

Expression, purification and X-ray characterization of residues 18–230 from the pneumococcal histidine triad protein A (PhtA) from *Streptococcus pneumoniae*

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

^aDivision of Infection and Immunity (IBLS), University of Glasgow, Joseph Black Building, University Avenue, Glasgow G12 8QQ, Scotland, and ^bDepartment of Chemistry, University of Glasgow, Joseph Black Building, University Avenue, Glasgow G12 8QQ, Scotland

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

A fragment of the *Streptococcus pneumoniae* PhtA gene product (residues 18–230) was cloned and overexpressed in *Escherichia coli*. The purified protein was crystallized using the sitting-drop vapour-diffusion technique. Crystals belong to the monoclinic space group C2, with unit-cell parameters $a = 62.19$, $b = 35.9$, $c = 72.54$ Å, $\beta = 90.01^\circ$. The crystals diffract X-rays to beyond 1.2 Å resolution.

Received 9 February 2004
Accepted 27 February 2004

1. Introduction

The rapid increase in the number of fully sequenced bacterial genomes, together with advances in computer-software technology and third-generation synchrotron beamlines for macromolecular crystallography, has allowed structural genomics to flourish. One of the foremost goals of structural genomics is to map the entire protein-folding space. This can be accomplished by determining the structures of a large number (15 000–20 000) of carefully selected proteins that show no significant sequence homology and are therefore likely to include the majority of unique protein folds (Vitkup *et al.*, 2001). This effort will expand knowledge of protein structure and facilitate the structural determination of other proteins. Large-scale genome-sequencing projects reveal that on average the functions of more than 50% of predicted proteins cannot be inferred from their amino-acid sequences alone. A structural approach has sometimes proven to be a valid way of deducing the molecular functions of hypothetical proteins (Zarembinski *et al.*, 1998; Hwang *et al.*, 1999; Teplova *et al.*, 2000; Schulze-Gahmen *et al.*, 2003; Christendat *et al.*, 2000; Murzin & Patthy, 1999; Oliver, 1996). The selection of proteins for structure determination is key to the structural genomics approach (Linial & Yona, 2000). We have chosen to work on the organism *Streptococcus pneumoniae* as it is a pathogenic bacteria with a known genome sequence, is easily manipulated both biochemically and genetically to allow the function of the protein to be studied and is the focus of intensive research by a large community. We have specifically selected surface-exposed proteins as deduced from the presence of a signal leader sequence. Here, we present work on the pneumococcal protein SP1175 (<http://www.tigr.org>) which in the TIGR4 strain is termed a conserved hypothetical protein and has in the R6 strain been annotated as the

pneumococcal histidine triad protein A (PhtA). The amino-acid sequence of these gene products contains three histidine triad motifs (HxxHxH). Within *S. pneumoniae* there are four related pneumococcal histidine triad proteins (PhtA, PhtB, PhtD and PhtE). These proteins are approximately 800 amino acids in length and are highly conserved within this organism. Previous work (Wizemann *et al.*, 2001) has shown that immunization using a fragment of PhtA (residues 18–230) protects mice from subsequent infection by several strains of *S. pneumoniae* (in contrast to other potential protein vaccines, which elicit a response but are strain-specific). We present here the preliminary crystallization and X-ray diffraction studies on the fragment of PhtA.

2. Materials and methods

2.1. Protein expression and purification

A PCR product containing the coding region for part of the N-terminal domain (residues 18–230) of the SP1175 protein was cloned between the *Bam*HI and *Hind*III sites of the pQE-10 vector (Qiagen) in frame with an N-terminal His₆ tag. This vector was then transformed into *Escherichia coli* strain BL21 (DE3) and the cells were grown overnight at 310 K on LB agar plates. A single colony was picked to inoculate 200 ml LB medium containing antibiotic. Eight 1 l cultures of LB media were then each inoculated with 10 ml of an overnight culture and shaken at 200 rev min⁻¹ at 310 K until OD₆₀₀ = 0.6 was reached. Expression of the His₆-tagged fusion protein was induced with 1 mM isopropyl- β -D-thiogalactoside (IPTG). Following induction, growth continued for 4 h at 310 K. The cells were centrifuged at 5000g and the pellet was resuspended in buffer A containing 200 mM NaCl, 25 mM Tris and 4 mM imidazole pH 7.5. Cells were lysed by sonication on ice for 6 × 30 s. The cell lysate was centrifuged at 20 000g

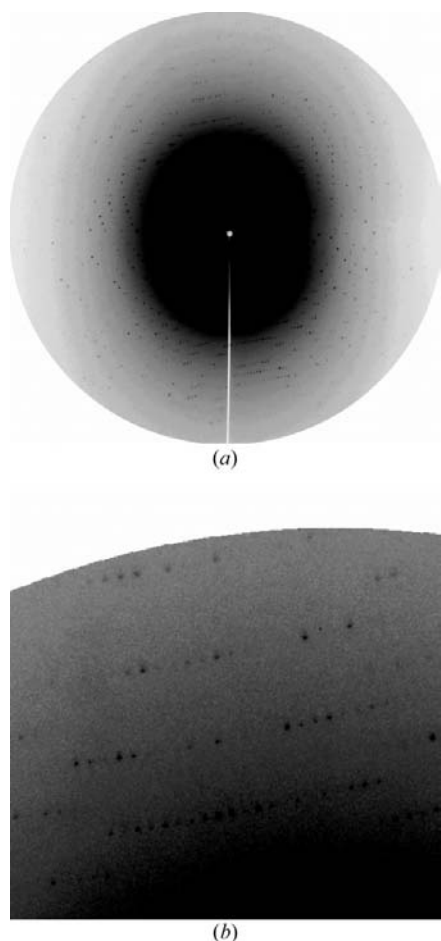


Figure 1
Diffraction image from SP1175. (a) Full image plate (b) close-up darkened image to show diffraction to the edge of the plate corresponding to 1.20 Å resolution.

for 1 h and the supernatant loaded onto an Ni-affinity column and eluted with a linear gradient from 0 to 50% buffer B (200 mM NaCl, 25 mM Tris and 1 M imidazole pH 7.5). Gel filtration using a Sephadex 26/60 column into buffer A resulted in pure protein.

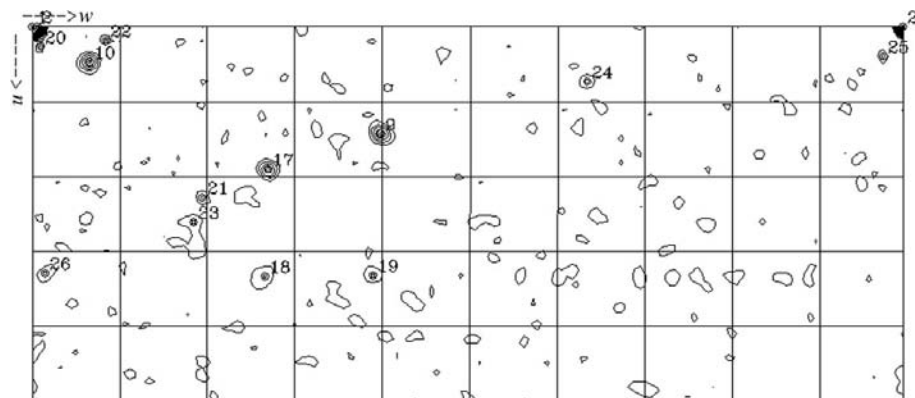


Figure 2
Anomalous difference Patterson map of the N-terminal domain of PhtA. The Harker section at $y = 0.0$, showing the positions of the three peaks. Peaks are contoured at 2.0σ ; the peak heights of the three major peaks are greater than 30σ .

Table 1
Data-collection statistics for SP1175.

Values in parentheses are data for the highest resolution shell.

Source	XRD-1, ELETTRA, Trieste
Wavelength (Å)	1.0
Resolution (Å)	74.0–1.2 (1.26–1.2)
Space group	C2
Unit-cell parameters (Å, °)	$a = 62.19, b = 35.9,$ $c = 72.54, \beta = 90.01$
R_{sym}	0.047 (0.175)
$\langle I/\sigma(I) \rangle$	8.1 (2.6)
Total No. reflections	593315
No. unique reflections	46930
Average redundancy	7.6 (7.2)
Completeness (%)	93.9 (93.9)

2.2. Crystallization and data collection

The protein was concentrated to 5–8 mg ml⁻¹ (calculated from the theoretical extinction coefficient of 0.910 M⁻¹ cm⁻¹) and crystallized using the sitting-drop vapour-diffusion technique, mixing equal volumes (2 µl) of protein and reservoir solutions to form the drop. Crystallization trials included the sparse-matrix screens Wizard I, Wizard II, Cryo I and Cryo II from Emerald Biostructures, and Crystal Screen, Crystal Screen II and the Grid Screens from Hampton Research. A single crystal grew at 293 K over 12 months from the Ammonium Sulfate Grid Screen condition B4 (1.6 M ammonium sulfate and 0.1 M HEPES pH 7.0) with approximate dimensions of 80 × 60 × 20 µm. The crystal was flash-cooled in a nitrogen stream at 100 K with dried paraffin oil as a cryoprotectant (Riboldi-Tunnicliffe & Hilgenfeld, 1999).

Diffraction data were collected from this crystal at the XRD-1 beamline at ELETTRA using a MAR CCD detector and a crystal rotation of 1° per frame. The wavelength was set at 1.0 Å to achieve the maximum flux during the experiment. A summary of the data-collection statistics is given in Table 1. The data were processed

and reduced using the program *MOSFLM* (v.6.2.2; Leslie, 1992) and scaled using *SCALA* (Evans, 1993) and *TRUNCATE* (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

The crystal belongs to the monoclinic space group C2, with unit-cell parameters $a = 62.19, b = 35.9, c = 72.54$ Å, $\alpha = 90.0, \beta = 90.01, \gamma = 90.0^\circ$. The calculated Matthews coefficient (V_M) for a monomer in the asymmetric unit is 1.80 Å³ Da⁻¹ with 33% of the unit-cell volume occupied by solvent (Matthews, 1968). The crystal diffracted to better than 1.2 Å resolution (Fig. 1).

As this family of proteins is known to bind metal ions, the data were processed to keep the Friedel pairs separate. The anomalous Patterson map contained three large peaks in the Harker section, which we assume arise from Zn atoms (Fig. 2). The positions, occupancy and temperature factors of these atoms are being refined in *SHARP* (de La Fortelle & Bricogne, 1997) to obtain estimates of phases.

We would like to thank the beamline staff at XRD-1, ELETTRA or help and support during data collection. This work was supported by EC grant QLRK2-2000-00542 and a grant from the BBSRC.

References

- Christendat, D. *et al.* (2000). *Nature Struct. Biol.* **7**, 903–908.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.
- Hwang, K. Y., Chung, J. H., Kim, S.-H., Ham, Y. S. & Cho, Y. (1999). *Nature Struct. Biol.* **6**, 691–696.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–493.
- Leslie, A. G. W. (1992). *Int. CCP4/ESF-EAMCB Newsl. Protein Crystallogr.* **26**, 27–33.
- Linial, M. & Yona, G. (2000). *Prog. Biophys. Mol. Biol.* **73**, 297–320.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Murzina, A. G. & Patthy, L. (1999). *Curr. Opin. Struct. Biol.* **9**, 359–362.
- Oliver, S. G. (1996). *Nature (London)*, **379**, 597–600.
- Riboldi-Tunnicliffe, A. & Hilgenfeld, R. (1999). *J. Appl. Cryst.* **32**, 1003–1005.
- Schulze-Gahmen, U., Pelaschier, J., Yokota, H., Kim, R. & Kim, S.-H. (2003). *Proteins*, **50**, 526–530.
- Teplava, M., Tereshko, V., Sanishvili, R., Joachimak, A., Bushueva, T., Anderson, W. F. & Egli, M. (2000). *Protein Sci.* **9**, 2557–2566.

Vitkup, D., Melamund, E., Moulton, J. & Sander, C. (2001). *Nature Struct. Biol.* **8**, 559–565.

Wizemann, T. M., Heinrichs, J. H., Adamou, J. E., Erwin, A. L., Kunsch, C., Choi, G. H., Barash, S. C., Rosen, C. A., Masure, H. R.,

Tuomanen, E., Gayle, A., Brewah, Y. A., Walsh, W., Barren, P., Lathigra, R., Hanson, M., Langermann, S., Johnson, S. & Koenig, S. (2001). *Infect. Immun.* **69**, 1593–1598.

Zarebinski, T. I., Hung, L.-W., Mueller-Dieckmann, H.-J., Kim, K.-K., Yokota, H., Kim, R. & Kim, S.-H. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 15189–15193.