Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Macromolecular crystal data phased by negative-stained electron-microscopy reconstructions

Stefano Trapani,[a,b]* Guy Schoehn,[c,d,e,f] Jorge Navaza[c,d,e] and Chantal Abergel[g]

[a]Université de Montpellier 1; Université de Montpellier 2; CNRS; UMR 5048, Centre de Biochimie Structurale, 29 Rue de Navacelles, 34080 Montpellier, France, [b]Inserm, Unité 554, Montpellier, France, [c]CEA, Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075, Grenoble, France, [d]CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075, Grenoble, France, [e]Université Joseph Fourier, Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075, Grenoble, France, [f]Unit for Virus Host Cell Interactions, UMI 3265, 6 Rue Jules Horowitz BP 181, 38042 Grenoble, France, and [g]Information Génomique et Structurale, CNRS-UPR2589, Institut de Microbiologie de la Méditerranée - IFR 88, 163 Avenue de Luminy, 13288 Marseille, France

Correspondence e-mail: stefano.trapani@cbs.cnrs.fr

Received 18 December 2009
Accepted 22 January 2010

PDB Reference: *C. albicans* type II dehydroquinase, 3kip.

The combination of transmission electron microscopy with X-ray diffraction data is usually limited to relatively large particles. Here, the approach is continued one step further by utilizing negative staining, a technique that is of wider applicability than cryo-electron microscopy, to produce models of medium-size proteins suitable for molecular replacement. The technique was used to solve the crystal structure of the dodecameric type II dehydroquinase enzyme from *Candida albicans* (~190 kDa) and that of the orthologous *Streptomyces coelicolor* protein.

1. Introduction

Electron microscopy (EM) and macromolecular crystallography (MX) both allow the retrieval of three-dimensional images (electron-density maps) of biological structures. On the one hand, transmission EM data provide relatively easy access to three-dimensional reconstructions of large biological complexes at low resolution. However, gains in resolution require a lot of effort owing to the radiation-sensitivity and conformational heterogeneity of the samples. Standard values range between 30 and 10 Å, although a number of structures, especially those of molecules with high internal symmetry, are known at higher resolution (better than 4 Å). On the other hand, MX delivers atomic or quasi-atomic resolution structures (1–3 Å). The limiting factors for this technique are the production of diffracting crystals and the phasing of the diffraction spectra, making it less appropriate for large structures. As a consequence EM and MX are complementary techniques; nowadays, they are often combined to address increasingly complex biological assemblies.

In one approach, the resolution of EM reconstructions is improved by fitting atomic structures produced by MX into EM maps (Wriggers *et al.*, 1999; Navaza *et al.*, 2002), thus providing pseudo-atomic information about subdomain interactions in large molecular assemblies such as ribosomes, viruses and microtubules. There have been considerable advances in fitting atomic level structures directly to EM reconstructions (Cheng *et al.*, 2009). In the reverse approach, EM reconstructions are used to solve MX structures by molecular replacement (MR; Dodson, 2001; Murshudov *et al.*, 2008).

Usually, MR procedures use atomic models of similar structures from the PDB to find the arrangement of the unknown molecules within the crystal (Rossmann, 1972). The correct positions and orientations of the models inside the unit cell are exhaustively searched for using scoring functions based on the agreement between the calculated and the

experimental diffraction amplitudes. Once the models have been positioned, theoretical phases can be calculated; they are relevant up to a resolution that depends on the similarity of the search models.

Alternatively, three-dimensional EM reconstructions can be used as MR probes provided that there is sufficient resolution overlap between the X-ray and EM data. Low-resolution X-ray data are often missing owing to instrumental settings (beam-stop size) or to strategies aimed at collecting high-resolution data at the expense of overloading low-resolution data. Although it is possible to measure data at quite low resolution, the primary difficulties with this type of phasing is that low-resolution X-ray terms are affected by bulk-solvent corrections, which are often not accurately defined. Again, once the search EM map has been positioned, theoretical phases can be calculated up to the EM reconstruction resolution. Phase-extension procedures (see, for example, Vellieux & Read, 1997; Terwilliger, 2000) are then needed to solve the structure at the resolution provided by the X-ray diffraction data. This strategy has been successfully applied to the study of rather large and highly symmetrical (icosahedral) viral particles (Rossmann, 1995; Duquerroy *et al.*, 2009; Zubieta *et al.*, 2005); to our knowledge it has never been used for particles smaller than 300 kDa for which cryo-EM is inappropriate.

In this article, we demonstrate that this strategy can also be used to study smaller symmetrical proteins using negative-stained EM reconstructions as MR probes. The EM/MR approach is particularly appropriate for the study of homomultimers because the internal symmetry both increases the size of the molecule, making it amenable to EM analysis, and provides the noncrystallographic symmetry (NCS) needed to extend the EM-based phases to higher resolution.

We have successfully applied the method to solve the crystal structures of type II dehydroquinase (DHQ) from *Candida albicans* (*Ca*) and *Streptomyces coelicolor* (*Sc*), two homologous dodecameric proteins of about 190 kDa that share 48% identity over 131 amino acids. While the *Ca*DHQ crystal structure is new, we used the already published *Sc*DHQ diffraction data (PDB entry 1v1j; Frederickson *et al.*, 2004) to validate the procedure.

Type II DHQs are enzymes that are involved in the shikimate pathway, a central biosynthetic route which uses erythrose 4-phosphate and phosphoenolpyruvate to produce chorismate, a precursor for the synthesis of aromatic amino acids, folic acid, ubiquinone and many other aromatic compounds. This pathway is found in bacteria, higher plants and fungi (Roberts *et al.*, 2002). *Ca*DHQ is a dodecameric protein and was studied as part of a structural genomics project (PROFUN) aimed at discovering new antifungal compounds.

2. Materials and methods

2.1. *Ca*DHQ cloning, expression and purification

PCR amplification was performed as described previously (Santini *et al.*, 2008) using primers with specific sequences for the *Ca*DHQ gene preceded by 5′-CATCACCATCAATTG (direct primer) and 5′-TCACCATCCAATTG (reverse primer) and purified *C. albicans* genomic DNA as template (*C. albicans* strain NIH 3147; ATCC No. MYA-2876D). Gene cloning was performed using the ligation-independent cloning method and a specific expression vector pSF-04 (Protein'eXpert; Santini *et al.*, 2008). This cloning procedure allowed the encoding gene to be placed in frame with an N-terminal His_6 tag followed by GHHHQL corresponding to the N-terminal annealing sequence and a C-terminal fragment QLDGDLEAA corresponding to a linker for the expression of the target fused to GFP (Santini *et al.*, 2008).

Expression screening was performed using a standard procedure (Abergel *et al.*, 2003). As a result, the plasmids carrying the *Ca*DHQ gene were overexpressed in *Escherichia coli* Rosetta in 1 l flasks containing TB medium at 298 K for 6 h after induction with 500 µ*M* IPTG at $OD_{600\,nm}$ = 0.5. After centrifugation, the pellets were resuspended in 50 m*M* NaH_2PO_4, 300 m*M* NaCl pH 8.0 (buffer *A*) with 5% glycerol and 0.1% Triton X-100, sonicated and centrifuged again.

The cleared lysate was applied onto a 5 ml HiTrap Chelating Column (GE Healthcare) charged with Ni^{2+} and equilibrated with buffer *A*. The column was washed with ten column volumes of buffer *A*, ten column volumes of buffer *A* containing 25 m*M* imidazole and five column volumes of buffer *A* containing 50 m*M* imidazole at a flow rate of 1 ml min^{-1}. Elution was performed with a linear gradient over seven column volumes from 50 to 500 m*M* imidazole. The fractions corresponding to the elution of *Ca*DHQ with 150–200 m*M* imidazole were run on a desalting column (Fast Desalting Column HR 10/10, Pharmacia) and analysed by mass spectroscopy and N-terminal Edman sequencing. After purification, the fractions contained at least 97% pure protein in 50 m*M* TEM buffer (50 m*M* Tris, 1 m*M* EDTA, 1 m*M* β-mercaptoethanol) pH 7.5.

2.2. *Ca*DHQ crystallization

The recombinant *Ca*DHQ protein was concentrated to 15.4 g l^{-1} in 50 m*M* TEM buffer pH 7.5 using a centrifugal filter device (Ultrafree Biomax 10K, Millipore, Bedford, Massachusetts, USA). Precipitation experiments were carried out at 293 K using various precipitating agents (*i.e.* ammonium sulfate, PEG, NaCl and MPD) at various pH values (*i.e.* 5, 6, 7, 8 and 9) to determine the optimal protein concentration for crystallization. Screening for crystallization conditions was performed on 3 × 96-well crystallization plates (Greiner), which were loaded with a 1 µl sitting drop per condition at 293 K using an eight-needle dispensing robot (Tecan WS 100/8 workstation modified for our needs). The 576 crystallization conditions tested included in-house-designed (Abergel *et al.*, 2003) and commercially available solution sets (Structure Screen from Molecular Dimensions Ltd and Wizard Screens from Emerald BioStructures). After refinement of the first crystallization condition, the best crystals were obtained using 10–20%(*w*/*v*) PEG 8000, 0.2 *M* lithium sulfate, 0.1 *M* HEPES buffer pH 7.0.

2.3. *Ca*DHQ X-ray diffraction and data collection

A previously described evaporation protocol (Abergel, 2004) was used to improve the crystal diffraction from almost nothing to a resolution of 3 Å. Crystals were soaked for 15–30 min in reservoir solution containing 10% glycerol as a cryoprotectant and were collected in Hampton Research loops, flash-frozen to 105 K in a cold nitrogen-gas stream and subjected to X-ray diffraction. Diffraction images were collected on the Swiss Light Source X06SA beamline. *MOSFLM* (Leslie, 1992) and *SCALA* (Evans, 1997) from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) were used for diffraction-image processing, scaling and data reduction (Table 1).

Table 1
X-ray data-collection and refinement statistics for *Ca*DHQ (PDB entry 3kip).

Values in parentheses are for the highest resolution shell.

Data collection	
Space group	$P2_12_12$
Unit-cell parameters (Å, °)	a = 159.10, b = 308.11, c = 97.15, α = 90.0, β = 90.0, γ = 90.0
Resolution (Å)	70.68–2.95 (3.11–2.95)
$R_{\rm merge}$ (%)	13.6 (53.9)
$\langle I/\sigma(I)\rangle$	4.3 (1.4)
Completeness (%)	99.9 (100.0)
Multiplicity	6.4 (6.5)
Observations	647771 (94461)
Refinement	
Resolution (Å)	70.68–2.95 (3.03–2.95)
No. of reflections	101228
$R_{\rm work}/R_{\rm free}$ (%)	20.3/24.5 (35.0/40.1)
No. of atoms	
Protein	26792
Ligand/ion	184
Water	168
B factors (Å^2)	
Protein	59.2
Ligand/ion	83.1
Water	45.7
R.m.s. deviations	
Bond lengths (Å)	0.025
Bond angles (°)	2.3

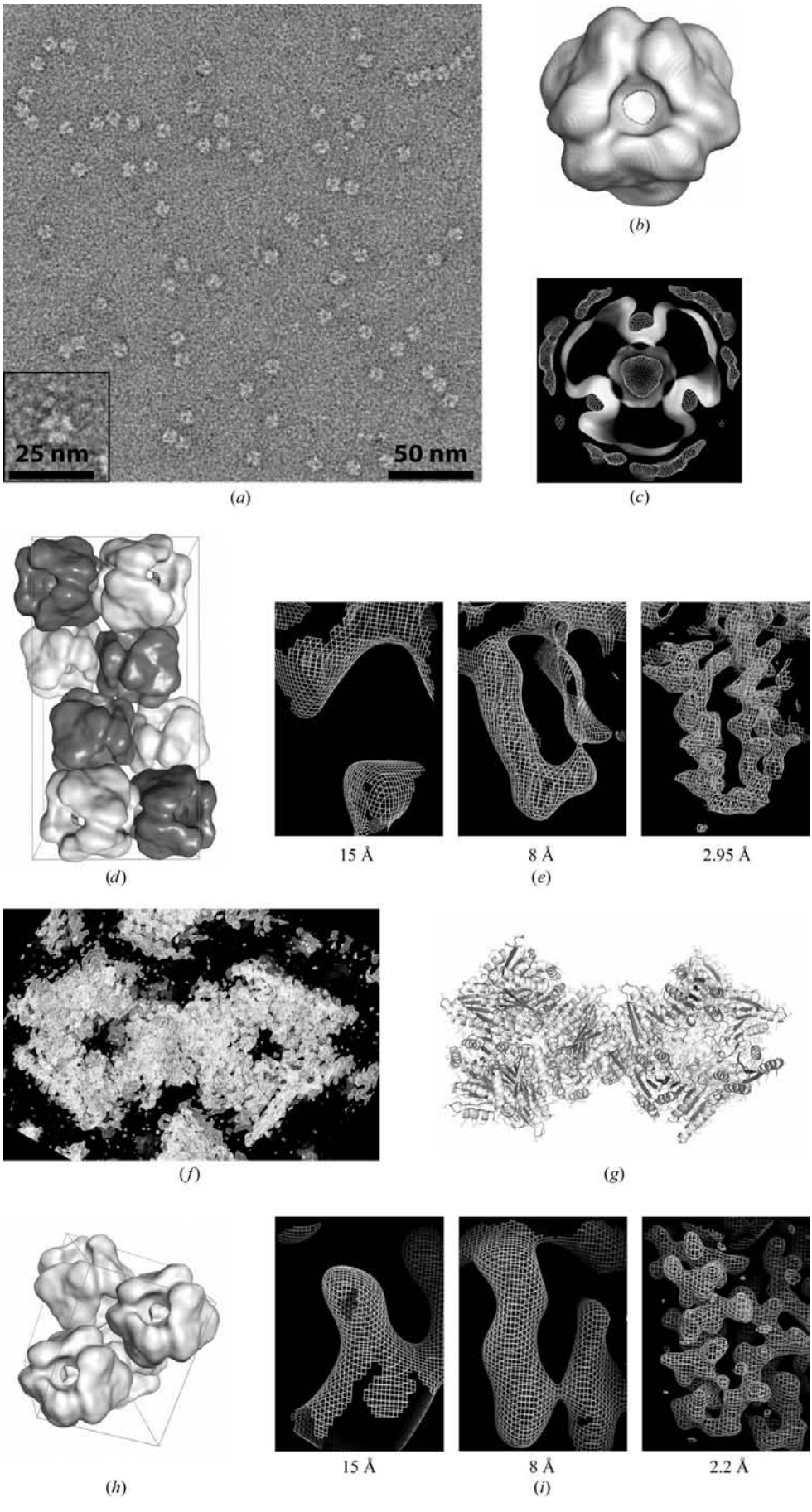
2.4. Electron microscopy

*Ca*DHQ samples at about 0.1 mg ml^{-1} in 50 m*M* TEM buffer pH 7.5 were negatively stained with 1% methylamine vanadate ('NanoVan', Nanoprobes Inc., Stony Brook, New York, USA). Samples were imaged under low-dose conditions at a magnification of 50 000 at 100 kV using a Jeol 1200EX-II electron microscope. Images were recorded on Kodak SO163 EM films and high-quality negatives were digitized with a 7 µm sampling step on a Z/I Imaging PhotoScan (1.4 Å on the sample's scale). The image-analysis procedure was started by manually selecting 12 635 particles in 128 × 128 pixel squares from eight micrographs using *X*3*D* (Conway & Steven, 1999). For each micrograph we computed the power spectrum and used a full search method to determine the best defocus and the best amplitude contrast by applying a root-mean-square method (SUMPS and CTFZEROS; Conway & Steven, 1999). In all micrographs the amplitude contrast was ∼25% and the defocus values ranged between 1 and 2.5 µm. These images were then corrected for the contrast transfer function (CTF) as described by Conway & Steven (1999). The fully deconvoluted images (used for the final reconstruction) and the phase-flipped images (used for the projection-matching cycle) were band-pass filtered between 200 and 8 Å, and between 200 and 15 Å, respectively. Both stacks of images were subsequently transferred to *SPIDER* (Frank *et al.*, 1996) and normalized to the same mean and standard deviation values.

We used the phase-flipped set of particles to start the reconstruction. A starting model was created using one centred raw image, highlighting a clear threefold axis. This single image was back-projected using threefold symmetry in a 128 × 128 × 128 box and the height of the reconstruction was limited to 1.5 times the apparent diameter of the particle (Franzetti *et al.*, 2002). This starting model was then used in the projection-matching method implemented in *SPIDER* (Franzetti *et al.*, 2002; Frank *et al.*, 1996). After 15 cycles of refinement (projection of the model, comparison with the raw images, class averaging, three-dimensional reconstruction by back-projection) and by imposing a threefold symmetry, the tetrahedral shape of the reconstruction became obvious. Owing to the tetrahedral shape of the threefold reconstruction and since the projections of the tetrahedral structure fitted very well in the raw images, we imposed full tetrahedral symmetry on the reconstruction (point group 32). This model was projected in 190 orientations equally distributed in the tetrahedral asymmetric unit and ten extra cycles of refinement were performed using the same projection-matching method as described above. 60% of the images (7200) that had the highest correlation coefficients with the projections of the model were included in the reconstruction. The final reconstruction was calculated using the full CTF deconvoluted set of images. The resolution limit was estimated by randomly splitting the data into two halves and by calculating reconstructions from each half of the data set. These reconstructions were then compared by Fourier shell correlation (FSC) and gave resolutions of 15.6 and 14.6 Å for a 0.5 and a 0.3 correlation cutoff, respectively. For the isosurface representation, we included the correct molecular weight of the molecule using an average protein density of 0.84 Da Å^{-3}.

2.5. Molecular replacement

We used the *AMoRe* software (Navaza, 1994) for molecular replacement using the electron-density map produced by the above procedure and the crystallographic data to 15 Å resolution. Negative-stained three-dimensional reconstructions have significant negative regions at the surface of the particles (shown as meshed contours in Fig. 1*c*) where the heavy-metal salts are deposited. In order to match the average density in the crystal, the values of the scattering density function in the surface regions and beyond were set to the average background value of the reconstruction. The high-resolution part of the model's spectrum was enhanced by dividing the Fourier coefficients by an appropriate Gaussian temperature factor (2000 Å^2). The modified EM reconstruction was used as the search model for rotation and translation searches followed by



rigid-body refinement. *Ca*DHQ crystals contain two independent molecules. No contrasted solution appeared for the first molecule. In order to assess many of the one-body putative solutions, the information provided by the self-rotation function was used. For each position of the first molecule, the orientation of the second molecule was restricted to that obtained by application of the self-rotation to the first molecule orientation. The resolution range was tested and the 25–15 Å resolution range gave the best results. In this way, a contrasted two-body solution appeared for the *Ca*DHQ crystal (correlation coefficient of 40.6%, compared with 36.4% for the first wrong position). For *Sc*DHQ a unique dodecamer had to be positioned and MR was straightforward using all data between 33 and 15 Å resolution, providing a contrasted solution with a correlation coefficient of 59.3% after rigid-body refinement (the correlation coefficient of the first wrong position was 50.1%).

Figure 1
Structure-determination process. (*a*) Negative-stained electron micrograph of the *Ca*DHQ protein. Inset: detailed view of one particle in the same orientation as in (*b*). (*b*) Three-dimensional EM reconstruction of the tetrahedral *Ca*DHQ dodecamer at 15 Å resolution. (*c*) Contour levels of a section of the *Ca*DHQ three-dimensional EM reconstruction; the meshed patches (-1σ) correspond to regions where the staining agent is deposited; the continuous isosurface ($+1\sigma$) corresponds to the molecular envelope. (*d*) Packing of the two independent dodecameric molecules in the *Ca*DHQ crystal. (*e*) Detailed view of the electron density of the *Ca*DHQ molecule in the initial, intermediate and final steps of phase extension. (*f*) View of the final electron-density map of *Ca*DHQ at 2.95 Å resolution (covering approximately one asymmetric unit of the crystal). (*g*) Result of the automatic tracing of the *Ca*DHQ structure. (*h*) Packing of the *Sc*DHQ dodecamer corresponding to the MR solution. (*i*) Detailed view of the corresponding electron-density maps during phase extension.

2.6. Phase extension

Theoretical X-ray data phases and amplitudes to 15 Å resolution were calculated for *Ca*DHQ and *Sc*DHQ based on the EM model and the MR positions. Phase extension to higher resolution was achieved by iterated density-modification procedures comprising NCS map averaging, solvent flattening, NCS-operator refinement and mask refinement as described below. We used a combination of programs from the *CCP*4 (Collaborative Computational Project, Number 4, 1994; Agarwal, 1978; Ten Eyck, 1973; Read, 1986) and *RAVE* (Jones, 1992; Kleywegt & Jones, 1994, 1996, 1999) packages.

The NCS operators relating one DHQ dodecamer to itself (proper NCS) or to other dodecamers (improper NCS) within the same crystal asymmetric unit were calculated by composing the symmetry operations of the tetrahedral EM reconstruction (the reference orientation) with the model rotation/translation operators obtained by *AMoRe*. For *Ca*DHQ (two independent dodecamers in the crystal) this gave 12 proper NCS operators and one improper NCS operator and for *Sc*DHQ (only one dodecamer per crystal asymmetric unit) this gave 12 proper NCS operators.

Molecular masks for one dodecamer within the crystal were generated after visual inspection of the initial ($F_{\rm calc}$, $\varphi_{\rm calc}$) Fourier maps. The molecular boundaries were defined (using the program *MAPMAN*) by choosing the lowest possible density contour that still showed globular particles that were not connected to their crystallographic neighbours and by wiping out (using the program *MAMA*) small unconnected blobs in the solvent region. In order to avoid accidental density flattening within protein regions, the large central holes of the masks were filled in and the masks were further smoothed, taking care to avoid overlaps between neighbouring particles. After an initial phase-extension procedure to the highest resolution available (2.95 and 2.2 Å for *Ca*DHQ and *Sc*DHQ, respectively), improved masks were generated. The new masks were based on visual inspection of correlation maps obtained by calculating for each sampled point the correlation between the electron density around the sampled point with the density around NCS-related points (using the program *COMA*). The phase-extension procedure was then repeated starting at 8 Å and using the refined masks.

For density averaging and solvent flattening ($2DF_{\rm obs} - mF_{\rm calc}$) difference Fourier maps were used (grid spacing of 1.4 and 1.1 Å for *Ca*DHQ and *Sc*DHQ, respectively). Firstly, the electron density related by improper NCS operators, if present, was averaged within the mask of one dodecamer. The averaged density within the mask was then further averaged following the 12 proper NCS operators. The averaged density within the mask was then expanded to a whole unit cell by improper NCS and crystallographic symmetry. The density in the remaining regions was flattened to a mean background value.

The NCS operators were regularly refined during the phase-extension procedure by optimizing the correlation between

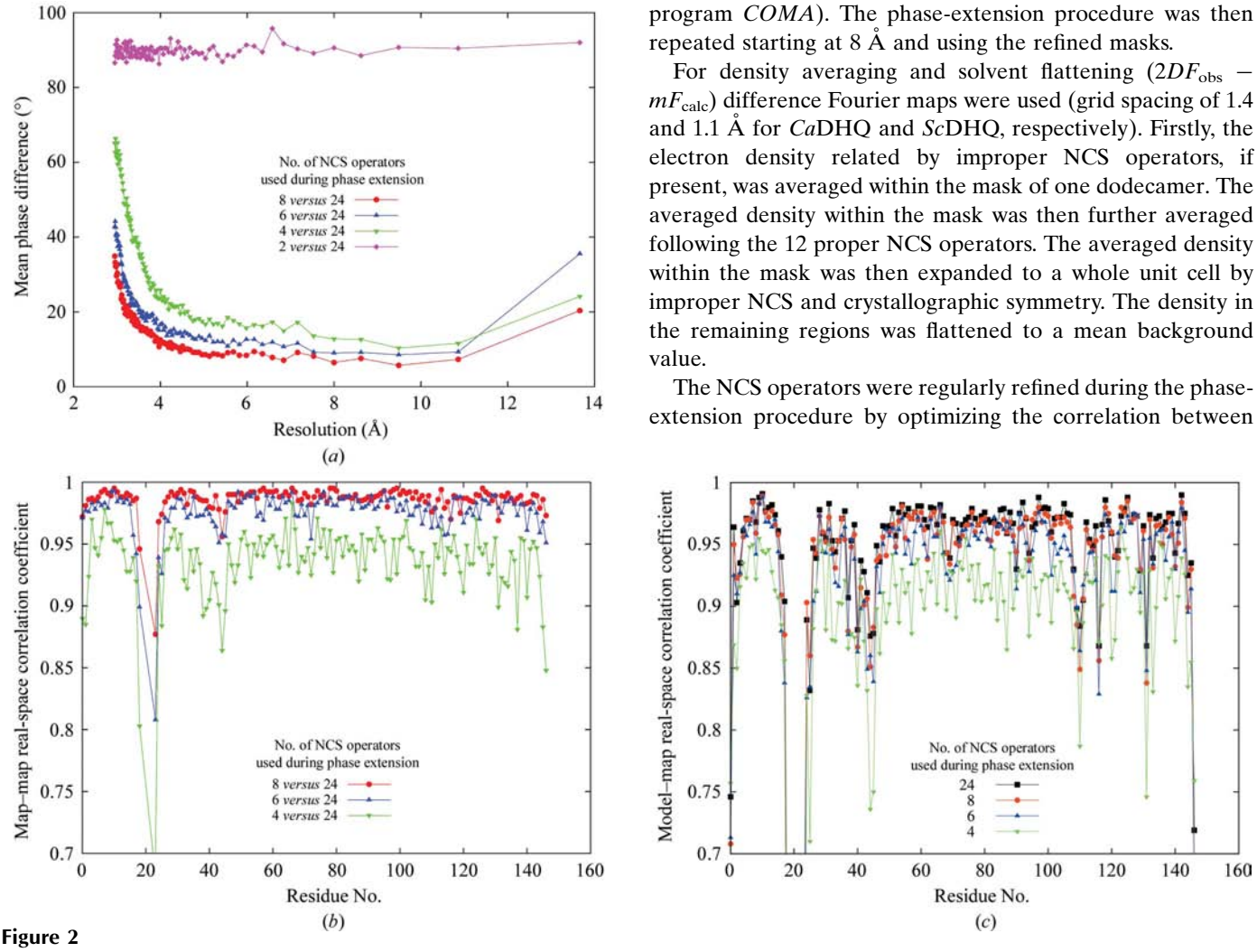


Figure 2
Analysis of the phase-extension results obtained for *Ca*DHQ by applying different sets of NCS operators. (*a*) Distribution of phase differences as a function of resolution; the reference phase set is that obtained by fully applying the 24-fold NCS of *Ca*DHQ. (*b*) Local real-space correlation between ($2DF_{\rm obs} - mF_{\rm calc}$) maps calculated at 2.95 Å resolution; the reference map corresponds to the reference phase set in (*a*). (*c*) Local real-space correlation between difference Fourier maps and the refined *Ca*DHQ model. Real-space correlation in (*b*) and (*c*) is calculated on a per-residue basis over a region corresponding to one *Ca*DHQ monomer (chain *A*). Similar plots were obtained for the remaining 23 monomers present in one crystal asymmetric unit.

Table 2
Analysis of phase-extension results obtained for *Ca*DHQ by applying different sets of NCS operators.

NCS redundancy	No. of cycles per resolution step	Interpretable map	Analysis of NCS-related density correlation (%)	*R* factor (%)	Real-space correlation (%)	
					With data phased by full NCS	With final refined model
24	17.2	Yes	85.7–93.4	54.1–35.3	100.0	74.0
8	28.6	Yes	87.2–94.3	51.1–35.3	97.3	72.9
6	47.4	Yes	87.1–94.9	51.4–33.7	95.1	73.0
4	132.0	Yes	89.1–94.6	47.5–34.4	91.2	69.5
2	174.6	No	91.3	41.6	0.0	0.0

the electron density within the mask and that of NCS-related regions (using the program *IMP*).

The phase-extension protocol consisted of iterated steps of (i) map calculation (using *SIGMAA* and *FFT*), (ii) refinement of the NCS operators (*IMP*), (iii) NCS averaging and solvent flattening (*AVE*) and (iv) structure-factor amplitude and phase calculation based on the modified map (*SFALL*) at the current resolution or at a slightly higher resolution (phase extension).

A rather conservative protocol was used with resolution steps of 1.167×10^{-2} Å^{-1} in reciprocal space (~0.7/mask radius) and iterated density averaging up to phase convergence (average phase shift of <0.01°) at each resolution step.

For *Ca*DHQ a total of 693 averaging cycles were necessary for a first extension from 15 to 2.95 Å. After mask refinement (see above) the whole procedure was repeated starting at 8.0 Å and required 354 averaging cycles. The total calculation time was about 4.75 h on a Linux PC with an Intel Xeon 3.2 GHz bi-processor and 6 Gb RAM. The mean number of iterations per resolution step was 17.2. The final NCS density correlation was 85.7% for the improper NCS operator and 93.2–94.4% for the proper NCS operators. The *R* factors were 54.1% for improper NCS and 35.3–38.5% for proper NCS.

For *Sc*DHQ the convergence of the procedure was slower, which was certainly a consequence of the lower number of NCS operations available. The first extension from 15 to 2.2 Å required 1178 iterations, while the second run, following mask refinement and starting at 8.0 Å, required 652 iterations. The total calculation time was 8.5 h and the mean number of iterations per resolution step was 23.7. The final NCS density correlation ranged between 90.1% and 94.7%, while the *R* factors were between 35.2% and 46.2%.

Further improvement of phase quality was achieved using the *RESOLVE* program (Terwilliger, 2000, 2003) during the subsequent automated model-building protocol.

In order to explore the NCS requirements for convergence of the phase-extension procedure, the set of proper NCS operators of *Ca*DHQ was reduced to that of point groups 222, 3, 2 and 1. The phase-extension protocol was then repeated, maintaining the intermolecular NCS operator. This resulted in an NCS redundancy of eight, six, four and two, respectively. In all but the last case the procedure converged to interpretable maps. The maps were compared (Table 2, Fig. 2) with the final refined model of *Ca*DHQ and with the map previously obtained by fully exploiting the 24-fold NCS. Real-space correlation coefficients were calculated (*MAPMAN*) both on the whole unit cell and on a per-residue basis.

2.7. *Ca*DHQ model building and refinement

An 80% complete polypeptide model of *Ca*DHQ was automatically built by *RESOLVE* (Terwilliger, 2000, 2003) using the 2.95 Å phased crystallographic data set. The iterative protocol included statistical density-modification procedures, model building, partial refinement (*REFMAC5*; Murshudov *et al.*, 1997) and model rebuilding using prime-and-switch composite OMIT maps. A total of eight main cycles of the protocol were run (7 d calculation time on our PC; see above), with the best model being obtained after cycle 5 (3231 of 4008 residues built, 3153 of them with complete side chains, R = 27.1%, $R_{\rm free}$ = 30.6%).

The model was further modified and completed by visual inspection of the maps, manual intervention (*Coot*; Emsley *et al.*, 2010) and automatic refinement of TLS parameters (one TLS group for each protein monomer), atomic coordinates and atomic *B* factors using a maximum-likelihood target with stereochemical and NCS restraints. The NCS restraints were gradually relaxed and completely eliminated in the final refinement steps. The final model (Table 1) had crystallographic residuals of R = 20.3% and $R_{\rm free}$ = 24.5%, had 96% of the modelled residues in allowed Ramachandran regions and comprised 86% of the expressed protein sequence.

The N-terminal and C-terminal expression tags were not visible in the electron-density maps. The region of residues 19–22 comprises a flexible loop which was built and refined for one of the 24 protein chains only (chain *I*). For the remaining chains the density was not sufficiently well defined for unambiguous fitting.

Electron density was systematically observed in a pocket situated along the threefold-symmetry axes of the dodecamers. This density was interpreted as belonging to a Tris molecule, a species that is present in the crystallization solution. This is consistent with the observed presence of Tris molecules at these sites in six crystal structures of DHQ from other organisms resolved at resolutions better than 1.8 Å.

Electron density was also observed at the phosphate-binding site (Asn77, His103, Ile104, Thr105 and Tyr23) of several *Ca*DHQ monomers. Density was most clear in chain *I* and was interpreted as a sulfate ion (another species that is present in the crystallization solution). Sulfate ions were positioned in 23 of the 24 phosphate-binding sites present in the asymmetric unit. In the remaining site the density was too weak (probably owing to low occupancy of the site and/or the presence of water molecules instead of a sulfate ion) and was not modelled.

Another sulfate ion was found at the interface between the two contacting dodecameric particles.

The *Ca*DHQ experimental structure factors and the refined crystal structure have been deposited with the Protein Data Bank (accession code 3kip). The *Ca*DHQ EM map has been deposited at the European Bioinformatics Institute (accession code EMD-1680).

3. Results and discussion

The *Ca*DHQ protein was produced using a previously described strategy (Santini *et al.*, 2008). Using crystals optimized for diffraction studies, a complete data set was obtained in the 70.68–2.95 Å resolution range (Table 1). The *Sc*DHQ diffraction data set (33–2.2 Å) was retrieved from the PDB (entry 1v1j).

For the EM part, the three-dimensional structure of dodecameric *Ca*DHQ was determined at 15 Å resolution using 7200 single images from eight different electron micrographs of samples negatively stained with methylamine vanadate (Fig. 1*a*). The reconstruction was obtained by the projection-matching method (Franzetti *et al.*, 2002; Frank *et al.*, 1996). After 15 cycles of refinement, tetrahedral symmetry was imposed for ten extra cycles. The resolution of the three-dimensional EM reconstruction (Figs. 1*b* and 1*c*), obtained using only half of the selected single images of the dodecamer, was estimated by Fourier shell correlation to be about 15 Å, which is a rather high resolution for the negative-staining technique.

The three-dimensional EM reconstruction of *Ca*DHQ was used to phase the diffraction data of both *Ca*DHQ and *Sc*DHQ. The MR solution of *Ca*DHQ (Fig. 1*d*) was rather difficult to obtain. A contrasted solution was eventually found by working with data in the 25–15 Å resolution shell and forcing the orientations of the two independent dodecamers to satisfy the rotational NCS. The correlation coefficient of the solution was 40.6% (36.4% for the first wrong position). In the case of *Sc*DHQ the unique independent molecule was easily positioned by MR (Fig. 1*h*), leading to a contrasted solution with a correlation coefficient of 59.3% (50.1% for the first wrong position).

The MR solutions defined the molecular boundaries (masks) of the dodecamers within the crystals (Figs. 1*d* and 1*h*). Based on the tetrahedral symmetry of the three-dimensional EM model, the rigid-body operators relating equivalent regions within the molecular boundaries were defined (24 and 12 NCS transformations for *Ca*DHQ and *Sc*DHQ, respectively). An iterative procedure of phase extension from 15 Å to the maximum available resolution was then carried out using density-modification techniques, including NCS map averaging, solvent flattening and intermediate steps in which the molecular masks and the NCS operators were refined. The convergence of the procedure was faster for *Ca*DHQ, which was certainly a consequence of the intermolecular NCS.

Figs. 1(*e*) and 1(*i*) show details of the electron-density maps obtained at different stages of the phase-extension process and highlight the dramatic enhancement of the final maps. The map calculated at 2.95 Å resolution for *Ca*DHQ (Fig. 1*f*) was sufficiently accurate to build 80% of the polypeptide model (Fig. 1*g*) using automated procedures (Terwilliger, 2000, 2003). Further refinement with minor manual interventions yielded an 86% complete model (expression tags were not visible) with crystallographic residual factors R = 20.3% and $R_{\rm free}$ = 24.5% and with 96% of the modelled residues in allowed Ramachandran regions (Table 1). For *Sc*DHQ, our 2.2 Å resolution map was in very good agreement with the refined atomic model deposited in the PDB. It is worth noting that the final high-resolution maps (Figs. 1*e*, 1*f* and 1*i*) are exclusively based on experimental data and are not biased by any pre-established atomic model.

To our knowledge, this is the first case in which a three-dimensional reconstruction based on negative staining has successfully been used to solve the structure of a small-sized protein (17 kDa) organized as a multimer. Even more interestingly, the three-dimensional EM reconstruction of *Ca*DHQ also permitted the solution of the crystal structure of the related *Sc*DHQ molecule (48% identity over 131 residues, 2.1 Å r.m.s.d. based on C^{α} superimposition) owing to the low divergence between the two structures at the resolution of the EM map.

Negative staining may cause particle distortion and limit the resolution of the reconstruction, but it provides contrasted EM images for particles the size of DHQ. As proven in this study, it can yield usable reconstructions for MR and MX data phasing of homologous crystal structures.

NCS is indispensable for phase extension and is present in oligomeric structures, where it corresponds to the internal symmetry of the often large particles. However, the crystal structures of large nonsymmetrical molecules may also be studied provided that there are multiple copies of the same molecule within the asymmetric unit to supply the intermolecular NCS. As previously noted by Rossmann (1995), a high NCS redundancy is not a strict requirement for convergence and accuracy of the phase-extension procedure. Indeed, we have observed that just four NCS operators are sufficient to produce interpretable maps for *Ca*DHQ (Table 2, Fig. 2) and that an NCS redundancy of six leads to maps of a quality comparable to that obtained by exploiting the full 24-fold NCS of the *Ca*DHQ crystals.

Our results suggest that EM reconstructions obtained by negative staining, a technique that is of wider applicability than cryo-EM, are suitable models for medium-sized proteins (greater than 100 kDa) provided that complete low-resolution diffraction data are available. Although it is possible to measure data at quite low resolution, the standard models for bulk-solvent correction need to be improved in order to render the very low-resolution data (lower than 35 Å) exploitable.

This project was supported by a grant from the French Ministry of Research and Education (No. 02 L 0194), by a grant from the Human Frontier Science Program (RGP0026/2003HFSP) and by the Centre National de la Recherche Scientifique (CNRS). We thank Dr Vincent Monchois and

Protein'eXpert for the gift of the expression vector and the IGS team for providing the *Ca*DHQ material. We also wish to thank Dr Ezequiel Panepucci at the SLS for helpful assistance on the X06SA beamline and Dr Sandra Jeudy for her help in the refinement process.

References

Abergel, C. (2004). *Acta Cryst.* D**60**, 1413–1416.
Abergel, C. *et al.* (2003). *J. Struct. Funct. Genomics*, **4**, 141–157.
Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
Cheng, L., Zhu, J., Hui, W. H., Zhang, X., Honig, B., Fang, Q. & Zhou, Z. H. (2009). *J. Mol. Biol.*, doi:10.1016/j.jmb.2009.12.027.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Conway, J. F. & Steven, A. C. (1999). *J. Struct. Biol.* **128**, 106–118.
Dodson, E. J. (2001). *Acta Cryst.* D**57**, 1405–1409.
Duquerroy, S., Da Costa, B., Henry, C., Vigouroux, A., Libersou, S., Lepault, J., Navaza, J., Delmas, B. & Rey, F. A. (2009). *EMBO J.* **28**, 1655–1665.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). *J. Struct. Biol.* **116**, 190–199.
Franzetti, B., Schoehn, G., Hernandez, J.-F., Jaquinod, M., Ruigrok, R. W. H. & Zaccai, G. (2002). *EMBO J.* **21**, 2132–2138.
Frederickson, M., Roszak, A. W., Coggins, J. R., Lapthorn, A. J. & Abell, C. (2004). *Org. Biomol. Chem.* **2**, 1592–1596.
Jones, T. A. (1992). *Proceedings of the CCP4 Study Weekend. Molecular Replacement*, edited by E. J. Dodson, S. Gover & W. Wolf, pp. 91–105. Warrington: Daresbury Laboratory.
Kleywegt, G. J. & Jones, T. A. (1994). *Proceedings of the CCP4 Study Weekend. From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Kleywegt, G. J. & Jones, T. A. (1999). *Acta Cryst.* D**55**, 941–944.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Murshudov, G., von Delft, F. & Ballard, C. (2008). *Acta Cryst.* D**64**, doi:10.1107/S0907444907058714.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J., Lepault, J., Rey, F. A., Álvarez-Rúa, C. & Borge, J. (2002). *Acta Cryst.* D**58**, 1820–1825.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Roberts, C. W. *et al.* (2002). *J. Infect. Dis.* **185**, *Suppl.* 1, S25–S36.
Rossmann, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon & Breach.
Rossmann, M. G. (1995). *Curr. Opin. Struct. Biol.* **5**, 650–655.
Santini, S., Monchois, V., Mouz, N., Sigoillot, C., Rousselle, T., Claverie, J. M. & Abergel, C. (2008). *BMC Struct. Biol.* **8**, 33.
Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Vellieux, F. M. D. & Read, R. J. (1997). *Methods Enzymol.* **277**, 18–53.
Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). *J. Struct. Biol.* **125**, 185–195.
Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). *Mol. Cell*, **17**, 121–135.