d. in Fig. 5) from

the correct orientation to permit the successful construction of a trimeric model in the subsequent steps. Counterintuitively, the best orientation is not the highest ranked cross-rotation peak. During the successive steps, the correct solution is found at the top positions of the search-function peaks. However, the correct solution is only marginally higher than the next peak and therefore is not detectable until the very last step. At higher resolution (5.5 Å, data not shown), no solution could be found with this model.

Better results were obtained using some of the normal-mode models. In particular, for normal mode No. 7, four of the 11 tested normal amplitudes (amplitudes 6, 7, 9 and 10) led to an MR solution. For two of these successful models, the advantage of normal-mode analysis appeared evident, as a significant contrast between the correct solution and the false peaks could be seen at 7.0 Å resolution before the last MR step (Fig. 5*b*). The solution could be found at 5.5 Å resolution (data not shown), but without contrast until the last step.

The whole set of models corresponding to a set of amplitudes looks like an NMR ensemble. We used this ensemble for normal mode 7 as a search model. In this case the solution appeared in the first position, although with no significant contrast with false solutions.

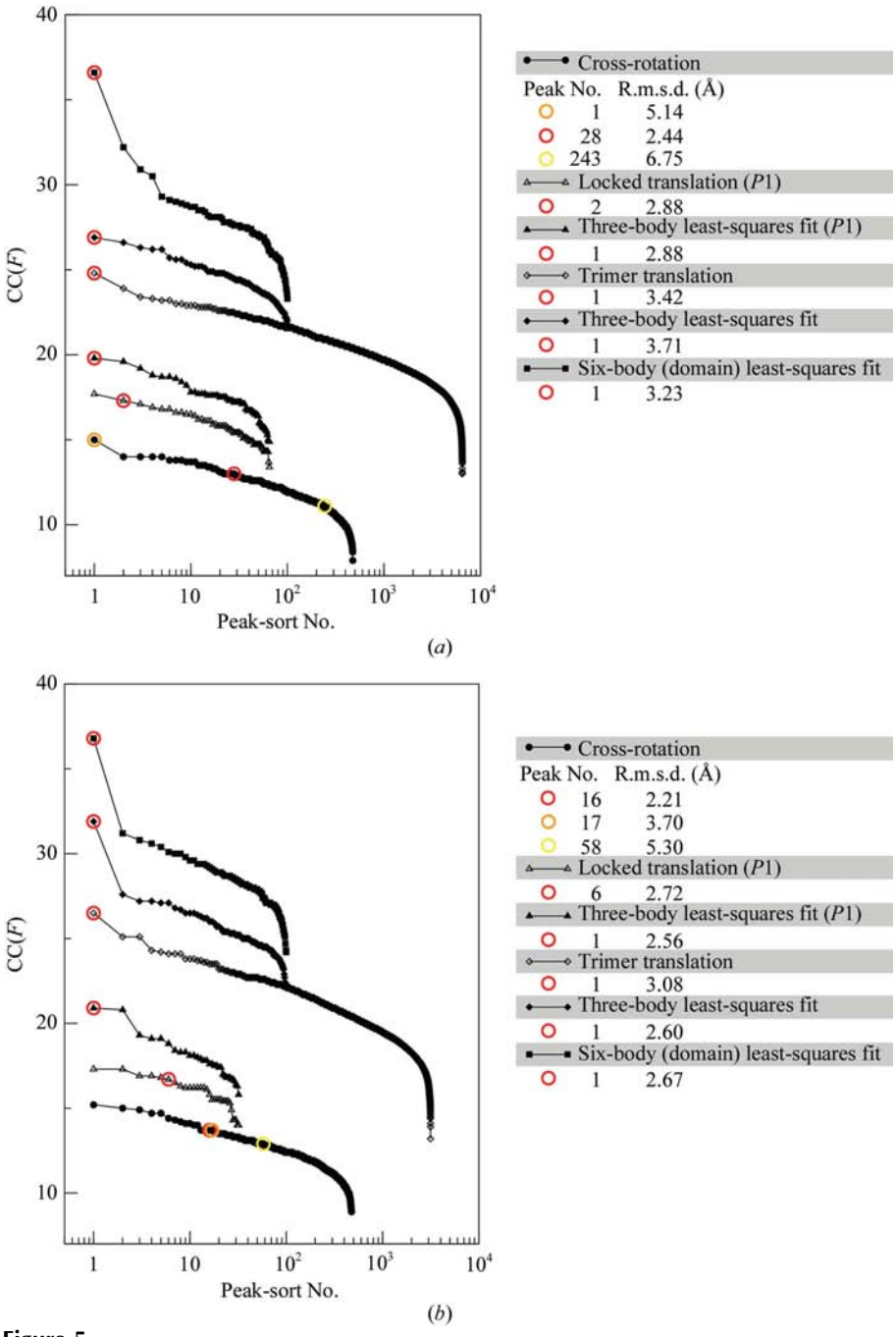

Figure 5
Results of the locked translation-based MR procedure at 7.0 Å resolution. The graphs show the search-function peaks sorted by reflection-amplitude correlation [CC(*F*)]. The positions closest to the correct solution are highlighted by circles in the graphs. (*a*) AtGs model. (*b*) NMA model corresponding to amplitude 9 of normal mode 7.

4.3. Building the trimer combining NCS with EM data

In our third MR strategy, we employed trimeric models based on three-dimensional EM reconstructions obtained from micrographs of negatively stained trimers.

4.3.1. Direct use of the EM map as MR model. Owing to the extremely low number of experimental reflections in the overlapping region of the EM and X-ray data resolution ranges (72 reflections up to 20 Å), the direct use of the EM map as a search model was impracticable. In order to assess the validity of the EM reconstruction and verify that the lack of low-resolution reflections was the only reason why EM data could not be used, the refined PaGS structure was used in *REFMAC*5 (Murshudov *et al.*, 1997), which takes into account the contribution of the bulk solvent, to calculate a complete low-resolution X-ray diffraction data set. The EM reconstruction, which had never 'seen' the refined PaGS structure, was then tested as a search model in *AMoRe* and the MR solution could be identified. We can thus conclude that the missing low-resolution data prevents the successful use of EM data to seek a proper solution in MR.

4.3.2. Using atomic models fitted into the EM map. A trimeric model was constructed by fitting the AtGS structure into the EM map (Fig. 6) using the program *URO* (Navaza *et al.*, 2002) and imposing threefold symmetry. The

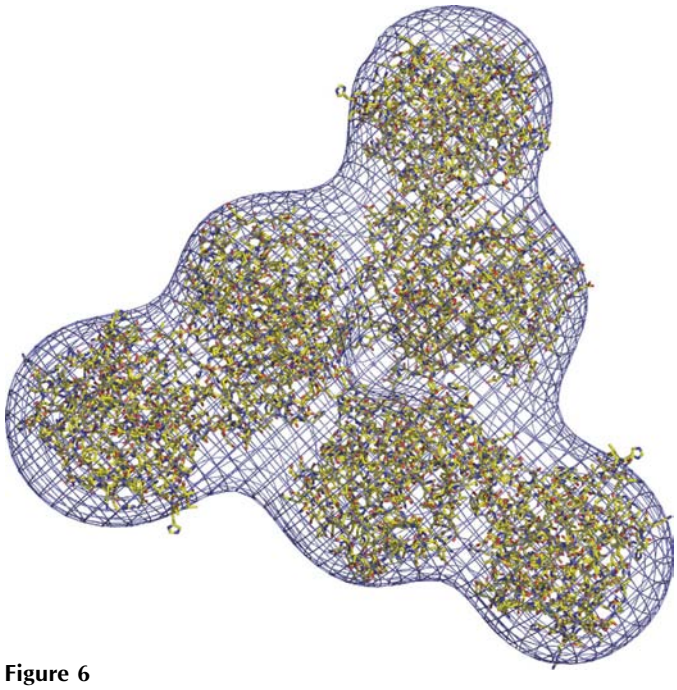

Figure 6
Trimeric model based on the monomeric AtGS structure fitted in the EM reconstruction imposing threefold symmetry. The figure was generated using *PyMOL* (DeLano, 2002).

fitting resulted in a correlation coefficient of 96% between the EM reconstruction and the electron density of the trimeric model and an R factor of 33.4% at 25 Å resolution. Despite the rather low resolution of the EM data, the visual similarity of the EM trimeric model to the solved PaGS structure was rather good (5.45 Å r.m.s.d. for equivalent main-chain atoms). Nevertheless, this model was not accurate enough to give an MR solution. Several attempts were carried out, using an upper resolution limit varying from 20 to 7 Å, without success.

Given the low resolution of the EM reconstruction, the minimum of the optimization criterion used during the EM fit is expected to be relatively flat, so that inevitable errors in the positional parameters may affect the quality of the fitted model as an MR probe. Thus, an alternative procedure was devised in which constraints based on crystallographic data were imposed. The combined information from the NCS (already used in the EM reconstruction) and the cross-rotation orientations was taken into account. The EM map was initially oriented in order to align its threefold axis with the crystal NCS axis (both directions of the threefold axis were tested). Trimers based on the AtGS monomer were then fitted into the map by translating the AtGS model previously oriented according to the cross-rotation peaks, applying the threefold symmetry and rotating the EM map around the NCS

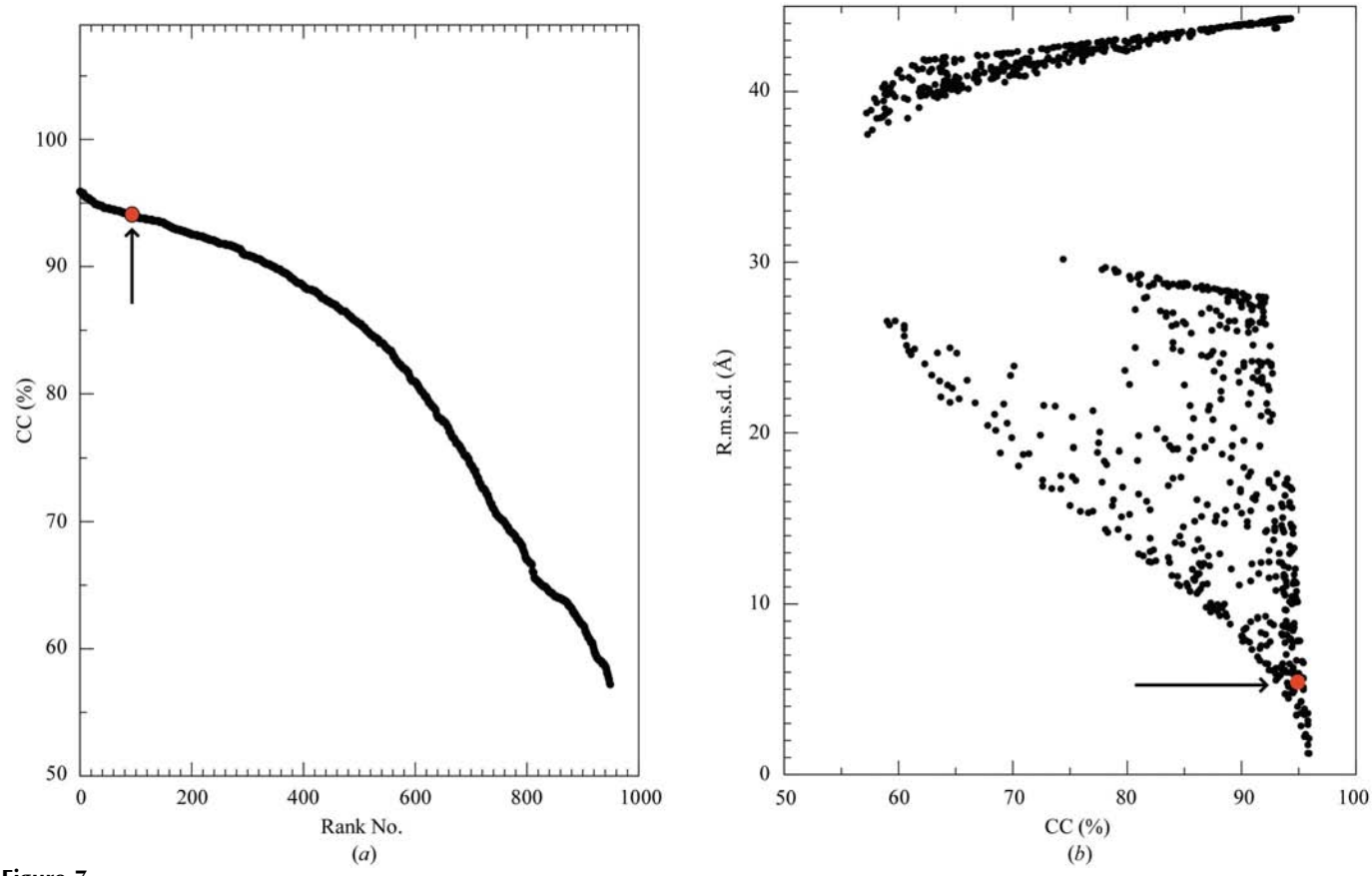

Figure 7
Results of the fitting of the AtGS atomic model in the PaGS trimer EM reconstruction at 25 Å resolution. A set of 950 putative orientations obtained from the crystallographic cross-rotation function were used as constraints during the trimer-generation and fitting process. (*a*) Correlation coefficients between the electron densities of the 950 constrained trimers and the EM map. The point corresponding to the trimeric model closest to the PaGS structure is highlighted. (*b*) R.m.s.d.s between the constrained trimers and the unconstrained trimer *versus* the correlation coefficient.

axis. The trimeric models thus obtained satisfied both the NCS and the cross-rotation orientations of the monomeric subunit. Results of the EM fit are shown in Fig. 7(*a*). In Fig. 7(*b*), the r.m.s.d.s between the fully unconstrained trimer and the constrained trimers are plotted *versus* the constrained-fit correlation coefficients.

The generated trimers were already oriented in the crystal frame and could be used directly for an MR translational search. The trimeric model corresponding to the most accurate cross-rotation orientation, highlighted in Fig. 7, has 5.4 Å r.m.s.d. with the unconstrained trimer and 3.3 Å with the PaGS crystal structure. This model could be successfully used for MR, where only the translational step and the rigid-body refinement of the six domains were needed. Calculations using NMA models were not performed, although one would expect improved results. The procedure described here to generate oriented trimeric models presents the advantage of being almost two orders of magnitude faster than the locked-translation approach described in §3.2 (CPU time: 0.15 *versus* 6.7 s per cross-rotation orientation on average in the PaGS case). Moreover, many of the generated trimers could be discarded based on a correlation coefficient cutoff or, more safely, on an r.m.s.d. cutoff (see Fig. 7*b*).

5. Conclusions

The crucial role of normal-mode analysis and the locked translation function to build an oligomeric model suited for the MR technique has been demonstrated for PaGS, a particularly difficult structure given its oligomeric state in addition to the internal flexibility of the monomers. However, our study suggests that for those cases the most promising strategy is that where information from NCS, NMA and EM is used to generate EM-based oligomeric models as the starting model for MR calculations. The first step in this approach is the combination of NCS information with electron-microscopy data of the oligomeric protein in solution to obtain an EM reconstruction. Specific computational methods could then be used to extract from this reconstruction the orientations and relative positions of each molecule in the oligomer in order to build the atomic model of the complex used later on to seek an MR solution.

Our results with simulated low-resolution X-ray data that allow overlap with EM data also suggest that EM reconstructions obtained by negatively staining the particles, a technique of wider applicability than cryo-EM, are suitable models for medium-size proteins (greater than 140 kDa), provided that complete low-resolution diffraction data are available. We are currently extending this approach to difficult crystallographic projects corresponding to oligomeric proteins, for which we will collect full low-resolution data sets in order to combine them with our EM reconstructions through our MR strategy.

Given the high cost of protein production, especially for mammalian targets, and the difficulties in producing reproducible batches of post-translationally modified proteins often of high pharmaceutical value, it may be valuable to consider using the first produced crystals as much as possible by applying combinatorial strategies in order to solve structures of predictable fold using MR.

The utilities developed to take into account NCS orientational constraints and to calculate the locked-translation function depend on the stand-alone version of the *AMoRe* package. They are available free of charge upon request.

This work was funded by Human Frontier Science Program RGP0026/2003. We acknowledge the use of the Marseille-Génopole bioinformatics platform supported by the French Genome Research Network (RNG).

References

Baker, T. S. & Cheng, R. H. (1996). *J. Struct. Biol.* **116**, 120–130.
Belmap, D. M., Conway, J. F. & Heymann, J. B. (2004). *PFT2 and EM3DR2.* http://www.niams.nih.gov/rcn/labbranch/lsbr/software/pft2_em3dr2/pft2_em3dr2.htm.
Blow, D. M. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 263–268. Dordrecht: Kluwer Academic Publishers.
Buschiazzo, A., Ugalde, J. E., Guerin, M. E., Shepard, W., Ugalde, R. A. & Alzari, P. M. (2004). *EMBO J.* **23**, 3196–3205.
Chothia, C. & Lesk, A. M. (1986). *EMBO J.* **5**, 823–826.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.
Dodson, E. J. (2001). *Acta Cryst.* D**57**, 1405–1409.
Hegerl, R. (1996). *J. Struct. Biol.* **116**, 30–34.
Hegerl, R. & Altbauer, A. (1982). *Ultramicroscopy*, **9**, 109–116.
Heymann, J. B. (2001). *J. Struct. Biol.* **133**, 156–169.
Horcajada, C., Guinovart, J. J., Fita, I. & Ferrer, J. C. (2006). *J. Biol. Chem.* **281**, 2923–2931.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). *J. Struct. Biol.* **128**, 82–97.
Murshudov, G., Vagin, A. & Dodson, E. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Navaza, J., Lepault, J., Rey, F. A., Alvarez-Rua, C. & Borge, J. (2002). *Acta Cryst.* D**58**, 1820–1825.
Navaza, J., Panepucci, E. H. & Martin, C. (1998). *Acta Cryst.* D**54**, 817–821.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Shi, J., Blundell, T. & Mizuguchi, K. (2001). *J. Mol. Biol.* **310**, 243–257.
Suhre, K. & Sanejouand, Y.-H. (2004*a*). *Nucleic Acids Res.* **32**, W610–W614.
Suhre, K. & Sanejouand, Y.-H. (2004*b*). *Acta Cryst.* D**60**, 796–799.
Tong, L. (2001). *Acta Cryst.* D**57**, 1383–1389.
Urzhumtseva, L. & Urzhumtsev, A. (2002). *J. Appl. Cryst.* **35**, 644–647.