Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

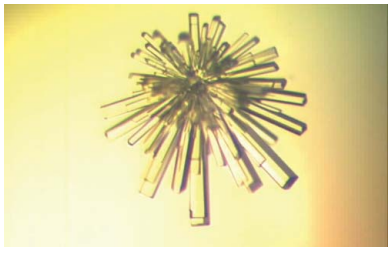
Crystallization and preliminary structure analysis of CobE, an essential protein of cobalamin (vitamin B_{12}) biosynthesis

Jitka Vévodová,[a] Ross M. Graham,[b] Evelyne Raux,[b] Martin J. Warren[b] and Keith S. Wilson[a]*

[a]Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, England, and [b]School of Biological Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, England

Correspondence e-mail: keith@ysbl.york.ac.uk

Received 23 January 2005
Accepted 3 March 2005
Online 1 April 2005



CobE, a protein implicated in vitamin B_{12} biosynthesis, from *Pseudomonas aeruginosa* has been overexpressed in *Escherichia coli*, purified and crystallized using hanging-drop vapour diffusion. The crystals belong to the primitive orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 31.86$, $b = 41.07$, $c = 87.41$ Å. The diffraction extends to a resolution of 1.9 Å. There is one molecule per asymmetric unit and the estimated solvent content is 35%. SeMet-labelled CobE has been prepared and crystallizes under the same conditions as the native protein with diffraction to 1.7 Å. The anomalous measurements will be used for phasing.

1. Introduction

Vitamin B_{12} (cobalamin) is a modified tetrapyrrole that belongs to the same family of metallo-prosthetic groups as haem, chlorophyll, sirohaem and coenzyme F_{430} (Warren & Scott, 1990). Modified tetrapyrroles are synthesized *via* a branched biosynthetic pathway, with 5-aminolevulinic acid (ALA) the first common intermediate and uroporphyrinogen III the first branch-point step in the pathway. The structural complexity of vitamin B_{12} is reflected in its biosynthesis, which remains one of nature's most convoluted metabolic pathways, requiring around 30 enzymes for the complete *de novo* construction of the coenzyme form (Warren *et al.*, 2002). For cobalamin biosynthesis, two distinct yet similar routes exist, known as the oxygen-dependent (aerobic) and oxygen-independent (anaerobic) pathways (Roessner *et al.*, 2001; Warren *et al.*, 2002). The aerobic pathway is found in bacteria such as *Psuedomonas denitrificans* and *P. aeruginosa*, whereas the more common anaerobic route is found in the archaea and eubacteria such as *Salmonella enterica* and *Bacillus megaterium*. The enzymes of the aerobic pathway are given the prefix Cob, whereas those of the anaerobic pathway are termed Cbi.

Although the anaerobic pathway to cobalamin has yet to be completely elucidated, all the intermediates in the aerobic pathway have now been elucidated and the enzyme activities determined. However, even in the aerobic pathway there are some proteins that are known to be required for cobalamin biosynthesis yet for which no function has been assigned. One of these proteins is CobE, a comparatively small protein with a molecular weight of about 15 kDa. An orthologue of this protein, CbiG, is also found in the anaerobic pathway. CobE displays about 25% identity with the C-terminal region of CbiG, a 350-amino-acid protein for which no function has yet been ascribed. Genetic studies have shown that deletion of either CbiG or CobE from their respective pathways leads to a drastic reduction in cobalamin biosynthesis (less than 1% of the levels found in control strains).

As part of a coordinated effort to establish the function of CobE, we have sought to determine the structure of this protein by X-ray crystallography. In this paper, we report the crystallization and initial characterization of CobE.

2. Material and methods

2.1. Cloning of *cobE*

The region of the *P. aeruginosa* genome corresponding to open reading frame (ORF) PA2947 (*cobE*; http://www.pseudomonas.com)

was amplified from PA01 genomic DNA by the polymerase chain reaction (PCR) with primers specific for the 5′ and 3′ ends of the ORF. The primers, 5′-GCCATATGCCCCTGCCCATCCCCAG-3′ and 5′-GCGGATCCATGAAATAGACGGTC-3′, contained *Nde*I and *Bam*HI sites, respectively. Cycling conditions were initial denaturation 368 K for 3 min, followed by 35 cycles of 369 K for 30 s, 331 K for 30 s and 345 K for 55 s with a final extension period of 345 K for 10 min. Following PCR, the 455 bp product was ligated into the pGEM-T(easy) cloning vector (Promega), excized by double digestion with *Nde*I and *Bam*HI and subcloned into the expression vector pET14b (Novagen), which encodes an N-terminal hexahistidine tag. Sequence analysis (MWG Biotech) of *cobE::pET14b* confirmed that ORF PA2947 (*cobE*) had been cloned in frame into pET14b.

2.2. Protein production

The vector *cobE::pET14b* was transformed into the *Escherichia coli* BL21(DE3) host strain harbouring the pLyS plasmid for overexpression. Cells were grown overnight at 310 K in 5 × 5 ml Luria–Bertrani (LB) medium containing 100 µg ml^{-1} carbenicillin. The stationary phase culture was inoculated into 5 × 500 ml LB medium containing 100 µg ml^{-1} carbenicillin. The culture was grown at 310 K to an optical density OD_{600} of 0.7, induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 m*M* and growth was continued for a further 5 h at 303 K. The cells were harvested by centrifugation at 7000*g* for 25 min at 277 K and resuspended in lysis buffer: 20 m*M* Tris–HCl pH 8.5 containing 0.2 *M* NaCl, 1 m*M* β-mercaptoethanol and EDTA-free Complete Protease Inhibitor (Roche).

To express SeMet-labelled protein, the *cobE::pET14b* vector was transformed into the Rb-competent *E. coli* B834 host strain. The protein-production strategy was very similar to that for the native protein using the protocol described by Ramakrishnan *et al.* (1993). Briefly, cells were grown overnight at 310 K in 2 × 5 ml LB medium containing 100 µg ml^{-1} carbenicillin. After centrifugation, the cell pellet was washed with M9 Minimal medium and then resuspended in 2 × 500 ml of the same medium supplemented with a vitamin solution, trace elements and 100 µg ml^{-1} carbenicillin. The 19 L-amino acids repressing methionine biosynthesis were added together with L-SeMet (4 mg ml^{-1} each). The culture was grown at 310 K to an optical density OD_{600} of 0.7 and induced with 1 m*M* IPTG for 5 h at 303 K.

2.3. Purification and His-tag cleavage

The same purification protocol was used for both native and SeMet-modified proteins. After addition of EDTA-free Complete Protease Inhibitor (Roche), the cells were sonicated (Soniprep 150) on ice with 5 × 30 s bursts and 2 min cooling intervals. Lysozyme (1 mg ml^{-1}) was added to assist in cell lysis. The cell debris was removed by centrifugation at 14 000*g* for 25 min at 277 K. The CobE-containing supernatant was further clarified using 0.45 µl filters (Sartorius) and applied onto a HiTrap Chelating HP 5 ml column (Pharmacia) for purification. The column was washed with buffer containing 20 m*M* Tris–HCl pH 8.5, 0.2 *M* NaCl and 1 m*M* β-mercaptoethanol. The protein was eluted from the column with a 10–500 m*M* imidazole gradient. Protein-containing fractions were determined by SDS–PAGE and CobE was concentrated to about 7 mg ml^{-1}. After dialysis to remove imidazole, the N-terminal His_6 tag was cleaved by digestion with thrombin overnight at room temperature. Approximately 10 units of thrombin were added per milligram of protein. The cleaved His tag was removed using a HiTrap Chelating HP 5 ml column with the same buffer as for the first purification step. The fractions containing CobE were concentrated to a volume of 2 ml and loaded onto a Superdex 75 gel-filtration column which had been previously equilibrated with 20 m*M* Tris–HCl buffer pH 8.5 containing 0.2 *M* NaCl and 10 m*M* DTT. The protein eluted as a single peak and the relevant fractions were pooled and concentrated.

2.4. Crystallization

Purified CobE was concentrated to 12 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.5 containing 0.2 *M* NaCl and 10 m*M* DTT. Sitting-drop vapour diffusion was used to screen the crystallization conditions. Initial crystallization screening was conducted using Crystal Screen 1, Crystal Screen 2, PEG/Ion Screen (Hampton Research) and Clear Strategy Screen (Brzozowski & Walton, 2001), but no crystals were obtained, only precipitate or clear solutions. Subsequently, hanging-drop vapour diffusion was successfully applied to explore further those conditions that had led to the most promising types of precipitate, as well as a range of different pH and types and concentration of precipitant. Drops containing 1 µl protein solution and 1 µl reservoir solution were equilibrated against 500 µl reservoir solution at 291 K.

2.5. Data collection and processing

Prior to data collection, crystals were mounted in a nylon CryoLoop (Hampton Research), immersed in cryoprotectant mother liquor supplemented with 27% glycerol and flash-frozen. Flash-frozen crystals were placed in a 100 K nitrogen-gas stream (Oxford Cryosystems Cryostream cooler) and native and SeMet-derivative data sets were collected at ESRF beamline ID14-1 at 100 K using an ADSC Quantum 4 detector.

The intensity data were indexed, integrated and scaled using the *HKL* programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). The data-collection statistics are summarized in Table 1.

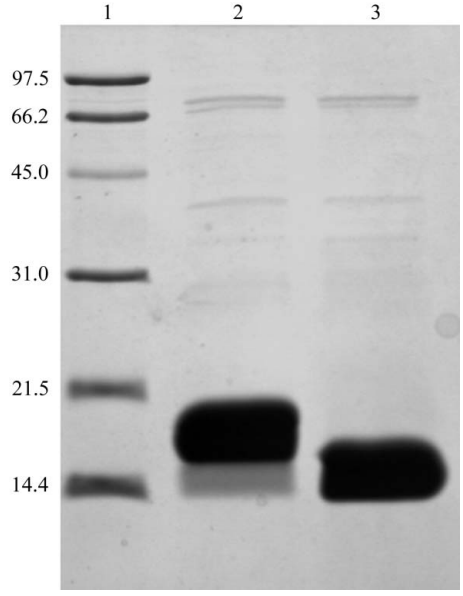


Figure 1
15% SDS–PAGE gel showing purified *P. aeruginosa* CobE (lane 2) and the same protein after cleavage of the His_6 tag (lane 3). The low-molecular-weight markers are shown (lane 1; values in kDa). There are some minor higher molecular-weight contaminants visible on this highly loaded gel, but the purity of CobE is at least 99% and the protein sample was successfully used for crystallization.

Table 1
Statistics of X-ray data collection.

Values in parentheses indicate statistics for the last resolution shell.

Data set	Native	Se, single wavelength
Wavelength (Å)	0.9340	0.9340
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 31.86, b = 41.07, c = 87.41, $\alpha = \beta = \gamma$ = 90	a = 31.73, b = 41.27, c = 88.00, $\alpha = \beta = \gamma$ = 90
Total No. of reflections	164671	888211
No. of unique reflections	9752	13347
Resolution (Å)	50–1.9 (1.93–1.9)	45–1.7 (1.76–1.7)
$R_{\rm merge}$	0.075 (0.341)	0.051 (0.486)
Completeness (%)	98.3 (73.0)	97.3 (96.0)
Redundancy	7.0 (3.4)	12.5 (11.8)
Rotation (°)	0.5	0.5
$\langle I/\sigma(I)\rangle$	26.8 (6.4)	30.4 (7.1)
Mosaicity	0.53	1.2
No. of images	380	720
Detector distance (mm)	180	135
$V_{\rm M}$† (Å^3 Da^{-1})	1.9	
Molecules per AU	1	

† Matthews (1968).

3. Results and discussion

The cloning of *cobE* into a suitable expression plasmid has facilitated the large-scale production of recombinant CobE protein, which was subsequently purified by Ni-chelate chromatography and gel filtration. The purity of the protein was validated by SDS–PAGE (Fig. 1) and MALDI mass spectrometry and the protein was judged to be 99% pure. After cleavage of the His tag using thrombin, the protein was successfully crystallized. Rod-like crystals of dimensions suitable for X-ray diffraction experiment were obtained from 0.1 *M* MES buffer pH 6.9, 2 *M* ammonium sulfate and 5% dioxane (Fig. 2). Clusters of rods of average dimensions 0.4 × 0.2 × 0.1 mm appeared in the drops after 5–10 d using 12 mg ml^{-1} protein solution in 20 m*M* Tris–HCl buffer pH 8.5 containing 0.2 *M* NaCl and 10 m*M* DTT. The best quality crystals diffracted X-rays to a resolution of 1.9 Å (Fig. 3). Data-collection statistics are shown in Table 1.

Since CobE shows no sequence homology to any other protein other than its orthologue CbiG in the PDB, it is planned to obtain experimental phases using the anomalous diffraction measurements or by multiple isomorphous replacement using either the L-SeMet derivative of CobE or heavy-atom soaks. Initial heavy-atom soaks have been unsuccessful, as the CobE crystals lost their diffraction power upon soaking or showed an unacceptable reduction in diffraction quality. SeMet-labelled CobE has been prepared and the expected presence of a single SeMet (corresponding to the N-terminal residue of native CobE) was confirmed by electrospray mass spectrometry.

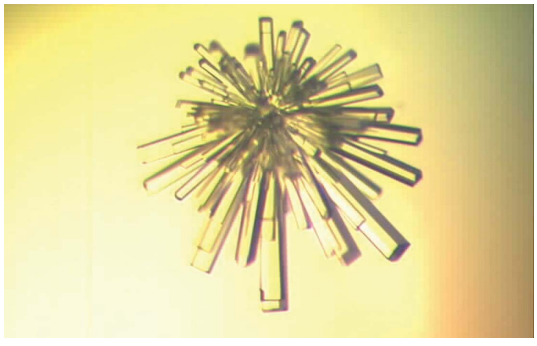


Figure 2
Crystals of CobE grown from 0.1 *M* MES buffer pH 6.9, 2 *M* ammonium sulfate and 5% dioxane. The rods have average dimensions of 0.4 × 0.2 × 0.1 mm.

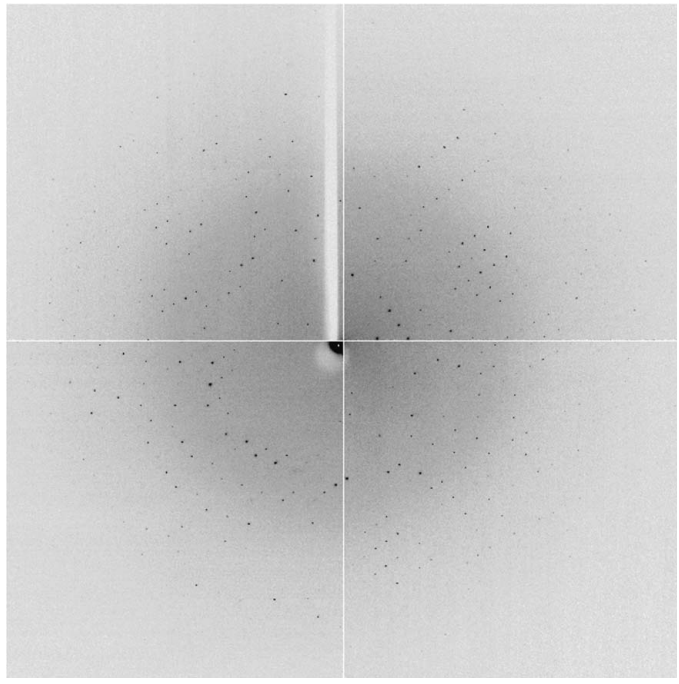


Figure 3
A representative 0.5° rotation image of data collected from a crystal of CobE using an ADSC Quantum 4 detector on beamline ID14-1 at the ESRF synchrotron in Grenoble. The edge of the image corresponds to a resolution of 1.9 Å.

The SeMet-containing protein crystallized under the same conditions as the native protein (0.1 *M* MES buffer pH 6.9, 2 *M* ammonium sulfate and 5% dioxane) and the crystals show similar diffraction quality, belong to the same space group $P2_12_12_1$ with essentially identical unit-cell parameters and diffract to a resolution of 1.7 Å. A single-wavelength data set has been collected on beamline ID14-1 at the ESRF, France. The data are 97.3% complete and $R_{\rm merge}$ is 0.051. The data-collection statistics are summarized in Table 1. Peaks corresponding to a single selenium site have been identified in the anomalous difference Patterson map, but the phases calculated from this do not appear to be of sufficient quality to build the structure. A three-wavelength MAD data set is now planned. Structure solution using the anomalous SeMet data is expected to be successful.

We acknowledge the support of the Wellcome Trust through grants 060822/Z/00/Z to KSW and MW. We also thank the York Biology Technology Facility for access to the Electrospray Mass Spectrometer and the ESRF for synchrotron-radiation facilities.

References

Brzozowski, A. M. & Walton, J. (2001). *J. Appl. Cryst.* **34**, 97–101.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. (1993). *Nature (London)*, **362**, 219–223.
Roessner, C. A., Santander, P. J. & Scott, A. L. (2001). *Vitam. Horm.* **61**, 267–297.
Warren, M. J., Raux, E., Schubert, H. L. & Esculante-Semerena, J. C. (2002). *Nat. Prod. Rep.* **19**, 390–412.
Warren, M. J. & Scott, A. I. (1990). *Trends Biochem. Sci.* **15**, 486–491.