Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

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, England

‡ These authors contributed equally.

Correspondence e-mail: s.j.matthews@imperial.ac.uk

Received 11 August 2010
Accepted 16 October 2010

PDB Reference: CsgC, 2xsk.

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 derivatized crystals suitable for structural determination. Unlike in previous reports, a standard autotrophic expression strain was used and only single-wavelength anomalous dispersion (SAD) data were required for successful phasing. The level of SeCys/SeMet incorporation was estimated by mass spectrometry to be about 80%. The native protein crystallized in two different crystal forms (form 1 belonging to space group $C222_1$ and form 2 belonging to space group $C2$), which diffracted to 2.4 and 2.0 Å resolution, respectively, whilst Se-derivatized protein crystallized in space group $C2$ and diffracted to 1.7 Å resolution. The Se-derivatized crystals are suitable for SAD structure determination using only the anomalous signal derived from the SeCys 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.

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, the precise function of which is currently mysterious. Loss of CsgC results in aberrant fibre formation arising from apparent structural changes in the fibre subunit protein CsgA (Gibson *et al.*, 2007). Currently, it is not known whether 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.

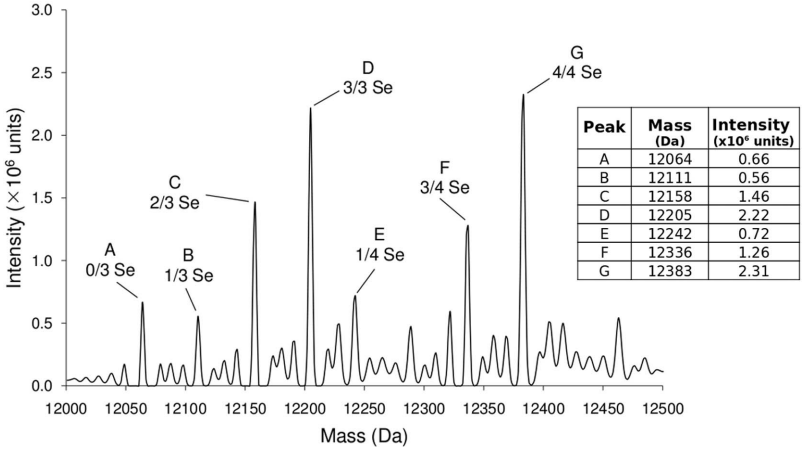
Here, we report the expression, purification and crystallization of recombinant CsgC. Apart from its N-terminal methionine, CsgC (102 residues) contains one methionine, which is predicted to be solvent-exposed and is hence expected to be of little use in selenomethionine (SeMet) multi-wavelength anomalous dispersion (MAD) phasing approaches. CsgC also contains two conserved cysteines that are capable of intramolecular disulfide 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 has 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 (Thépaut *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 the 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 and coworkers both consist of ∼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 of obtaining phase information. It is important to note that our data extend to 1.7 Å resolution, whereas previously MAD data were only collected to 2.4 Å resolution (Sanchez *et al.*, 2002), 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 *Escherichia coli* strains, therefore enabling wider and more general use.



Peak	Mass (Da)	Intensity (x10⁶ units)
A	12064	0.66
B	12111	0.56
C	12158	1.46
D	12205	2.22
E	12242	0.72
F	12336	1.26
G	12383	2.31

Figure 1
Mass spectrum of SeCys/SeMet CsgC. Main peaks are labelled and the level of selenium labelling is indicated for each peak: for example 3/4 corresponds to three sulfurs out of a possible four have been substituted by selenium. The presence of sodium and nickel adducts (not labelled) within the sample complicates the spectrum, but evaluation of Se incorporation was possible. The table inset shows intensities and the respective calculated masses (Da) for the main peaks (non-adduct species) from which estimated incorporation was calculated.

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 *Nco*I and *Xho*I sites. The eight-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 medium supplemented with 30 µg ml^{-1} kanamycin at 310 K with shaking. Expression was induced with 1 m*M* IPTG and the cells were harvested after 4 h by centrifugation at 4000*g* for 15 min.

Labelled CsgC for crystallography (SeCys/SeMet) was expressed in a standard minimal medium consisting of 6.4 g Na_2HPO_4, 1.5 g KH_2PO_4, 0.25 g NaCl, 0.5 g NH_4Cl, 2 m*M* $MgCl_2$, 10 µ*M* $FeSO_4$, 10 µ*M* $CaCl_2$, 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 was grown overnight in LB. A 1 ml aliquot was then centrifuged at 4000*g* for 5 min and the cell pellet was resuspended in 1 ml minimal medium before adding to 1 l minimal medium. Cells were grown to an OD_{600} of >0.6, at which point 1.6 g serine, 1 g leucine, 0.4 g alanine, glutamate, glutamine, arginine and glycine, 0.25 g aspartate, 0.1 g lysine, threonine, phenylalanine, asparagine, histidine, proline, tyrosine and tryptophan, 50 mg isoleucine and valine, 0.1 g selenomethionine and 0.1 g selenocysteine were added. The supplemented medium was incubated at 310 K for 15 min before being cooled to 291 K. Expression of labelled CsgC was induced by the addition of 0.5 m*M* IPTG and the cells were harvested as described above after 16 h.

Harvested cells were resuspended and incubated for 20 min in lysis buffer [50 m*M* sodium phosphate pH 7.8, 300 m*M* sodium chloride, 10 m*M* imidazole, 50 µg ml^{-1} lysozyme, 1.25 kU Benzonase Nuclease (Novagen), 1 m*M* $MgCl_2$ and Complete Mini protease inhibitor (Roche)]. Lysis was performed by a cell disrupter (Constant Systems) and insoluble material was removed by centrifugation at 18 000 rev min^{-1} for 25 min. The supernatant was incubated with Ni^{2+}–NTA resin (Generon) for 20 min and then washed in 20 column volumes (CV) of wash

Table 1
Data-collection and phasing statistics.

	SeCys/SeMet	Crystal form 1	Crystal form 2
X-ray source	PROXIMA 1, SOLEIL	In-house	PROXIMA 1, SOLEIL
Wavelength (Å)	0.9792	1.5418	1.1271
Space group	*C*2	$C222_1$	*C*2
Unit-cell parameters (Å, °)	*a* = 79.4, *b* = 36.2, *c* = 39.6, β = 117.1	*a* = 37.0, *b* = 80.5, *c* = 70.6, β = 90.0	*a* = 79.49, *b* = 36.43, *c* = 39.52, β = 117.45
Resolution range (Å)	32.20–1.70 (1.80–1.70)	20.0–2.40 (2.45–2.40)	33.80–1.99 (2.04–1.99)
No. of images	285 [1° oscillations]	226 [1° oscillations]	360 [1° oscillations]
Observations (total/unique)	64039/21425	18583/4326	47415/13038
Multiplicity	3.0	4.3	3.7
Completeness (%)	98.8 (98.3)	100.0 (100.0)	95.9 (84.0)
$I/\sigma(I)$	16.8 (2.6)	23.2 (6.2)	14.4 (4.6)
$R_{\rm merge}$† (%)	3.5 (41.2)	5.3 (15.8)	5.8 (20.5)
$R_{\rm meas}$ (%)	4.3 (50.1)	—	6.9 (24.1)
Molecules in the asymmetric unit	1	1	1
Solvent content (%)	41.0	41.0	41.0
Substructure solution (*SHELX*)			
Anomalous signal	8.1–1.1		
d″/sig	1.7 at 2.5–2.3 Å		
Low to high resolution	Resolution cutoff for phasing		
CC (anom)	98.1		
No. of sites	4	—	—
CC (all/weak)	46.1/31.0	—	—
PATFOM	76.1	—	—
Phasing: *Phaser*			
FOM	0.43	—	—
Density modification: *PARROT*			
FOM	0.75	—	—

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, 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*.

buffer (50 m*M* sodium phosphate pH 7.8, 300 m*M* sodium chloride, 30 m*M* imidazole). The bound CsgC protein was eluted after incubation for 10 min in 5 CV elution buffer (wash buffer with 300 m*M* imidazole). The fractions containing CsgC were concentrated to less than 1 ml using a Vivaspin device with a 3 kDa cutoff and loaded onto a Superdex 75 16/60 gel-filtration column (GE Healthcare) equilibrated in 10 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl. The purified protein was >99% pure as judged by SDS–PAGE. In general, we regularly obtained ∼10 mg unlabelled CsgC from 1 l LB medium. Yields of SeCys/SeMet CsgC were lower at around 5 mg per litre.

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

Native or labelled CsgC protein was concentrated to approximately 10 mg ml^{-1} and crystallized overnight at room temperature in a variety of different conditions from commercial screens set up using a Mosquito robot (100 nl protein:100 nl reservoir) using the sitting-drop vapour-diffusion method. The conditions that yielding optimal diffraction patterns for SeCys/SeMet CsgC (oxidized) consisted of 25%(*w*/*v*) PEG 4000, 0.2 *M* ammonium sulfate, 0.1 *M* sodium acetate pH 4.6 and 18%(*v*/*v*) MPD, which also acted as a cryoprotectant (see Supplementary Fig. 1*a*[1]). Native CsgC crystallized in both the reduced (with free cysteines) and oxidized (with disulfide bonded cysteines) states, referred to here as native crystal forms 1 and 2, respectively. Form 1 crystals grew in 25%(*w*/*v*) PEG 4000, 30%(*v*/*v*) ethylene glycol (Supplementary Fig. 1*b*[1]), whereas form 2 crystals grew in 20%(*w*/*v*) PEG 3350, 0.2 *M* ammonium nitrate. Native crystal forms 1 and 2 were cryoprotected by briefly incubating the crystals in 20% glycerol or paraffin oil, respectively.

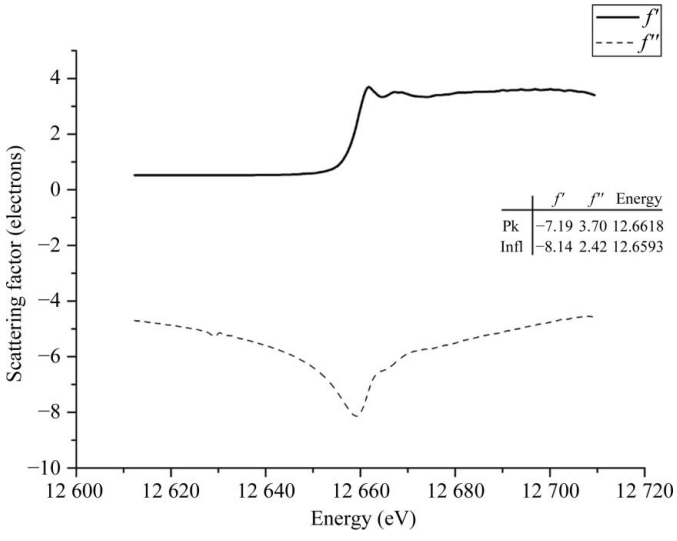


Figure 2
Se *K* edge fluorescence scan of an SeCys/SeMet CsgC crystal performed on PROXIMA 1. The plot shown is the result of analysis of the raw data with *CHOOCH* (Evans & Pettifer, 2001).

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 100 K in-house at a wavelength of 1.5418 Å using a Saturn detector with 1° oscillation per image and was processed using *d*TREK* (Pflugrath, 1999) (see Table 1 for details). Native crystal form 2 diffraction data to 2.0 Å resolution and SeCys/SeMet-CsgC diffraction data to 1.7 Å resolution (Supplementary Fig. 1*c*) were collected at 100 K at PROXIMA 1, SOLEIL, Gif-sur-Yvette using an ADSC 315r detector with 1° oscillation per image and were processed using *XDS* (Kabsch, 2010) as detailed in Table 1. For native data collection at PROXIMA 1 a wavelength of 1.1271 Å was

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DZ5216). Services for accessing this material are described at the back of the journal.

used, whilst Se-derivative data were collected at the Se *K* edge at 0.9792 Å, as determined after performing a fluorescence scan around the Se edge (Fig. 2).

The data collected from SeCys/SeMet-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 *CCP*4 program suite (Collaborative Computational Project, Number 4, 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 *REFMAC*5 (Vagin *et al.*, 2004), respectively, to determine the final SeCys/SeMet CsgC model. Refinement statistics are given in Table 2.

Table 2
Refinement statistics for the SeCys/SeMet CsgC model.

Resolution range (Å)	26.60–1.70 (1.74–1.70)
R†	0.239
R_{free}‡	0.285
No. of atoms	
Total	777
Protein	729
Water	32
Other	11
Average B factor (Å^2)	
Overall	28.9
Protein (main chain/side chain)	28.2 (26.5/30.9)
Water	32.5
Other	45.7
R.m.s.d.	
Bond lengths (Å)	0.021
Bond angles (°)	2.128
Ramachandran plot statistics (% of residues)	
Most favoured	94.0
Additionally allowed	6.0
Generously allowed	0.0
Disallowed	0.0
PDB code	2xsk

† $R = \sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$. ‡ R_{free} was calculated from a randomly selected 5% of unique reflections that were omitted from structure refinement.

3. Results

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

Mass-spectrometric analysis of the purified derivatized 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 the 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 nonterminal methionine (Fig. 1), as determined from the relative peak intensities when compared with the overall signal measured. Moreover, we estimate that 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 SeCys/SeMet-labelled crystal performed on the PROXIMA 1 beamline around the Se *K* absorption edge, which indicates a peak ($f'' = 3.7\ e^-$) 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 to be consistently above 1 to approximately 2.0 Å resolution (Fig. 3). This indicated that the anomalous data recorded in the SAD experiment might be sufficient for 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. Upon visual inspection, the two strong positions were shown to be close together, indicating that a diselenide bond had been formed (Fig. 4*a*). These peaks 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. 4).

Analysis of the resulting anomalous difference maps indicates that only two contiguous strong peaks (above 4.5σ) are readily observed. The peak intensity heights and Se-atom substructures confirm that they correspond to a diselenide bridge (Fig. 4*a*). 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 summarized in Table 1.

The electron-density maps obtained in *Phaser* were used for iterative cycles of model building and refinement, and a final model of SeCys/SeMet CsgC was obtained. This model has

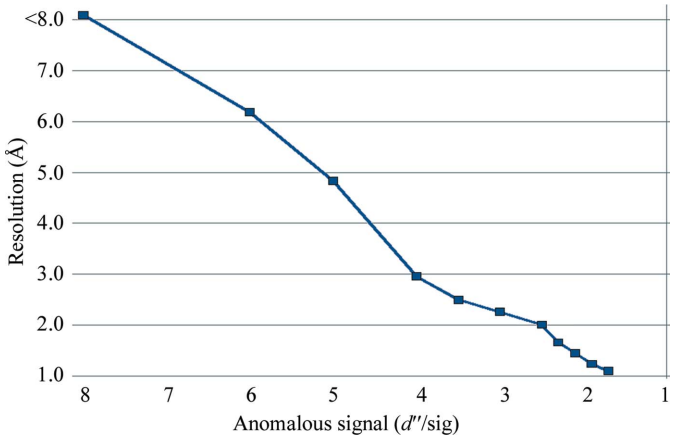


Figure 3
Anomalous signal variation as a function of resolution as calculated using *SHELX*.

subsequently allowed determination of the structures of the native crystal forms of CsgC by molecular replacement, which is under way.

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 a consequence of 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 derivatization under the conditions used. Platinum and gold compounds were also screened, with a 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, although within those limits, CsgC is a borderline case (Hendrickson & Ogata, 1997; Hendrickson, 1999). As predicted, the only methionine present is located in a solvent-exposed position within the crystal structure and thus is too disordered to provide a good anomalous signal (Fig. 4*b*). Based on these predictions, we sought to improve the phasing power by also substituting the conserved cysteines Cys29/Cys31 with SeCys.

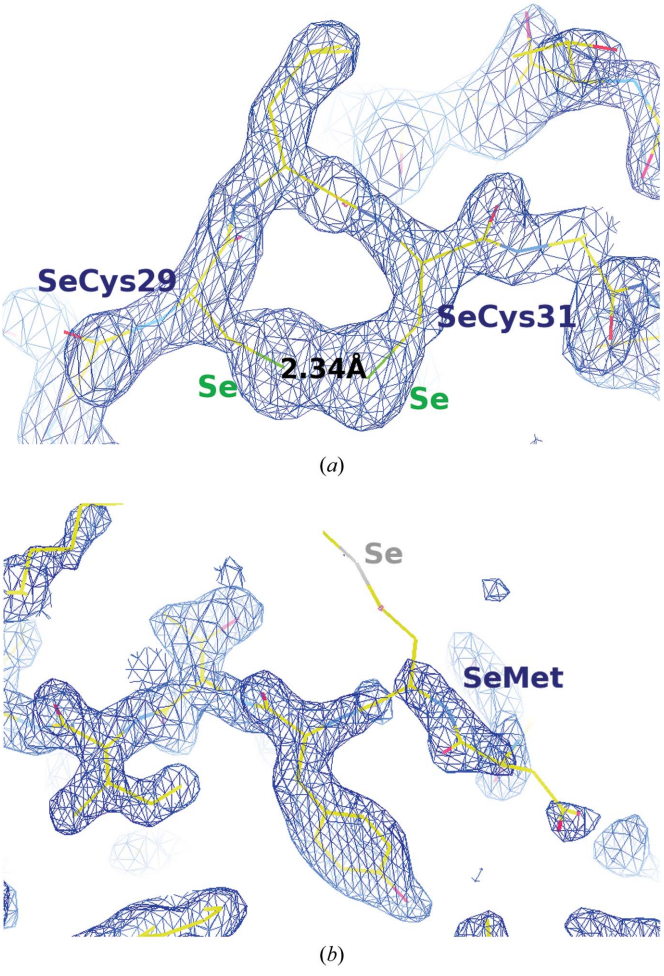


Figure 4
Close-up views of SeCys/SeMet CsgC model details: (*a*) diselenide bond and (*b*) SeMet site. Electron-density maps (blue chicken wire) are contoured at 1.5σ. Figures were created from *Coot* (Emsley & Cowtan, 2004).

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 providing an easier, quicker and less expensive approach. Firstly, the preferred expression strain for the production of unlabelled protein may be employed for SeCys and/or SeMet labelling. This point is of particular importance when expression of the target protein requires certain genetic features that are 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 when 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 crystallizability (Heinz & Matthews, 1994; Boulter *et al.*, 2003; Banatao *et al.*, 2006). Diselenides are intrinsically more stable than disulfides, as indicated by a recent measurement for oxidized glutathione with an increase in thermodynamic stability of 29 kJ mol^{-1} (Beld *et al.*, 2007). Formation of a diselenide bond is more energetically favourable compared with a disulfide owing to the lower pK_a of SeCys (5.24) *versus* Cys (8.7) and the 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, particularly if resolution and crystal quality are not a limiting factor, calculation of phase information by SAD methods is preferred over the MAD approach since it reduces radiation damage to the crystal and hence minimizes 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.

We thank the staff at the PROXIMA1 beamline at the SOLEIL Synchrotron (Gif-sur-Yvette, France), 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 Jeremy Moore for data collection and analysis. We are grateful for the expertise of Dr James Ault in the collection and analysis of mass spectrometry data.

References

Alteri, C. J., Xicohtencatl-Cortes, J., Hess, S., Caballero-Olin, G., Giron, J. A. & Friedman, R. L. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 5145–5150.
Banatao, D. R., Cascio, D., Crowley, C. S., Fleissner, M. R., Tienson, H. L. & Yeates, T. O. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 16230–16235.
Barnhart, M. M. & Chapman, M. R. (2006). *Annu. Rev. Microbiol.* **60**, 131–147.
Beld, J., Woycechowsky, K. J. & Hilvert, D. (2007). *Biochemistry*, **46**, 5382–5390.
Boulter, J. M., Glick, M., Todorov, P. T., Baston, E., Sami, M., Rizkallah, P. & Jakobsen, B. K. (2003). *Protein Eng. Des. Sel.* **16**, 707–711.
Chapman, M. R., Robinson, L. S., Pinkner, J. S., Roth, R., Heuser, J., Hammar, M., Normark, S. & Hultgren, S. J. (2002). *Science*, **295**, 851–855.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Epstein, E. A., Reizian, M. A. & Chapman, M. R. (2009). *J. Bacteriol.* **191**, 608–615.
Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
Gibson, D. L., White, A. P., Rajotte, C. M. & Kay, W. W. (2007). *Microbiology*, **153**, 1131–1140.
Heinz, D. W. & Matthews, B. W. (1994). *Protein Eng. Des. Sel.* **7**, 301–307.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Huber, R. E. & Criddle, R. S. (1967). *Arch. Biochem. Biophys.* **122**, 164–173.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Larsen, P., Nielsen, J. L., Dueholm, M. S., Wetzel, R., Otzen, D. & Nielsen, P. H. (2007). *Environ. Microbiol.* **9**, 3077–3090.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Muller, S., Senn, H., Gsell, B., Vetter, W., Baron, C. & Bock, A. (1994). *Biochemistry*, **33**, 3404–3412.
Nelson, J. W. & Creighton, T. E. (1994). *Biochemistry*, **33**, 5974–5983.
Nenninger, A. A., Robinson, L. S. & Hultgren, S. J. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 900–905.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.
Sanchez, J. F., Hoh, F., Strub, M. P., Aumelas, A. & Dumas, C. (2002). *Structure*, **10**, 1363–1370.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Singh, R. & Whitesides, G. M. (1991). *J. Org. Chem.* **56**, 6931–6933.
Strub, M. P., Hoh, F., Sanchez, J. F., Strub, J. M., Bock, A., Aumelas, A. & Dumas, C. (2003). *Structure*, **11**, 1359–1367.
Thépaut, M., Valladeau, J., Nurisso, A., Kahn, R., Arnou, B., Vivès, C., Saeland, S., Ebel, C., Monnier, C., Dezutter-Dambuyant, C., Imberty, A. & Fieschi, F. (2009). *Biochemistry*, **48**, 2684–2698.
Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Cryst.* D**60**, 2184–2195.
Zhang, K.-Y., Cowtan, K. & Main, P. (1997). *Methods Enzymol.* **277**, 53–64.