Published in final edited form as:

Acta Crystallogr D Biol Crystallogr. 2011 January ; 67(Pt 1): 8–13. doi:10.1107/S0907444910042022.

Extending the usability of the phasing power of diselenide bonds: SeCys SAD phasing of CsgC using a non-auxotrophic strain

Paula S. Salgado, Jonathan D. Taylor, Ernesto Cota, and Steve J. Matthews*

Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

Abstract

The CsgC protein is a component of the curli system in *Escherichia coli*. Reported here is the successful incorporation of selenocysteine (SeCys) and selenomethionine (SeMet) into recombinant CsgC, yielding derivatised crystals suitable for structural determination. Unlike in previous reports, a standard, autotrophic expression strain was used and only single anomalous diffraction (SAD) data were required for successful phasing. The level of SeCys/SeMet incorporation was estimated by mass spectrometry to be about 80%. Native protein crystallised in two different crystal forms (C222₁, form 1 and C2₁, form 2) both diffracting to 2.4 Å, whilst Se-derivatised protein crystallised in C2₁ and diffracted to 1.7 Å. The Se-derivatised crystals are suitable for SAD structure determination using only anomalous signal derived from the selenocysteine residues. These results extend the usability of SeCys labelling to more general and less favourable cases, rendering it a suitable alternative to traditional phasing approaches.

Keywords

selenocysteine; SAD; non-auxotrophic *E. coli* strains; CsgC protein

1. Introduction

1.1. CsgC

A wide variety of bacteria produce extracellular amyloid fibres that promote biofilm formation and mediate adherence to biological and synthetic surfaces (Alteri *et al.*, 2007, Barnhart & Chapman, 2006, Larsen *et al.*, 2007). Fibre assembly occurs in a highly controlled manner involving several accessory proteins (Chapman *et al.*, 2002, Epstein *et al.*, 2009, Nenninger *et al.*, 2009). One such protein is CsgC, whose precise function is currently mysterious. Loss of CsgC results in aberrant fibre formation due to apparent structural changes in the fibre subunit protein CsgA (Gibson *et al.*, 2007). Currently, it is not known if CsgC directly chaperones CsgA or whether it interacts with other parts of the secretion machinery to ensure proper fibre biosynthesis. Intriguingly, CsgC is expressed at very low levels and is completely absent from many bacterial species that produce amyloid fibres (Gibson *et al.*, 2007). Thus, the importance and mechanism of action of CsgC remain unexplained. Here we report the expression, purification and crystallisation of recombinant CsgC. Apart from its N-terminal methionine, CsgC (102 residues) contains one methionine, predicted to be solvent-exposed and hence expected to be of little use in selenomethionine (SeMet) MAD phasing approaches. CsgC also contains two conserved cysteines capable of

*s.j.matthews@imperial.ac.uk.

intramolecular disulphide bonding. We describe below the use of a selenocysteine (SeCys) labelled form that paved the way to solving the structure of this intriguing curli biosynthesis component.

1.2. Phasing with selenium-containing amino acid residues

SeMet incorporation has become the method of choice for obtaining phases in the crystallographic analysis of proteins of unknown structure via anomalous dispersion analysis (Hendrickson, 1991). Use of SeCys sites was in theory expected to be as effective; however, the difficulty of labelling target proteins limited its use to proteins that have native SeCys side-chains (Hendrickson, 1991). Since then, to our knowledge, there have only been two successful structure determinations using artificially introduced SeCys residues as anomalous signal sources for MAD (Sanchez *et al.*, 2002) or MIRAS (Thepaut *et al.*, 2009) phasing. In these studies, the protein was expressed in a BL21 (DE3) *cys* auxotroph, cultured in a defined minimal medium supplemented with cysteine or SeCys, as appropriate. The level of Se-incorporation in the purified recombinant protein was in excess of 90% (Sanchez *et al.*, 2002). The same authors subsequently demonstrated that double-labelling proteins with SeCys and SeMet effectively improves phasing power such that SAD data alone can be sufficient for phasing in optimal cases (Strub *et al.*, 2003).

In this study we followed a similar approach. Importantly, we were able to achieve sufficient SeCys/SeMet incorporation using the widely-used parent, autotrophic BL21 (DE3) strain. Despite the use of a protocol that allows incorporation of both SeCys and SeMet, for CsgC, only the SeCys anomalous centres contributed to the overall phasing power. Although CsgC and the cathelicidin motif solved by Sanchez *et al.* are both ~100 residues, CsgC has 50% fewer cysteines (n=2) available for labelling. Notably, we were able to calculate phases for CsgC using SAD data alone. Thus, our results show that labelling cysteine side-chains in proteins is an attractive alternative or supplementary method to obtain phase information. It is important to note that our data extends to 1.7 Å, whilst previously MAD data was only collected to 2.4 Å, which might be related to the successful SAD approach in this case. Furthermore, this work extends the method to proteins that cannot be expressed in cysteine or methionine auxotrophic *E. coli* strains, therefore enabling wider and more general use.

2. Materials and Methods

2.1. Protein expression, purification and crystallization

The *csgC* gene was amplified from *E. coli* genomic DNA by PCR and ligated into pET28a using the *NcoI* and *XhoI* sites. The 8-residue signal peptide was omitted from the recombinant protein and the C-terminus was fused to a hexahistidine tag (LEHHHHHH). Unlabelled CsgC was expressed in BL21 (DE3) cells grown in LB media supplemented with 30 µg/ml kanamycin at 37 °C with shaking. Expression was induced with 1 mM IPTG and the cells were harvested after four hours by centrifugation at 4000 g for 15 minutes.

Labelled CsgC for crystallography (SeCys/SeMet) was expressed in a standard minimal medium containing 6.4 g Na₂HPO₄, 1.5 g KH₂PO₄, 0.25 g NaCl, 0.5 g NH₄Cl, 2 mM MgCl₂, 10 µM FeSO₄, 10 µM CaCl₂, 2 g glucose, 1 ml vitamins and 1 ml micronutrients per litre. Starting from a single-colony picked from an LB-Agar plate, a starter culture in LB was grown overnight. A 1 ml aliquot was then centrifuged at 4000 g for 5 minutes and the cell pellet resuspended in 1 ml minimal medium, before adding to 1 L of minimal medium. Cells were grown to OD₆₀₀ > 0.6 at which point 1.6 g of serine, 1 g of leucine, 0.4 g of alanine, glutamate, glutamine, arginine and glycine, 0.25 g of aspartate, 0.1 g of lysine, threonine, phenylalanine, asparagine, histidine, proline, tyrosine and tryptophan, 50 mg of isoleucine and valine, 0.1 g of Se-methionine and 0.1 g of Se-cysteine were added. The

supplemented medium was incubated at 37 °C for 15 minutes before being cooled to 18 °C. Expression of labelled CsgC was induced by addition of 0.5 mM IPTG and the cells harvested as described above after 16 hours.

Harvested cells were resuspended and incubated for 20 minutes in lysis buffer (50 mM sodium phosphate pH 7.8, 300 mM sodium chloride, 10 mM imidazole, 50 µg/ml lysozyme, 1.25 kU Benzonase Nuclease (Novagen), 1 mM MgCl₂, Complete-mini protease inhibitor (Roche)). Lysis was performed by a cell disrupter (Constant Systems) and insoluble material removed by centrifugation at 18,000 rpm for 25 minutes. The supernatant was incubated with Ni²⁺-NTA resin (Generon) for 20 minutes and then washed in 20 column volumes (CV) of Wash buffer (50 mM sodium phosphate pH 7.8, 300 mM sodium chloride, 30 mM imidazole). The bound CsgC protein was eluted after incubation for 10 minutes in 5 CV of Elution buffer (Wash buffer with 300 mM imidazole). The fractions containing CsgC were concentrated to less than 1 ml using a Vivaspin device with a 3 kDa cut-off and loaded onto a Superdex 75 16/60 gel filtration column (GE Healthcare) equilibrated in 10 mM Tris-HCl pH 7.5, 150 mM NaCl. The purified protein was >99% pure as judged by SDS-PAGE. In general we regularly obtain ~10 mg unlabelled CsgC from one litre of LB medium. Yields of SeCys/SeMet-CsgC were lower, around 5 mg/L.

Mass spectrometry of the SeCys/SeMet derivative was carried out on a Q-TOF mass spectrometer (Fig. 1a) at the Mass Spectrometry Facility, Astbury Centre for Structural Molecular Biology, University of Leeds, UK. Data analysis was performed by Dr James Ault.

Native or labelled CsgC protein was concentrated to approximately 10 mg/ml and crystallised overnight at room temperature in a variety of different conditions within commercial screens, set up using a Mosquito robot (100nl:100nl protein:reservoir) using the sitting drop vapour diffusion method. The conditions yielding optimal diffraction pattern for SeCys/SeMet-CsgC (crystal form 1) contained 25% [w/v] PEG 4000, 0.2 M ammonium sulphate, 0.1 M sodium acetate pH 4.6, and 18% [v/v] MPD, which also acted as a cryoprotectant (Fig. 2a). For crystal form 2 optimal data was obtained with 20% [w/v] PEG 3350, 0.2 M ammonium nitrate (Fig. 2b). The best SeCys/SeMet crystals were grown in 25% [w/v] PEG 4000, 30% [v/v] ethylene glycol (Fig. 3a). Paraffin oil was used as a cryoprotectant for collection of native crystal form 2 and derivatised diffraction data.

2.2. Data collection, processing and phasing

For native crystals belonging to form 1, a data set to 2.4 Å resolution was collected at 100K, in-house, wavelength 1.542 Å, using a Saturn detector with 1° oscillation per image and processed using D*trek (Pflugrath, 1999) (see Table 1 for details). Crystal form 2 diffraction data to 2.4 Å and SeCys/SeMet diffraction data to 1.7 Å resolution (Fig. 3b), were collected at 100K at PROXIMA1, SOLEIL, Paris, using an ADSC 315r detector with 1° oscillation per image and processed using XDS (Kabsch, 2010), as detailed in Table 1. For native data collection at PROXIMA1, a wavelength of 1.127 Å was used, whilst Se derivative data were collected at the Se *K* edge, at 0.979 Å, determined after performing a fluorescence scan around the Se edge (Fig xx).

Data collected from SeMet/SeCys-labelled CsgC were processed and scaled with XDS (Kabsch, 2010) and used to determine the selenium substructure with the SHELX suite of programs (Sheldrick, 2008). Initial phases were then calculated with Phaser (McCoy *et al.*, 2007) and modified with PARROT (Zhang *et al.*, 1997), within the CCP4 program suite (CCP4, 1994), in order to obtain interpretable electron density maps. Iterative cycles of model building and refinement were carried out using COOT (Emsley & Cowtan, 2004) and REFMAC5 (Vagin *et al.*, 2004) respectively.

3. Results

3.1. Incorporation of Se containing amino acids in non-auxotrophic *E. coli* strains

Mass spectrometry analysis of the purified derivatised protein indicates a mixture of species, reflecting intermediate levels of SeCys/SeMet incorporation and incomplete processing of the N-terminal methionine (Fig. 1). The mass measurements indicate that all species in the sample contain an intramolecular diselenide bridge.

Comparative analysis of peak heights within the mass spectrum indicates a high incorporation level of selenium. Indeed, over 80% of the sample contained both cysteine residues fully substituted, as well as the non terminal methionine (Fig. 1). Moreover, we estimate only a small fraction (less than 5%) of the protein in the sample remains completely unsubstituted (Fig. 1). Similar levels of SeCys incorporation (75-80%) have been observed when using an auxotrophic strain (Muller *et al.*, 1994).

These high incorporation levels are consistent with X-ray fluorescence analysis of a SeMet/SeCys-labelled crystal performed at the PROXIMA1 beam line, around the Se *K* absorption edge which indicates a strong peak ($f'' = 6.2 e^-$) at the absorption edge (Fig. 2).

3.2. Substructure Determination by SAD methods

Data processing estimated the overall anomalous signal-to-noise ratio to be 1.6, and consistently above 1 to approximately 2.0 Å. This indicates that sufficient anomalous data had been recorded in a SAD experiment to allow the determination of the selenium substructure present in the crystals. SHELXC was used to prepare the data for selenium substructure determination with SHELXD, which identified two possible, strong positions and two extra weaker positions. Upon visual inspection, the two strong positions were shown to be close together, indicating a diselenide bond had been formed (Fig. 4). Moreover, it was clear that the two weaker positions were symmetry-related sites, and subsequent steps were carried out using a substructure with only the first two Se sites. These were iteratively refined and used for initial phase calculations within Phaser (McCoy *et al.*, 2007), followed by density modification protocols carried out with PARROT (Zhang *et al.*, 1997), to yield readily interpretable electron density maps (Fig. 3a).

Analysis of the anomalous maps indicates that only two contiguous strong peaks (above 4.5σ) are readily observed (Fig. 3b). From the peak intensity height and the selenium atom substructure, it corresponds to a diselenide bridge. Notably, no further significant peaks are observable and the selenium substructure does not seem to contain a site that could correspond to the SeMet side-chain. Clearly, this residue makes little or no contribution to the overall phasing power. The final phasing statistics are summarised in Table 1.

The electron density maps obtained are currently under analysis and model building and refinement is under way. The model of SeCys/SeMet-labelled CsgC should subsequently allow determination of the structures of the native crystal forms of CsgC by molecular replacement.

4. Conclusions

Initial attempts to solve the structure of CsgC using heavy atom derivatives yielded no solutions. Several mercuric compounds were tested but no anomalous signal was detected (data not shown). Presumably, this is due to the fact that the cysteine residues are not present as free sulfhydryl groups, but instead can form a disulfide bond and are therefore not accessible for derivatisation under the conditions used. Platinum and gold compounds were also screened, with similar lack of success (data not shown).

Since traditional SeMet-based phasing methods are only generally applicable to proteins with at least one SeMet per 75-100 residues, CsgC is a borderline case (Hendrickson & Ogata, 1997, Hendrickson, 1999). Moreover, its only methionine present was predicted to be in a solvent exposed loop and therefore likely to be too disordered to provide good anomalous signal. Analysis of the refined SeCys/SeMet-labelled CsgC structure confirmed this too be true (Fig. 4a). Our results confirm that SeMet labelling alone would not have been sufficient for phasing. Therefore, we sought to improve the phasing power by also substituting the conserved cysteines C29/C31 with selenocysteine.

Previous reports detailing SeCys labelling and successful phasing specified the use of an auxotrophic strain and defined minimal medium to ensure adequate incorporation (Sanchez *et al.*, 2002, Strub *et al.*, 2003). Our study found that such strains are perhaps unnecessary. This approach confers a wider usability to this phasing method, as well as savings in time and expense. Firstly, the preferred expression strain for production of unlabelled protein may be employed for SeCys and/or SeMet labelling. This point is of particular importance where expression of the target protein requires certain genetic features absent from BL21 (DE3) *cys* and other such auxotrophs, including tight expression control and co-expressed components that improve solubility and/or folding and reduce toxicity. Secondly, double-labelling with SeCys/SeMet may circumvent the need to introduce non-native methionines via site-directed mutagenesis, a common strategy where a target lacks sufficient viable methionine sites. Moreover, in cases where the target protein is devoid of sulfur-containing side-chains, one is not limited to engineering in large, bulky methionines. Instead, cysteines may be introduced for the purpose of labelling. Indeed, careful provision of SeCys sites may actually increase tertiary or quaternary protein structure stability, with concomitant improvements in crystallisability (Heinz & Matthews, 1994, Boulter *et al.*, 2003, Banatao *et al.*, 2006). Diselenides are intrinsically more stable than disulphides, as indicated by a recent measurement for oxidised glutathione with an increase in thermodynamic stability of 7 kcal/mol (Beld *et al.*, 2007). Formation of a diselenide bond is more energetically favourable compared to a disulphide due to the lower pK_a of SeCys (5.24) *vs.* Cys (8.7) and greater nucleophilicity of selenium (Huber & Criddle, 1967, Singh & Whitesides, 1991, Nelson & Creighton, 1994). Thus, judicious use of SeCys may provide better diffracting crystals and a reliable route to obtaining phases.

Another key aspect of our approach refers to the fact that phases for CsgC could be calculated using the SAD method from the anomalous signal derived from the diselenide bond. The one previous report of phase calculation using SeCys sites alone utilized MAD methods (Sanchez *et al.*, 2002). Where possible, calculation of phase information by SAD methods is preferred over the MAD approach since it reduces radiation damage to the crystal and hence minimises errors (Rice *et al.*, 2000). Our study confirms the efficacy of SeCys SAD towards phase calculation, even when non-auxotrophic strains are used for protein expression.

As larger and more complex protein structures are studied, the availability of simple labelling methods that deviate only slightly from native protein production conditions become increasingly valuable. Our study shows that phase calculation using SeCys sites alone and SAD methodology is an effective route to structure determination with or without accompanying SeMet sites. It represents an attractive alternative to the much more widely-used SeMet/MAD approach, extending previous reports to more general and less favourable situations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank staff at PROXIMA1, SOLEIL, Paris, particularly Andy Thompson and Pierre Legrand for their expert help with data collection. The authors would also like to acknowledge the Imperial College Centre for Structural Biology and J. Moore for data collection and analysis

References

- Alteri CJ, Xicohtencatl-Cortes J, Hess S, Caballero-Olin G, Giron JA, Friedman RL. *Proc Natl Acad Sci U S A*. 2007; 104:5145–5150. [PubMed: 17360408]
- Banatao DR, Cascio D, Crowley CS, Fleissner MR, Tienson HL, Yeates TO. *Proc Natl Acad Sci U S A*. 2006; 103:16230–16235. [PubMed: 17050682]
- Barnhart MM, Chapman MR. *Annu Rev Microbiol*. 2006; 60:131–147. [PubMed: 16704339]
- Beld J, Woycechowsky KJ, Hilvert D. *Biochemistry*. 2007; 46:5382–5390. [PubMed: 17419591]
- Boulter JM, Glick M, Todorov PT, Baston E, Sami M, Rizkallah P, Jakobsen BK. *Protein Eng*. 2003; 16:707–711. [PubMed: 14560057]
- CCP4. *Acta Crystallographica Section D*. 1994; 50:760–763.
- Chapman MR, Robinson LS, Pinkner JS, Roth R, Heuser J, Hammar M, Normark S, Hultgren SJ. *Science*. 2002; 295:851–855. [PubMed: 11823641]
- Emsley P, Cowtan K. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132. [PubMed: 15572765]
- Epstein EA, Reizian MA, Chapman MR. *J Bacteriol*. 2009; 191:608–615. [PubMed: 19011034]
- Evans G, Pettifer RF. *J. Appl. Cryst*. 2001; 34:82–86.
- Gibson DL, White AP, Rajotte CM, Kay WW. *Microbiol*. 2007; 153:1131–1140.
- Heinz DW, Matthews BW. *Protein Eng*. 1994; 7:301–307. [PubMed: 8177878]
- Hendrickson WA. *Science*. 1991; 254:51–58. [PubMed: 1925561]
- Hendrickson WA. *Journal of Synchrotron Radiation*. 1999; 6:845–851.
- Hendrickson WA, Ogata CM. *Methods in Enzymology*. 1997; 276:494–523.
- Huber RE, Criddle RS. *Arch Biochem Biophys*. 1967; 122:164–173. [PubMed: 6076213]
- Kabsch W. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:125–132. [PubMed: 20124692]
- Larsen P, Nielsen JL, Dueholm MS, Wetzel R, Otzen D, Nielsen PH. *Environ Microbiol*. 2007; 9:3077–3090. [PubMed: 17991035]
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *J Appl Crystallogr*. 2007; 40:658–674. [PubMed: 19461840]
- Muller S, Senn H, Gsell B, Vetter W, Baron C, Bock A. *Biochemistry*. 1994; 33:3404–3412. [PubMed: 8136378]
- Nelson JW, Creighton TE. *Biochemistry*. 1994; 33:5974–5983. [PubMed: 8180227]
- Nenninger AA, Robinson LS, Hultgren SJ. *Proc Natl Acad Sci U S A*. 2009; 106:900–905. [PubMed: 19131513]
- Pflugrath JW. *Acta Crystallogr D Biol Crystallogr*. 1999; 55:1718–1725. [PubMed: 10531521]
- Rice LM, Earnest TN, Brunger AT. *Acta Crystallogr D Biol Crystallogr*. 2000; 56:1413–1420. [PubMed: 11053839]
- Sanchez JF, Hoh F, Strub MP, Aumelas A, Dumas C. *Structure*. 2002; 10:1363–1370. [PubMed: 12377122]
- Sheldrick GM. *Acta Crystallogr A*. 2008; 64:112–122. [PubMed: 18156677]
- Singh R, Whitesides GM. *Journal of Organic Chemistry*. 1991; 56:6931–6933.
- Strub MP, Hoh F, Sanchez JF, Strub JM, Bock A, Aumelas A, Dumas C. *Structure*. 2003; 11:1359–1367. [PubMed: 14604526]
- Thepaut M, Valladeau J, Nurisso A, Kahn R, Arnou B, Vives C, Saeland S, Ebel C, Monnier C, Dezutter-Dambuyant C, Imberty A, Fieschi F. *Biochemistry*. 2009; 48:2684–2698. [PubMed: 19175323]
- Vagin AA, Steiner RS, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. *Acta Cryst*. 2004; D60:2284–2295.

Zhang KY, Cowtan K, Main P. *Methods Enzymol.* 1997; 277:53–64. [PubMed: 18488305]

Synopsis

We report the successful incorporation of SeCys and SeMet into a recombinant protein, yielding derivatised crystals suitable for phasing. This is the first report of such double derivatisation using non-auxotrophic *E. coli* strains and phasing was carried out by SeCys SAD.

Figure 1.

Mass spectrum of SeCys/SeMet-CsgC. Peaks labelled with green text correspond to full length CsgC whereas those with red text correspond to a truncated form lacking the N-terminal methionine. The level of selenium labelling is indicated for each peak: for example 3/4 corresponds to three sulphurs out of a possible four have been substituted by selenium. The presence of sodium (Na⁺) and nickel (Ni²⁺) adducts within the sample complicates the spectrum.

(a)

Figure 2.

Se K edge fluorescence scan of selenocysteine/selenomethionine CsgC crystal, performed on PROXIMAL. The plot shown is the result of CHOOCH analysis (Evans & Pettifer, 2001) of the raw data.

(a)

(b)

Figure 3.
Anomalous signal variation in function of resolution, as calculated from SHELX.

Figure 4. Close-up views of CsgC model details: (a) diselenide bond and (b) SeMet site. Electron density map (blue chicken-wire) is contoured at 1.5σ . Figures were created from COOT (Emsley & Cowtan, 2004).

Table 1

Data collection and phasing statistics

	SeCys/SeMet	Crystal form 1	Crystal form 2
X-ray source	PROXIMA1, SOLEIL	in-house	PROXIMA1, SOLEIL
Wavelength (Å)	0.979	1.542	1.127
Space Group	C2 ₁	C222 ₁	C2 ₁
Unit Cell (a,b,c [Å]; β[°])	79.4, 36.2, 39.6; 117.1	37.0, 80.5, 70.6; 90.0	79.490, 36.430, 39.520; 117.45
Resolution range (Å)	32.20 - 1.70 (1.80 - 1.70)	20.0 - 2.40 (2.45 - 2.40)	33.80 - 2.39 (2.45 - 2.39)
N. of images	285		360
Observations total/unique	64039/21425	18583/4326	28097/7646
Redundancy	3.0	4.3	3.7
Completeness (%)	98.8 (98.3)	100.0 (100.0)	97.0 (75.4)
I/σI	16.8 (2.6)	23.2 (6.2)	18.4 (7.8)
R _{merge} * (%)	3.5 (41.2)	5.3 (15.8)	5.3 (15.8)
R _{meas} (%)	4.3 (50.1)	-	6.0 (18.6)
Molec. in A.U.	1	1	1
Solvent content (%)	41.0	41.0	41.0
Substructure solution - SHELX			
Anomalous signal (d''/sig)	8.1 - 1.1 1.7 at 2.5-2.3Å		
(low to high resol.)	(resl. cut off for phasing)		
CC (anom)	98.1		
No sites	4	-	-
CC all /weak	46.1 / 31.0	-	-
PATFOM	76.1	-	-
Phasing - PHASER			
FOM	0.43		
DM - PARROT			
FOM	0.75		-

* $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th observation of reflection hkl and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl .

Table 2

Refinement statistics for SeCys/SeMet CsgC model

	SeCys/SeMet
Resolution range (Å)	26.60 - 1.70 (1.74 - 1.70)
R^+	0.239
R_{free}^{++}	0.285
N. atoms	
total	777
protein	729
water	32
other	11
Average B	
overall	28.85
Protein (main/side)	28.21 (26.48/30.94)
water	32.53
other	45.74
R.m.s.d.	
Bond length	0.021
Bond angles	2.128
Ramachandran plot statistics (% residues)	
Most favoured	94.0
Additionally allowed	6.0
Generously allowed	0.0
Disallowed	0.0
PDB accession code	2XSK

$$^+R = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}$$

$^{++}R_{\text{free}}$ was calculated from a randomly selected 5% of unique reflections that were omitted from structure refinement.