research papers

Ultralow-resolution *ab initio* phasing of filamentous proteins: crystals from a six-Ig fragment of titin as a case study

Alexandre Urzhumtsev,[a,b]*
Eleonore von Castelmur[c] and
Olga Mayans[c,d]

[a]IGBMC, CNRS-INSERM-ULP, 1 rue Laurent Fries, 67404 Illkirch, France, [b]Department of Physics, Faculty of Science and Technologies, Nancy-University, 54506 Vandoeuvre-les-Nancy, France, [c]Division of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland, and [d]School of Biological Sciences, Biosciences Building, University of Liverpool, Crown Street, Liverpool L69 7ZB, England

Correspondence e-mail:
sacha@igbmc.u-strasbg.fr

Low-resolution diffraction data (resolution below 12 Å) from crystals of a filamentous six-Ig fragment of titin, I65–I70, were used in *ab initio* phasing with the aim of calculating its lattice packing and molecular envelope. Filamentous molecules, characterized by marked anisometry and idiosyncratic crystal lattices, have not been addressed before using this methodology. In this study, low-resolution phasing (19–122 Å) successfully identified the region of the unit cell occupied by the molecule. Phase extension to a higher resolution (12 Å) yielded regions of high density that corresponded either to the positions of individual Ig domains or to zones of dense intermolecular contacts, hindering the identification of individual domains and the interpretation of electron-density maps in terms of a molecular model. This problem resulted from the acutely uneven packing of the molecules in the crystal and it was further accentuated by the presence of partially disordered regions in the molecule. Addition of low-resolution reflections with phases computed *ab initio* to those obtained experimentally using MIRAS improved the initial electron-density maps of the atomic model, demonstrating the generic utility of low-resolution phases for the structure-elucidation process, even when individual molecules cannot be resolved in the lattice.

1. Introduction

Filamentous proteins are intrinsically flexible and have a poorly defined long-range order, which often renders them unsuitable for crystallographic analysis either by preventing crystal growth or by yielding crystalline formations of low diffraction quality. An examination of crystal lattices for cases in which successful structure elucidation has been achieved shows that the anisometry of these molecules results in two types of frequently observed packing arrangements: (i) tight lateral associations of molecules with only minimal interstitial bulk solvent or (ii) loose arrangements of crisscrossed molecules that delineate vast solvent channels (Fig. 1). In the latter, crystallographic contacts involve only a small fraction of the molecule, so that most of it remains free from interactions and suspended in solution. Owing to the inherent flexibility of filamentous proteins, the unrestrained fraction often exhibits crystalline disorder to a greater or lesser extent. In brief, these high-solvent lattices tend to be characterized by an alternation of well defined areas rich in crystal contacts, which are usually coincident with primary crystallographic axes, and loose 'bridging' regions of poor crystallographic definition.

Table 1
Statistics for low-resolution X-ray data from I65–I70.

Space group $P6_522$, unit-cell parameters $a = b = 140.15$, $c = 164.32$ Å. The Matthews coefficient V_M of 3.8 Å^3 Da^{-1} corresponds to a solvent content of 67%.

	Whole data set	Highest resolution shell
Resolution zone (Å)	122.0–8.0	9.0–8.0
Unique reflections	1176	334
R_{merge}	7.5	8.3
Multiplicity†	22.8	14.7
Completeness (%)	100.0	100.0
$I/\sigma(I)$	31.9	19.6

† Data were collected in three sweeps that extended to different resolutions, thus the multiplicity for the whole set is noticeably higher than that for the highest resolution shell.

The phasing of filamentous proteins can also be affected by the specific characteristics of these molecules. For example, molecular-replacement approaches can be unsuccessful because of long-range structural deviations along the molecular axis. Thus, it is desirable to broaden the palette of methods that are applicable to the determination of these structures. To this effect, we have investigated the potential of *ab initio* ultralow-resolution phasing for the calculation of a molecular envelope that could serve as an initial source of phases in the elucidation process.

The current work is an exploratory *ab initio* study of a six-Ig fragment, I65–I70, from the muscle filament titin. This protein is composed of ~300 serially connected Ig and FnIII domains and acts as a molecular spring in the sarcomere (reviewed by Granzier & Labeit, 2004). The crystal structure of I65–I70 has recently been elucidated at 3.3 Å resolution using MIRAS (von Castelmur *et al.*, 2008). Diffraction of poor quality and limited maximal resolution (ranging from 6.0 to 3.3 Å for individual crystals) prevented the steady progress of experimental phasing in this case. Thus, we decided to explore

independently and in parallel the capabilities of ultralow-resolution *ab initio* phasing in aiding the process. Obviously, evaluation of the results could only be carried out *a posteriori* once the atomic model had been completed. Even though we anticipated that the composition of I65–I70 by Ig globular domains would make the molecule methodologically more tractable than the strongly anisometric motifs of minimal cross-sections, such as coiled coils, the final analysis showed that I65–I70 posed an extreme challenge to this technique. This text discusses the source of the difficulties encountered and the extent to which viable results could be obtained. Even though the *ab initio* calculations did not allow us to make conclusions on molecular features, they led us to identify with reasonable accuracy the fraction of the unit cell occupied by the protein fragment. This information, which is particularly suitable for exploitation in density-modification approaches, indicates that *ab initio* low-resolution phasing can be productively applied to this class of proteins, albeit with limitations.

2. Data and tools

2.1. Ultralow-resolution diffraction data and molecular characteristics of I65–I70

The cloning, expression and crystallization of domains I65–I70 (amino acids 7946–8511) from rabbit soleus titin (human titin at TrEMBL Q8WZ42) have been reported previously (Marino *et al.*, 2005; von Castelmur *et al.*, 2008). In brief, crystals belonged to space group $P6_522$, with unit-cell parameters $a = b = 140.15$, $c = 164.32$ Å and one molecule per asymmetric unit ($V_M = 3.8$ Å^3 Da^{-1}; 67% solvent content). Low-resolution diffraction intensities (122–8 Å resolution) were recorded on beamline X06SA at SLS (Villigen) independently from those finally used in structure elucidation,

(a)

(b)
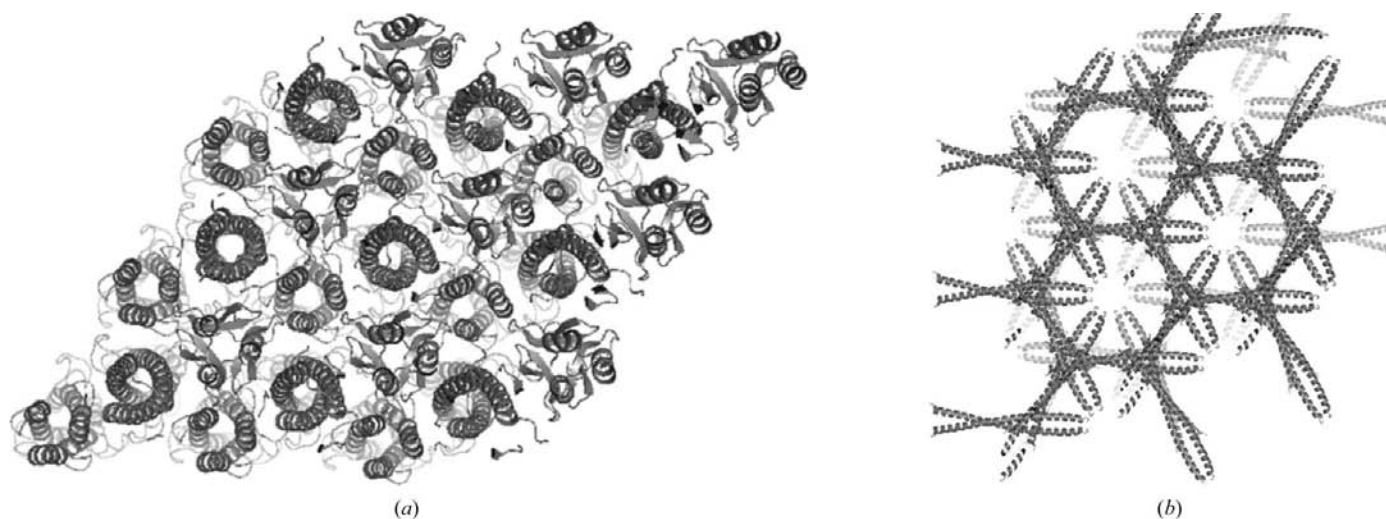
Figure 1
Crystallographic lattices characteristic of filamentous proteins. (a) Tightly packed lattice exemplified by a fragment of fibritin that includes a coiled-coil motif and a C-terminal foldon (PDB code 1aa0; Strelkov *et al.*, 1996). The space group is *P3*, with an estimated solvent content of 36% (V_M = 2.35 Å^3 Da^{-1}). (b) Loose crisscrossed crystal packing illustrated by a coiled-coil fraction from lamin A in space group $P6_522$ (PDB code 1x8y; Strelkov *et al.*, 2004). The estimated solvent content is 71% (V_M = 4.2 Å^3 Da^{-1}). V_M is the Matthews coefficient (Matthews, 1968).

which originated from a different crystal. Data were collected at a crystal-to-detector distance of 475 mm on a MAR CCD 165 detector at $\lambda = 0.9537$ Å. The images corresponded to non-overlapping 1° oscillation steps, amounting to a total of 100° rotation. Three passes were carried out with decreasing exposure times and using an increasingly attenuated beam to ensure optimal recording of strong reflections. Data were processed using the *XDS/XSCALE* suite (Kabsch, 1993). Table 1 lists the data-processing statistics.

I65–I70 is composed of six Ig domains that share an average pairwise sequence identity of ∼35%. Individual domains are small (composed of about 100 amino acids) and have an elliptical shape; their axial dimensions are approximately $45 \times 30 \times 20$ Å. The interdomain linker sequences were predicted to consist of three, three, zero, zero and zero residues in length based on sequence data (Marino *et al.*, 2005). This suggested that the N-terminal half of the fragment has a high potential for flexibility and conformational heterogeneity, while the C-terminal part may be better defined structurally. Previous small-angle X-ray scattering (SAXS) data and an electron-microscopy analysis of the end-to-end molecular distance distribution of I65–I70 indicated that it is semi-rigid and predominantly adopts a semi-extended arrangement in solution, with an estimated length of about 23 nm (Marino *et al.*, 2005). However, it should be borne in mind that parameters corresponding to molecular averages and estimated in solution are not necessarily indicative of the overall molecular conformation of I65–I70 in a crystalline lattice.

2.2. Phasing methods

Diffraction data from native crystals were recorded to 3.3 Å resolution. Repeated attempts to elucidate the structure by molecular replacement or MAD on SeMet-labelled crystals were unsuccessful. Therefore, experimental phasing using heavy-atom derivatives was instead pursued. Because this became a demanding task with slow progress, we initiated a parallel search for the overall conformation and packing of
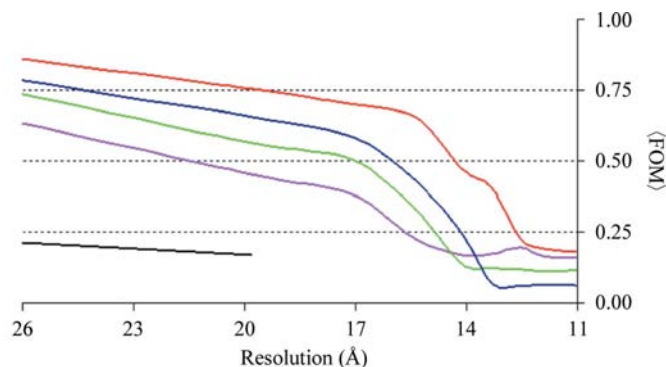

Figure 2
Figure of merit and effective resolution of *ab initio* phase sets. The mean figure of merit ⟨FOM⟩ is shown as a function of resolution. Phase sets were obtained by selecting maps at resolutions of 19 Å (set FAM19; black), 16 Å (set conn16; magenta), 15 Å (set conn15; green), 14 Å (set conn14; blue) and 12 Å (set conn12; red). The resolution at which ⟨FOM⟩ falls below 0.5 can be considered as the effective value.

I65–I70 in the crystal using ultralow-resolution data with *ab initio* phasing.

We applied two previously reported low-resolution *ab initio* phasing methods: the *FAM* (Lunin *et al.*, 1995, 1998) and image-connectivity approaches (Lunin *et al.*, 2000). A reminder of these methods is given in Appendix *A*. The study of I65–I70 was a challenge for both methodologies, but especially for the connectivity-based approach. Filamentous proteins can present unusual lattice arrangements (Fig. 1) in which locally dense packing makes it scarcely possible to recognize individual molecules in contact regions in low-resolution electron-density maps. As a consequence, it is difficult to predict a correct scoring rule for the selection of best phase variants based on the interpretability of connectivity in these maps. Thus, the *FAM* approach was chosen as being the most suitable for the initial stages of *ab initio* phasing in this project.

The programs *SF2CNS* (Urzhumtsev & Urzhumtseva, 2002) and *PyMOL* (DeLano, 2002) were used to calculate and display the Fourier maps in this study.

3. Results of *ab initio* phasing

3.1. Initial phasing with the *FAM* method

The *FAM* phasing procedure depends on two key parameters: the composition of the models (*i.e.* the number of spheres in the asymmetric unit of the cell) and the number of reflections used. The result of the *FAM* method is not the position of the spheres, but a set of phases and their figures of merit. Thus, it should be emphasized that the positions of spheres for the best *FAM* models do not necessarily correspond to the centres of molecules or domains (Lunin *et al.*, 1995). The larger the number of spheres, the better the model shape can be approximated and the more closely the experimental structure factors may be fitted. Increasing the number of spheres (each characterized by the three coordinates of its geometrical centre) requires increasing the number of structure-factor magnitudes used for phasing in order to avoid over-parametrization. In the case of I65–I70, initial models were composed of six spheres, which were explored using all 90 reflections with resolution below 21 Å.

Two other important parameters in *FAM* phasing are the *B* value and the minimal distance D_{min} between the centers of the positioned spheres. The parameter *B* represents the size of the Gaussian isotropic peak of the electron density for each *FAM* sphere. To find suitable values for these parameters, we screened them systematically using several short series generations each consisting of 1000 models with given *B* and D_{min} values. Those series containing a large number of models whose calculated structure-factor magnitudes bore a relatively high correlation to the experimental values determined the optimal values of *B* and D_{min}. This analysis estimated $B = 10\,000$ Å^2 and $D_{min} = 25$ Å; the latter is consistent with the smallest dimension of the Ig domain (about 20–25 Å).

Selected *B* and D_{min} values were used to calculate electron-density maps for I65–I70 by three consecutive iterations of *FAM*. The phase set $\{\varphi_{FAM21}\}$ selected from the first iteration

extending to 21 Å resolution was treated using clustering techniques (Lunin *et al.*, 1990). However, since the main cluster gave an image that was basically identical to that obtained by the averaging of all selected phase sets, clustering was not performed in subsequent iterations and simple averaging was instead applied. The correct sign of the density cannot be calculated directly by *FAM*; therefore, we applied a generalized maximum-likelihood technique proposed by Lunin *et al.* (1998) and Petrova *et al.* (2000) (further details are given in Appendix *A*). After a generic first cycle with models generated uniformly across the unit cell, two further iterations (data resolution 122–19 Å; resulting in phase set $\{\varphi_{FAM19}\}$) were directed to improve the molecular envelope by generating new sphere models inside the envelope obtained from the previous iteration and defined by a volume of $0.8 \times V_{cell}$ (see Appendix *A* for the choice of the density-cutoff level in defining molecular envelopes). The values of all other parameters in these cycles were kept as in the initial cycle. In subsequent iterations, the number of spheres was increased

from 6 to 12, D_{min} was decreased from 25 to 15 Å, the *B* value was decreased slightly from 10 000 to 8000 Å^2 and the envelope volume was reduced. The formal resolution of the newly obtained phase set was increased stepwise up to 16 Å. However, the fine refinement of parameters and the more detailed 12-sphere *FAM* models did not result in a significant image improvement and thus *FAM* phasing was interrupted at the phase set $\{\varphi_{FAM19}\}$ after the third iteration.

3.2. Phase extension with connectivity

The low-resolution image of I65–I70 computed with the primary *FAM* phase set $\{\varphi_{FAM19}\}$ was not interpretable in terms of individual domains. To improve the image, we applied a connectivity-based phase-extension procedure in which data resolution was gradually increased to 16, 15, 14 and 12 Å (phase sets $\{\varphi_{conn16}\}$, $\{\varphi_{conn15}\}$, $\{\varphi_{conn14}\}$ and $\{\varphi_{conn12}\}$, respectively). Fig. 2 shows the variation of the mean figure of merit with resolution for the intermediate and final phase sets in this study. Since the loss of phase information at increasing resolution is reflected by a corresponding decrease in the associated mean figure of merit, the effective resolution of a given phase set is lower than its nominal resolution and can then be considered as that where the mean figure of merit falls below 0.5 (Lunin & Woolfson, 1993). In particular, it should be noted that the effective resolution of the final set $\{\varphi_{conn12}\}$ is approximately 15 Å rather than the calculated 12 Å.

The initial expectation in this study was that a correct low-resolution Fourier synthesis would reveal local maxima near the centre of each individual Ig domain. For this reason, we formulated the topological requirement of identifying six independent globules of similar size (one globule per domain), each with 12 copies (the number of symmetry copies in the space group), as a rule for map selection during phase extension and improvement. Indeed, at a resolution of 19 Å the molecular envelope defined by the unit-cell volume $0.30 \times V_{cell}$ in $\{\varphi_{conn12}\}$ maps was more detailed than a similar envelope for $\{\varphi_{FAM19}\}$ maps at the same resolution (Figs. 3*a* and 3*b*). At a resolution of 15 Å, $\{\varphi_{conn12}\}$ maps were too noisy and when applying higher cutoff levels, for example $0.15 \times V_{cell}$, the molecular envelopes became split into separate globules (Fig. 3*c*). All efforts to interpret these maps in terms of six linked Ig domains failed. In retrospect, the high density peaks corresponded to regions of tight crystallographic contacts and not to individual Ig domains. This complication primarily arose
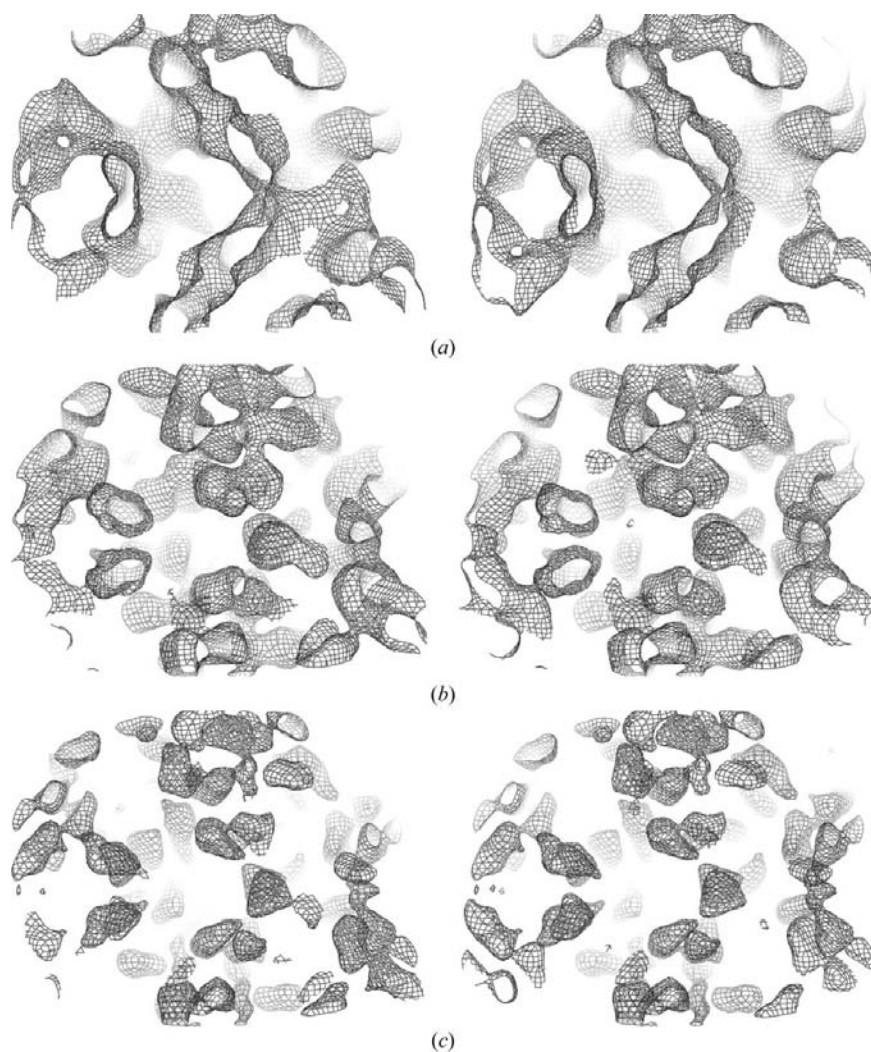

Figure 3
Fourier synthesis calculated with *ab initio* phase sets. Electron-density maps corresponding to (*a*) phase set $\{\varphi_{FAM19}\}$, resolution 19 Å, $0.30 \times V_{cell}$, contour level 0.6σ, (*b*) phase set $\{\varphi_{conn12}\}$, resolution 19 Å, $0.30 \times V_{cell}$, contour level 0.4σ, (*c*) phase set $\{\varphi_{conn12}\}$, resolution 15 Å, $0.15 \times V_{cell}$, contour level 0.8σ.

from the crystal lattice features of I65–I70, a phenomenon of low-resolution phasing that has been observed previously for globular proteins (see, for example, Lunin *et al.*, 1995).

4. *A posteriori* comparison with the atomic model

4.1. Atomic model and molecular packing

The crystal structure of I65–I70 has been elucidated at 3.3 Å resolution using MIRAS independently of this *ab initio* study (von Castelmur *et al.*, 2008). In agreement with SAXS and EM data (Marino *et al.*, 2005; von Castelmur *et al.*, 2008), I65–I70 adopts a semi-extended conformation with an end-to-end distance of ~212 Å (280 Å contour length). The molecule is
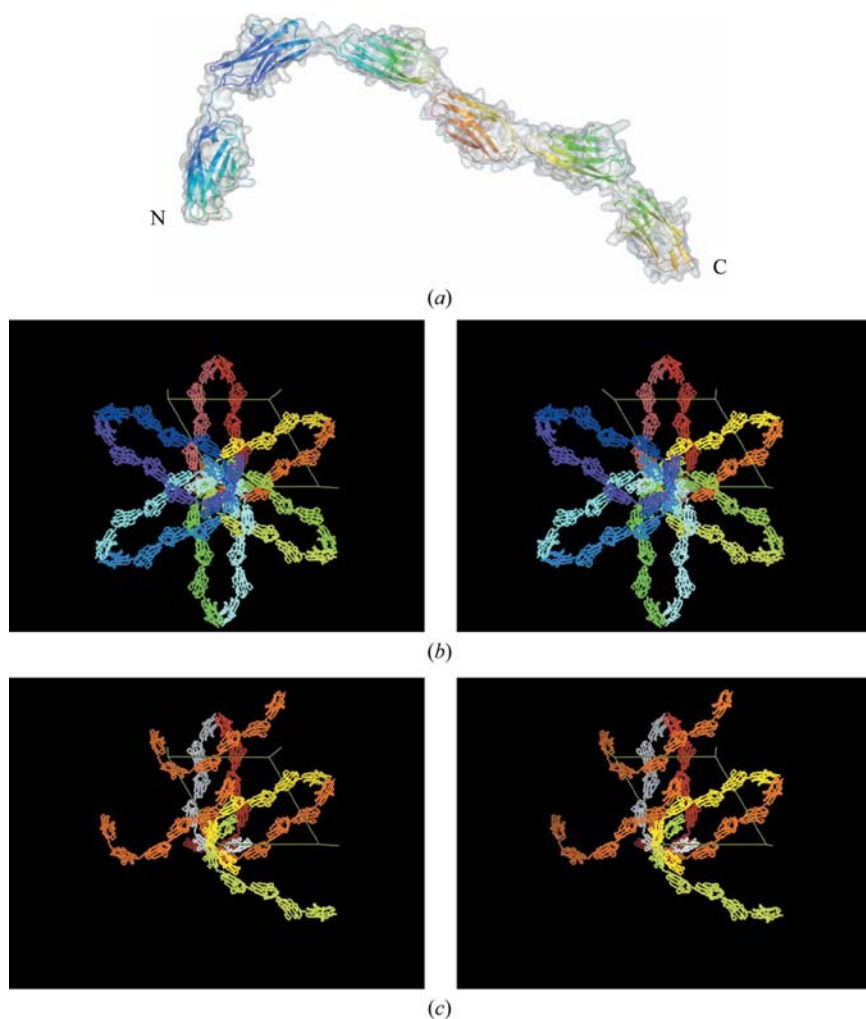

(a)

(b)

(c)

Figure 4
Crystallographic lattice of I65–I70 and molecular contacts. (a) Crystal structure of I65–I70 (PDB code 3b43, von Castelmur *et al.*, 2008); the molecular cartoon is coloured in a blue-to-red gradient according to model temperature factors, where blue represents the lowest values and thus indicates well ordered regions of the molecule; the molecular envelope calculated from the atomic model is shown in grey. (b) Stereoview of 12 molecules around the sixfold axis forming a left-handed helix; the colour, from red to blue, reflects the *z* position of the centre of the molecule. (c) Example of molecular contacts: an I65–I70 molecule (white) in contact with as many as three identical crystallographic copies (orange) that belong to three neighbouring unit cells (a stereoview); the 'head' of the molecule in white is inserted between the symmetry copies forming a left-handed molecular helix (in yellow and bottom-right in orange).

bent at its N-terminal 'head' composed of two Ig domains, while the C-terminal half has a straight appearance (Fig. 4a). In the crystal, each molecule is roughly contained in a plane normal to the *z* axis, in which the bent 'heads' wrap around the sixfold axis and the 'tails' extend away from it. The 'heads' stack laterally upon each other along the *z* axis, building a continuous intermolecular β-sheet that results in the formation of an infinite left-handed helix within the crystal (Fig. 4b). In this arrangement, the N-terminal 'heads' form a densely packed region with numerous intermolecular contacts. In contrast, the molecular 'tails' form a loosely packed region in the form of 'bridges' that connect the dense 'columns' centred at the sixfold axes. The unrestrained 'tails' of the molecule suffered from poor crystalline order (probably owing to chain dynamics), further increasing the contrast between the vast continuous zones of high density and the rest of the unit cell.

4.2. Analysis of low-resolution maps calculated with model phases

In order to analyze the origin of the difficulties in interpreting low-resolution maps, we calculated structure factors $F_{mod}\exp(i\varphi_{mod})$ from the atomic model and applied a bulk-solvent correction using the flat-mask model (Jiang & Brünger, 1994) and bulk-solvent values of $k_{bulk} = 0.32$ e Å^{-3} and $B_{bulk} = 70$ Å^2 (Fokine & Urzhumtsev, 2002; Afonine *et al.*, 2005). The phases φ_{cryst} of the resulting structure factors $F_{cryst} \exp(i\varphi_{cryst})$ were a better approximation to the unknown phases φ_{obs} associated with F_{obs} than the phases φ_{mod} merely calculated from the model. Three Fourier syntheses were then calculated at a resolution of 19 Å with coefficients $F_{mod}\exp(i\varphi_{mod})$, $F_{obs}\exp(i\varphi_{mod})$ and $F_{obs}\exp(i\varphi_{cryst})$. In the first synthesis, calculated using $F_{mod}\exp(i\varphi_{mod})$, the envelope with a unit-cell volume $0.30 \times V_{cell}$, roughly corresponding to the protein fraction suggested by the value of the Matthews coefficient V_M, included the atomic model reasonably well (only about 20% of the model was missed). At high cutoff levels, for example corresponding to $0.02 \times V_{cell}$, the peaks of the synthesis generally corresponded to the centres of individual Ig domains. However, syntheses calculated from experimental F_{obs} and model phases (φ_{mod} or φ_{cryst}) showed a deterioration of the image. The envelopes of a volume near $0.3 \times V_{cell}$ missed more atoms than the model synthesis, practically 40% for the synthesis with $F_{obs}\exp(i\varphi_{cryst})$. At higher cutoff levels, the density at the centres of individual Ig positions progres-

sively split or disappeared. In addition, spurious peaks appeared at the interface between Ig domains belonging to the same or different molecules.

A more detailed inspection of density peaks was performed in order to examine whether they presented the expected connectivity features, *e.g.* one peak per domain. This would mean the presence of six peaks of similar volume (individual Ig domains), each with 12 copies (reflecting the space-group symmetry). However, the Fourier syntheses calculated with the 'best possible' set of coefficients $F_{\mathrm{obs}}\exp(i\varphi_{\mathrm{cryst}})$ at a resolution of 19, 15 and 12 Å and explored using different cutoff levels did not reveal the desired features. In other words, even the best possible phases φ_{cryst} did not lead to molecular envelopes with optimal coverage of the model and with readily interpretable characteristics in terms of individual domains. Thus, we concluded that an interpretation of low-resolution maps in terms of a molecular structure of I65–I70 was unfeasible, even when using very accurate phases and at a resolution higher than that actually applicable to *ab initio* phasing in this study.

4.3. Comparison of results

The primary Fourier synthesis $\rho_{\mathrm{FAM19}}(\mathbf{r})$ obtained by *ab initio* phasing was that calculated with the coefficients $m_{\mathrm{FAM19}}F_{\mathrm{obs}}\exp(i\varphi_{\mathrm{FAM19}})$ at 19 Å resolution using experimental structure-factor magnitudes weighted by the corresponding figures of merit. In order to perform a comparison with the model, an optimal relative choice of the unit-cell origin had to be made as well as a decision on whether density $\rho_{\mathrm{FAM19}}(\mathbf{r})$ was to be flipped or not. Visual inspection as well as numerical analysis of two allowed origin shifts, namely $(0, 0, 0)$ and $(0, 0, \frac{1}{2})$, in both $\rho_{\mathrm{FAM19}}(\mathbf{r})$ and $-\rho_{\mathrm{FAM19}}(\mathbf{r})$ unambiguously indicated that a best superposition required shifting the origin of the *ab initio* calculated image onto the point $(0, 0, \frac{1}{2})$ but that density flipping was not required, confirming the likelihood-based choice of the sign of the density.

A superposition of the atomic model with $\rho_{\mathrm{FAM19}}(\mathbf{r})$ shifted by $(0, 0, \frac{1}{2})$ revealed that a molecular envelope of volume 0.30 $\times$ V_{cell} showed reasonably continuous molecular packing (Fig. 5a) and covered half of the model. The enlargement of the envelope volume to 0.6 $\times$ V_{cell} reduced the missed part of the model to 20%. The imperfection in the model coverage is not surprising since maps calculated using the model phases φ_{cryst} also failed to fully enclose the molecule (see §4.2). A synthesis calculated using final refined phases, $m_{\mathrm{conn12}}F_{\mathrm{obs}}\exp(i\varphi_{\mathrm{conn12}})$, at 19 Å showed that at the same cutoff level the molecular envelope became split into globules (Fig. 5b). However, these did not entirely correspond to individual Ig domains and this new envelope did not improve model coverage. At the higher resolution of 15 Å and using a cutoff volume of 0.15 $\times$ V_{cell} the envelope did not systematically correlate with individual domains either (Fig. 5c). In summary, *ab initio* phasing in this study succeeded in correctly identifying the volume of the unit cell occupied by the protein fraction but could not resolve individual molecules within the lattice or individual domain components within the molecule.

5. Low-resolution-based image improvement

During the current study, it became evident that the crystallographic case of I65–I70 was extremely unfavourable for low-resolution *ab initio* phasing. It was not possible to identify individual molecular envelopes, as was the case, for example, in Lunin *et al.* (2001), Fokine, Morales *et al.* (2003) and Müller *et al.* (2006), where the molecules or their core parts were more compact and had less pronounced intermolecular contacts. However, even the identification of the fraction of the unit-cell volume occupied by the ensemble of molecules (*i.e.* the protein mask) can be helpful in structure elucidation. For example, it can aid MIR-based experimental phasing, which traditionally ignores structure factors at a resolution below approximately 15–30 Å. For I65–I70, experimental phases became available in the resolution range 4.2–18.0 Å (von Castelmur *et al.*, 2008). However, the mean figure of merit of these phases was low for resolutions lower than 17 Å. We substituted the poorly defined MIRAS phases for *ab initio* phases in the resolution range 17–18 Å and completed the diffraction data down to 122 Å resolution using the low-resolution *ab initio* set including the calculated phases. Such composite map calculation, mixing phase information, can be performed at early stages of structure investigation. In the case of I65–I70, a Fourier synthesis with coefficients

$$\begin{cases} m_{\mathrm{conn12}}F_{\mathrm{obs}}\exp(i\varphi_{\mathrm{conn12}}), & 122.0\ \text{Å} > d > 17.0\ \text{Å} \\ m_{\mathrm{MIRAS}}F_{\mathrm{obs}}\exp(i\varphi_{\mathrm{MIRAS}}), & 17.0\ \text{Å} > d > 4.2\ \text{Å} \end{cases}$$

substantially improved the starting images derived from the experimental phases alone, exhibiting an enhanced solvent–protein delimitation (Fig. 6). This confirmed that the *ab initio* low-resolution phases contained correct and useful structural information, even when this was scarcely recognizable by visual inspection. Thus, it could be inferred that phase sets obtained through the combination of low-resolution *ab initio* approaches and other phasing methods might facilitate significantly the structure-elucidation process by aiding density-modification protocols (Podjarny *et al.*, 1981; Urzhumtsev, 1991). This would particularly overcome bottlenecks in those cases in which only phases of poor quality were obtained by conventional means.

6. Discussion

The six-Ig fragment I65–I70 from titin was extremely unfavourable for low-resolution *ab initio* studies. Given that low-resolution *ab initio* phasing has been applied successfully to a number of globular proteins, we saw the need to (i) understand the parameters that had hindered the application of the technique in this case and (ii) evaluate to what extent these parameters were generic for filamentous proteins and likely to affect the overall applicability of the technique to this class of proteins. In I65–I70, the irregularity of the crystal packing, the multi-domain composition, the large number of symmetry copies in the unit cell and the numerous intermolecular contacts of each copy worked together to compromise the success of the technique. Nonetheless, the essential difficulty

in this study lay in our inability to predict the 'correctness' of the *ab initio* phase sets and thus to implement suitable scoring functions. Current phase-evaluation protocols largely rely on

Figure 5
Stereoview of the superposition of the atomic model with the *ab initio* phased Fourier syntheses. The same view is shown as in Fig. 3. (*a*) Phase set $\{\varphi_{\text{FAM19}}\}$, resolution 19 Å, $0.30 \times V_{\text{cell}}$, contour level 0.6σ, (*b*) phase set $\{\varphi_{\text{conn12}}\}$, resolution 19 Å, $0.30 \times V_{\text{cell}}$, contour level 0.4σ, (*c*) phase set $\{\varphi_{\text{conn12}}\}$, resolution 15 Å, $0.15 \times V_{\text{cell}}$, contour level 0.8σ.

Figure 6
MIRAS-phased syntheses at 4.2 Å resolution; the volume of the selected region was $0.15 \times V_{\text{cell}}$. The same view is shown as in Figs. 3 and 5. (*a*) Experimental phases only, no phase information below 18 Å, contour level 0.8σ. (*b*) Experimental phases at 4.2–17.0 Å and phases from the *ab initio* set $\{\varphi_{\text{conn12}}\}$ at 17.0–122 Å, contour level 0.9σ.

the interpretability of the resulting maps in terms of individual molecular features and lattice connectivity, often by visu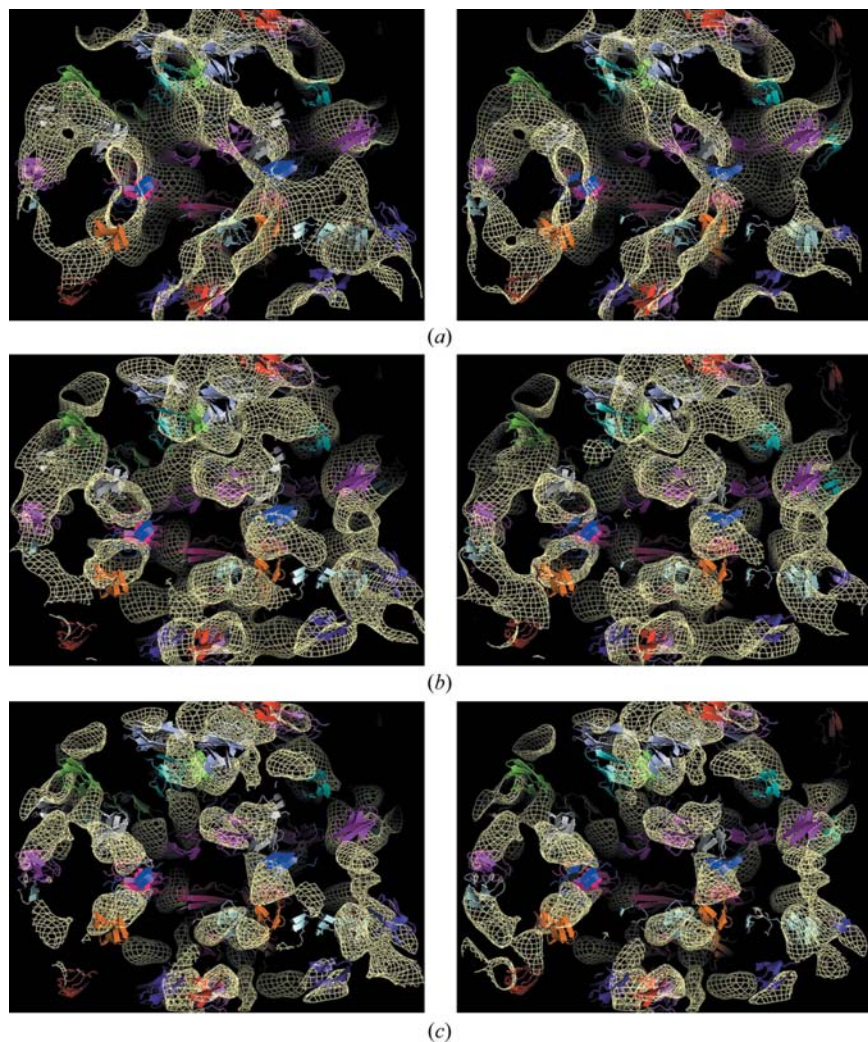al inspection. In the case of filamentous proteins, such map interpretation at ultralow resolutions can be unworkable. The reason for this is the small dimensions (particularly in cross-section) of individual molecules (or domains) and their tendency to form tight intermolecular contacts where the assembly area is large in relation to the molecular volume. This distorts the resulting images to the extent that lattice features strongly dominate any individual shapes. In the case of I65–I70, the interpretability of the molecular envelope in terms of structural features was precluded by the inability to identify individual Ig domains. The grounds for this appeared to be the dense molecular packing of multiple domains near the sixfold axis, which caused individual molecular envelopes not to be resolved and the peaks in the syntheses to become displaced from the domain centres toward interdomain positions. This complicated enormously the scoring of resulting phase sets, affected confidence estimations and led to the impossibility of defining correct phase-selection criteria that would allow extension of the phases to resolution limits suitable for the recognition of individual domains. An *a posteriori* analysis of Fourier syntheses calculated from model phases showed that the difficulty of map interpretation was high even when maps had been computed using phases derived from the model. Nevertheless, even under these circumstances, *ab initio* phasing was able to predict the phases for the lowest resolution reflections with reasonable accuracy, leading to the determination of the region of the unit cell occupied by the ensemble of I65–I70 molecules, *i.e.* the protein mask, and resulting in a significant improvement of experimental maps.

In summary, this study suggests that filamentous molecules are likely to be challenging targets for low-resolution *ab initio* methodologies, partly because of the lack of voluminous domains that are well resolved in space and of simple identification and partly because of the peculiarities of their atypical crystal lattices (Fig. 1*b*). This highlights the need to implement scoring parameters that incorporate lattice considerations into these calculations. However, even if the position of individual molecules within the unit cell may not be determined using this methodology, it can still lead to
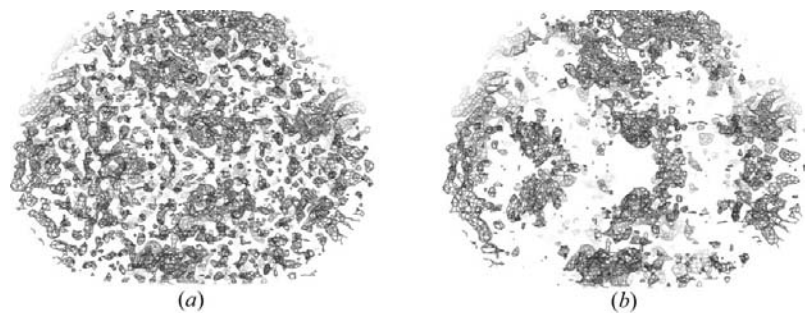
the successful determination of a lattice mask that can aid significantly in the structure-elucidation process of these molecules by improving the initial map quality.

APPENDIX A
Main features of low-resolution *ab initio* phasing

A1. *FAM* and connectivity-based *ab initio* phasing methods

The *FAM* protocol (Lunin *et al.*, 1995, 1998) generates multiple random models each composed of a small number of large spherical scatterers ('few-dummy-atoms models'). For each model, structure factors are calculated and their magnitudes are compared with the experimental values. Typically, several hundred sets exhibiting high correlation are selected. The corresponding phase values are then either averaged over all the sets or processed using clustering techniques (Lunin *et al.*, 1990). This procedure gives one or a few sets of phase values and corresponding figures of merit for each reflection. These are then associated with the experimental structure-factor magnitudes to compute a Fourier synthesis.

The protocol of 'image-connectivity' applies phase selection based on the connectivity features of high-density regions in Fourier syntheses. These syntheses are calculated using experimental structure-factor magnitudes and randomly generated phase sets (Lunin *et al.*, 2000). Here again one or a few possible phase sets are obtained per reflection by averaging selected phases. The two methods, which are complementary to a certain extent, have been shown to be capable of phasing several hundred of the lowest resolution structure factors and have been applied efficiently in combination to the *ab initio* phasing of large macromolecular complexes, such as, for example, LDL (Lunin *et al.*, 2001).

A2. Tools for *a posteriori* map scoring and analysis

The calculation of a conventional correlation coefficient is unsuitable for the comparative assessment of low-resolution electron-density maps. This is because the main goal of low-resolution phasing is to reproduce the molecular envelope or shape of the target protein and a parameter that merely quantitates the difference in low and high density values is irrelevant to molecular-shape definition. Thus, to evaluate the results of this study, we used the similarity of the envelopes as a function of their volume. For a given Fourier synthesis, an envelope is defined as a set of grid points whose value is superior to a given cutoff level. Such a level is usually quoted in different units according to the map resolution. Studies at subatomic resolution use the absolute scale of e Å^{-3}, while resolutions common in conventional macromolecular crystallography (1–4 Å) traditionally define cutoff levels in σ values. Neither of these units is convenient for low-resolution studies, where the sharpness of the maps varies strongly from one synthesis to another, especially when varying the resolution or weighting function. As a consequence, the same σ level may produce envelopes with significantly different volume that are difficult to compare directly. To make the comparison adequate, we chose cutoff levels such that a given volume

fraction of the unit cell V_{cell} was selected consistently across tests.

A3. General features of low-resolution images

The experience derived from conventional crystallographic studies at resolutions near 1–4 Å cannot be applied to the evaluation of low-resolution Fourier syntheses in a straightforward manner. The previous section illustrates the necessity of switching from σ units to fractional volumes, which indirectly reflects the fact that low-resolution maps contain a much weaker signal. It should be considered that low-resolution envelopes cover the molecular chains as well as the space between them, making the volume of such envelopes larger than the sum of strictly atomic volumes. While Fourier maps at usual resolutions are often displayed at the level of 1–2σ, this would produce unreasonably small envelopes in low-resolution phasing. Another important feature of low-resolution envelopes is that they cannot represent sharp molecular features well and thus cannot cover all macromolecular atoms (unless the fractional volume approaches 1) even when the synthesis is calculated with exact structure-factor values.

There is a common belief that low-resolution Fourier syntheses represent molecular envelopes when the cutoff level is relatively low and the centres of the molecules when this level is high. However, density peaks are often shifted from molecular centres toward regions of close intermolecular contacts. Moreover, owing to the relatively small number of reflections used in the calculation, a small change in phases can significantly modify the image and essentially the position of the peaks. It should also be borne in mind that bulk-solvent correction decreases the contrast between two neighbouring peaks and can result in their merging. Together with the difficulty of establishing reliable and unbiased assessment parameters, the shifting of features complicates the evaluation of results enormously.

Finally, it should be noted that an increase in the resolution of the Fourier maps is rarely able to resolve the arising conflicts. On the contrary, an increase in resolution from 20–25 to 10–12 Å often makes maps even less suitable for visual inspection since they stop showing molecular-envelope features but are not yet capable of revealing secondary-structure elements.

A4. Phase-ambiguity problems and their solution

There are several possible transformations of an electron density $\rho(\mathbf{r})$ that do not change the structure-factor magnitudes but that do affect the calculated phases and images. During structure elucidation, choices of corresponding parameters have to be made *a priori* and arbitrarily, although this does not preclude the phasing process. One of these transformations refers to the shift $\mathbf{t}$ of the origin permitted for a given space group (Lunin & Lunina, 1996), which modifies electron density according to $\rho(\mathbf{r}) \rightarrow \rho(\mathbf{r} - \mathbf{t})$. This does not change the image itself, but should be taken into account during map alignment in comparative processes. In space

groups $P6_522/P6_122$ the only permitted shift is $\mathbf{t} = \frac{1}{2}\mathbf{c}$, where $\mathbf{c}$ is the corresponding base vector of the unit cell.

The enantiomer density transformation $\rho(\mathbf{r}) \rightarrow \rho(-\mathbf{r})$ conserves the connectivity features of the Fourier maps and the magnitudes of structure factors and substitutes the phases φ by $-\varphi$. In the current study, since $P6_522$ and $P6_122$ are enantiomeric space groups, this means that phasing cannot identify the true space group. It is not possible to resolve this ambiguity at low resolution and high-resolution images are required to reveal the 'hand' through the features of the secondary-structure elements. Thus, for this *ab initio* study we arbitrarily chose space group $P6_522$. This was shown at a later stage to be the correct symmetry. Although this facilitated *a posteriori* comparison with the model, it did not influence the phasing process itself.

Finally, a density transformation $\rho(\mathbf{r}) \rightarrow -\rho(\mathbf{r})$ keeps the magnitudes, changes the phases by π and changes the envelopes for a given cutoff level. Again, in contrast to Fourier maps at conventional resolutions, the overall features of the flipped map at ultralow resolution are very similar to those of the direct map. This might make it difficult to visually identify the correct sign of the image. In order to assess which of the two molecular envelopes is more probable, $\rho(\mathbf{r})$ or $-\rho(\mathbf{r})$, a generalized maximum-likelihood technique can be used (Lunin *et al.*, 1998; Petrova *et al.*, 2000). If two (or several) possible molecular envelopes are known, this technique randomly generates multiple models consisting of the same number of dummy atoms inside each of the envelopes. These models are composed of a relatively large number of identical scatterers of a size (determined by the *B* value) much smaller than that of spheres used for *FAM* phasing. For each model, the correlation of its structure-factor magnitudes with the experimental data is calculated. The envelope for which the number of generated models with high correlation is largest is considered to be the most likely to be correct. Since phasing at low resolution requires a complete set of data at a given resolution, such a control set now includes reflections of a resolution slightly higher than that used for phasing. Thus, the correlation is calculated for structure factors excluded from previous calculations.

References

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850–855.

Castelmur, E. von, Marino, M., Svergun, D. I., Kreplak, L., Labeit, D., Ucurum-Fotiadis, Z., Konarev, P. V., Urzhumtsev, A., Labeit, S. & Mayans, O. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 1186–1191.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.

Fokine, A., Morales, R., Contreras-Martel, C., Carpentier, P., Renault, F., Rochu, D. & Chabriere, E. (2003). *Acta Cryst.* D**59**, 2083–2087.

Fokine, A. & Urzhumtsev, A. (2002). *Acta Cryst.* D**58**, 1387–1392.

Granzier, H. L. & Labeit, S. (2004). *Circ. Res.* **94**, 284–295.

Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.

Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.

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

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

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

Lunin, V. Y., Lunina, N. L., Ritter, S., Frey, I., Berg, A., Diederichs, K., Podjarny, A. D., Urzhumtsev, A. & Baumstark, M. W. (2001). *Acta Cryst.* D**57**, 108–121.

Lunin, V. Y., Lunina, N. L. & Urzhumtsev, A. G. (2000). *Acta Cryst.* A**56**, 375–382.

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

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

Marino, M., Svergun, D. I., Kreplak, L., Konarev, P. V., Maco, B., Labeit, D. & Mayans, O. (2005). *J. Muscle Res. Cell Motil.* **26**, 355–365.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Müller, J. J., Lunina, N. L., Urzhumtsev, A., Weckert, E., Heinemann, U. & Lunin, V. Y. (2006). *Acta Cryst.* D**62**, 533–540.

Petrova, T. E., Lunin, V. Y. & Podjarny, A. D. (2000). *Acta Cryst.* D**56**, 1245–1252.

Podjarny, A. D., Schevitz, R. W. & Sigler, P. B. (1981). *Acta Cryst.* A**37**, 662–668.

Strelkov, S. V., Schumacher, J., Burkhard, P., Aebi, U. & Herrmann, H. (2004). *J. Mol. Biol.* **343**, 1067–1080.

Strelkov, S. V., Tao, Y., Rossmann, M. G., Kurochkina, L. P., Shneider, M. M. & Mesyanzhinov, V. V. (1996). *Virology*, **219**, 190–194.

Urzhumtsev, A. G. (1991). *Acta Cryst.* A**47**, 794–801.

Urzhumtsev, A. & Urzhumtseva, L. (2002). *J. Appl. Cryst.* **35**, 750.