short communications

Molecular replacement in the 'twilight zone': structure determination of the non-haem iron oxygenase NovR from *Streptomyces spheroides* through repeated density modification of a poor molecular-replacement solution

**Sascha Keller,[a] Florence Pojer,[b]‡
Lutz Heide[b] and David M.
Lawson[a]***

[a]Department of Biological Chemistry, John Innes
Centre, Norwich NR4 7UH, England, and
[b]Universität Tübingen, Pharmazeutische
Biologie, Auf der Morgenstelle 8,
72076 Tübingen, Germany

‡ Present address: The Jack H. Skirball Center
for Chemical Biology and Proteomics, The Salk
Institute for Biological Studies, 10010 North
Torrey Pines Road, La Jolla, CA 92037, USA.

Correspondence e-mail:
david.lawson@bbsrc.ac.uk

Crystals of recombinant NovR (subunit MW = 29 924 Da; 270 amino acids), a non-haem iron oxygenase from *Streptomyces spheroides*, were grown by vapour diffusion. The protein crystallized in space group $C2$, with unit-cell parameters $a = 86.69$, $b = 139.38$, $c = 100.82$ Å, $\beta = 101.18°$. Native data were collected to a resolution of 2.1 Å from a single crystal at a synchrotron and a molecular-replacement solution was obtained using the program *AMoRe*. The starting phase information was very poor and did not permit model building. Phases were subsequently improved using a combination of fourfold averaging and very gradual phase extension in the program *DM* to yield an interpretable map. NovR belongs to a novel class of non-haem iron oxygenases that share sequence similarity with class II aldolases. It is predicted to perform two consecutive oxidative decarboxylation steps in the biosynthesis of the prenylated hydroxy-benzoic acid moiety of the aminocoumarin antibiotic novobiocin.

1. Introduction

Although estimates vary, as a general rule of thumb, for molecular replacement to have a high probability of success the template structure must share at least 35% sequence identity with the target structure over a significant portion of the molecule. For identities in the range 20–35%, the so-called 'twilight zone' (Rost, 1999), the fold may be conserved but the differences between the template and target structures are frequently too great to be handled well by standard molecular-replacement protocols (Claude *et al.*, 2004). Furthermore, the likelihood of success tends to decrease with an increasing number of target molecules in the crystallographic asymmetric unit. Even if molecular replacement is apparently successful, the resultant model is likely to be difficult to refine because of the poor accuracy of the phases. In these cases, additional sources of phase information (*e.g.* from a heavy-atom derivative) may be required to complete the structure solution. Alternatively, where the situation permits, density-modification procedures may help to yield interpretable electron-density maps. Particularly relevant to this study is the technique of local symmetry averaging as reviewed by Kleywegt & Read (1997).

We report here the crystallization and structure solution of NovR from *Streptomyces spheroides*, which is predicted to be a bifunctional non-haem iron oxygenase based on very high sequence identity (95%) to the known protein CloR (Pojer *et al.*, 2003). Initial poor-quality phases were obtained by molecular replacement using a template structure sharing only 21% sequence identity with NovR. An interpretable electron-density map was subsequently obtained through a combination of fourfold averaging and phase extension. NovR is predicted to be responsible for two consecutive oxidative decarboxylation steps in the biosynthesis of the prenylated hydroxy-benzoic acid moiety of novobiocin, which is an aminocoumarin antibiotic (Pojer *et al.*, 2003).

2. Materials and methods

2.1. Protein expression and purification

The *novR*-coding sequence (UniProtKB/TrEMBL entry Q9L9F0) was amplified by PCR using the *10-9C* cosmid (Steffensky *et al.*, 2000) containing the complete novobiocin-biosynthetic gene cluster as a template. The PCR product was cloned into pET100/D-TOPO (Novagen), resulting in an expression construct encoding for NovR with an enterokinase-cleavable N-terminal hexahistidine tag. This added a further 36 residues to the native protein (with sequence MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDHPFT), giving a total deduced molecular weight of 34 049 Da. In order to maximize the yield of soluble protein, the plasmid containing the *novR* gene was cotransformed with plasmid pGroESL-911 encoding the GroES/GroEL chaperone system (Ichetovkin *et al.*, 1997) into BL21(DE3) (Studier & Moffatt, 1986). For protein expression, 10 ml of an overnight culture of these cells was used to inoculate 1 l Luria–Bertani medium containing 100 mg carbenicillin and 5 mg tetracycline. The cells were grown at 310 K to an OD_{600} of around 0.4. The culture was then cooled to 293 K, induced by the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 1 mM and shaken for an additional 16 h at 293 K. The cells were harvested by centrifugation in a Sorvall Evolution centrifuge (10 min, 7000 rev min^{-1}, 277 K, SLC-6000 rotor) and stored at 253 K until further purification.

The cell pellet was subsequently resuspended in buffer A (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 30 mM imidazole) and lysed by two passes through a French press (6.9 MPa, 277 K). The cell debris was removed by centrifugation in a Sorvall RC5C centrifuge (30 min, 19 000 rev min^{-1}, 277 K, SS34 rotor) and the supernatant was applied onto a pre-equilibrated 5 ml Ni^{2+}-charged Hi-Trap Chelating HP column (GE Healthcare). The column was then washed with ten column volumes (CV) of buffer A and the bound protein was eluted over 10 CV in a linear gradient to 1 M imidazole (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 1 M imidazole). The fractions containing the NovR protein (confirmed by SDS–PAGE) were pooled and dialysed at 277 K overnight against 2 l buffer B (50 mM Tris–HCl pH 8.0, 150 mM NaCl, 10 mM EDTA). For further purification, the protein was applied onto a Superdex 200 HiLoad HP gel-filtration column (GE Healthcare) pre-equilibrated in 50 mM Tris–HCl pH 8.0, 150 mM NaCl. The elution fractions containing NovR were pooled and concentrated to around 10 mg ml^{-1} using an Ultrafree 10 kDa cutoff concentrator (Millipore). At this stage, the purified protein was stored in 100 µl aliquots at 193 K.

Prior to crystallization of NovR, the N-terminal His tag was cleaved using enterokinase (Invitrogen). The cleavage site lies after the final Lys of the tag, leaving just five residues of the linker (*i.e.* DHPFT). The resulting 275-residue protein, which we shall refer to as DHPFT-NovR, would have a calculated molecular weight of 30 521 Da. For cleavage, the NovR protein aliquots were diluted to 1 mg ml^{-1} in 50 mM Tris–HCl pH 8.0, 1 mM $CaCl_2$ and incubated for 16 h at 293 K in the presence of enterokinase at a concentration of 5.5 U per milligram of NovR protein. The completeness of the cleavage reaction was confirmed by SDS–PAGE, after which the enterokinase was removed with EK-Away Resin (Invitrogen) according to the manufacturer's guidelines. The NovR preparation was then re-buffered into 50 mM Tris–HCl pH 8.0, 150 mM NaCl and concentrated to around 10 mg ml^{-1} using an Ultrafree 10 kDa cutoff concentrator (Millipore). Dynamic light scattering (DLS) was used to monitor the solution properties of the protein. For this purpose, approximately 30 µl of sample was centrifuged through a 0.1 µm Ultrafree filter (Millipore) to remove particulate material before

introduction into a 12 µl microsampling cell. The latter was then inserted into a Dynapro-MSTC molecular-sizing instrument at 293 K (Protein Solutions Inc.). A minimum of 15 scattering measurements were taken and the resulting data were analysed using the *DYNAMICS* software package (Protein Solutions Inc.).

2.2. Crystallization and X-ray data collection

Crystallization screening trials were carried out by vapour diffusion in a sitting-drop format using 96-well Greiner plates and a variety of in-house and commercially available screens (Hampton Research and Nextal) at a constant temperature of 291 K. Drops consisted of 1 µl protein solution mixed with 1 µl well solution and the well volume was 100 µl; the protein concentration was approximately 10 mg ml^{-1}. Trials were performed both with and without the NovR reaction product, 3-dimethylallyl-4-hydroxybenzoic acid (also referred to as ring A), at a concentration of 1 mM and in some cases the protein was supplemented with 50 µM ferrous ammonium sulfate and 2 mM ascorbic acid. Improved crystals were subsequently obtained by refining the successful conditions from the initial screens and adapting them to a hanging-drop format using 24-well VDX plates (Hampton Research). In this case, the well volume was 1 ml and the protein:precipitant ratio was either 2 µl:1 µl, 1 µl:1 µl or 1 µl:2 µl.

Prior to cryogenic data collection, crystals were given a brief soak (<30 s) in cryoprotectant, which corresponded to the mother liquor with the addition of 25%(*v/v*) ethylene glycol in place of an equivalent volume of buffer. Crystals were routinely transferred from one solution to another and were ultimately mounted for X-ray data collection using cryoloops (Hampton Research). Crystals were flash-cooled by plunging into liquid nitrogen and stored prior to transport to the synchrotron. For data collection, a single crystal was transferred to the goniostat on station PX10.1 at the Daresbury Synchrotron Radiation Source using Hampton Research cryotools and was maintained at 100 K with a Cryojet cryocooler (Oxford Instruments). Diffraction data were recorded using a MAR 225 CCD detector (X-ray Research) with the wavelength set to 1.488 Å. The diffraction data were integrated using *MOSFLM* (Leslie, 2006) and scaled using *SCALA* (Evans, 2006). At the outset, approximately 5% of the reflections were set aside for the calculation of 'free' crystallographic R factors (Brünger, 1993).

Further analysis of the data was performed using programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994), including *SFCHECK* (Vaguine *et al.*, 1999) and *MOLREP* (Vagin & Teplyakov, 2000), which were used specifically to detect the presence of noncrystallographic symmetry.

2.3. Structure solution

A search of the Protein Data Bank (PDB; Berman *et al.*, 2000) using the NovR sequence yielded a number of potential structural homologues that could serve as molecular-replacement templates. However, even the best of these were in the so-called 'twilight zone' of sequence identity, *i.e.* in the range 20–35% (Rost, 1999). The closest match was with L-fuculose-1-phosphate aldolase from *Escherichia coli*, which exists as a homotetramer with subunits related by fourfold (C_4) symmetry. Since there were several structures for this enzyme, we chose the highest resolution structure as the basis for our template, namely that of the S71Q mutant (PDB code 1e4c), which had been determined at 1.66 Å resolution (Joerger *et al.*, 2000). Nevertheless, the sequence identity reported by the PDB was only 25% calculated over 159 aligned residues, which represents just 59% of the native NovR sequence. An extended alignment was subse-

Table 1
Summary of X-ray data for NovR.

Values in parentheses are for the outer resolution shell.

Wavelength (Å)	1.488
Resolution range (Å)	44.46–2.10 (2.21–2.10)
Unique reflections	65200 (7664)
Completeness (%)	95.5 (77.3)
Redundancy	3.5 (2.5)
R_{merge}†	0.062 (0.321)
$\langle I/\sigma(I)\rangle$	15.3 (3.0)
Wilson B value (Å^2)	33.2

† $R_{merge} = \sum(|I_j - \langle I_j\rangle|)/\sum\langle I_j\rangle$, where I_j is the intensity of an observation of reflection j and $\langle I_j\rangle$ is the average intensity for reflection j.

quently produced manually with reference to comparisons of secondary-structural elements in the aldolase structure with those predicted from the NovR sequence using the *PSIPRED* server (http://bioinf.cs.ucl.ac.uk/psipred/; McGuffin *et al.*, 2000). This new alignment covered residues 25–238 (or 79%) of the NovR sequence, giving a sequence identity of 21%. In order to generate the best possible search model, the sequence of the aldolase monomer structure was 'mutated' to that of NovR, with reference to this sequence alignment, using the program *CHAINSAW* (Collaborative Computational Project, Number 4, 1994; Schwarzenbacher *et al.*, 2004), and any non-aligned regions were deleted from the structure. In addition, all solvent molecules, metal ions and other ligands were stripped from the model. A template for a NovR tetramer was generated from the monomer template by applying the fourfold symmetry that generates the corresponding biological unit of the aldolase structure. Molecular replacement was performed with the program *AMoRe* (Navaza, 1994) using both monomer and tetramer templates. The initial phase estimates derived from the molecular-replacement solution were improved using the program *DM* (Cowtan, 1994) and automated model building was performed using *ARP/wARP* (Perrakis *et al.*, 1999). The molecular-graphics program *Coot* (Emsley & Cowtan, 2004) was used to display structures and inspect electron-density maps and *REFMAC*5 (Murshudov *et al.*, 1997) was used for rigid-body and restrained refinement.

3. Results

NovR was overexpressed and purified with an approximate yield of 5 mg of protein from 1 l culture after tag cleavage and was judged to be greater than 98% pure by SDS–PAGE analysis. DLS analysis gave
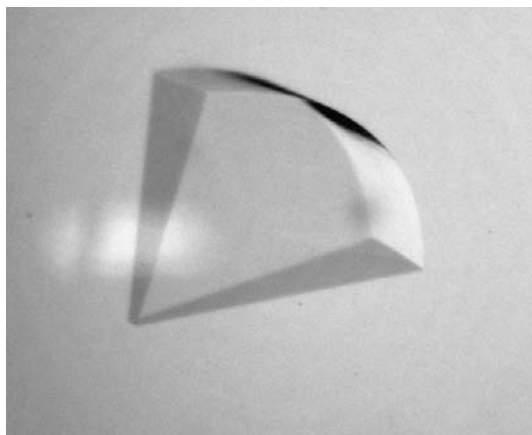
Figure 1
Single crystal of *S. spheroides* NovR, approximately 500 × 400 × 100 µm in size.

a monomodal distribution, with a relatively low polydispersity of 25.6%. From these results, the molecular size was estimated at 113 kDa, close to the value expected for a DHPFT-NovR tetramer (*i.e.* 122.1 kDa).

Although crystals were obtained from several distinct but related conditions, only those grown from 100 m*M* HEPES pH 7.5, 1.5 *M* Li$_2$SO$_4$ proved to be suitable for structure solution. No crystals were obtained when ring A, the NovR product, was absent. These conditions were optimized to give large wedge-shaped crystals with dimensions of up to 500 × 400 × 100 µm (Fig. 1) by mixing 2 µl protein solution, containing 50 µ*M* ferrous ammonium sulfate, 2 m*M* ascorbic acid and 1 m*M* ring A, with 1 µl reservoir solution.

Native X-ray data were collected from a single NovR crystal and a total of 225 0.8° oscillation images were recorded to 2.1 Å resolution. The space group was established as *C*2, with unit-cell parameters $a = 86.69$, $b = 139.38$, $c = 100.82$ Å, $\beta = 101.18°$. After merging, the data set was 95.5% complete to 2.1 Å, with an overall R_{merge} of 0.062. A summary of data-collection statistics is given in Table 1. Estimation of the content of the asymmetric unit suggested that two to five DHPFT-NovR subunits were possible, giving a solvent content in the range 74.9–37.2%, although a tetramer was the most likely based on previous findings (*i.e.* DLS). A tetramer would give a crystal-packing parameter (V_M) of 2.45 Å^3 Da^{-1}, with a corresponding solvent content of 49.8% (Matthews, 1968).

Analysis of the data set using *SFCHECK* did not find any evidence of pseudo-translational symmetry, indicating that none of the monomers in the asymmetric unit were related by translational symmetry alone. However, a self-rotation function calculated on data in the resolution range 10.0–5.0 Å using *MOLREP* showed a clear noncrystallographic fourfold axis at $\omega = 44.7$, $\varphi = 25.3°$ with a peak height of 39% of the origin peak height.

Molecular replacement using the monomer template created by *CHAINSAW* did not give any convincing solutions. However, the tetramer template yielded a top solution for a single tetramer per asymmetric unit that gave sensible packing in the unit cell after the application of crystallographic symmetry and, importantly, oriented the tetramer such that its fourfold axis was consistent with the noncrystallographic fourfold axis identified by the self-rotation function. However, after rigid-body refinement of the individual subunits to 2.1 Å resolution in *REFMAC*5, the 'working' (R_{work}) and 'free' (R_{free}) crystallographic R factors were disappointingly high at 0.542 and 0.544, respectively, and the figure of merit (FOM) was very low at 0.123. Not surprisingly, the resultant electron-density maps calculated from these phases were poor and difficult to interpret, with many regions displaying very little connectivity that was consistent with the molecular-replacement solution.

Attempts to rebuild the preliminary model against these maps did not result in lower R factors after refinement; moreover, fourfold averaging using standard *DM* protocols did not improve the situation. We then reasoned that the molecular-replacement solution was probably a reasonable approximation to the final structure at low resolution. With this in mind, we ran ten cycles of fourfold averaging with solvent flattening in *DM* using just the data to 5 Å resolution. This improved the FOM at this resolution from 0.283 to 0.666. The resultant map showed connectivity for most of the main chain present in the model, as well as some regions of weaker density that could perhaps be attributed to missing parts of the structure. Re-running the same *DM* protocol, but this time with 100 cycles, gave noticeable improvements in the map and a corresponding increase in the FOM to 0.718. We then decided to couple the fourfold averaging with a very gradual phase extension from the 'good' 5 Å resolution phases up to the 2.1 Å resolution limit. Since the subunits were related by

purely rotational, or proper, noncrystallographic symmetry (NCS), we were able to use an averaging mask that covered the whole tetramer, thereby avoiding the requirement to precisely define the subunit boundaries within the tetramer. In order to correct errors in the initial NCS operators and mask, we opted to refine these parameters periodically during the *DM* run. After 1000 cycles of averaging and phase extension (with solvent flattening), the overall FOM was 0.792 at 2.1 Å resolution.

A 2.1 Å resolution F_{obs} electron-density map calculated using this phase set was readily interpretable. Notably, there was density for regions of the structure that were absent from the molecular-replacement solution and in some cases the density indicated that substantial shifts (up to 5 Å for C^α atoms) were required for some existing parts of the model. However, rather than manually rebuild the molecular-replacement solution, we opted to use the *ARP/wARP* procedure for automated model building. This was successful in docking a total of 966 residues to the sequence, thereby accounting for 89% of the 1080 amino acids possible in the native tetramer (Fig. 2). The corresponding R_{work} and R_{free} values were 0.195 and 0.259, respectively, at this stage. A summary of the structure-solution procedure is given in Fig. 3.

4. Discussion

Initially, we planned to use the anomalous signal of the active-site iron to help resolve the NovR structure, since it was predicted to be a non-haem iron oxygenase (Pojer *et al.*, 2003). To this end, we performed an EXAFS scan across the *K* X-ray absorption edge of iron (approximate wavelength 1.743 Å) on station PX10.1 at Daresbury. However, no discernable edge was apparent and thus we abandoned this method of solving the structure. After we had solved the structure, the metal content of a similarly prepared sample was determined using inductively coupled plasma analysis by Southern Water Scientific Services (Brighton, England), as described elsewhere (Eady *et al.*, 1972). To our surprise, the predominant metal was nickel (0.9 mol per subunit of NovR) and iron was not detected at all. The nickel had presumably leached from the nickel-affinity column and this observation should serve as a cautionary note to others using metal-affinity chromatography to purify metalloproteins. The data presented here were collected at a wavelength of 1.488 Å, this being the wavelength that gives the highest beam intensity on station PX10.1, and is coincidentally very close to the nickel *K* edge. With this in mind, we calculated an anomalous difference Fourier map using the *ARP/wARP* phases and this showed a number of significant peaks throughout the structure. The majority of these were associated with S atoms, but four peaks (*i.e.* one per subunit) coincided with sites corresponding to those of the active-site Zn atoms in the aldolase structure (Fig. 2). These were assumed to be nickel ions, although the peak

heights were similar to those of the sulfurs. *A posteriori* attempts to solve the structure by SAD phasing using these nickel sites in the program *SHELXE* (Sheldrick, 2002) were unsuccessful (even using the 1000-cycle *DM* protocol starting from 5 Å resolution SAD phases), suggesting that we had collected the data just on the low-energy side of the nickel edge where the magnitude of f'' falls off rapidly, although a more redundant data set might have proved to be more useful. Since the crystals apparently contained no ordered Fe atoms, it is not clear why adding ferrous ammonium sulfate should give improved crystals.

Despite the low sequence identity between the target and template structures, molecular replacement was successful and this was undoubtedly helped by the conservation of the quaternary structure. This enabled us to perform searches using a tetramer template that contained 79% of the C^α atoms expected in the asymmetric unit, thereby effectively reducing the problem to a 'one molecule per asymmetric unit' case. Nevertheless, the differences between the model generated by *ARP/wARP* and the tetramer template (and by definition, the molecular-replacement solution) were quite substantial: a superposition using *LSQKAB* (Kabsch, 1976) gave a root-
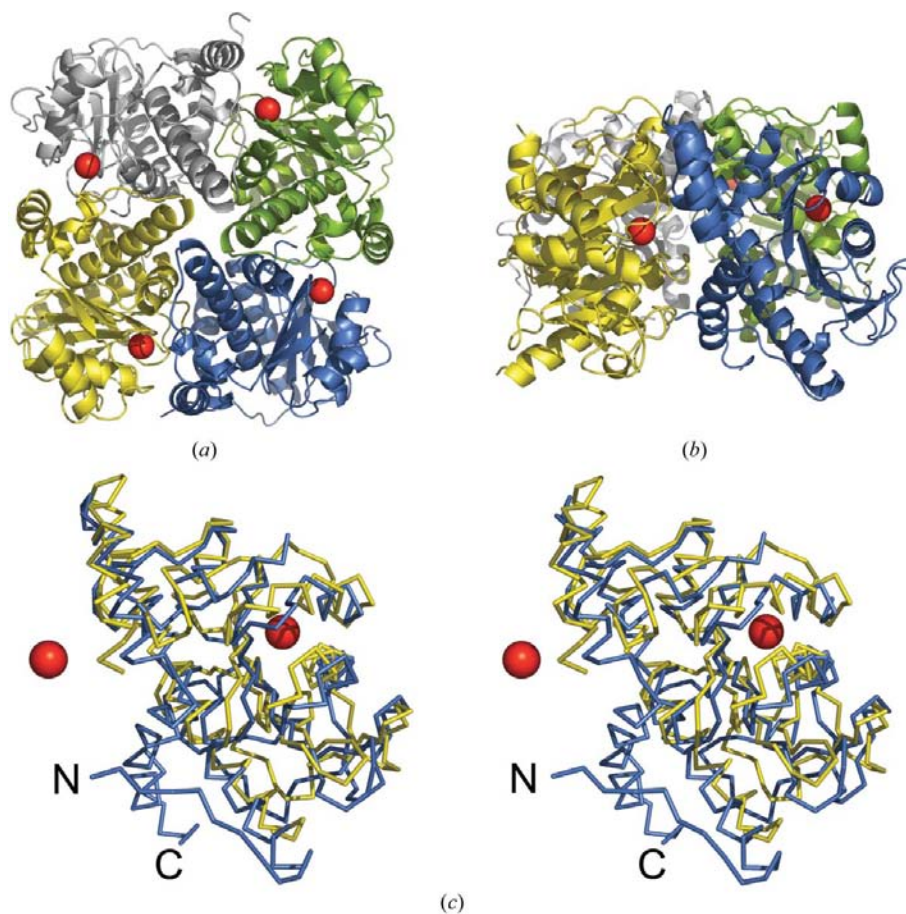
Figure 2
Preliminary structure of NovR. Ribbon representations showing (*a*) top and (*b*) side views of the model built by *ARP/wARP* with individual subunits shown in different colours. The red spheres indicate putative nickel sites found by inspection of an anomalous difference Fourier map, which correspond to the zinc sites of the template aldolase structure. (*c*) Stereographic projection showing overlaid (not least-squares superposed) C^α traces of single subunits from the molecular-replacement solution (yellow) and the *ARP/wARP* model (blue). This corresponds to the blue subunit depicted in (*a*) and (*b*) and has the same orientation as that shown in (*b*). Note in particular the additional structure at the N- and C-termini of the *ARP/wARP* model and the significant shifts in some of the secondary-structural elements. The red spheres show the two metal ions associated with this subunit. Figures were generated using *PyMOL* (DeLano, 2002).

mean-square-displacement (r.m.s.d.) of 3.07 Å over 801 equivalent C$^\alpha$ atoms and there were significant rotations in the individual subunits (Table 2). In comparison, a recent study suggested that for successful molecular replacement a superposition of template and target structures should give an r.m.s.d. value for C$^\alpha$ atoms not exceeding 2.5 Å (Schwarzenbacher *et al.*, 2004). We subsequently showed that the *CHAINSAW* 'mutation' procedure had not been a crucial factor, since an equivalent polyalanine tetramer template containing 18% fewer atoms gave the same top solution in *AMoRe*.

In previous studies where local symmetry averaging was used to improve a poor set of starting phases, the necessity of starting at low resolution was demonstrated (Gaykema *et al.*, 1986; Valegard *et al.*, 1990; Braig *et al.*, 1994; Kleywegt & Read, 1997). Moreover, the importance of using a small step size in phase extension has also been shown (Arnold *et al.*, 1987; Kleywegt & Read, 1997; Vellieux & Read, 1997; Cowtan & Zhang, 1999) and can be explained by the G function or interference function. Briefly, this function describes how structure factors influence (or interfere with) one another as the result of density-modification procedures and applies not only to those that are related by noncrystallographic symmetry (in the case of averaging), but also to those reflections that are close together in reciprocal space. Consequently, in each cycle of phase extension useful phase information can only be transferred to newly included reflections close to the current phasing resolution limit. Thus, for successful phase extension, higher resolution reflections must be introduced gradually. To test the effect of the step size on our phase-extension procedure, we decided to estimate the minimum number of cycles required to produce a phase set that was sufficiently accurate to enable successful automated model building. To this end, we re-ran *DM* several times with exactly the same protocol, but with different numbers of cycles, starting from 100, with increments of 100 for each successive run. The step size was determined automatically by *DM*, with fewer reflections being introduced into the density-modification procedure at each cycle as the number of cycles increased. All runs gave apparently good FOM values, the lowest being 0.693 for 100 cycles, showing only a gradual increase for each additional 100 cycles. Nevertheless, visual inspection of map quality with reference to the *ARP/wARP* model revealed a marked change from very poor maps for runs of up to 300 cycles to a map that was largely consistent with the model after 400 cycles. Even so, only the phase set produced after 500 *DM* cycles was good enough for automated model building in *ARP/wARP*. This highlights the danger of relying entirely on FOM values as a measure of successful phasing, as they tend to be overestimated by density-modification programs, especially when many cycles of density modification are involved. Indeed, it is notable that the FOM value after *ARP/wARP* (0.738), where the phases were very good, is actually slightly lower than that after 400 cycles of *DM* (0.745), where the phases were not good enough for auto-building.

As alternative measures of phasing quality, the program *OVERLAPMAP* (Brändén & Jones, 1990) was used to calculate map correlation coefficients (mapCC) between maps produced at the various stages and a map produced using the *ARP/wARP* phases and the program *PHISTATS* (Collaborative Computational Project, Number 4, 1994) was used to calculate the mean phase change ($\Delta\varphi$) *versus* the *ARP/wARP* phases. Both quantities showed the largest changes on going from 300 to 400
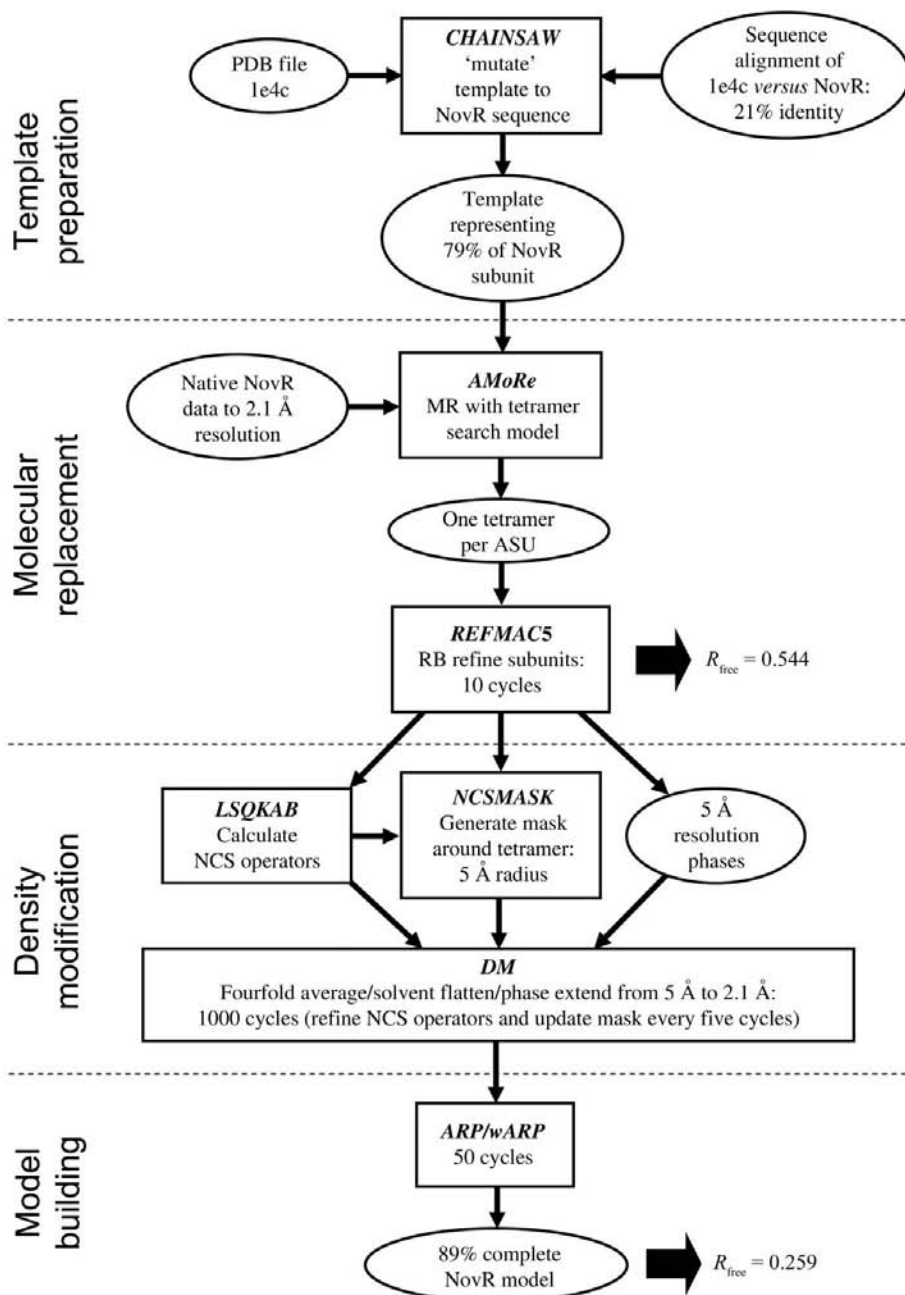
Figure 3
Flow diagram outlining the procedure used to solve the NovR structure. Program tasks are shown in rectangles and inputs/outputs are shown in ellipses. MR, molecular replacement; ASU, asymmetric unit; RB, rigid body; NCS, noncrystallographic symmetry. R_{free} values are calculated to 2.1 Å resolution.

Table 2
Comparison of NovR models at various stages of structure solution.

REFMAC versus AMoRe, *AMoRe* solution compared before and after rigid-body refinement in *REFMAC*5; *ARP/wARP versus AMoRe*, *ARP/wARP* model compared with *AMoRe* solution. Values were obtained by superposition using the program *LSQKAB*, where rotation is the spherical polar χ angle required to superpose models, centroid is the distance between the centroids of the two models and r.m.s.d. is the root-mean-square displacement (based on a minimum of 194 equivalent residues superposed per subunit or 801 per tetramer).

Subunit	Parameter	*REFMAC versus AMoRe*	*ARP/wARP versus AMoRe*
A	Rotation (°)	0.54	6.68
	Centroid (Å)	0.54	0.90
	R.m.s.d. (Å)	0	2.76
B	Rotation (°)	1.57	6.65
	Centroid (Å)	0.49	0.56
	R.m.s.d. (Å)	0	2.79
C	Rotation (°)	2.63	6.84
	Centroid (Å)	0.39	1.11
	R.m.s.d. (Å)	0	2.76
D	Rotation (°)	2.46	7.35
	Centroid (Å)	0.34	0.68
	R.m.s.d. (Å)	0	2.76
Tetramer (ABCD)	Rotation (°)	0.56	1.59
	Centroid (Å)	0.27	0.44
	R.m.s.d. (Å)	0.58	3.07

Table 3
Summary of phasing for NovR.

na, not applicable. FOM, figure of merit; mapCC, map correlation coefficient calculated *versus* map after *ARP/wARP*; Δφ, mean phase difference *versus* phases from *ARP/wARP*; residues fitted, number of residues docked into sequence by *ARP/wARP*.

Phase set	FOM	mapCC	Δφ (°)	Residues fitted
AMoRe solution	0.123	0.176	85.9	na
DM, 100 cycles	0.693	0.177	87.0	0
DM, 200 cycles	0.715	0.199	86.0	0
DM, 300 cycles	0.724	0.182	85.9	0
DM, 400 cycles	0.745	0.616	61.7	0
DM, 500 cycles	0.780	0.780	46.8	970
DM, 1000 cycles	0.792	0.793	45.5	966
ARP/wARP	0.738	na	na	na

cycles, consistent with the significant visual improvement in map quality. Further large changes were observed after 500 cycles corresponding to the transition from a reasonable map to one that was good enough for successful auto-building. Whilst mapCC and Δφ are clearly superior measures of phasing quality than the FOM, since both quantities are calculated with reference to an accurate phase set, neither measure can be used sensibly until the structure has effectively been solved. All phasing statistics are summarized in Table 3 and the improvement in phases is illustrated graphically in Fig. 4. From inspection of the latter, it is clear that there was no significant improvement in the quality of the phases beyond 500 cycles.

Tête-Favier and coworkers showed that significant deviations from NCS can severely hamper the averaging process (Tête-Favier *et al.*, 1993). In our case, there were only very minor deviations from exact C_4 symmetry: in pairwise superpositions of subunits from the *ARP/wARP* model, the maximum r.m.s.d. in C^α atoms was only 0.42 Å and the rotation angles were 90 ± 0.24 and $180 \pm 0.19°$ for adjacent and opposite subunits within the tetramer, respectively. It is notable that
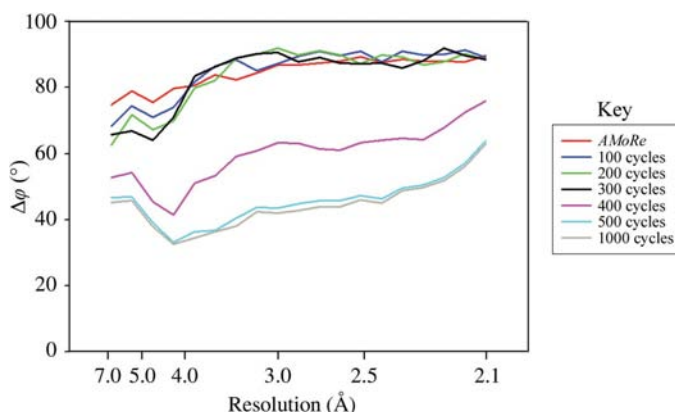
Figure 4
Plots illustrating the differences between the phase set determined after *ARP/wARP* (the most accurate) and the phase sets calculated from the *AMoRe* model and those obtained after running *DM* for various numbers of cycles. Δφ values were averaged in shells of resolution and are plotted against $1/d^2$ along the x axis, where d is the resolution at the higher end of each range (*i.e.* the lower number). For convenience this axis is labelled with the corresponding resolutions.

the orientation of the fourfold axis was already well defined from the molecular-replacement solution, as the value determined subsequently from the *ARP/wARP* model differed from this by less than 2°. However, the orientations of the individual subunits changed significantly within the tetramer: they rotated by up to 2.63° during rigid-body refinement in *REFMAC*5 and by approximately 7° (with respect to the molecular-replacement solution) after the averaging procedure (Table 2). To test the effect of not refining these operators, we re-ran the 1000-cycle *DM* protocol with this option turned off. Although the resultant FOM was apparently good (0.799), *ARP/wARP* was unable to dock any sequence. When the refined NCS matrices after *DM* were compared from the original 1000-cycle run and the 100-cycle run, they were found to be essentially the same (Eulerian angles differing by no more than 0.7° and translations differing by no more than 0.6 Å), suggesting that a very large number of *DM* cycles was not required for successful refinement of the symmetry operators.

The averaging mask for NovR was calculated directly from the molecular-replacement solution and a border of 5 Å was used in an attempt to compensate for missing side chains. When overlaid on the *ARP/wARP* model, the vast majority of the main-chain atoms were within the mask, with the exception of the N- and C-terminal arms that were missing from the molecular-replacement model (Fig. 2c). Although the refined mask output at the end of *DM* enclosed more of the *ARP/wARP* model, it was not significantly different from the initial mask. To test the effect of not updating the mask, this option was not implemented in a further re-run of *DM*. Again the FOM was high (0.802), but this time *ARP/wARP* was still able to dock 964 residues into the sequence.

In conclusion, we have solved the structure of NovR from *S. spheroides* at 2.1 Å resolution using a multi-cycle density-modification procedure involving fourfold averaging and phase extension, starting from a poor set of molecular-replacement phases at 5 Å resolution. Although fourfold averaging should give only a modest improvement in the signal-to-noise ratio (by a factor of approximately $N^{1/2}$, where N is the number of independent copies), we were fortunate in being able to derive a reasonably complete averaging mask and a well defined NCS fourfold axis directly from the molecular-replacement solution. Subsequent investigation showed that it was important to use a small step size in the phase-extension procedure, necessitating the use of at least 500 cycles of density modification, and that refinement of the NCS operators was essential. Final refinement and analysis of the NovR structure will be reported elsewhere.

short communications

References

Arnold, E., Vriend, G., Luo, M., Griffith, J. P., Kamer, G., Erickson, J. W., Johnson, J. E. & Rossmann, M. G. (1987). *Acta Cryst.* A**43**, 346–361.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). *Nature (London)*, **371**, 578–586.

Brändén, C. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.

Claude, J.-B., Suhre, K., Notredame, C., Claverie, J.-M. & Abergel, C. (2004). *Nucleic Acids Res.* **32**, W606–W609.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

Cowtan, K. D. & Zhang, K. Y. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.

DeLano, W. L. (2002). *The PyMOL User's Manual.* San Carlos, CA, USA: DeLano Scientific.

Eady, R. R., Smith, B. E., Cook, K. A. & Postgate, J. R. (1972). *Biochem. J.* **128**, 655–675.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Gaykema, W. P., Volbeda, A. & Hol, W. G. (1986). *J. Mol. Biol.* **187**, 255–275.

Ichetovkin, I. E., Abramochkin, G. & Shrader, T. E. (1997). *J. Biol. Chem.* **272**, 33009–33014.

Joerger, A. C., Mueller-Dieckmann, C. & Schulz, G. E. (2000). *J. Mol. Biol.* **303**, 531–543.

Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.

Kleywegt, G. J. & Read, R. J. (1997). *Structure*, **5**, 1557–1569.

Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.

McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). *Bioinformatics*, **16**, 404–405.

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

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.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Pojer, F., Kahlich, R., Kammerer, B., Li, S.-M. & Heide, L. (2003). *J. Biol. Chem.* **278**, 30661–30668.

Rost, B. (1999). *Protein Eng.* **12**, 85–94.

Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. (2004). *Acta Cryst.* D**60**, 1229–1236.

Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.

Steffensky, M., Muhlenweg, A., Wang, Z. X., Li, S. M. & Heide, L. (2000). *Antimicrob. Agents Chemother.* **44**, 1214–1222.

Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189**, 113–130.

Tête-Favier, F., Rondeau, J. M., Podjarny, A. & Moras, D. (1993). *Acta Cryst.* D**49**, 246–256.

Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.

Valegard, K., Liljas, L., Fridborg, K. & Unge, T. (1990). *Nature (London)*, **345**, 36–41.

Vellieux, F. M. D. & Read, R. J. (1997). *Methods Enzymol.* **277**, 18–53.