

Neil G. Paterson,^{a*} Alan Riboldi-Tunnicliffe,^a Timothy J. Mitchell^b and Neil W. Isaacs^a

^aDepartment of Chemistry and WestCHEM, Glasgow Biomedical Research Centre (GBRC), University of Glasgow, 120 University Place, Glasgow G12 8TA, Scotland, and ^bDivision of Infection and Immunity (IBLS), Glasgow Biomedical Research Centre (GBRC), University of Glasgow, 120 University Place, Glasgow G12 8TA, Scotland

Correspondence e-mail:
neison@chem.gla.ac.uk

Received 31 May 2006
Accepted 7 June 2006

Purification, crystallization and preliminary X-ray diffraction analysis of RafE, a sugar-binding lipoprotein from *Streptococcus pneumoniae*

Streptococcus pneumoniae contains a large number of sugar-transport systems and the system responsible for raffinose uptake has recently been identified. The substrate-binding protein component of this system shares strong sequence homology with the multiple sugar metabolism substrate-binding protein MsmE from *S. mutans* and contains a lipoprotein-attachment site at cysteine residue 23. A truncated form (residues 24–419) of RafE from *S. pneumoniae* was cloned and overexpressed in *Escherichia coli*. Native and selenomethionine-labelled protein have been crystallized in the hexagonal space group $P6_122$. Diffraction data have been successfully phased to 2.90 Å using Se SAD data and model building is in progress.

1. Introduction

Streptococcus pneumoniae is a major human pathogen that mainly affects the young, elderly and immunocompromized populations. Pneumococcal infection is estimated to kill over one million children under the age of five annually, with the majority of these deaths occurring in developing countries (World Health Organization, 1999).

Genomic analysis of *S. pneumoniae* TIGR4 reveals a high level of carbohydrate utilization in comparison to other species (*Haemophilus influenzae* and *Neisseria meningitidis*) that colonize the human upper respiratory tract (Tettelin *et al.*, 2001). Over 30% of the transport systems found in the pneumococcus genome are predicted to be sugar transporters, with a high proportion of these being ATP-binding cassette (ABC) transporters. Bacterial ABC transporters play an important role in organism virulence and also have immunogenic potential (Garmory & Titball, 2004).

S. pneumoniae TIGR4 open reading frame SP1897 encodes a protein, designated RafE, that is predicted to be the substrate-binding component of an ATP-binding cassette (ABC) transport system (Rosenow *et al.*, 1999). It shares 60.3% sequence identity (76.7% similarity) with MsmE from *S. mutans*, the substrate-binding domain of an ABC transport system responsible for the uptake of multiple sugars including raffinose, melibiose and isomaltotriose (Russell *et al.*, 1992). In common with substrate-binding proteins from other Gram-positive bacteria (Gilson *et al.*, 1988), MsmE contains a lipid-attachment site (conserved in RafE), which allows anchoring of the protein to the cell membrane by cleavage upstream of cysteine residue 23 and attachment of a fatty-acid lipid (Sutcliffe *et al.*, 1993). It appears that sequence differences between RafE and MsmE result in different substrate specificities, with RafE only capable of raffinose transport (Rosenow *et al.*, 1999). We present here preliminary crystallization and X-ray diffraction studies of a truncated form of RafE.

2. Materials and methods

2.1. Cloning

A polymerase chain reaction (PCR) product containing the coding region for a truncated form of RafE (residues 24–419) was cloned between the *Bam*HI and *Hind*III sites of the pQE-10 vector (Qiagen) with an in-frame N-terminal His₆ tag and linker (MRGSHHH-HHHTDP). Transformation was carried out into *Escherichia coli*



Table 1

Data-collection statistics for native and selenomethionine-labelled RafE.

Values in parentheses are for the highest resolution shell.

Data set	Native	SeMet (peak)
X-ray source	14.2 SRS	14.2 SRS
Wavelength (Å)	0.9790	0.9792
Space group	<i>P6₂22</i>	<i>P6₂22</i>
Unit-cell parameters (Å, °)		
<i>a</i> = <i>b</i>	145.10	144.54
<i>c</i>	226.98	224.08
α = β	90.0	90.0
γ	120.0	120.0
Resolution limits (Å)	48.34–3.65 (3.78–3.65)	29.51–2.90 (3.00–2.90)
No. of observations	337049	668591
No. of unique observations	16328	57961
Average redundancy	20.64 (21.12)	11.54 (11.56)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
$\langle I/\sigma(I) \rangle$	13.1 (7.1)	15.2 (5.4)
$R_{\text{merge}}^{\dagger}$	0.113 (0.413)	0.083 (0.417)

$$\dagger R_{\text{merge}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$$

strain BL21 (DE3) and the cells were grown overnight on LB-agar plates containing 50 $\mu\text{g ml}^{-1}$ ampicillin at 310 K. Single colonies were used to inoculate overnight cultures.

2.2. Expression and purification

2.2.1. Native RafE. Cultures for induction (8 \times 11) were each inoculated with 1 ml of a 10 ml 310 K overnight LB culture containing 50 $\mu\text{g ml}^{-1}$ ampicillin and grown at 310 K. Isopropyl thio- β -D-galactoside (IPTG) was added to a final concentration of 1 mM at $A_{600} = 0.6$. After 16 h growth at 310 K, the cells were harvested by centrifugation at 3500g and 277 K for 15 min. Cells were resuspended in lysis buffer [50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 \times EDTA-free protease-inhibitor tablet (Roche) per 25 ml] and lysed by sonication (Status US200 with TT13 tip, 10 \times 30 s bursts at 100% power). Cell debris was pelleted by centrifugation at 8000g and 277 K for 15 min and any residual cell debris was removed by a further centrifugation of the supernatant at 40 000g and 277 K for 20 min. Supernatant was loaded onto a Ni-NTA (Qiagen) column previously equilibrated with 50 mM Tris-HCl pH 8.0, 300 mM NaCl and washed with this buffer until A_{280} was constant. Protein was eluted with a gradient to 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 500 mM imidazole. Eluted fractions containing RafE were pooled and concentrated at 2000g and 277 K to 60 mg ml⁻¹ (Amicon, 10 kDa molecular-weight cutoff), calculated using a theoretical extinction coefficient of 80 790 $M^{-1} \text{cm}^{-1}$ (Gasteiger *et al.*, 2005). Gel filtration was performed using a 100 ml Superdex 75 (Amersham) column equilibrated with 10 mM Tris-HCl pH 8.0, 50 mM NaCl and a load volume of 1 ml. Retardation of the protein through the column was observed, possibly owing to interaction with the cross-linked agarose-dextran matrix, and this allowed selection of an active form of the protein. Eluted protein was judged to be over 95% pure by SDS-PAGE analysis. The protein was then concentrated to 20 mg ml⁻¹ by centrifugation at 2000g and 277 K (Amicon, 10 kDa molecular-weight cutoff).

2.2.2. Selenomethionine-labelled RafE. Sequence analysis of RafE showed 11 non-terminal methionine residues and a molecular weight of 46.7 kDa, making it a good candidate for phasing by selenomethionine labelling. Selenomethionine-labelled protein was prepared using methionine-biosynthesis inhibition (Van Duyne *et al.*, 1993). A 10 ml 310 K overnight LB culture containing 50 $\mu\text{g ml}^{-1}$ ampicillin and 25 $\mu\text{g ml}^{-1}$ kanamycin was prepared and gently pelleted by centrifugation at 2000g and 300 K for 5 min. This pellet

was resuspended in 10 ml M9 minimal medium pre-warmed to 310 K and used to inoculate selenomethionine growth media comprising M9 minimum medium plus sterile filtered MgSO₄ (1 mM), glucose [0.4% (w/v)], thiamine [0.00005% (w/v)] and FeSO₄ (15 μM). Cells were grown at 310 K to an A_{600} of 0.3, at which point L-selenomethionine (Acros Organics; 50 mg l⁻¹), L-leucine (50 mg l⁻¹), L-lysine (100 mg l⁻¹), L-isoleucine (50 mg l⁻¹), L-phenylalanine (100 mg l⁻¹), L-threonine (100 mg l⁻¹) and L-valine (50 mg l⁻¹) were added, followed 15 min later by 1 mM IPTG. Cells were grown to $A_{600} = 1.0$ and then harvested by centrifugation at 3500g and 277 K for 15 min. Purification of selenomethionine-labelled RafE was carried out as previously described, with the addition of reducing agent at each stage. Lysis and Ni-NTA buffers were supplemented with 10 mM β -mercaptoethanol. Following elution from the Ni-NTA column, 10 mM DTT was added to the protein and used in all subsequent buffers. The retardation on the Superdex-75 column previously described was also observed for selenomethionine-labelled RafE. Incorporation of selenomethionine was checked by MALDI-TOF mass spectroscopy and found to be 100%.

2.3. Crystallization and data collection

2.3.1. Native RafE. Crystallization trials were performed using the sitting-drop vapour-diffusion method at 295 K with a protein concentration of 20 mg ml⁻¹ in 10 mM Tris-HCl pH 8.0, 50 mM NaCl. Initial screening was conducted using Hampton Research 24-well Cryschem Plates, a drop volume of 1.5 μl protein and 1.5 μl reservoir solution and a reservoir volume of 750 μl . Drops were mixed by aspiration. Screening was conducted using Hampton Research Crystal Screens 1 and 2 (Jancarik & Kim, 1991), Emerald Biosystems Wizard 1 and 2 and Cryo 1 and 2, Molecular Dimensions Structure Screen 1 and 2 and a wide range of in-house conditions.

Initial screening revealed single crystals in Crystal Screen 2 condition No. 14 and Structure Screen 2 condition No. 29. Both conditions are comprised of 2.0 M ammonium sulfate, 0.2 M sodium/potassium tartrate, 0.1 M sodium citrate pH 5.6. Crystals appeared after 3 d and continued growing for several months, but showed relatively poor diffraction, with a resolution limit of around 4.5 Å. Optimizations using Additive Screens 1, 2 and 3 (Hampton Research) with the original condition yielded crystals in 43 of the 72 conditions. These were all flash-cooled in liquid nitrogen using dry paraffin oil as a cryoprotectant (Riboldi-Tunnicliffe & Hilgenfeld, 1999). Owing to the large number of crystals, they were screened at the SRS Daresbury, beamline 14.2, using the Actor robotic system (Rigaku MSC).

**Figure 1**

Crystals of selenomethionine-labelled RafE. The dimensions of the largest crystal are approximately 400 \times 400 \times 250 μm .

Divalent cations proved the best additives, in particular MgCl_2 (10 mM), CaCl_2 (10 mM), CoCl_2 (10 mM), CuCl_2 (10 mM) and CdCl_2 (10 mM), with diffraction visible to 3.65 Å. Data were collected from a crystal grown in 2.0 M ammonium sulfate, 0.1 M Tris-HCl, 0.2 M sodium/potassium tartrate and CdCl_2 (10 mM) at the SRS Daresbury, beamline 14.2, using a MAR CCD detector and a crystal oscillation of

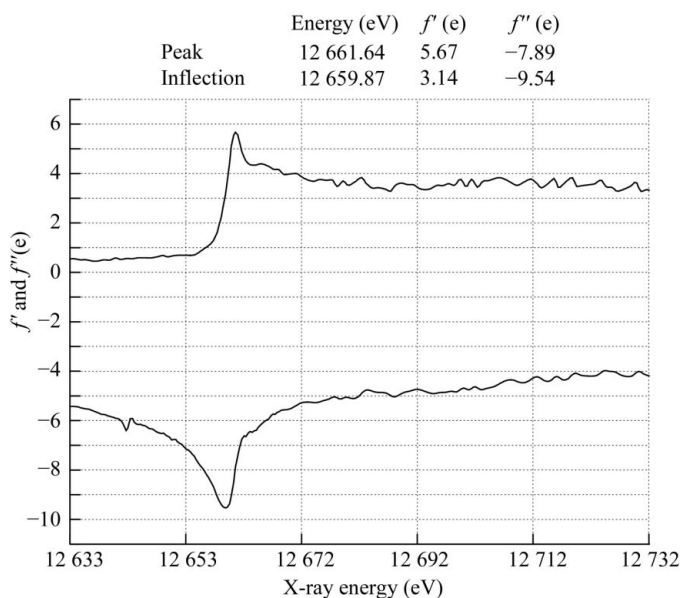


Figure 2 Scattering plot obtained from selenomethionine-labelled crystals. Figure produced using *CHOOCH* (Evans & Pettifer, 2001).

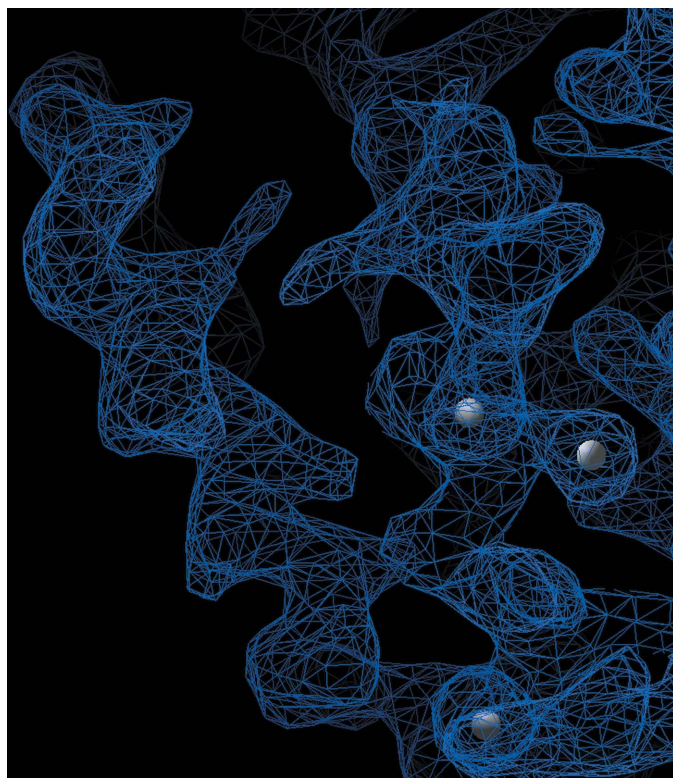


Figure 3 Experimental electron-density maps obtained from selenomethionine phasing contoured at 1.5σ . Figure produced using *Coot* (Emsley & Cowtan, 2004). Located Se-atom positions are indicated by grey spheres.

1° per frame. A summary of the data-collection statistics is given in Table 1. The data were processed and scaled using the *d*TREK* package (Pflugrath, 1999) with systematic absences indicating a space group of either $P6_122$ or $P6_522$, with unit-cell parameters $a = b = 145.10$, $c = 226.98$ Å, $\alpha = \beta = 90.0$, $\gamma = 120.0^\circ$. The number of molecules in the asymmetric unit was also ambiguous at this stage, with either two ($V_M = 3.76$ Å³ Da⁻¹, 67.33% solvent content), three ($V_M = 2.51$ Å³ Da⁻¹, 50.99% solvent content) or four ($V_M = 1.88$ Å³ Da⁻¹, 34.66% solvent content) molecules per asymmetric unit.

2.3.2. Selenomethionine-labelled RafE. Optimization of the initial crystallization condition discovered for the native protein yielded crystals that appeared identical to those previously obtained. Adjusting the conditions to 1.8 M ammonium sulfate, 0.1 M sodium citrate pH 5.1 and 0.2 M MgCl_2 yielded the best-looking crystals (Fig. 1) and these were flash-cooled as previously described. An XAFS scan at the SRS Daresbury, beamline 14.2, showed a strong signal at the selenium edge and following analysis with *CHOOCH* (Evans & Pettifer, 2001; Fig. 2), diffraction data were collected at the peak wavelength of 0.9792 Å. Fig. 2 shows the importance of collecting an accurate fluorescence scan from a selenomethionine-labelled protein crystal, as collection at the theoretical selenium absorption edge of 0.9795 Å (12657.9 eV) would have missed the actual absorption edge of this crystal and resulted in very low anomalous signal. Data-collection statistics are shown in Table 1. These data were processed and scaled using *d*TREK* and 33 potential Se sites were found using *SnB* (Blessing & Smith, 1999; Smith *et al.*, 1998; Turner *et al.*, 1998; Weeks & Miller, 1999) with a clear bimodal distribution of correct and incorrect solutions. The top 14 of these sites with peak height greater than 6σ were subjected to maximum-likelihood heavy-atom parameter refinement using *SHARP* (Bricogne *et al.*, 2003) and an additional two sites were identified from subsequent residual map analysis. Solvent flattening was performed using *SOLOMON* (Abrahams & Leslie, 1996) and indicated an optimal solvent content of 69.8%, suggesting the presence of two molecules per asymmetric unit, and produced readily interpretable maps (Fig. 3) that confirm the space group as $P6_122$. Maps generated in $P6_522$ were uninterpretable.

3. Concluding remarks

Both native and selenomethionine-labelled RafE from *S. pneumoniae* have been crystallized in space group $P6_122$ with two molecules in the asymmetric unit. Phasing using selenomethionine SAD data has been successful and has produced good-quality electron-density maps, although it is clear that in the asymmetric unit one molecule is well ordered while the other shows significant disorder in a number of regions. Of the 16 selenium sites found, 11 are from the well ordered molecule and show clear density. Model building is currently in progress and it appears that RafE shares the same periplasmic binding protein fold as seen in other ABC substrate-binding protein structures (Murzin *et al.*, 1995).

We would like to thank the beamline staff at station 14.2, SRS for help and support during data collection. This work was made possible by a studentship awarded to NGP by the Biotechnology and Biological Sciences Research Council (BBSRC).

References

- Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* **D52**, 30–42.
- Blessing, R. H. & Smith, G. D. (1999). *J. Appl. Cryst.* **32**, 664–670.

- Bricogne, G., Vonnrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* **D59**, 2023–2030.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
- Garmory, H. S. & Titball, R. W. (2004). *Infect. Immun.* **72**, 6757–6763.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomic Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa, NJ, USA: Humana Press.
- Gilson, E., Alloing, G., Schmidt, T., Claverys, J. P., Dudler, R. & Hofnung, M. (1988). *EMBO J.* **7**, 3971–3974.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
- Pflugrath, J. W. (1999). *Acta Cryst.* **D55**, 1718–1725.
- Riboldi-Tunnicliffe, A. & Hilgenfeld, R. (1999). *J. Appl. Cryst.* **32**, 1003–1005.
- Rosenow, C., Maniar, M. & Trias, J. (1999). *Genome Res.* **9**, 1189–1197.
- Russell, R. R., Aduse-Opoku, J., Sutcliffe, I. C., Tao, L. & Ferretti, J. J. (1992). *J. Biol. Chem.* **267**, 4631–4637.
- Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A. & Blessing, R. H. (1998). *Acta Cryst.* **D54**, 799–804.
- Sutcliffe, I. C., Tao, L., Ferretti, J. J. & Russell, R. R. (1993). *J. Bacteriol.* **175**, 1853–1855.
- Tettelin, H. *et al.* (2001). *Science*, **293**, 498–506.
- Turner, M. A., Yuan, C. S., Borchardt, R. T., Hershfield, M. S., Smith, G. D. & Howell, P. L. (1998). *Nature Struct. Biol.* **5**, 369–376.
- 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.
- World Health Organization (1999). *Wkly Epidemiol. Rec.* **74**, 177–183.