Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

3-Carboxy-*cis,cis*-muconate lactonizing enzyme from *Neurospora crassa*: MAD phasing with 80 selenomethionines

Michael C. Merckel,[a] Tommi Kajander,[a] Ashley M. Deacon,[b] Andrew Thompson,[c] J. Günter Grossmann,[d] Nisse Kalkkinen[a] and Adrian Goldman[a]*

[a]Institute of Biotechnology, PO Box 56, University of Helsinki, FIN-00014 Helsinki, Finland, [b]SSRL, MS99, 2575 Sand Hill Road, Menlo Park, CA 94025, USA, [c]EMBL Grenoble Outstation, 38024 Grenoble, France, and [d]CLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England

Correspondence e-mail: adrian.goldman@helsinki.fi

Received 17 October 2001
Accepted 7 February 2002

PDB Reference: CMLE, 1jof, r1jofsf.

The structure of 3-carboxy-*cis,cis*-muconate lactonizing enzyme from *Neurospora crassa* was determined at 3.0 Å resolution. Phase information was derived from a multi-wavelength anomalous dispersion (MAD) experiment conducted at three wavelengths using crystals of fully substituted selenomethionine protein. However, the structure determination was not routine owing to the relatively poor quality of the diffraction data and the large number of twofolds in the unit cell. Eventually, 80 selenium sites were identified by the combined use of direct methods and real-space map interpretation. This represents one of the largest selenium substructures solved and used for phasing. Some of the difficulties in the structure determination and the methods used to address them are discussed.

1. Introduction

The muconate lactonizing enzymes (MLEs) are a part of the β-ketoadipate pathway in soil microorganisms. They convert aromatic compounds into citric acid cycle intermediates. This pathway consists of two branches, the catechol (MLEs; EC 5.5.1.1) and protocatechuate (3-carboxy-*cis,cis*-MLEs or CMLEs; EC 5.5.1.2) branches. The lactonizing enzymes can be divided into three evolutionarily distinct classes (Mazur *et al.*, 1994). Bacterial MLEs catalyze *syn* addition, require an Mn^{2+} cofactor and have an α/β-barrel as the catalytic domain (Kirby *et al.*, 1975; Goldman *et al.*, 1987), whereas bacterial CMLEs catalyse an *anti* addition, require no metal cofactor for activity and have been shown to be related to class II fumarases (Williams *et al.*, 1992). In addition, both eukaryotic MLEs and CMLEs (Mazur *et al.*, 1994) have no metal requirement and the stereochemical and regiochemical course of the reaction is (i) the opposite to that of bacterial CMLEs and (ii) identical to that of bacterial MLEs (Mazur *et al.*, 1994). The *N. crassa* CMLE is likely to have a structure unrelated to the other MLE structures and thus presents a novel metal-independent cycloisomerization motif. Also, it does not have sequence similarity to any protein of known structure or function and hence represents a group of previously uncharacterized proteins.

The CMLE monomer has a molecular weight of 41 kDa and is composed of 365 amino acids. Solution studies have shown the enzyme to most likely be a homotetramer (Mazur *et al.*, 1994). Crystals were obtained in space group $P2_12_12_1$, with unit-cell parameters $a = 92.1$, $b = 159.7$, $c = 236.6$ Å (Glumoff *et al.*, 1996). Based on the solvent content, there are probably two tetramers per asymmetric unit (the Matthews coefficient $V_M = 2.64$ Å^3 Da^{-1} corresponds to a solvent content of 53.0%; Matthews, 1968). Flotation measurements of the crystal

density in a Ficoll gradient supported this model (Glumoff *et al.*, 1996). The crystals diffracted under cryogenic conditions to 2 Å resolution, but flash-cooling was irreproducible and always resulted in crystals with a high mosaicity. The large unit cell combined with the high mosaicity restricted the best obtainable resolution to 2.5 Å. Heavy-atom derivatization was unsuccessful, partly owing to difficulties in measuring data anywhere other than at a synchrotron source. However, the recent application of direct methods to the determination of large selenomethionine substructures, in particular with the program *Shake-and-Bake* (*SnB*; Weeks & Miller, 1999), has resulted in considerable success (Deacon & Ealick, 1999). This prompted us to attempt the phasing of SeMet-CMLE by MAD.

The structure of CMLE was determined using phases from a three-wavelength MAD data set collected from a single crystal. Initially, 60 of 80 possible selenium sites were found with the program *SnB*. However, the phases calculated from these sites did not yield an interpretable electron-density map. The phases were improved with NCS averaging. Manual real-space interpretation of the solvent-flattened unaveraged 4 Å resolution electron-density maps and the correlation between sites from the *SnB* solution were used to identify seven NCS operators and generate a complement of 80 selenium sites. Refinement of these sites followed by NCS averaging resulted in a traceable 3.0 Å resolution Fourier map. Independent data from SAXS experiments were used to confirm the correctness of the initial maps.

2. Materials and methods

2.1. Protein expression and purification

CMLE was expressed in the methionine-auxotrophic *Escherichia coli* BL21 (DE3) B834 using the New Minimal Medium (Hädener *et al.*, 1993) and 50 mg l^{-1} selenomethionine and with the non-auxotroph method (Van Duyne *et al.*, 1993; Doublié, 1997). Substituted CMLE was purified similarly to native CMLE (Mazur *et al.*, 1994), with the exception that only 1 m*M* EDTA was used and 5 m*M* β-mercaptoethanol (βME) was added to prevent proteolysis and oxidation. In addition, we replaced histidine–HCl buffer with MES in the purification. Removal of salt after the 75% $(NH_4)_2SO_4$ cut was carried out by overnight dialysis against 20 m*M* MES pH 6.0, 1 m*M* EDTA, 5 m*M* βME; anion exchange was performed with a Pharmacia XK-Q-Sepharose column.

2.2. Mass spectroscopy

Purified CMLE was lyophilized, desalted by reverse-phase chromatography and analyzed by electrospray mass spectrometry (Q-TOF, Micromass). A 470 Da difference between the native and SeMet-labelled protein indicated full substitution of ten methionines per monomer.

2.3. Crystallization, data collection and processing

Compared with the native protein, SeMet-CMLE crystallized under the same conditions of 100 m*M* PIPES pH 5.7, 1.56 *M* $(NH_4)_2SO_4$ (Glumoff *et al.*, 1996), but more rapidly (native protein typically crystallized in three weeks to several months, whereas the SeMet-substituted form crystallized in one to two weeks). Data were collected from flash-cooled crystals; we found it best to use very quick soaks in the cryoprotectant solution instead of the several minutes time previously reported (Glumoff *et al.*, 1996). It was later found that small crystals could also in some cases be flash-cooled in a nitrogen stream directly from the crystallization mother liquor [*i.e.* using 1.56 *M* $(NH_4)_2SO_4$ as the sole cryoprotectant]. All diffraction data were measured at beamline BM14 (BL 19) at the ESRF. The crystal orientation was set to utilize mirror symmetry by allowing the collection of Friedel mates on the same image, which helped to reduce systematic errors. The best diffracting SeMet crystal had a high mosaicity of 0.82° (Table 1), which was typical of CMLE crystals and allowed the collection of 3.0 Å resolution data in the time available. All data were processed with the *HKL* suite (Otwinowski & Minor, 1997). It is interesting to note that this crystal had a slightly different unit-cell size (Table 1) compared with previous studies and this turned out to be important for solving the structure (see below).

Table 1
Data-collection and phasing statistics for the CMLE crystal from which the phases were determined (note altered unit cell; see text).

Values in parentheses are for the highest resolution shell. For phase refinement, data were used only to 4 Å, where the completeness for all three wavelengths was at least 84%. f' is the wavelength at which there is a minimum in the Kramers–Kronig integration of the X-ray absorption scan of the selenium absorption edge. f'' is the wavelength at the fluorescence signal maximum in the X-ray absorption scan at the selenium absorption edge.

	Peak f''	Inflection f'	Remote†
Space group	$P2_12_12_1$		
Unit-cell parameters (Å)	a = 90.4, b = 152.1, c = 247.1		
λ (Å)	0.9786	0.9795	0.8856
Resolution range (Å)	25–3.0	25–3.2	25–3.2
No. of observations	1091016	800714	757161
No. of unique reflections	71878	51788	48429
Mosaicity (°)	0.82	0.82	0.82
Average $I/\sigma(I)$	18.5 (6.6)	20.3 (7.7)	19.6 (8.3)
R_{sym} (%)	6.3 (11.2)	4.7 (8.2)	4.6 (7.2)
Completeness (%)	89.8 (60.6)	87.6 (53.1)	83.3 (63.5)
R_{cullis}	0.72 (anomalous)	0.62/0.54 (acentric/centric)	
Phasing power	1.52	2.00	
FOM‡ (4.0 Å, *MLPHARE*)	0.56		
FOM (3.0 Å *DM*, fourfold NCS averaged)	0.83		
CC§, NCS averaging	0.819/0.840		

† Used as native data set in *MLPHARE*. ‡ Figure of merit. § Correlation coefficient for averaging of each tetramer.

2.4. Solution X-ray scattering

For the X-ray scattering study, native CMLE was prepared in 20 m*M* Tris, 5 m*M* βME buffer solution at pH 7.5. Experiments were performed with protein solutions of concentrations between 0.5 and 45 mg ml^{-1} on station 2.1 (Towns-Andrews *et al.*, 1989) at the SRS Daresbury Laboratory using a position-sensitive multiwire proportional counter (Lewis, 1994). Sample-to-detector distances of 7.5 and 1.5 m and the

X-ray wavelength of $\lambda = 1.54$ Å allowed the coverage of momentum-transfer intervals of $0.002s$ Å^{-1} × 0.016 Å^{-1} and $0.005s$ Å^{-1} × 0.082 Å^{-1}, respectively. The modulus of the momentum transfer is defined as $s = 2\sin\theta/\lambda$, with 2θ being the scattering angle. The s range was calibrated using an oriented specimen of wet rat-tail collagen (based on a diffraction spacing of 670 Å). Reduction of scattering data, analysis and shape reconstruction followed previously described procedures (Grossmann *et al.*, 2001). This included the calculation of the radius of gyration, volume, the intraparticle distance distribution function $p(r)$ and the restoration of the molecular shape of CMLE. The latter was characterized with spherical harmonics up to the seventh order assuming D_2 symmetry for the tetramer, which is acceptable considering the information content of the experimental data.

2.5. Phasing

The anomalous differences from the peak-wavelength data set were processed with the *DREAR* suite accessed through the *SnB* interface to generate difference-normalized structure-factor values (diffEs). The 2400 largest diffEs were used to generate 24 000 invariants at 3.5 Å resolution. Trials were run simultaneously from five different random seeds, producing almost 2000 trials. Two solutions with significant reduction in $R_{\rm min}$ (Hauptman, 1991) were identified. Sites from the two best solutions were cross-examined to find identical or symmetry-related peaks and verify a consistent set of coordinates. Subsets of the solutions were also tested to see if they reproduced the reduction in $R_{\rm min}$. The resulting set of selenium sites were refined with *MLPHARE* (Collaborative Computational Project, Number 4, 1994). The three wavelengths of the MAD data set were treated as MIR, with the remote wavelength ($\lambda = 0.8856$ Å) as native. Dispersive differences were calculated from the inflection point and anomalous differences from the peak data set. Density modification was carried out with *DM* (Collaborative Computational Project, Number 4, 1994). Map interpretation and model building were performed with the program *O* (Jones *et al.*, 1991).

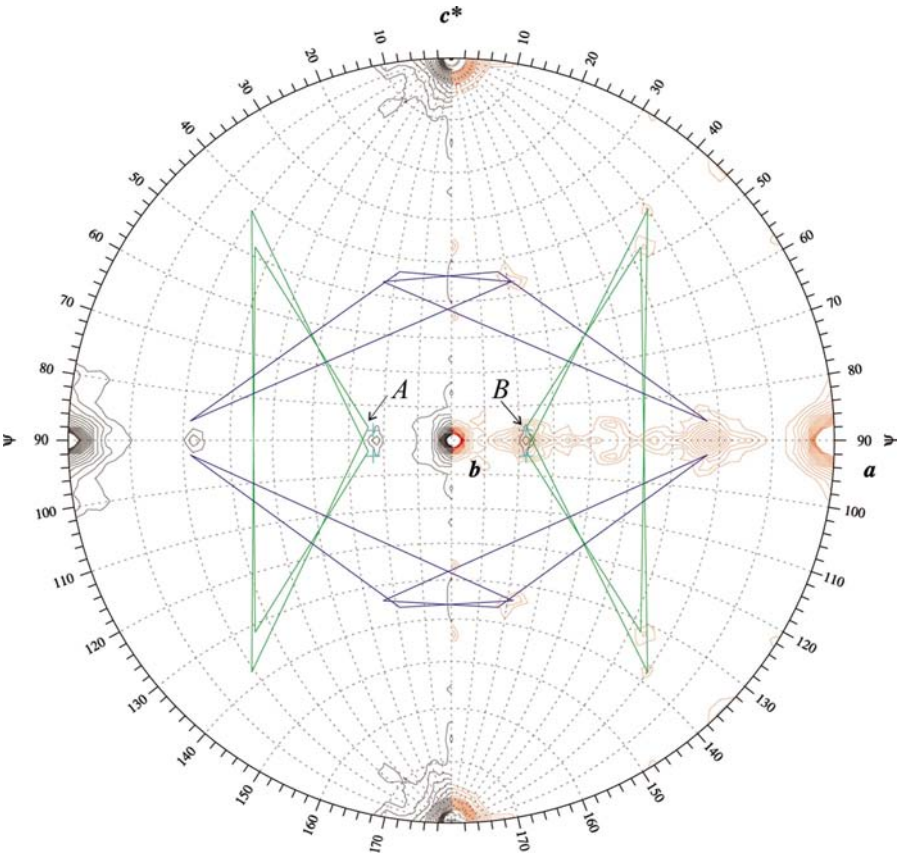

Figure 1
Self-rotation function plots for observed structure factors. ψ measures the declination from the c^* axis and is plotted as lines of latitude; φ measures the angle between the a and b axes. Although only one quadrant would suffice, all four are plotted to make the relationship between model dyads and correlation peaks clearer. The left-hand side (in black) shows the experimental correlation map and the right-hand side the same correlation map calculated using F_cs. Four green triangles (relating tetramers 1) and four blue triangles (relating tetramers 2) connect mutually orthogonal local twofold axis directions in the unit cell. Cyan crosses mark the positions of the twofold κ axes between two tetramers. The positions of the peaks found by the program *FindNCS* are marked *A* (one of the apices of a green triangle) and *B* (one of the cyan crosses).

3. Results and discussion

3.1. Data collection

The crystallographic analysis of CMLE was hampered by poorly diffracting crystals with high mosaicity and poor isomorphism that were also difficult to derivatize. The SeMet-CMLE crystal used for the MAD experiment was actually unique. It represented a slightly different crystal type (Table 1) with altered unit-cell parameters compared with all other previously studied crystals, which have unit-cell parameters $a = 92.3$, $b = 160.6$, $c = 237.5$ Å (Kajander *et al.*, 2002). Thus far, it is the only crystal with these unit-cell parameters. This turned out to be very important, as we could not obtain a solution for the selenium sites from *SnB* using a second MAD data set collected to 3 Å resolution from the dominant crystal type, even though it had a better signal (data not shown). For this latter data set, the completeness at 3 Å was >95% for all wavelengths. We do not understand why *SnB* failed for the latter case, but suspect that it may be because of the high number of sites with similar coordinates owing to the directions of the twofold axes (Fig. 1).

Self-rotation functions of the data sets did not allow a clear identification of the twofold NCS-symmetry axes (Fig. 1). The final 3.0 Å structure was used to calculate

the expected positions of the twofold peaks and several of these fall in regions where no peaks were observed in the self-rotation function calculated from the experimental data (Fig. 1). In addition, other peaks overlap. Also, inspection of native Patterson function plots did not reveal any non-origin peaks indicative of translational NCS.

3.2. Solution X-ray scattering

X-ray scattering data were collected to determine the low-resolution shape envelope of CMLE in solution, based on spherical harmonics expansion. The data analysis yielded the following geometrical parameters: R_g (radius of gyration) = 38.7 Å ± 1%, D_{max} (maximum particle dimension) = 109 Å ± 3% and V (particle volume) = 280 000 Å^3 ± 5%. These results are consistent with CMLE being a tetramer in solution (Fig. 2) and confirmed earlier gel-filtration studies (Mazur *et al.*, 1994). Most importantly, the SAXS envelope showed very similar features to the solvent mask calculated from the MAD experimental phases and was thus indicative of the correctness of the initial 4.0 Å F_o maps (see below and Fig. 2). As reliable scattering data were collected to a Bragg resolution of 12 Å, scattering-pattern simulations were performed against known tetramer structures of similar molecular mass from the Protein Data Bank (Bernstein *et al.*, 1977) consisting of non-homologous monomers with similar molecular mass. Intriguingly, the Gβ protein tetramer (Sondek *et al.*, 1996; PDB code 1tbg, excluding the γ-chain) was amongst the three best scoring structures (J. G. Grossmann *et al.*, in preparation). The Gβ protein is the only one of these three candidates which has a β-sheet-dominated secondary structure as does CMLE. In fact, it is also a seven-bladed β-propeller structure (Kajander *et al.*, 2002). However, molecular replacement using 1tbg did not provide a solution. Interestingly, the angles of the 'X-shaped' D_2 tetramer seem to be slightly different in the solution model of CMLE compared with the crystal structure. This possibly reflects the flexibility of the tetramer, as observed from the r.m.s.d. between the C$^\alpha$ atoms of the two tetramers in the asymmetric unit (see below), and the slight differences between the solution conditions for the SAXS experiment and those in the crystal.

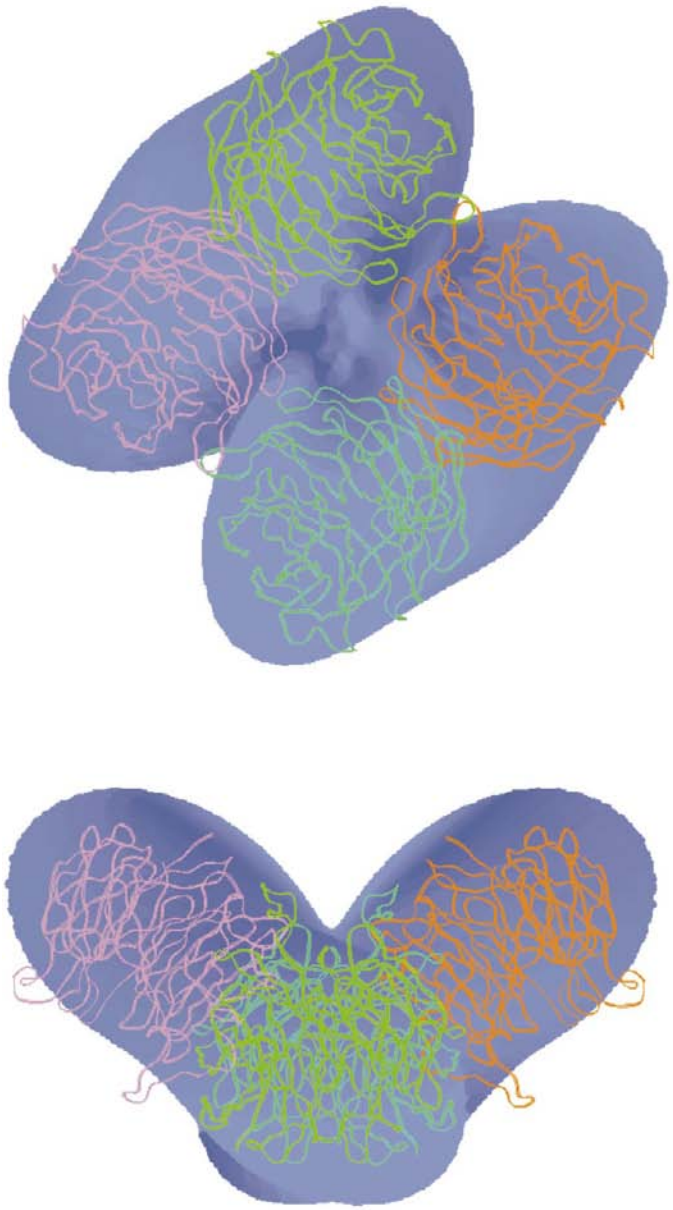

Figure 2
Restored molecular envelope of CMLE in solution deduced from X-ray solution-scattering data for up to the seventh order of harmonics (blue) and superimposed onto the ribbon structure of the current model (monomers are highlighted in different colours). Two orthogonal orientations are displayed.

3.3. Initial phasing

The peak anomalous difference data were scaled and normalized with the *DREAR* suite as default in the *SnB* program (see above). Of five separate *SnB* runs two indicated possible solutions, with a bimodal minimal function (R_{min}) distribution for the trials. A total of 1967 trials were processed, with a success rate of 2 out of 1967. The best two trials had final R_{min} values of 0.463 and 0.457, compared with a mean value of R_{min} = 0.495 for the unsuccessful trials. The coordinates from the two best independent solutions were compared in order to pick identical or symmetry-related coordinates (most were found to be related by $x + \frac{1}{2}, y, \frac{1}{2} - z$). Furthermore, when subsets of the sites from one of the above solutions were fed back into *SnB*, the search still converged on the same full solution (even with only a single correct peak the search converged rapidly). These comparisons yielded 60 consistent sites, which were refined with *MLPHARE* at 4.0 Å resolution. Of the 60 sites, 57 refined stably and these were independently refined again. The sign of the solution was determined by inspection of the initial maps after solvent flattening. The correct sign exhibited a lower free R factor for solvent flattening [as defined in *DM* (Collaborative Computational Project, Number 4, 1994), http://www.ccp4.ac.uk/dist/html/dm.html] and also gave sensible solvent masks with clear solvent-molecule boundaries and reasonable arrangement of

monomer masks. The maps also showed a recurring feature, a dimple, in the monomer mask. The quality of the maps was, however, not good enough to allow tracing (this was complicated further by the fact that the CMLE structure is an all-β structure with no α-helices that could possibly be traced at 4 Å resolution; Kajander *et al.*, 2002).

Knowing that up to seven non-crystallographic twofold axes could possibly exist in the asymmetric unit (three per tetramer and one inter-tetramer), self-rotation Patterson functions were calculated. These, however, did not show significant peaks above noise level (Fig. 1), as was also found in the dominant crystal type (Glumoff *et al.*, 1996). Consequently, these could not be used to obtain the NCS operators and we therefore attempted to use the selenium sites from the *SnB* solution to determine the position of the operators.

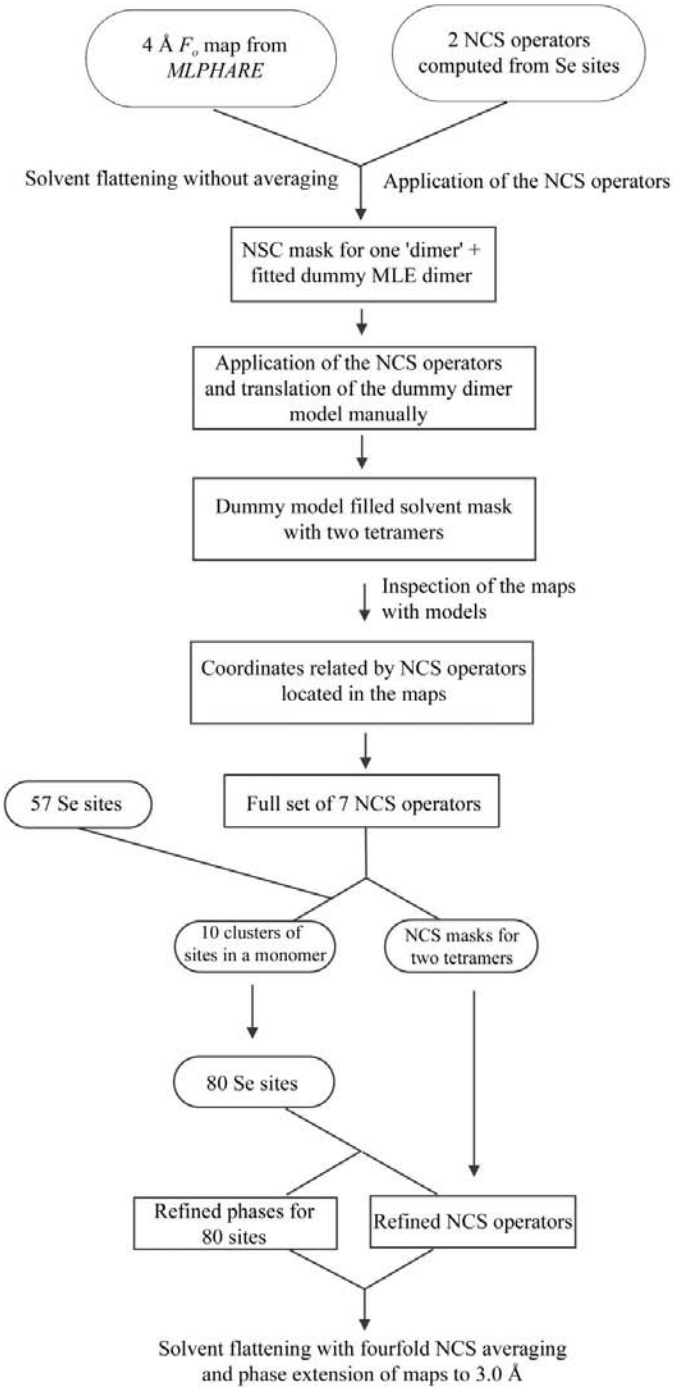

Figure 3
A flow chart for the CMLE structure solution with NCS averaging (see text for details).

3.4. Real-space NCS determination and phasing with averaging.

We attempted to determine the NCS relations from the Se sites and the real-space features of the initial 4.0 Å resolution unaveraged, solvent-flattened electron-density maps (Fig. 3). Inspection of the 57 sites obtained from *SnB* did not reveal any obvious symmetry relations because of the incompleteness of the model (57/80 sites) and the large unit-cell size. A purely computational approach was attempted by running the program *FindNCS* (Lu, 1999) with the 57 Se sites. Although the program did not complete, even with parameter adjustments and run times of over one week on a Digital UNIX Alpha 433au workstation, it was possible to sort and extract two NCS operators ($A, \varphi \simeq 68, \varphi \simeq 86^\circ$ and $B, \varphi \simeq 114, \varphi \simeq 87^\circ$; Fig. 1) from the log files.

Correlation maps were calculated with *COMA* (Kleywegt & Jones, 1999) to obtain masks for density averaging. The *A* operator produced a mask that was consistent with two monomers forming a dimer; it corresponded well to the solvent mask produced by *DM*. The *B* operator did not produce a good mask for reasons that became clear later (see below). As the *DM* molecular envelope had a toroidal shape, an appropriately shaped dummy model, the MLE α/β-barrel (Goldman *et al.*, 1987), was used to manually fill the solvent mask. The model was manually fitted to the region of the mask next to the local twofold (*A*) that yielded a dimer correlation mask and yielded a first (*A*) dimer. A second, unrelated, dummy 'dimer' was generated by applying the *B* twofold to the *A* dimer. $P2_12_12_1$ crystal symmetry was then applied to the pair of dimers created by the *A* and *B* twofolds: this generated a lattice with gaps in it. It was then clear that the *A* dimer could be manually fitted into an adjacent unoccupied region of the mask to generate an A–A' tetramer. Applying the *B* twofold again to the A' dimer generated a fourth dimer: a dimer belonging to the second tetramer in the asymmetric unit (Fig. 3). Application of $P2_12_12_1$ symmetry now generated a non-overlapping model which completely filled the masked region. The only manual bias introduced was in the fitting of the initial monomer and the translation and fitting of the first 'dimer' to generate a tetramer. In addition, the position and orientation of the non-crystallographic twofolds perpendicular to the *A* twofold were then determined by inspection before refinement using *O* and it became apparent that the F_o correlation map was very flat because of the many special positions of the local twofolds (Fig. 1; see below). Using the model for orientation allowed close inspection of the initial solvent-flattened MAD map. It was eventually possible to identify electron-density features that were twofold related.

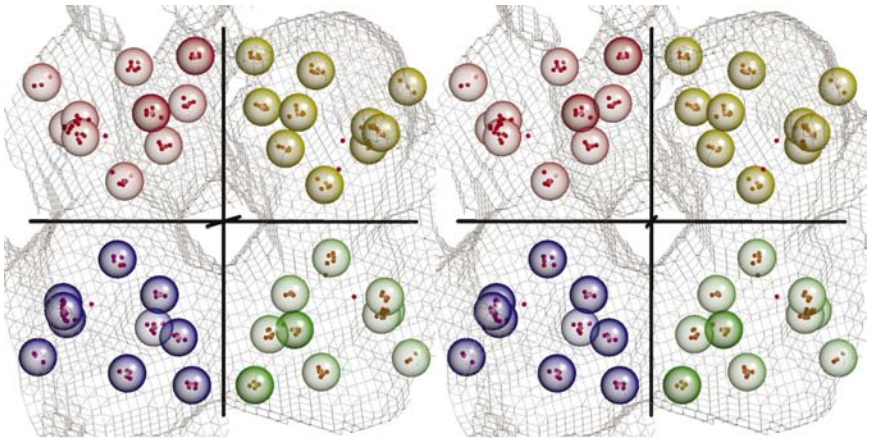

Figure 4
The clustering of Se sites. The clusters (see text) are shown in stereo within a tetramer mask shown in fishnet; large transparent spheres represent centroids for each sites after refinement, coloured according to the monomer to which they belong, and small spheres represent the Se positions expanded using local and crystallographic symmetry. Some of the Se sites from the initial solution that were wrong can be seen outside the large spheres. The vectors for the refined local twofolds are shown in black.

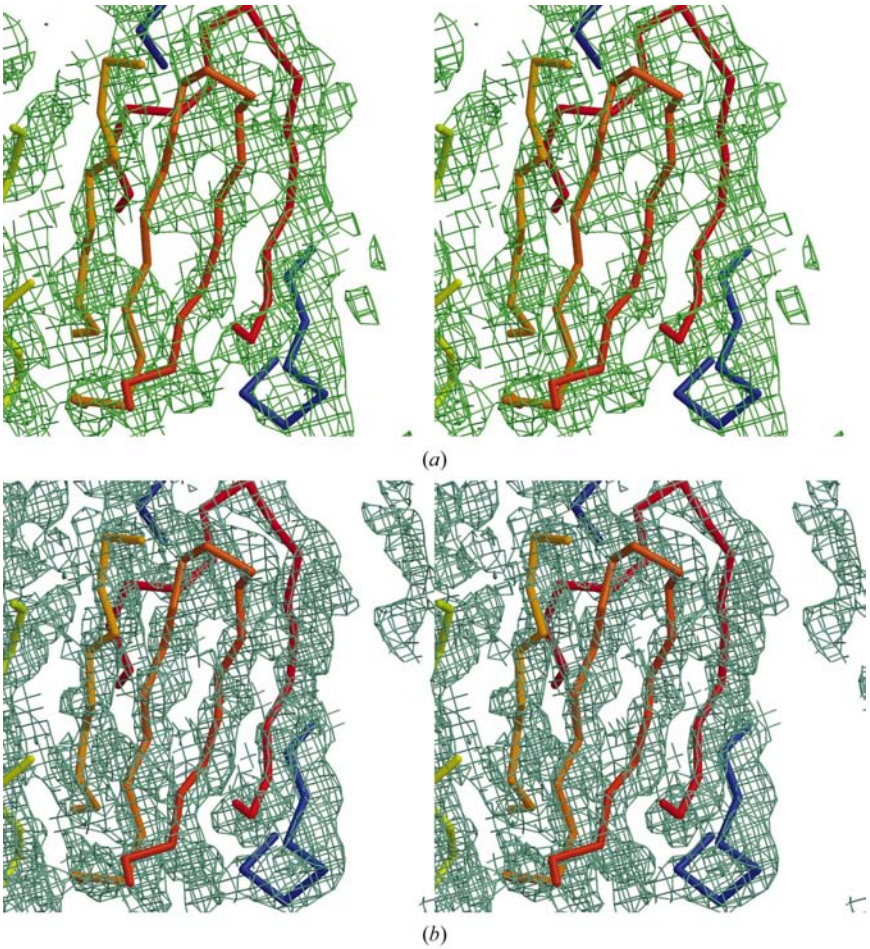

Figure 5
Representative electron-density maps. (*a*) Initial 4 Å solvent-flattened map and (*b*) solvent-flattened, NCS-averaged 3 Å map, both with C^{α} trace of the model built at 3 Å resolution.

The use of the solvent mask and dummy model made it possible to inspect directly for symmetries, which could not be performed in the initial map. Coordinates were determined for symmetry-related features defining each twofold and NCS operators were calculated from these in *O*. These operators were used to generate NCS masks for two D_2 tetramers using *COMA* and *MAMA* (Kleywegt & Jones, 1999).

We have also, retrospectively, examined the positions of the twofold peaks in order to understand why the rotation function was so featureless and for clues as to why attempts to compute the NCS operators from the initial 60 Se sites failed. It is clear that the *A* peak corresponds to a local twofold that is part of one of the two tetramers in the asymmetric unit and the *B* peak corresponds to the intertetramer twofold (Fig. 1). Even the F_c-based rotation function is essentially flat except for what appear to be noise features along the $\psi = 90^\circ$ directions (red; Fig. 1). The reason for this becomes clear upon inspecting the twofolds. Because the *A*-type and *B*-type twofolds (intramolecular and intermolecular, respectively) are almost parallel and close to $\psi = 90^\circ$, the positions of the other non-crystallographic twofolds are also close. As a result, all of the twofold peaks close to $\psi = 90^\circ$ collapse, both in the F_o and in the F_c correlation functions, to peaks exactly along $\psi = 90^\circ$. In addition, the peaks that are *not* along $\psi = 90^\circ$ are then too small to be seen in the F_o map; even in the F_c map, some of them (*e.g.* in the blue triangle) coalesce into one peak. We also believe that this may explain why *FindNCS* (Lu, 1999) failed to find all the local twofolds in a week of CPU time: there are 6320 interatomic vectors in the full selenium structure, with little distinction between them as to which twofold relates them. Alternatively, it may be that the program would not have been able to find a solution within limits of finite calculation time as the problem of finding the operators is NP-complete.

A complete set of 80 selenium sites for phasing was created by expanding the original 57 sites with $P2_12_12_1$ symmetry and all seven twofold NCS operators. (This created $2^7 \times 57$ sites altogether inside the unit cell, as two of the local twofolds and one of the 2_1 operators are redundant.) All the sites were displayed simultaneously on the graphics and, as we hoped, this created

80 clusters of related sites within the asymmetric unit (Fig. 4). All correct sites formed clusters within the solvent masks of the two tetramers; the 80 unique clusters per asymmetric unit were consistent with ten selenomethionines per monomer (Fig. 4). For sites where a selenium had not been identified initially, coordinates were taken from the centroid of the cluster of Se atoms resulting from the superimposition. From an analysis of the 57 sites expanded to fill the asymmetric unit, it became apparent that 56 of 57 sites constituted the 80 clusters and therefore one site was shown to be false and was eliminated.

Table 2
The *SnB* sites ordered by peak height tabulated to show methionine and monomer position in final model.

False sites found (in two independent solutions) by *SnB* v2.0: peaks 18/25 (Mol3), 55/104 (outside), 56/57 (Mol2), 61/70 (Mol4), 68/48 (Mol1) and 69/51 (Mol6). ASA = relative solvent accessibility from *NACCESS*. Mol1–4 = first tetramer; Mol5–8 = second tetramer. *S* = secondary-structure element.

Met	Mol1	Mol2	Mol3	Mol4	Mol5	Mol6	Mol7	Mol8	Total for residue	ASA	*S*
6	59/49	62/46	—	13/12	30/33	—	—	—	4	1.0	Strand
45	25/16	28/39	—	—	2/5	15/7	3/ 2	16/26	6	0.2	Strand
59	—	63/61	32/40	26/30	—	—	—	70/44	4	37.2	Turn
150	—	57/36	4/1	44/41	—	72/72	14/17	34/32	6	4.3	Strand
200	6/4	9/11	—	—	31/28	—	27/27	29/22	5	13.1	Strand
211	—	65/43	43/18	12/8	45/50	1/3	3./29	11/14	7	0.0	Strand
228	41/47	38/21	—	52/53	21/62	36/38	—	—	5	14.0	Turn
269	—	7/6	33/42	43/64	—	—	20/13	35/52	5	0.0	Strand
324	—	—	17/15	5/9	48/54	10/19	58/56	49/35	6	0.0	Strand
359	—	8/23	24/20	23/24	—	53/31	40/59	19/10	6	8.9	Turn
Total for monomer	4	9	6	8	6	6	7	8	56		

It is worth noting that all more conventional approaches, including calculating anomalous difference Fourier maps using phases based on the initial 57 Se-atom solution, did not produce any clear additional peaks, while the approach we took here allowed us to proceed and solve the structure. We then successfully refined all 80 sites with *MLPHARE* (Collaborative Computational Project, Number 4, 1994). Subsequent inspection of the model revealed that these matched the positions of the Se atoms in the SeMet residues in the protein sequence. The NCS operators were refined at this point with the program *IMP* (Kleywegt & Jones, 1999). Solvent flattening with independent fourfold NCS averaging of the tetramers and phase extension to 3.0 Å resolution was performed with *DM* (Collaborative Computational Project, Number 4, 1994). Full eightfold averaging was not possible, as the two tetramers in the asymmetric unit were apparently not identical but differed slightly in their inter-monomer angles (as was seen in the final refined model; Kajander *et al.*, 2002). However, the resulting map was dramatically improved compared with the 4.0 Å unaveraged solvent-flattened map and allowed most of the polypeptide to be traced (Fig. 5). Sequence identification was simplified by using the selenium positions. This initial model was used to find a molecular-replacement solution in the dominant crystal type, which had slightly differing unit-cell parameters but diffracted to higher resolution. The model was refined to 2.5 Å resolution and the structure, a β-propeller, will be reported elsewhere (Kajander *et al.*, 2002).

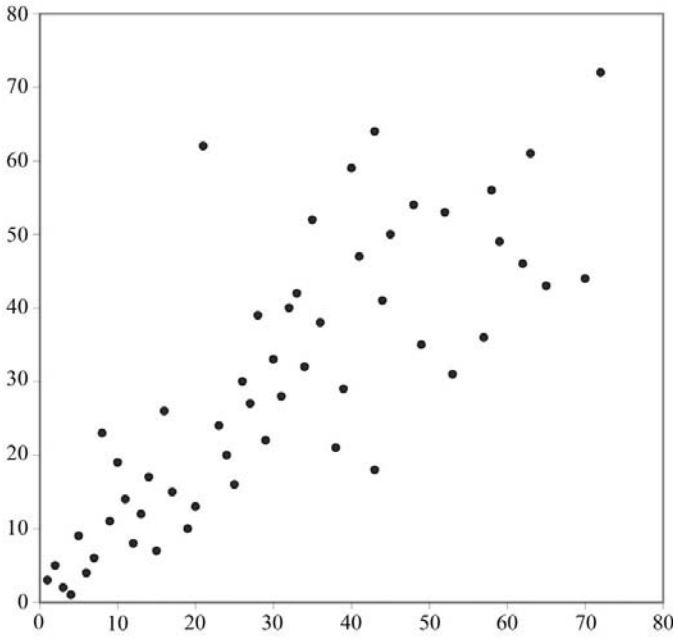

Figure 6
Correlation between the peak heights (given as the order of the peak in each solution) of equivalent coordinates in the two best *SnB* solutions. (Peak-order numbers are given for two solutions on the *x* and *y* axes, while each data point represents a Se position.)

The distribution of the 57 peaks found by *SnB* was analyzed for the two best solutions. There was no observed correlation between Se positions and peak height (Table 2). There was, however, clear correlation in the peak order of related peaks between the two runs (Fig. 6). Overall, it appears that simply cross-checking between two solutions is not enough to fully validate that a peak is correct, because false positives still occur. Of the 60 sites initially found, 56 turned out to be real and 24 sites were not found. Consequently, the generation and application of local symmetry elements by hand was crucial in two ways. Firstly, it allowed us to validate the initial maps, showing that application of the two local twofolds found computationally and by map inspection could be used to construct a sensible protein model. Secondly, it allowed us to identify all of the Se sites and thus improve our phasing; all other attempts left 30% of the sites not found. The difficulty of finding the Se sites is not, as far as we have been able to determine, related to our data-collection strategies; the unit cell is relatively large for a MAD experiment and the high non-isotropic mosaicity meant that data were always systematically absent at high angle. (For instance, although native crystals diffract to better than 2 Å, we have only been able to refine a 2.5 Å structure owing to spot overlaps.) In conclusion, we believe that our work has shown the necessity for imagi-

native holistic use of all available information, including manual inspection of electron-density maps, in solving large structures by MAD. This will become more important as MAD is used to tackle structures larger than CMLE, with 80 Se sites and 320 kDa in the asymmetric unit. As such, we offer this study as a potential roadmap for future investigators.

We thank Anna-Karoliina Ahonen for assistance with data collection, Raija Andersen for technical assistance, and Petri Vähäkoski and Markus Tujula for computer support. This work was supported by the Finnish Academy, Grants 69520 and 63252.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **256**, 590–600.

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

Deacon, A. M. & Ealick, S. E. (1999). *Structure*, **7**, 161–166.

Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.

Glumoff, T., Helin, S., Mazur, P., Kozarich, J. W. & Goldman, A. (1996). *Acta Cryst.* D**52**, 221–223.

Goldman, A., Ollis, D. L. & Steitz, T. A. (1987). *J. Mol. Biol.* **194**, 143–153.

Grossmann, J. G., Sharff, A. J., O'Hare, P. & Luisi, B. (2001). *Biochemistry*, **40**, 6267–6274.

Hädener, A., Matzinger, P. K., Malashkevich, V. N., Louie, G. V., Wood, S. P., Oliver, P., Alefounder, P. R., Pitt, A. R., Abell, C. & Battersby, A. R. (1993). *Eur. J. Biochem.* **211**, 615–624.

Hauptman, H. A. (1991). *Crystallographic Computing 5: From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 324–332. Oxford University Press.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kajander, T., Merckel, M. C., Thompson, A., Deacon, A. M., Mazur, P., Kozarich, J. W. & Goldman, A. (2002). In the press.

Kirby, G. W., O'Laughlin, G. J. & Robins, D. J. (1975). *J. Chem. Soc. Chem. Commun.*, pp. 402–403.

Kleywegt, G. J. & Jones, T. A. (1999). *Acta Cryst.* D**55**, 941–944.

Lewis, R. (1994). *J. Synchrotron Rad.* **1**, 43–53.

Lu, G. J. (1999). *J. Appl. Cryst.* **32**, 365–368.

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

Mazur, P., Henzel, W. J., Mattoo, S. & Kozarich, J. W. (1994). *J. Bacteriol.* **176**, 1718–1728.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E. & Sigler, P. B. (1996). *Nature (London)*, **379**, 369–374.

Towns-Andrews, E., Berry, A., Bordas, J., Mant, G. R., Murray, P. K., Roberts, K., Sumner, I., Morgan, J. S., Lewis, R. & Gabriel, A. (1989). *Rev. Sci. Instrum.* **60**, 2346–2349.

Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.

Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.

Williams, S. E., Woolridge, E. M., Ransom, S. C., Landro, J. A., Babbitt, P. C. & Kozarich, J. W. (1992). *Biochemistry*, **31**, 9768–9776.