Solving a 300 kDa multimeric protein by low-resolution MAD phasing and averaging/phase extension

The structure of the conjugative coupling protein TrwBΔN70 from Escherichia coli plasmid R388 was solved using two crystal forms. This large multimeric membrane protein of 437 residues per monomer is involved in cell-to-cell single-strand DNA transfer. Diffraction data to 2.4 Å were available from trigonal crystals obtained from ammonium sulfate and to 2.5 Å from monoclinic crystals grown from tartrate. A single tantalum bromide (Ta₆Br₂²⁺) derivative of the trigonal form, which presented a protein hexamer with C₆ local symmetry in the asymmetric unit, was used in a three-wavelength MAD experiment to achieve 4.5 Å resolution for initial phases. Sixfold averaging and phase extension increased the effective phasing resolution and eventually produced a straightforwardly traceable electron-density map. The monoclinic structure was solved by molecular replacement, i.e. a hexamer of the trigonal form was used as a search model. Two such hexamers are present in the asymmetric unit.

1. Introduction

The main issue in protein crystallography is the determination of phases for the measured structure-factor amplitudes. The most widely used technique for novel structures has historically been isomorphous replacement, which consists of the introduction of atoms or ions with high scattering power into the otherwise unchanged crystal lattice, yielding detectable changes in the intensities of the diffraction patterns. This technique is susceptible to non-isomorphism between the crystal lattices of the native compound and the heavy-ion/atom derivative prepared. However, the use of multi-wavelength anomalous diffraction (MAD), which has increased with the availability of wavelength-tuneable synchrotron beam time and the emergence of crystal cryo-cooling techniques, has reduced this problem (González et al., 1999; Terwilliger, 1997).

Although experimental approaches have been developed to incorporate modified amino acids containing ‘heavy-atom’ scatterers [mainly selenomethionine (Budisa et al., 1995; Doublié, 1997)] into proteins efficiently, a large or small number of methionine residues in the asymmetric unit of crystals of large macromolecular assemblies hinders detection or renders poor phasing power, respectively. Conventional derivative preparation may also fail, since heavy ions/atoms are not easily localized within large cells owing to the excessive number of sites. In these cases, bulky ionic clusters of high scattering power can be useful and detectable using low-resolution data.

This is valid for tantalum bromide (Ta₆Br₂²⁺), a regular octahedron of a 4.3 Å radius, successfully employed in a...
number of structural analyses (Knäblein et al., 1997). This compound introduces significant changes in the diffraction pattern of large unit cells and is formed of ions with absorption edges both with significant anomalous signals appropriate for MAD experiments and energies that can be reached at most tunable synchrotron beamlines ($L_{III}$ edge at 1.25 Å for Ta, $K$ edge at 0.95 Å for Br). It displays an intrinsic deep green colour that indicates putative derivatization. If the cluster does not adopt a unique conformation in the crystal lattice, the exact position of the 18 constituting ions is not determinable and so the derivative has to be treated as a single point scatterer. In these cases, the useful phasing high-resolution limit seems to be 6 Å (Knäblein et al., 1997).

We applied this derivatizing agent to solve the structure of a soluble variant of TrwB, a basic (pI = 10) integral membrane plasmid protein encoded by the $dtr$ region of E. coli plasmid R388 made up of 507 residues (Llosa et al., 1994; Moncalián et al., 1999). Sequence analysis localized the transmembrane part in the first 70 N-terminal residues (two transmembrane helices and a small periplasmic domain in between). Therefore, TrwB was overproduced and purified as a soluble fragment lacking these residues (TrwB$_N$70).

Here, we report the crystallization and the detailed structure determination protocol of 437-residue TrwB$_N$70 by MAD and subsequent averaging/phase extension.

2. Methods

2.1. Purification, crystallization and derivative preparation

TrwB$_N$70, produced by overexpression and purified as described in Moncalián et al. (1999), was provided by G. Moncalián and F. de la Cruz. The protein was crystallized by the hanging-drop vapour-diffusion method using Linbro plates and Hampton Research Screens to obtain initial conditions.

The best conditions consisted of equivolumetric drops (4 µl:4 µl) of protein solution (9 mg ml$^{-1}$) and 1.5 M ammonium sulfate, 0.1 M NaCl, 0.1 M HEPES pH 7.5 as precipitant agent solution, which yielded well shaped trigonal crystals belonging to space group $P3_121$ (unit-cell parameters $a = b = 151.3$, $c = 258.2$ Å). In parallel, when 0.9 M potassium/sodium tartrate, 0.1 M HEPES pH 7.5 was used as precipitant, the same trigonal crystals appeared in the drops together with a second, monoclinic ($P2_1$) crystal form (unit-cell parameters $a = 107.4$, $b = 153.4$, $c = 162.5$ Å, $\beta = 94.2^\circ$; see Fig. 1). Some of the monoclinic crystals presented a second, smaller monoclinic cell (unit-cell parameters $a = 95.2$, $b = 155.8$, $c = 104.7$ Å, $\beta = 113.1^\circ$). N-terminal sequencing and mass spectrometry of carefully washed and dissolved crystals revealed the expected molecular weight and sequence. Prior to data collection, cryoprotecting strategies consisted of stepwise increasing amounts of glycerol (5–20%; crystals obtained from ammo-
Table 1
Data collection and processing.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Native sulfate complex</th>
<th>Native</th>
<th>Ta₆Br²⁺ (f_max)</th>
<th>Ta₆Br²⁺ (f_min)</th>
<th>Ta₆Br²⁺ (remote)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group P₃[21]</td>
<td>P₃[21]</td>
<td>P₃[21]</td>
<td>0.844</td>
<td>1.053</td>
<td>1.254</td>
</tr>
<tr>
<td>Wavelength λ (Å)</td>
<td>72.6–2.40</td>
<td>2.64–2.50</td>
<td>3.32–3.13</td>
<td>3.13–3.00</td>
<td>3.13–3.00</td>
</tr>
<tr>
<td>No. of measurements</td>
<td>1205165</td>
<td>728622</td>
<td>260537</td>
<td>317859</td>
<td>317424</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9</td>
<td>97.3</td>
<td>93.1</td>
<td>99.7</td>
<td>95.8</td>
</tr>
<tr>
<td>Anomalous completeness (%)</td>
<td>5.7</td>
<td>5.4</td>
<td>9.1</td>
<td>7.5</td>
<td>6.9</td>
</tr>
<tr>
<td>Rmerge(%)</td>
<td>9.0</td>
<td>9.9</td>
<td>3.7</td>
<td>6.8</td>
<td>5.5</td>
</tr>
<tr>
<td>Last resolution shell (Å)</td>
<td>2.53–2.40</td>
<td>2.64–2.50</td>
<td>3.32–3.13</td>
<td>3.32–3.13</td>
<td>3.13–3.00</td>
</tr>
<tr>
<td>Average multiplicity</td>
<td>9.0</td>
<td>6.2</td>
<td>5.2</td>
<td>4.5</td>
<td>4.7</td>
</tr>
</tbody>
</table>

*The high-resolution data collection was limited by the diameter of the detector and the large cell constants.

A high-resolution cutoff of 6 Å, which allowed each cluster site to be considered as a single point scatterer, was applied. The three reduced data sets were prepared with XPREP and dispersive and anomalous sharpened difference Patterson maps were calculated. The program produced a unique reduced AF-like data set which allowed us to determine three cluster positions by SHELXS that were consistent with the Patterson maps.

2.3. Data analysis, structure solution and refinement

Self-rotation calculations were performed with ncode = 1 using data in the 10–3.5 Å range and a vector radius of 37 Å with the program GLRF (Tong, 1993). The orthogonalization convention chosen was AXABZ (Tong, 1993). Initial cluster sites were determined with XPREP/ SHELXS from the SHELX97 suite (Sheldrick, 1997; Sheldrick et al., 1993; Sheldrick & Schneider, 1997) using the data corresponding to the three-wavelength MAD experiment.
3. Results and discussion

3.1. Self-rotation calculations

The putatively high number of protomers in the asymmetric units of both crystal types suggested high-order local symmetry. Indeed, self-rotation calculations indicated a local sixfold axis at polar angles $\Psi = 30^\circ$, $\Phi = 112^\circ$, $\kappa = 60^\circ$ (height 4.0$\sigma$) for the trigonal crystals and $\Psi = 54^\circ$, $\Phi = 85^\circ$, $\kappa = 60^\circ$ (6.5$\sigma$) and $\Psi = 54^\circ$, $\Phi = 91^\circ$, $\kappa = 60^\circ$ (6.2$\sigma$) for monoclinic data. At the same orientation, the associated twofold and threefold axes appeared as the only additional peaks. These results are consistent with the presence of six monomers displaying local $C_6$ symmetry (hexamer) in the trigonal a.u. and two hexamers in the monoclinic a.u. (see Fig. 2). Accordingly, a putative selenomethionine derivative of TrwB would have entailed localizing 60 selenium sites (ten methionine residues per protomer, see TrEMBL access code Q04230) in the trigonal form or 120 in the monoclinic space group. Although this is not an impossible task with the methods available currently, it remains complicated.

3.2. Initial phasing

In order to determine the initial phases, a MAD experiment was performed with the $\text{Ta}_{6}\text{Br}_{2}^{12+}$ derivative of the trigonal crystal form. The intense green colour of the soaked (initially transparent and colourless) crystals contrasted with the pale colouring of the drops and was the first evidence of derivatization. Three data sets were recorded after determination of the appropriate wavelengths to optimize the anomalous signal. A total of six sites were found and reined and phases were computed for the resolution range 40–4.5 Å; the overall mean figure of merit (f.o.m.; definition in Table 2) was 0.63 (see Table 2). However, these sites did not follow the local $C_6$ symmetry found with the self-rotation function calculation (see Fig. 3a). Although the whole cluster was treated as a single point scatterer, this figure was lower but still significant in the 5.1–4.5 Å (0.49) resolution shell, in contrast to the reported limit of usefulness of 6 Å (Knäβlein et al., 1997).

3.3. Density modification, averaging and phase extension

A subsequent density-modification step (35–4.5 Å; f.o.m. = 0.79) assuming a solvent content of 59% (six molecules per a.u.) rendered an $F_{\text{obs}}$ electron density that was not traceable (see Fig. 3a) but allowed us to build a monomer mask because the protomer boundaries were clearly visible. A molecule of similar size to TrwBΔN70 (procarboxypeptidase A, see Gomis-Rüth et al., 1995) was initially fitted into the density. This operation was repeated for each monomer to determine

---

**Table 2**

MAD phasing statistics.

<table>
<thead>
<tr>
<th>Data set</th>
<th>$\text{Ta}<em>{6}\text{Br}</em>{2}^{12+}$ ($f_{\text{max}}$)</th>
<th>$\text{Ta}<em>{6}\text{Br}</em>{2}^{12+}$ ($f_{\text{min}}$)</th>
<th>$\text{Ta}<em>{6}\text{Br}</em>{2}^{12+}$ (remote)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy ion cluster sites§ $\dagger$</td>
<td>0.546, 0.181, 0.650</td>
<td>0.28</td>
<td>3.56</td>
</tr>
<tr>
<td>0.792, 0.279, 0.535</td>
<td>0.29</td>
<td>4.50</td>
<td>0.0</td>
</tr>
<tr>
<td>0.204, 0.966, 0.807</td>
<td>0.24</td>
<td>2.91</td>
<td>0.0</td>
</tr>
<tr>
<td>0.287, 0.701, 0.293</td>
<td>0.10</td>
<td>1.35</td>
<td>0.0</td>
</tr>
<tr>
<td>0.245, 0.641, 0.744</td>
<td>0.09</td>
<td>1.19</td>
<td>0.0</td>
</tr>
<tr>
<td>0.105, 0.374, 0.419</td>
<td>0.06</td>
<td>0.73</td>
<td>0.0</td>
</tr>
<tr>
<td>Resolution range used for phasing (Å)</td>
<td>–</td>
<td>40–4.5</td>
<td>–</td>
</tr>
<tr>
<td>Mean figure of merit (f.o.m.) $\dagger$</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phasing power (acentric)$\dagger$</td>
<td>1.31</td>
<td>–</td>
<td>1.70</td>
</tr>
<tr>
<td>$R_{\text{calc}}$ (acentric)$\dagger$</td>
<td>0.79</td>
<td>–</td>
<td>0.72</td>
</tr>
<tr>
<td>$R_{\text{calc}}$ (anomalous)</td>
<td>0.75</td>
<td>0.81</td>
<td>0.87</td>
</tr>
</tbody>
</table>

† The data set at the inflection point ($f_{\text{max}}$) was taken as reference. § Fractional cell coordinates. § Occupancy and anomalous occupancy are on arbitrary scales. $\dagger$ f.o.m. = $\sum |F_{\text{hkl}}| - |F_{\text{hkl}}|_{\text{calc}}$/$\sum |F_{\text{hkl}}|$. 

---

**Research papers**

_Gomis-Rüth & Coll • TrwBΔN70_ 803

preliminary parameters for the local sixfold axis and the associated approximate non-crystallographic symmetry (NCS) operators (see Table 2). These values were consistent with self-rotation calculations (see above). A second density-modification step with density averaging and phase extension under refinement of these NCS operators increased the fom to 0.75 (50–3.0 Å). An inspection of the electron density around each cluster site revealed an almost spherical blob that hindered the precise determination of the position of the 18 constituting ions. Thus, phasing using the full resolution of the \( Ta_{6}Br_{9}^{2+} \)-derivatized crystal could not be applied. The refined NCS operators were used in the last step: the original experimental MAD phases to 4.5 Å were directly used with the native sulfate complex structure-factor amplitudes and a last density-modification step was performed under averaging, phase extension and NCS-operator refinement starting at 5 Å resolution and finally encompassing the 50–2.7 Å resolution range (final overall f.o.m. = 0.67; see Table 2). Although the trigonal cells of the native sulfate complex and the derivative displayed a noteworthy anisomorphism (\( c = 258.2 \) versus 262.2 Å, a difference of 1.8%; see Table 1), this effect was not significant at 5 Å resolution, at which averaging and phase extension started. At this point, a straightforward traceable map was obtained (see Fig. 3b).

The electron density was interpreted and a starting model was built. Successive cycles of maximum-likelihood positional and temperature-factor refinement progressively using all data to the full resolution of 2.4 Å, NCS restraints, computation of phased-combined maps and manual modelling led to gradual completion of the model (see Figs. 3b and 3c). The refined model has been discussed elsewhere (Gomis-Rüth et al., 2001).

### 3.4. Molecular replacement to localize 40 000 atoms in the monoclinic native asymmetric unit and monoclinic crystal ambiguity

The native structure of the monoclinic crystal form was solved using data in the 15–4.5 Å resolution range by molecular replacement, which confirmed \( P2_1 \) as the correct space group. A whole TrwBΔN70 hexamer was used as a searching model and two clear solutions were obtained at

\[
\begin{align*}
\alpha &= 97.9, \quad \beta = 141.8, \quad \gamma = 293.6^\circ, \quad x = 0.2398, \\
y &= -0.0001, \quad z = 0.4666 \quad \text{and} \quad \alpha = 86.3, \quad \beta = 145.2, \quad \gamma = 340.7^\circ.
\end{align*}
\]

**Figure 3**

Electron-density maps corresponding to the trigonal crystal form. (a) Initial \( F_{\text{obs}} \) density contoured at 2\( \sigma \) calculated with experimental phases using data to 4.5 Å resolution after density modification but prior to averaging, viewed along the local sixfold axis. Red spheres denote the position of the six heavy-ion cluster sites. The unit cell displayed in the upper right corner orients the C6 macro-molecule cluster. (b) Interpretable \( F_{\text{obs}} \) density contoured at 1.2\( \sigma \) computed using experimental phases after density modification, averaging and phase extension to 2.7 Å. (c) Final \( \sigma_{\text{a}} \)-weighted (Read, 1986) \( 2F_{\text{obs}} - F_{\text{calc}} \) type map (1.2\( \sigma \) contour) to 2.4 Å resolution. The latter two maps further display the final refined model superimposed on the density as sticks.
providing TrwB short-cell crystals. The two hexamers found in the large monoclinic cell would be interconnected by one cell translation along the diagonals of the former, with equivalent hexamer per a.u. This small unit cell is related to the large one in such a way that the hexamer per a.u. This agrees with the presence of 12 monomers arranged as two hexamers per a.u. The model was refined after `fitting' [CC/Rfactor (for definition, see Navaza, 1994) equal to 67.4%/36.3%; second highest solution 41.2%/47.1%; calculated after positioning and rigid-body refinement with the self-rotation calculation (see Fig. 4). The transformation of the two molecules present in the large monoclinic cell corresponds to a rotation of \( \alpha = 95.2^\circ \), \( b = 155.8 \AA \), \( c = 104.7 \AA \), \( \beta = 113.1^\circ \) and therefore only one hexamer per a.u. This small unit cell is related to the large one in such a way that the \( a \) and \( c \) axes of the latter are the diagonals of the former, with equivalent \( b \) axes (see Fig. 4). The two hexamers found in the large monoclinic cell would be interconnected by one cell translation along the \( a \) axis in the short-cell crystals.

We thank Gabriel Moncalián and Fernando de la Cruz for providing TrwB\( \Delta \)N70, Robert Huber for supplying tantalum bromide and Isabel Usón for help with XPREP/SHELXS. We are also indebted to Rosa Pérez-Luque for assistance in crystallization experiments and to Ana González for synchrotron data collection. This study was supported by grants BIO2000-1659, PB98-1631 and 2FD97-0518 from the Ministerio de Educación y Cultura, Spain, and by grant 1999SGR188 and the Centre de Referència en Biotecnologia, both from the Generalitat de Catalunya. Support was also available from the ESRF and the European Commission Improving Human Potential Programme, Contract No. HPRI-1999-00017, and from EU TMR/LSF grant ERBFMGECT980134 to the EMBL Hamburg Outstation and ERBFMECT980133 to the EMBL Grenoble Outstation. The ESRF, Grenoble, France is acknowledged for the provision of synchrotron-radiation facilities.

Figure 4
Monoclinic ambiguity. Monoclinic crystals randomly presented two cells (see §3.4). The transformation of the two molecules present in the large-cell asymmetric unit corresponds to a rotation of \( \alpha \) and a translation of \( 93 \AA \). In the small unit cell, this rotation is reduced to a pure cell translation and both molecules belong to different unit cells.

\[
x = 0.7450, \ y = 0.9884, \ z = 0.9687 \quad (\text{where } \alpha, \beta \text{ and } \gamma \text{ are Eulerian angles and } x, y \text{ and } z \text{ are in fractional cell coordinates}) \quad \text{after `fitting' } [\text{CC/Rfactor} \text{ (for definition, see Navaza, 1994)} \text{ equal to } 67.4\%/36.3\%; \text{second highest solution } 41.2\%/47.1\%; \text{calculated after positioning and rigid-body refinement of both hexamers}]. \text{This agrees with the presence of } 12 \text{ monomers arranged as two hexamers per a.u. The model was inspected and the new electron-density-based differences with the trigonal structure were corrected. The structure has been described elsewhere (Gomis-Ruth et al., 2001). The two hexamers in the monoclinic a.u. are related by a quasi-pure (local) translation of \( 93 \AA \) and a slight relative rotation of \( 8^\circ \), in agreement with the two close peaks found with the self-rotation calculation (see §3.1), just separated by \( 6^\circ \) in \( \Phi \). This finding explains why some monoclinic crystals presented a smaller cell with unit-cell parameters \( a = 95.2, \ b = 155.8 \AA, \ c = 104.7 \AA, \ \beta = 113.1^\circ \) and therefore only one hexamer per a.u. This small unit cell is related to the large one in such a way that the \( a \) and \( c \) axes of the latter are the diagonals of the former, with equivalent \( b \) axes (see Fig. 4). The two hexamers found in the large monoclinic cell would be interconnected by one cell translation along the \( a \) axis in the short-cell crystals.

We thank Gabriel Moncalián and Fernando de la Cruz for providing TrwB\( \Delta \)N70, Robert Huber for supplying tantalum bromide and Isabel Usón for help with XPREP/SHELXS. We are also indebted to Rosa Pérez-Luque for assistance in crystallization experiments and to Ana González for synchrotron data collection. This study was supported by grants BIO2000-1659, PB98-1631 and 2FD97-0518 from the Ministerio de Educación y Cultura, Spain, and by grant 1999SGR188 and the Centre de Referència en Biotecnologia, both from the Generalitat de Catalunya. Support was also available from the ESRF and the European Commission Improving Human Potential Programme, Contract No. HPRI-1999-00017, and from EU TMR/LSF grant ERBFMGECT980134 to the EMBL Hamburg Outstation and ERBFMECT980133 to the EMBL Grenoble Outstation. The ESRF, Grenoble, France is acknowledged for the provision of synchrotron-radiation facilities.

References