STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

The 1.1 Å resolution structure of a periplasmic phosphate-binding protein from *Stenotrophomonas maltophilia*: a crystallization contaminant identified by molecular replacement using the entire Protein Data Bank

Ronan Keegan, David G. Waterman, David J. Hopper, Leighton Coates, Graham Taylor, Jingxu Guo, Alun R. Coker, Peter T. Erskine, Steve P. Wood and Jonathan B. Cooper

Acta Cryst. (2016). D**72**, 933–943

STRUCTURAL BIOLOGY

ISSN 2059-7983

The 1.1 Å resolution structure of a periplasmic phosphate-binding protein from *Stenotrophomonas maltophilia*: a crystallization contaminant identified by molecular replacement using the entire Protein Data Bank

Received 27 April 2016
Accepted 28 June 2016

Edited by J. Newman, Bio21 Collaborative Crystallisation Centre, Australia

Keywords: protein crystallography; periplasmic binding proteins; molecular replacement; phosphate-binding proteins; *Stenotrophomonas maltophilia*.

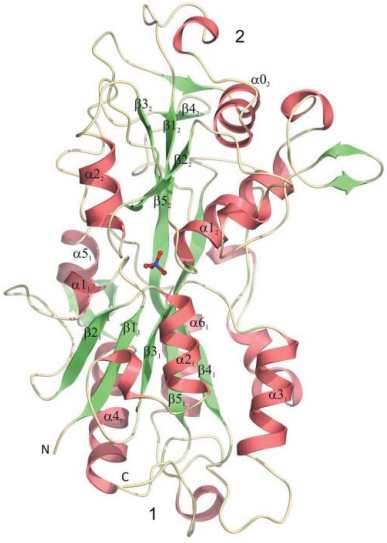
PDB reference: periplasmic phosphate-binding protein, 5jk4

Supporting information: this article has supporting information at journals.iucr.org/d

Ronan Keegan,[a,b,c] David G. Waterman,[a,b] David J. Hopper,[d] Leighton Coates,[e] Graham Taylor,[f] Jingxu Guo,[g] Alun R. Coker,[g] Peter T. Erskine,[g,h] Steve P. Wood[g] and Jonathan B. Cooper[g,h]*

[a]STFC, Rutherford Appleton Laboratory, Harwell Oxford, Didcot OX11 0FA, England, [b]CCP4, Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Oxford, Didcot OX11 0FA, England, [c]Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, England, [d]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth SY23 3DA, Wales, [e]Biology and Soft Matter Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA, [f]Wolfson Drug Discovery Unit, Centre for Amyloidosis and Acute Phase Proteins, UCL Division of Medicine (Royal Free Campus), Rowland Hill Street, London NW3 2PF, England, [g]Division of Medicine, UCL, Gower Street, London WC1E 6BT, England, and [h]Department of Biological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, England. *Correspondence e-mail: jon.cooper@ucl.ac.uk

During efforts to crystallize the enzyme 2,4-dihydroxyacetophenone dioxygenase (DAD) from *Alcaligenes* sp. 4HAP, a small number of strongly diffracting protein crystals were obtained after two years of crystal growth in one condition. The crystals diffracted synchrotron radiation to almost 1.0 Å resolution and were, until recently, assumed to be formed by the DAD protein. However, when another crystal form of this enzyme was eventually solved at lower resolution, molecular replacement using this new structure as the search model did not give a convincing solution with the original atomic resolution data set. Hence, it was considered that these crystals might have arisen from a protein impurity, although molecular replacement using the structures of common crystallization contaminants as search models again failed. A script to perform molecular replacement using *MOLREP* in which the first chain of every structure in the PDB was used as a search model was run on a multi-core cluster. This identified a number of prokaryotic phosphate-binding proteins as scoring highly in the *MOLREP* peak lists. Calculation of an electron-density map at 1.1 Å resolution based on the solution obtained with PDB entry 2q9t allowed most of the amino acids to be identified visually and built into the model. A *BLAST* search then indicated that the molecule was most probably a phosphate-binding protein from *Stenotrophomonas maltophilia* (UniProt ID B4SL31; gene ID Smal_2208), and fitting of the corresponding sequence to the atomic resolution map fully corroborated this. Proteins in this family have been linked to the virulence of antibiotic-resistant strains of pathogenic bacteria and with biofilm formation. The structure of the *S. maltophilia* protein has been refined to an *R* factor of 10.15% and an $R_{\rm free}$ of 12.46% at 1.1 Å resolution. The molecule adopts the type II periplasmic binding protein (PBP) fold with a number of extensively elaborated loop regions. A fully dehydrated phosphate anion is bound tightly between the two domains of the protein and interacts with conserved residues and a number of helix dipoles.



1. Introduction

The widely occurring Gram-negative bacterium *Stenotrophomonas maltophilia* is a relatively rare cause of human disease, but it is known to cause pulmonary infections in immunocompromised patients and lung cancer victims (Brooke, 2012). It is an obligate aerobe, and along with

Pseudomonas aeruginosa is known to co-colonize the lungs of cystic fibrosis sufferers. In the past, the bacterium has been given a range of names, including *P. maltophilia*, and as a 'maltophile' it has a marked preference for metabolizing maltose rather than glucose in the growth medium. It is becoming increasingly recognized as an antibiotic- and biocide-resistant pathogen which frequently colonizes hospital catheters and respiratory equipment, forming extensive biofilms (Deredjian *et al.*, 2016).

The uptake of phosphate anions by bacteria in conditions of phosphate starvation relies on phosphate-specific transport (Pst) proteins which form an ATP-binding cassette (ABC) transporter. This system includes a water-soluble, periplasmic binding protein (PstS) with a high affinity for phosphate ($K_d \simeq 0.2\ \mu M$; Webb *et al.*, 1992). Bound phosphate groups are donated to the transmembrane domains of the ABC transporter and hydrolysis of ATP by the intracellular ATP-binding domain powers the transport of the phosphate anion across the lipid bilayer against its concentration gradient. PstS acts as a periplasmic phosphate sensor, which together with the permease components of the transporter and a range of intracellular Pho proteins triggers changes in the expression levels of many genes involved in phosphate metabolism (Santos-Beneit, 2015). It is interesting that whilst PBPs are restricted to Gram-negative bacteria, Gram-positive bacteria, which lack an outer membrane, possess PBP homologues known as extracytoplasmic binding lipoproteins which are anchored in the cytoplasmic membrane. The more exposed nature of these proteins in Gram-positive bacteria means that they tend to be highly antigenic.

Periplasmic binding proteins (PBPs) constitute a superfamily of proteins which have a two-domain α/β-fold that is present in a wide range of protein families, including various enzymes, receptors and transporter proteins (Felder *et al.*, 1999). These molecules often lack significant sequence identity, but all adopt a conserved fold and have an underlying tendency to bind anionic ligands. PBPs exist for the recognition of monosaccharides and oligosaccharides, amino acids and polypeptides as well as inorganic nutrients and vitamins. Thus, PBPs act as the initial receptor for active transport of these molecules into the cell and are also involved in drug transport.

The two structural domains of these proteins are connected by flexible linker regions which act as a hinge, allowing relative movement of the two domains by as much as 70° to encapsulate the bound ligand (Felder *et al.*, 1999). This domain movement has been likened to a Venus flytrap and plays a pivotal role in signalling and transport as well as catalysis by the enzymes which possess this fold. Examples of proteins in the PBP superfamily include transcriptional regulators such as the LacI repressor and the extracellular domains of various neurotransmitter receptors. Many of these proteins are targets for active drug development and also have biotechnological applications, for example the familiar maltose-binding protein (MBP), which is used as an affinity tag in protein expression. MBP is classed as a type I PBP and like all members of this class has three crossovers between the two structural domains of the protein. Each domain of a type I PBP is formed by two separate segments from the N- and C-terminal halves of the polypeptide. In contrast, type II PBPs have only two crossovers between the domains. This arises by forming domain 1 from the N-terminal and C-terminal portions of the polypeptide, while domain 2 is formed exclusively by the middle segment. Structurally, the domains of both types of PBP consist of a central β-sheet flanked on one or both sides by α-helices. Mutagenesis and structural studies showed that the residues which interact with the transmembrane components are only brought close together by the closure of the two domains upon ligand binding (Martineau *et al.*, 1990; Sharff *et al.*, 1992). To ensure efficient sensing of the ligand, PBP molecules are present in much larger numbers than the transporter or chemotaxis receptors. Therefore, it is particularly important that only the closed, ligand-bound form of the PBP can activate the transmembrane components. A review of the structure and function of the PBP superfamily is given by Berntsson *et al.* (2010).

Phosphate-binding proteins belong to the type II class of PBPs and achieve high specificity by fully dehydrating the ligand to form a complex in which buried salt bridges and numerous hydrogen bonds are formed with the buried HPO_4^{2-} (P_i) dianion (Quiocho & Ledvina, 1996). The closely related bacterial sulfate-binding proteins entrap SO_4^{2-} ions by hydrogen bonding to the ligand in a similar manner, albeit without any ionizable residues taking part (Pflugrath & Quiocho, 1985). In the phosphate- and sulfate-binding PBPs, each O atom of the ligand is involved in multiple hydrogen bonds to amino acids of the protein, and many involve the protein main chain. These tight interactions are necessary for PBPs to distinguish between phosphate and sulfate which, owing to their marked difference in acid strength, can be discriminated on the basis of their protonation state. In the lower specificity PBPs, such as those which bind sugars, less intense hydrogen bonding to the ligand, usually involving protein side chains, is found and aromatic residues also play a key role in elegant ring-stacking interactions with the hydrophobic faces of the sugar (Quiocho & Ledvina, 1996).

The expression of bacterial phosphate-binding proteins is known to be increased by several orders of magnitude in phosphate starvation (Fischer *et al.*, 2006; Lewenza *et al.*, 2005; Madhusudhan *et al.*, 2003). These proteins have been linked to the virulence of antibiotic-resistant strains and to biofilm formation, which stems from the formation of large appendages on the bacteria that are rich in secreted phosphate-binding proteins (Díaz *et al.*, 2005; Mudrak & Tamayo, 2012; Neznansky *et al.*, 2014; O'May *et al.*, 2009; Zaborina *et al.*, 2008).

In the protein crystallization field it is sometimes found that protein impurities, which are present at levels of only a few percent, crystallize rather than the molecule of interest. Some examples include the bacterial stress proteins and bacterioferritins (see, for example, van Eerde *et al.*, 2006). Even when the target protein is affinity-purified, other molecules which have inherent affinity for the immobilized ligand can become concentrated along with the protein of interest, for example

the histidine-rich sequences of *Escherichia coli* SlyD and AcrB cause them to contaminate recombinant His-tagged proteins. In the case of the membrane protein AcrB, the problem is compounded by its ability to crystallize when present in picogram quantities (Psakis *et al.*, 2009). However, in favourable cases these crystallization 'accidents' have led to new structures being solved, such as yeast nicotinamidase (Hu *et al.*, 2005, 2006, 2007) and YcaC from both *E. coli* (Colovos *et al.*, 1998) and *P. aeruginosa* (Grøftehauge *et al.*, 2015). Crystallization contaminants are often only identified after X-ray data collection and processing, at which point molecular replacement with structures that are expected to have sequence similarity to the target protein proves unsuccessful. Sometimes the unit-cell dimensions are then found to match those of a known structure in the PDB and these suspicions can then be confirmed or refuted ultimately by refinement. In difficult cases, the molecule may be identified by dissolving the crystal and running an SDS–PAGE gel for liquid-chromatography mass-spectrometric (LC-MS) sequencing of tryptic peptides. In addition, molecular replacement may be performed using structures of common crystallization contaminants; failing this, heavy-atom methods may be the only option available for solving the structure of the impurity if its crystallization is reliable. To help to resolve problems of this sort, the SBGrid service (Stokes-Rees & Sliz, 2010) has been developed to perform molecular replacement using *Phaser* (McCoy *et al.*, 2007) with approximately 100 000 protein domains in the SCOP database (Murzin *et al.*, 1995) as potential search models.

Here, we report the structure determination of a PstS-like phosphate-binding protein from *S. maltophilia* which crystallized unexpectedly during efforts to grow crystals of another protein. Since the original crystals were no longer available for LC-MS sequence analysis, the constituent protein was identified and solved by a PDB-wide molecular-replacement search. The structure has been refined to *R*-factor and $R_{\rm free}$ values of 10.15 and 12.46%, respectively, at 1.1 Å resolution. The structure adopts the type II periplasmic binding protein fold and possesses a bound phosphate anion in the inter-domain cleft. We also describe the methodological approaches needed to process the atomic resolution data with the program *DIALS* and to solve the structure efficiently by molecular replacement using all entries in the Protein Data Bank as potential search models.

2. Methods

2.1. Crystallization

Crystals of the *S. maltophilia* protein were obtained during efforts to crystallize the enzyme 2,4′-dihydroxyacetophenone dioxygenase (DAD) from *Alcaligenes* sp. 4HAP, which had been expressed in *E. coli* and purified as described by Hopper & Kaderbhai (1999). The freeze-dried protein was dissolved in deionized water to give a concentration of 2.5 mg ml^{-1}, and crystals which eventually proved to be of the PBP impurity were obtained by the hanging-drop method in 15%(*w*/*v*) PEG 8K, 100 m*M* Tris–HCl pH 8.0. The droplet consisted of 5 µl protein solution, 5 µl well solution and 1 µl 100 m*M* sodium formate, a product of the reaction catalysed by DAD. A photomicrograph of the largest crystal, which was obtained after two years of growth at room temperature, is shown in Supplementary Fig. S1. The crystals were obtained in only one crystallization droplet and were transferred using loops to 10 µl of the well solution, to which glycerol was added in four 1 µl droplets to increase the cryoprotectant concentration to approximately 30%(*v*/*v*). The crystals were then mounted again in loops and flash-cooled in liquid ethane for storage in liquid nitrogen.

2.2. X-ray data collection

X-ray diffraction data were collected to almost 1.0 Å resolution (see Supplementary Fig. S2) using an ADSC Q4r detector on the ID14-EH2 beamline (Wakatsuki *et al.*, 1998) at ESRF, Grenoble, France with a fixed wavelength of 0.933 Å, some 14 years before completion of the structure analysis. A high-resolution pass was collected with a crystal-to-detector distance of 80 mm and, for each diffraction image, the crystal was rotated slowly through 1° while being exposed to the beam for 10 s. To record overloaded reflections, a low-resolution pass was then collected with a longer crystal-to-detector distance of 250 mm and shorter exposures of 1 s duration for each successive 2° rotation of the goniometer φ axis.

2.3. Data processing

Initially, the data were integrated with *MOSFLM* (Battye *et al.*, 2011) and scaled using *SCALA* (Evans, 2006) in the *CCP*4 program suite (Winn *et al.*, 2011) to a resolution of 1.1 Å. The scaling statistics and inspection of systematic absences indicated that the space group was $P2_1$, with unit-cell parameters a = 37.7, b = 77.9, c = 56.3 Å, β = 102.1°. Use of the *MATTPROB* server (Kantardjieff & Rupp, 2003) indicated that there were likely to be two molecules of DAD (which are approximately 20 kDa each) in the asymmetric unit. However, attempts at molecular replacement with various search models which had sequence homology to DAD failed to yield a solution, and when the structure of DAD was solved in a different crystal form (Beaven *et al.*, 2014; Keegan *et al.*, 2014) this also did not provide a convincing solution with the 1.1 Å resolution data set.

2.4. Data reprocessing

We then felt it necessary to check the original data processing in case any complications such as crystal twinning or incorrect space-group assignment had occurred. To reprocess the original diffraction images collected some 14 years prior, the data were read from DLT4 tape by Advanced Downloading Ltd, Putney, London, England. The high- and low-resolution passes were processed separately by *xia*2 (Winter, 2010), running the *DIALS* (Waterman *et al.*, 2013) pipeline, as available in *CCP*4 v.7.0. Spot-finding in *DIALS* is based on the method used by *XDS* (Kabsch, 2010), in which a

series of filters are applied to the images, resulting in a binary threshold map which identifies the strong pixels that may be part of a spot. Since the default spot-finding parameters in *DIALS* are currently optimized for Pilatus detectors, rather than CCDs, some optimization was required to reduce noise and spurious artefacts in the strong-spot list. This was undertaken interactively using *dials.image_viewer* (Supplementary Fig. S3) and the improved parameters were passed to *dials.find_spots* using the command-line options `kernel_size=2,2`, `sigma_background=10` and `sigma_strong=10`. In addition, to index the high-resolution pass correctly it was necessary to provide the known unit cell as a hint using the option `unit_cell=37.8,56.4,78.0,90,90,102`. Successful processing was also found to be possible by correcting the poor initial diffraction geometry using *dials.discover_better_experimental_model* followed by *dials.index*. In retrospect, we found that the beam centre in the image headers was wrong by approximately 1 mm, causing one of the unit-cell parameters to become doubled, and the use of *dials.check_indexing_symmetry* showed that the supposed beam centre for this solution was located at the (0, 1, 0) reflection.

The programs *dials.refine_bravais_settings* and *dials.reindex* were used to apply monoclinic constraints in the conventional setting and further geometry refinement was performed with *dials.refine*, fitting first a static model of the crystal followed by a scan-varying model (Waterman *et al.*, 2016). Integration of the images was perfomed by *dials.integrate* with default parameters, and subsequent scaling and merging of the intensity data was undertaken using *AIMLESS* (Evans, 2006; Evans & Murshudov, 2013).

2.5. Molecular replacement using the entire PDB

Initial attempts at molecular replacement (MR) with the reprocessed data set were performed using the previously solved DAD structure (PDB entry 4p9g), as before. Several search models derived from PDB entry 4p9g and its closest homologues were used, but none yielded any indication of possible success. Hence, it was concluded that this crystal probably did not contain what we expected. We next used the *AMPLE* (Bibby *et al.*, 2012) and *ARCIMBOLDO* (Rodríguez *et al.*, 2009) programs from the *CCP*4 suite to try and position fragment models such as helices using molecular replacement. With the high-resolution data even a small fragment correctly positioned should be enough of a 'seed' for the *SHELXE* (Sheldrick, 2010) density modification and C^{α}-tracing algorithm to potentially build up a more complete backbone of the target structure. However, all attempts at this failed, which may be a consequence of the relatively large unit cell and the lack of a heavier-than-average atom to 'seed' the *ab initio* phasing. We then speculated that it was a contaminant protein that may have already been solved and deposited in the PDB, as for example reported by Benini *et al.* (2004). With this hypothesis, we searched for known structures with similar unit-cell dimensions using the *nearest-cell* server (Ramraj *et al.*, 2012). However, this did not generate any candidates, suggesting that a more comprehensive approach was needed.

Finally, we considered performing MR using search models from all of the available PDB structures. To limit the computational overhead, we decided to invoke a quick rotation and translation search using *MOLREP* (Vagin & Teplyakov, 2010) with the resolution of the data limited to 3.0 Å. We also limited the number of top rotation-function (RF) peaks [as scored by RF/σ(RF)] that were sampled in the translation function to five. This gave us the advantage of being able to sample a large subset of the entire PDB relatively quickly using a multi-core cluster. We used a simple script that only sampled the first chain from each of the approximately 116 000 known PDB structures (at the time of running). In addition, to further speed up calculations only one copy of the model was searched for. No discrimination was made on the type of search model (DNA, RNA *etc.*) or on size or similarity to others in the PDB. This potentially resulted in a lot of unnecessary tests being performed, but with each *MOLREP* job running for an average of 38 s the effort required to eliminate any redundancy would have taken longer than to run the entire test. To complete the search took 20 h on the 60 cores of the CCP4 cluster.

When the test had completed, the log files indicated that several search models had clearly defined strong and outstanding peaks in the rotation search. The rotation search lacks the sensitivity of the translation search, but has the advantage of being independent of packing constraints and screw-axis assignment, so for this coarse-grained approach using unmodified search models it was deemed to be a more reliable indicator of success. The best-scoring search models were chain *A* from each of PDB entries 4f1u and its related structures 4f1v, 4f19, 4f18 and 2q9t, which are all high-resolution structures of phosphate-binding proteins. For example, with chain *A* of PDB entry 2q9t (Ahn *et al.*, 2007) the rotation function had a signal-to-noise ratio of 13.16 (which compared favourably with the highest noise peak of 4.92) and the corresponding translation function had a signal-to-noise ratio of 8.63 (which again compared well with the highest noise peak in the list of 2.99).

Further validation of the solution was carried out by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using chain *A* of PDB entry 2q9t as a search model. It produced a single solution with a translation-function *Z*-score of 13.7 and a log-likelihood gain of 286 after rigid-body refinement. The positioned model was then subjected to ten cycles of restrained refinement in *REFMAC*5 (Murshudov *et al.*, 1997, 2011), giving a final R and $R_{\rm free}$ of 0.43 and 0.45, respectively. Visual examination of the resulting electron-density map in *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010) showed a clear correlation between the model and the electron density, with strong difference map peaks that suggested improvements to the model (Supplementary Fig. S4). The unknown structure was later found to have a sequence identity of 44% and an r.m.s.d. of 1.3 Å to PDB entry 2q9t.

At this point it was clear that chain *A* of PDB entry 2q9t provided a good approximation for the phases of the unknown structure, but model building and completion would still present a problem owing to the lack of sequence information.

However, with data extending to 1.1 Å resolution, it was found that most of the side chains could be identified correctly by density modification using *SHELXE* (Sheldrick, 2010). This program generated a model of the protein main chain which included 338 residues (out of a final total of 373) and had a correlation coefficient to the data of 40.35%. Refining this backbone model against the data produced a difference map that gave clear density for many of the side chains. Using *Coot*, we initially added the larger, more obvious side chains such as Trp, Tyr and His. Further refinement of the improved model allowed more of the side chains to be added and the addition of the missing residues. This refinement and manual building procedure was cycled through until much of the model was complete, and a round of automatic rebuilding with *ARP/wARP* (Langer *et al.*, 2008) was then undertaken.

At this stage, some ambiguity in the sequence remained (*e.g.* Asp/Asn, Glu/Gln and Val/Thr) but it was correct enough to provide a good approximation to the target sequence. A *BLAST* search then indicated that the closest known sequence was that of a PstS-like protein from *S. maltophilia* strains R551-3 (UniProt ID B4SL31) or RA8 (UniProt ID M5D7J9), which have identical amino-acid sequence. Further inspection of the electron density showed that the sequence of this protein fitted the atomic resolution map extremely well and indicated how improvements could be made to the model. High sequence identity was found with a range of PBPs from other strains of *S. maltophilia*, but efforts to build in the sequence of at least one close homologue quickly stalled owing to poor fit to the density. Subsequent rebuilding of the completed model using *Coot* and refinement with *REFMAC*, *SHELX* (Sheldrick & Schneider, 1997; Sheldrick, 2011) and *PHENIX* (Adams *et al.*, 2010) was undertaken using default geometric restraints with each program. The structure was refined with anisotropic *B* factors, riding H atoms and 15 dual-occupancy side-chain groups. Figures of the structure were prepared using *CueMol* (http://www.cuemol.org/en) and the sequence alignment was prepared using *ALSCRIPT* (Barton, 1993). Bioinformatic analysis was performed using ExPASy (Gasteiger *et al.*, 2003), UniProt (2015) and KEGG (Kanehisa *et al.*, 2016). The structure and reflection data have been deposited in the Protein Data Bank (http://www.wwpdb.org) with accession code 5jk4. The original diffraction images are openly accessible at the DOI 10.5281/zenodo.49859.

For proteomic analyses, gel samples were reduced, carbamidomethylated and digested with trypsin using a standard protocol (Abdul-Salam *et al.*, 2006). Peptides were analysed using a LTQ-Velos mass spectrometer (Thermo Fisher Scientific) coupled to a nanoACQUITY UPLC (Waters) and were identified using the *Mascot* search engine (Matrix Science, London, England; http://www.matrixscience.com).

3. Results

The predicted molecular weight of the *S. maltophilia* phosphate-binding protein (38 338 Da) is consistent with at least one faint band visible on SDS–PAGE analysis of the original freeze-dried, purified DAD sample (see Supplementary Fig. S5), which initially suggested that the protein which crystallized may have originated from contamination by *S. maltophilia* during expression or purification. However, LC-MS sequencing of the proteins of this molecular weight which were present in the SDS–PAGE gel established that they were most likely of *E. coli* origin (Supplementary Table S1), indicating instead that *S. maltophilia* may have grown in the crystallization droplet and produced the phosphate-binding protein, which crystallized there.

Table 1
X-ray data-collection, processing and refinement statistics.

Values in parentheses are for the outer resolution shell.

Beamline	ID14-EH2
Wavelength (Å)	0.933
Space group	$P2_1$
Unit-cell parameters	
a (Å)	37.73
b (Å)	77.86
c (Å)	56.29
β (°)	102.08
Solvent content (%)	41.7
No. of molecules per asymmetric unit	1
Matthews coefficient (Å^3 Da^{-1})	2.1
Mosaic spread (°)	0.44
Resolution (Å)	44.94–1.10 (1.12–1.10)
Completeness (%)	94.1 (85.9)
$R_{\rm merge}$† (%)	7.6 (63.1)
$R_{\rm meas}$‡ (%)	8.5 (73.8)
$CC_{1/2}$§ (%)	99.8 (70.8)
Average $I/\sigma(I)$	9.7 (2.3)
Multiplicity	4.4 (3.7)
No. of observed reflections	533955 (20147)
No. of unique reflections	120995 (5469)
Wilson plot B factor (Å^2)	5.3
R factor (%)	10.15
Free R factor (%)	12.46
No. of reflections in test set	5935
R.m.s.d., bond lengths (Å)	0.006
R.m.s.d., bond angles (°)	0.961
Mean protein B factor (Å^2)	9.9

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. § Half-set correlation coefficient (Karplus & Diederichs, 2012).

3.1. Three-dimensional structure

Refinement of the *S. maltophilia* phosphate-binding protein structure at 1.1 Å resolution decreased the R factor and $R_{\rm free}$ to 10.15 and 12.46%, respectively. The refinement and other statistics are shown in Table 1, where it can be seen that the overall data completeness (94%) is perhaps slightly lower than would be desired. However, the situation is better at lower resolution: for example, the overall completeness increases to 96% at 1.3 Å and to 98% at 1.7 Å. Since this effect probably stems from the low symmetry of the crystal and the fact that the data were collected in a single pass, improving the completeness would require the collection of data from a re-oriented crystal. However, the expected difficulties in reproducing the crystallization of an impurity and the length of time required to obtain the original crystals (two years) suggested that it was better to focus on the available data set, which appeared to be of high quality by most other criteria. In addition, the electron density for the molecule was found to be

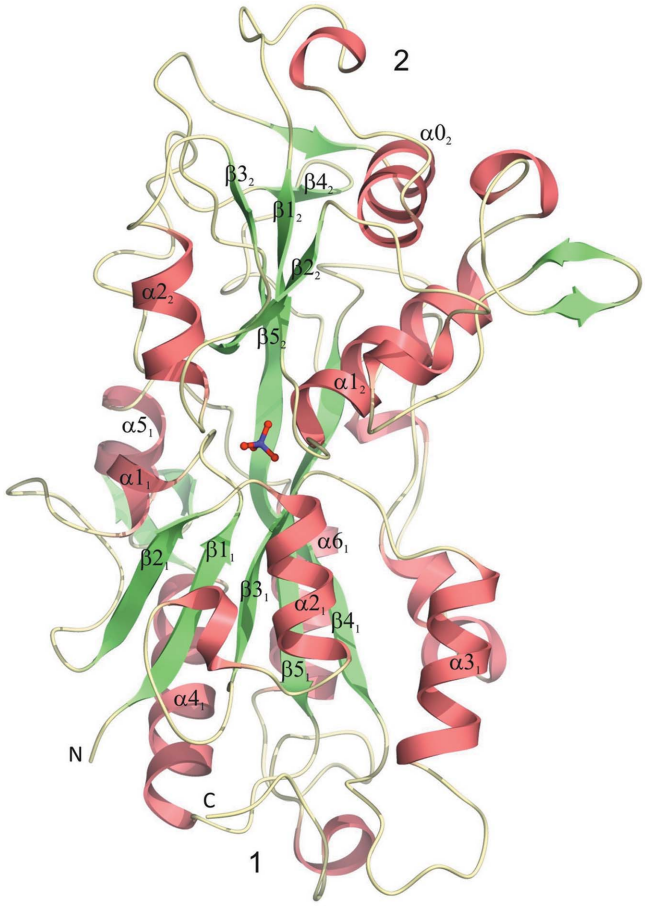


Figure 1
The tertiary structure of the *S. maltophilia* phosphate-binding protein. The secondary-structure elements are numbered to emphasize the topological twofold symmetry that exists between domains 1 (bottom) and 2 (top), with the domain number of each element indicated by the subscript. The phosphate group is drawn in the centre of the structure in ball-and-stick representation with the P atom coloured purple. The N- and C-termini are labelled at the bottom.

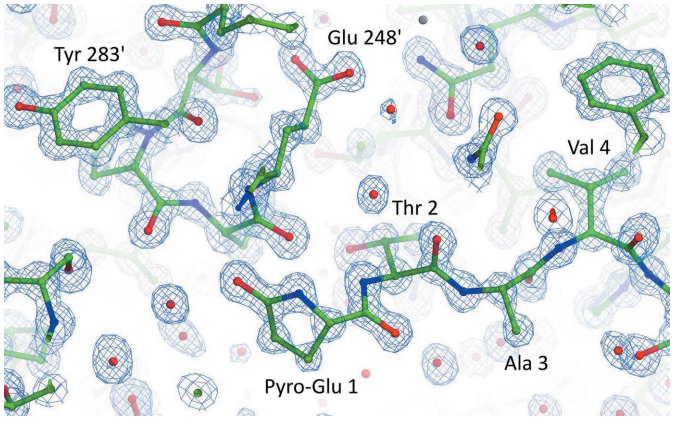


Figure 2
A sample of the $2F_o - F_c$ electron density for the *S. maltophilia* phosphate-binding protein at 1.1 Å resolution. The N-terminal pyro-glutamate residue (formed from glutamine) is shown at the bottom centre and the symmetry-related molecule with which it interacts is shown at the upper left (residue numbers indicated by primes). The electron density is contoured at 2.0 r.m.s.

extremely well defined throughout its length, with only half a dozen of the longer side chains being poorly defined out of 373 amino acids. According to the *MolProbity* criteria (Chen *et al.*, 2010), 98.4% of all residues in the model are in 'favoured' regions of the Ramachandran plot and all residues have 'allowed' main-chain conformations.

Given that the protein has been identified by 'visual' X-ray sequencing, it is possible that the poorly fitting side chains may actually be different amino acids. However, replacing these residues with those that appear to fit the electron density better and repeating the *BLAST* search still indicates that this molecule is most likely to be the *S. maltophilia* PstS-like protein. This suggests that the initial identification is correct and that any possible deviation from this sequence is most likely owing to natural variation within this bacterial species.

The overall structure is shown in Fig. 1 and, as an example of the map quality, Fig. 2 shows the N-terminal glutamine. This residue has very clearly undergone a cyclization reaction to form pyroglutamate, a condensation product which commonly

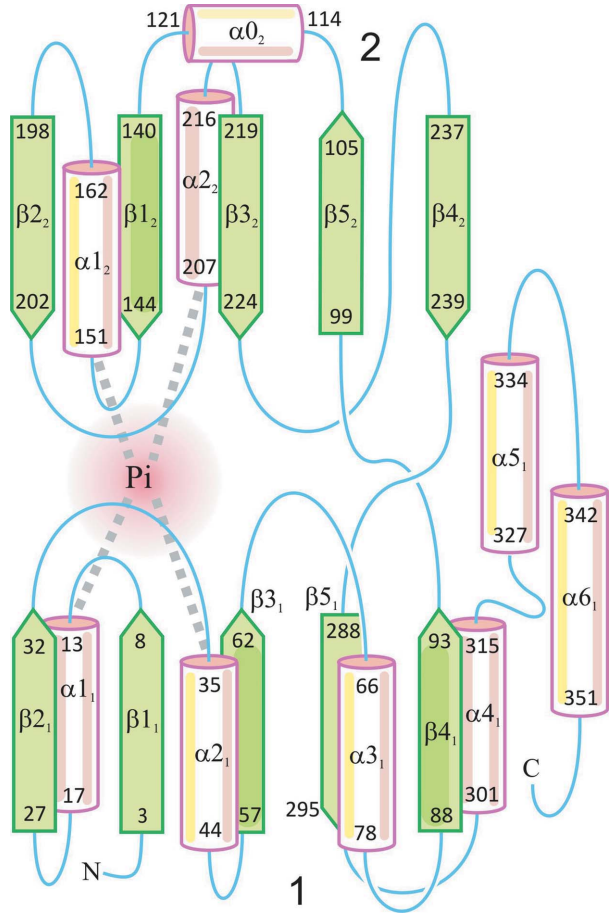


Figure 3
The topology of the *S. maltophilia* phosphate-binding protein. The numbers for residues forming each of the secondary-structure elements are shown at the termini of each strand and helix. The general position of the phosphate anion between domains 1 and 2 is indicated as 'Pi' and the helix dipoles which interact with it are shown as pale grey dashed lines. Note that the secondary-structure elements and loop regions are not drawn to scale and that some loops are particularly extended, for example the region between $\beta4_2$ and $\beta5_1$ is almost 50 residues in length.

occurs in proteins possessing N-terminal Glu or Gln residues. In this case, the N-terminal glutamine is generated by cleavage of the N-terminal signal peptide, which is 24 amino acids in length.

The topology is shown in Fig. 3, where the secondary-structure elements of this PBP superfamily member are numbered according to their positions within each of the domains. The general position of the phosphate anion bound between domains 1 and 2 is indicated and it can be seen that the N-termini of four α-helices ($\alpha1_1$, $\alpha2_1$, $\alpha1_2$ and $\alpha2_2$) are oriented towards it. The involvement of these helix dipoles is believed to be a major factor in the affinity of anion-binding PBPs for their ligands (He & Quiocho, 1993).

Given that only around 200 amino acids are required to form the essential type II PBP fold (shown in Fig. 3), the fact that the *S. maltophilia* protein is almost twice this length emphasizes that it has a number of substantial elaborations on the core structure shown. These mainly occur in the form of very extended loops, such as the region connecting $\beta5_2$ to $\beta1_2$, which includes a section of α-helix from residues 114 to 121 ($\alpha0_2$). This helix contains the residue Cys118 which forms a disulfide bridge to Cys163 in helix $\alpha1_2$. Another very extended loop region is that which crosses over from domain 2 to domain 1 by connecting $\beta4_2$ to $\beta5_1$. In addition, there is a curious projecting β-hairpin formed by 12 residues that are inserted in between helices $\alpha4_1$ and $\alpha5_1$, and towards the C-terminal end of the protein there is an extra α-helix $\alpha6_1$ which runs antiparallel to $\alpha4_1$. The final 20 residues at the C-terminal end of the protein form a largely irregular but very well defined region which in three dimensions appears to slot between strands $\beta4_1$ and $\beta5_1$. Here, Cys366, which is in the C-terminal irregular region, forms a disulfide bridge to Cys296 that immediately follows $\beta5_1$.

In general, PBPs contain a region of α-helix in the connection between $\beta3_2$ and $\beta4_2$ which maintains the topological twofold symmetry relating domains 1 and 2. However, this helix in domain 2, which would be topologically equivalent to $\alpha3_1$ in domain 1, is not present in the *S. maltophilia* protein and is replaced by an irregular loop region. In the other loops, there are several additional short helical regions which have a mixture of 3_{10}-helical and α-helical character at positions 127–131, 164–169, 188–192, 250–257, 262–272 and 362–367. Although some of the secondary-structure elements shown in Fig. 3 are reduced to being 2–3 residues in length in both domains of the protein, it appears that domain 1 adheres more strongly to the archetypal type II PBP fold than domain 2 does.

A sequence alignment of the *S. maltophilia* phosphate-binding protein with two related molecules from different bacteria is shown in Fig. 4. The second sequence shown has approximately 51% sequence identity and is a putative secreted alkaline phosphatase from *P. aeruginosa* (LapA; Ball *et al.*, 2002). However, the structure of the *S. maltophilia* protein does not reveal any bound metal ions, which are usually essential for alkaline phosphatase activity, suggesting that the *P. aeruginosa* molecule is also a phosphate-binding protein. The *S. maltophilia* protein has 44% sequence identity to its counterpart in *P. fluorescens* (Elias *et al.*, 2012) but only

Figure 4
A sequence alignment of the *S. maltophilia* protein with two similar bacterial phosphate-binding proteins. The amino acids are coloured according to the following scheme: acidic, red; basic, pale blue; neutral polar, green; hydrophobic, purple; cysteine, yellow; the structurally important residues Gly, Ala and Pro, white. The residues which interact directly with the phosphate are boxed and the disulfide pairing of the four invariant cysteines is indicated by the symbols * and + below. The secondary-structure elements of the protein are labelled and displayed with a similar colour scheme to that in Figs. 1 and 3.

around 28% identity to the more distantly related *E. coli* protein (Wang *et al.*, 1997; not shown). At 321 residues, the *E. coli* protein is somewhat shorter than the PBPs from *S. maltophilia* and *P. fluorescens*, which consist of 373 and 371 amino acids, respectively. These two proteins also have significant homology to the cDNA sequence of a reported human phosphate-binding protein (DING), although the gene for this protein does not appear to be present in the human genome (Morales *et al.*, 2006; Berthier, 2013), possibly suggesting that it might instead be of bacterial origin, as corroborated by codon-usage analysis (Lewis & Crowther, 2005). PstS from *Clostridium perfringens* has recently been analysed (Gonzalez *et al.*, 2014) and at 246 amino acids it is significantly shorter than even the *E. coli* protein. Structures are available for PstS from a range of bacteria, including *P. aeruginosa* (Neznansky *et al.*, 2014; 299 residues), *Yersinia pestis* (Tanabe *et al.*, 2007; 321 residues), *Vibrio cholerae* (PDB entry 1twy; 253 residues; New York SGX Research Center for Structural Genomics, unpublished work), *Lactobacillus brevis* (PDB entry 4ecf; 262 residues; Joint Center for Structural Genomics, unpublished work), *Mycobacterium tuberculosis* (Ferraris *et al.*, 2014; 348 residues), *Borreliella burgdorferi* (Brautigam *et al.*, 2014; 265 residues) and *Streptococcus pneumoniae* (PDB entries 4exl and 4lat; 274 residues; Center for Structural Genomics of Infectious Diseases, unpublished work). Overall, the sequence of the *S. maltophilia* PstS-like protein aligns quite poorly with its counterparts in these other organisms, although the sequence identity rises significantly above 30% in local regions corresponding to the main secondary-structure elements.

The structure of the *S. maltophilia* PBP superimposes on the *P. fluorescens* protein with an r.m.s.d. of 1.3 Å for 335 structurally equivalent residues (Fig. 5). The notably high value of this r.m.s.d. in spite of the reasonably high sequence identity of these two proteins (44%) could in principle be a reflection of either domain movement or large local structural differences between these two ligand-bound PBPs. The two structures diverge significantly in the following regions of the *S. maltophilia* protein: 17–32, 50–55, 78–84, 165–179, 192–198, 203–205, 228–230, 248–264 and 353–355. In addition, the *P. fluorescens* protein has quite large insertions between residues 242 and 243 and between residues 277 and 278 of the *S. maltophilia* sequence. Conversely, the loop region between residues 317 and 326 of the *S. maltophilia* protein is absent in *P. fluorescens* PstS. Since domains 1 and 2 appears to be oriented almost identically in both structures, the local differences, which are apparent in Fig. 5, are likely to be the dominant factor contributing to the high r.m.s.d.

3.2. Phosphate-binding site

Very close to the N-terminal end of the protein, the loop linking strand $\beta1_1$ and helix $\alpha1_1$ plays an important role in binding the phosphate anion. The main-chain amide N atom and the side-chain hydroxyl group of Ser10 form hydrogen bonds to the O1 atom of the phosphate group. In addition, the main-chain amide of the next residue in the chain (Leu11) donates a hydrogen bond to the O4 atom of the phosphate. These residues lie at the N-terminal end of helix $\alpha1_1$, where the helix dipole most likely stabilizes the negative charge on the phosphate group. The next residue to interact with the phosphate is Ser35, the amide and hydroxyl groups of which both hydrogen-bond to the O3 atom of the ligand. This serine residue lies at the N-terminal end of helix $\alpha2_1$. Further along the chain, the carboxyl group of Asp62 appears to interact very strongly with O4 of the phosphate by forming a short hydrogen bond to it. This has a donor-to-acceptor atom distance of 2.5 Å, which is within the range of a low-barrier hydrogen bond (Cleland *et al.*, 1998), suggesting that the proton between these two groups is shared to a greater extent than in a normal hydrogen bond. Full-matrix inversion with *SHELX* (Sheldrick & Schneider, 1997) following refinement using all reflections indicates that the standard deviations of the bond lengths in the binding site are in the region of 0.01 Å, suggesting that the structure has been determined with high accuracy and that the shortness of this hydrogen bond is likely to be a real effect. Another unusual feature of this aspartate is that its side chain adopts a rather unfavourable conformation in which the $C^{\beta}-C^{\gamma}$ bond is almost fully eclipsed with the main-chain $N-C^{\alpha}$ bond, suggesting that a strain mechanism may be involved in phosphate binding and release.

The next residue of the polypeptide to interact with the phosphate is the guanidinium group of Arg145 in domain 2. This residue forms charge-assisted hydrogen bonds to O1 and

Figure 5
A superposition of the *S. maltophilia* phosphate-binding protein with the successful search model from *P. fluorescens* (PDB entry 2q9t). The molecules are in approximately the same orientation as in Fig. 1 and are coloured green for *S. maltophilia* and yellow for *P. fluorescens*, with the phosphate groups for both shown in ball-and-stick representation.

Table 2
Hydrogen bonds at the phosphate-binding site.

In all cases the refinement e.s.d. for the donor-to-acceptor atom distance is 0.01 Å.

Hydrogen-bonded atom pairs	Donor–acceptor distance (Å)
Ser10 N···P_i O1	2.81
Ser10 O^{γ}···P_i O1	2.69
Leu11 N···P_i O4	3.04
Ser35 N···P_i O3	2.76
Ser35 O^{γ}···P_i O3	2.69
Asp62 $O^{\delta1}$···P_i O4	2.51
Arg145 $N^{\eta1}$···P_i O2	2.83
Arg145 $N^{\eta2}$···P_i O1	2.81
Ser149 O^{γ}···P_i O2	2.71
Gly150 N···P_i O3	2.79
Thr151 N···P_i O2	3.03
Thr151 $O^{\gamma1}$···P_i O2	2.68

O2 of the phosphate. Arg145 lies in the loop linking strand $\beta1_2$ and the N-terminal end of helix $\alpha1_2$, which is the next region of the protein to interact intensively with the phosphate group. Here, various main-chain amide and side-chain hydroxyl groups of the tripeptide Ser-Gly-Thr (residues 149–151) form a total of four hydrogen bonds to O2 and O3 of the phosphate (for full details, see Table 2 and Fig. 6). In total, O2 accepts four hydrogen bonds from donor groups in the protein, O1 and O3 both accept three and O4 appears to form only two hydrogen bonds. Interestingly, one of these involves the only carboxyl side chain in the binding site, that of Asp62, which is likely to be deprotonated at the pH of crystallization, especially in such close proximity to the phosphate group which is likely to be in the dianionic form. Thus, it is probable that this aspartate accepts a hydrogen bond from the single proton of the bound HPO_4^{2-} and this appears to be a low-barrier hydrogen bond (see above), suggesting that the proton is shared equally by both groups rather than being localized preferentially on one of them. The short strong hydrogen bond presumably stabilizes the unfavourable side-chain conformation of Asp62. Interestingly, this residue is thought to be the basis for the specificity of PstS proteins for phosphate relative to sulfate anions. Owing to its lack of a proton at physiological pH, sulfate would be unable to hydrogen-bond to the aspartate side chain. Conversely, the specificity of sulfate-binding proteins is thought to stem from their lack of charged residues at the binding site and the fact that none of the groups contacting the ligand would be able to accept a hydrogen bond from a phosphate anion. The complete dehydration of the ligand upon binding is clearly important to achieve this sort of specificity (Quiocho & Ledvina, 1996)

The residues forming the phosphate-binding site are strongly conserved. For example, in the *E. coli* and *P. fluorescens* proteins all of the above hydrogen-bonding side chains are conserved, with the exception of Ser10, which is replaced by a functionally analogous threonine in the other two species. Interestingly, a glycine residue at the N-terminal end of helix $\alpha2_2$, Gly207, is replaced by larger, polar side chain in the other organisms. This helix has its dipole oriented towards the phosphate-binding site but is at a greater distance than the other helices mentioned above and is likely to play a less important role in ligand binding. Gly207 is replaced by asparagine in *E. coli* and serine in *P. fluorescens*, the polar side chains of which are oriented towards the phosphate, although neither interact with it by hydrogen bonding. A number of water molecules occupy the pocket between Gly207 and the phosphate; indeed, these are the water molecules which come closest to the otherwise fully dehydrated phosphate.

4. Discussion

We have determined the structure of a periplasmic phosphate-binding protein from *S. maltophilia* as a result of it crystallizing unexpectedly as an impurity following the expression and purification of a completely unrelated protein. Others have reported the unexpected crystallization of phosphate-binding proteins (Morales *et al.*, 2006; Diemer *et al.*, 2008; Gai *et al.*, 2013), which we suggest is most likely to arise from the contamination of expression, purification or crystallization media by antibiotic-resistant bacteria, or possibly also from accidental incorporation of DNA encoding a bacterial PBP into the expression vector during the gene-cloning stages.

In our structural studies, a PDB-wide molecular-replacement search followed by refinement at atomic resolution allowed the unambiguous identification of the crystallization impurity reported here as a PstS-like protein from *S. maltophilia*. The sequence determined from the atomic resolution X-ray data indicate that it originates from either the R551-3 or the RA8 strain of *S. maltophilia.* Interestingly, the R551-3 strain is the second most common endophytic (*i.e.* plant-endosymbiotic)

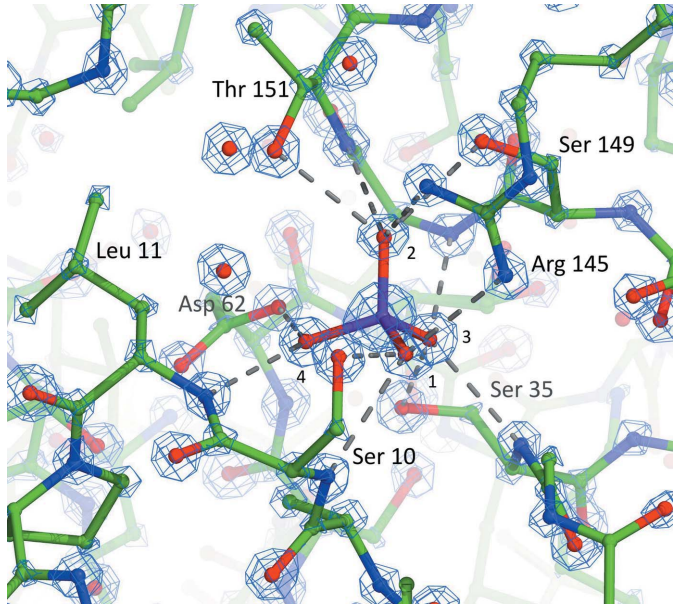


Figure 6
The phosphate-binding site. Residues forming hydrogen bonds to the phosphate anion are shown along with the $2F_o - F_c$ electron density contoured at 5.0 r.m.s. The P atom is coloured purple and the numbers of the phosphate O atoms, corresponding to the text, are shown. The phosphate hydrogen bonds are drawn as dashed lines. Note the unfavourable side-chain conformation of Asp62, which forms a low-barrier hydrogen bond to O4 of the phosphate. Further details of the phosphate hydrogen bonds are given in Table 2.

bacterium in the poplar tree *Populus trichocarpa* (Ryan *et al.*, 2009), suggesting that our structural studies may have agronomic application. In contrast, the RA8 strain was isolated from a sewage-treatment works in Karlsruhe (Adamek *et al.*, 2014). However, since the protein from these isolates has in excess of 90% sequence identity to those produced by human pathogenic strains, our structure is an excellent model of the protein from disease-causing strains of *S. maltophilia*.

Several studies have suggested that phosphate-binding proteins have an ultrastructural role in bacterial adhesion and biofilm stability as well as in phosphate scavenging (see, for example, Zaborina *et al.*, 2008). Accordingly, at the genetic level, the protein we have analysed has suggestive differences from the archetypal PstS protein, the gene for which resides in an operon together with genes for the associated transmembrane transporter components. In contrast, the coding region for the phosphate-binding PBP reported here lies adjacent to the gene for another PBP with 47% identity and both are within a cluster of genes for components of a type II secretion system, as also reported for *P. aeruginosa* (Ball *et al.*, 2002).

Given that the closest homologue of known structure (*P. fluorescens*) only has 44% sequence identity, the high resolution of the X-ray data set that we have obtained from the *S. maltophilia* protein has no doubt aided greatly in the structure analysis. The 1.1 Å resolution data have also allowed the definition of unusual features such as the pyroglutamate residue which is formed by cyclization of the N-terminal glutamine and the presence of a low-barrier hydrogen bond between Asp62 and the bound phosphate. Overall, the protein adopts the type II PBP fold with many extensive elaborations on the core topology, which is conserved in many protein families. The role of these elaborations remains to be determined, but it is possible that they allow interactions with other bacterial proteins involved in transport and biofilm formation. The phosphate anion is bound in between the two domains of the protein and is fully dehydrated in the complex, which involves many conserved hydrogen bonds and dipole interactions at the N-termini of several helical segments. The structure corroborates previous suggestions that the aspartate and arginine residues which interact with the bound anion (Asp62 and Arg145) play key roles in allowing the protein to discriminate phosphate from sulfate.

Since the use of PstS as a vaccine has been explored at the protein and DNA levels (see, for example, Tanghe *et al.*, 1999; Vyas *et al.*, 2003), knowledge of the structure of *S. maltophilia* PstS-like protein, and in particular its exposed, irregular and most likely highly antigenic regions, may now potentially facilitate vaccine design targeting this pathogen. In addition, knowledge of the three-dimensional structure of this protein will assist in defining its possible role in virulence and biofilm formation by multidrug-resistant pathogens, potentially with therapeutic applications.

Acknowledgements

Proteomics analysis was undertaken by Ms Lucia Di Vagno and Dr Nigel Rendell at the UCL Centre of Amyloidosis and Acute Phase Proteins. We acknowledge the BBSRC, UK for past financial support (reference B18665) and the ESRF (Grenoble, France) for synchrotron beam time and travel support. LC's contribution was sponsored by the Scientific User Facilities Division, Office of Basic Energy Sciences, US Department of Energy.

References

Abdul-Salam, V. B., Paul, G. A., Ali, J. O., Gibbs, S. R., Rahman, D., Taylor, G. W., Wilkins, M. R. & Edwards, R. J. (2006). *Proteomics*, **6**, 2286–2294.
Adamek, M., Linke, B. & Schwartz, T. (2014). *Microb. Pathog.* **67–68**, 20–30.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Ahn, S., Moniot, S., Elias, M., Chabriere, E., Kim, D. & Scott, K. (2007). *FEBS Lett.* **581**, 3455–3460.
Ball, G., Durand, É., Lazdunski, A. & Filloux, A. (2002). *Mol. Microbiol.* **43**, 475–485.
Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Beaven, G., Bowyer, A., Erskine, P., Wood, S. P., McCoy, A., Coates, L., Keegan, R., Lebedev, A., Hopper, D. J., Kaderbhai, M. A. & Cooper, J. B. (2014). *Acta Cryst.* F**70**, 823–826.
Benini, S., Degrassi, G., Krastanova, I., Lamba, D. & Venturi, V. (2004). *Acta Cryst.* D**60**, 1346.
Berntsson, R. P.-A., Smits, S. H. J., Schmitt, L., Slotboom, D.-J. & Poolman, B. (2010). *FEBS Lett.* **584**, 2606–2617.
Berthier, F. (2013). *Cell. Mol. Life Sci.* **70**, 3045–3056.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2012). *Acta Cryst.* D**68**, 1622–1631.
Brautigam, C. A., Ouyang, Z., Deka, R. K. & Norgard, M. V. (2014). *Protein Sci.* **23**, 200–212.
Brooke, J. S. (2012). *Clin. Microbiol. Rev.* **25**, 2–41.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cleland, W. W., Frey, P. A. & Gerlt, J. A. (1998). *J. Biol. Chem.* **273**, 25529–25532.
Colovos, C., Cascio, D. & Yeates, T. O. (1998). *Structure*, **6**, 1329–1337.
Deredjian, A., Alliot, N., Blanchard, L., Brothier, E., Anane, M., Cambier, P., Jolivet, C., Khelil, M. N., Nazaret, S., Saby, N., Thioulouse, J. & Favre-Bonté, S. (2016). *Res. Microbiol.* **167**, 313–324.
Díaz, M., Esteban, A., Fernández-Abalos, J. M. & Santamaría, R. I. (2005). *Microbiology*, **151**, 2583–2592.
Diemer, H., Elias, M., Renault, F., Rochu, D., Contreras-Martel, C., Schaeffer, C., Van Dorsselaer, A. & Chabriere, E. (2008). *Proteins*, **71**, 1708–1720.
Eerde, A. van, Wolterink-van Loo, S., van der Oost, J. & Dijkstra, B. W. (2006). *Acta Cryst.* F**62**, 1061–1066.
Elias, M., Wellner, A., Goldin-Azulay, K., Chabriere, E., Vorholt, J. A., Erb, T. J. & Tawfik, D. S. (2012). *Nature (London)*, **491**, 134–137.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Felder, C. B., Graul, R. C., Lee, A. Y., Merkle, H.-P. & Sadee, W. (1999). *AAPS PharmSci*, **1**, 7–26.
Ferraris, D. M., Spallek, R., Oehlmann, W., Singh, M. & Rizzi, M. (2014). *Proteins*, **82**, 2268–2274.
Fischer, R. J., Oehmcke, S., Meyer, U., Mix, M., Schwarz, K., Fiedler, T. & Bahl, H. (2006). *J. Bacteriol.* **188**, 5469–5478.

Gai, Z., Nakamura, A., Tanaka, Y., Hirano, N., Tanaka, I. & Yao, M. (2013). *J. Synchrotron Rad.* **20**, 854–858.
Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. (2003). *Nucleic Acids Res.* **31**, 3784–3788.
Gonzalez, D., Richez, M., Bergonzi, C., Chabriere, E. & Elias, M. (2014). *Sci. Rep.* **4**, 6636.
Grøftehauge, M. K., Truan, D., Vasil, A., Denny, P. W., Vasil, M. L. & Pohl, E. (2015). *Int. J. Mol. Sci.* **16**, 15971–15984.
He, J. J. & Quiocho, F. A. (1993). *Protein Sci.* **2**, 1643–1647.
Hopper, D. J. & Kaderbhai, M. A. (1999). *Biochem. J.* **344**, 397–402.
Hu, G., Taylor, A. B., McAlister-Henn, L. & Hart, P. J. (2005). *Acta Cryst.* F**61**, 486–488.
Hu, G., Taylor, A. B., McAlister-Henn, L. & Hart, P. J. (2006). *Acta Cryst.* F**62**, 709.
Hu, G., Taylor, A. B., McAlister-Henn, L. & Hart, P. J. (2007). *Arch. Biochem. Biophys.* **461**, 66–75.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 133–144.
Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. (2016). *Nucleic Acids Res.* **44**, D457–D462.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Keegan, R., Lebedev, A., Erskine, P., Guo, J., Wood, S. P., Hopper, D. J., Rigby, S. E. J. & Cooper, J. B. (2014). *Acta Cryst.* D**70**, 2444–2454.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Lewenza, S., Falsafi, R. K., Winsor, G., Gooderham, W. J., McPhee, J. B., Brinkman, F. S. & Hancock, R. E. (2005). *Genome Res.* **15**, 583–589.
Lewis, A. P. & Crowther, D. (2005). *FEMS Microbiol. Lett.* **252**, 215–222.
Madhusudhan, K. T., McLaughlin, R., Komori, N. & Matsumoto, H. (2003). *J. Basic Microbiol.* **43**, 36–46.
Martineau, P., Saurin, W., Hofnung, M., Spurlino, J. C. & Quiocho, F. A. (1990). *Biochimie*, **72**, 397–402.
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.
Morales, R. *et al.* (2006). *Structure*, **14**, 601–609.
Mudrak, B. & Tamayo, R. (2012). *Infect. Immun.* **80**, 1794–1802.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Neznansky, A., Blus-Kadosh, I., Yerushalmi, G., Banin, E. & Opatowsky, Y. (2014). *FASEB J.* **28**, 5223–5233.
O'May, G. A., Jacobsen, S. M., Longwell, M., Stoodley, P., Mobley, H. L. T. & Shirtliff, M. E. (2009). *Microbiology*, **155**, 1523–1535.
Pflugrath, J. W. & Quiocho, F. A. (1985). *Nature (London)*, **314**, 257–260.
Psakis, G., Polaczek, J. & Essen, L.-O. (2009). *J. Struct. Biol.* **166**, 107–111.
Quiocho, F. A. & Ledvina, P. S. (1996). *Mol. Microbiol.* **20**, 17–25.
Ramraj, V., Evans, G., Diprose, J. M. & Esnouf, R. M. (2012). *Acta Cryst.* D**68**, 1697–1700.
Rodríguez, D. D., Grosse, C., Himmel, S., González, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M. & Usón, I. (2009). *Nature Methods*, **6**, 651–653.
Ryan, R. P., Monchy, S., Cardinale, M., Taghavi, S., Crossman, L., Avison, M. B., Berg, G., van der Lelie, D. & Dow, J. M. (2009). *Nature Rev. Microbiol.* **7**, 514–525.
Santos-Beneit, F. (2015). *Front. Microbiol.* **6**, 402.
Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. (1992). *Biochemistry*, **31**, 10657–10663.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Sheldrick, G. M. (2011). *International Tables for Crystallography*, Vol. *F*, 2nd ed., edited by E. Arnold, D. M. Himmel & M. G. Rossmann, pp. 529–533. Dordrecht: Kluwer Academic Publishers.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Stokes-Rees, I. & Sliz, P. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 21476–21481.
Tanabe, M., Mirza, O., Bertrand, T., Atkins, H. S., Titball, R. W., Iwata, S., Brown, K. A. & Byrne, B. (2007). *Acta Cryst.* D**63**, 1185–1193.
Tanghe, A., Lefèvre, P., Denis, O., D'Souza, S., Braibant, M., Lozes, E., Singh, M., Montgomery, D., Content, J. & Huygen, K. (1999). *J. Immunol.* **162**, 1113–1119.
UniProt Consortium (2015). *Nucleic Acids Res.* **43**, D204–D212.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (2003). *Structure*, **11**, 765–774.
Wakatsuki, S., Belrhali, H., Mitchell, E. P., Burmeister, W. P., McSweeney, S. M., Kahn, R., Bourgeois, D., Yao, M., Tomizaki, T. & Theveneau, P. (1998). *J. Synchrotron Rad.* **5**, 215–221.
Wang, Z., Luecke, H., Yao, N. & Quiocho, F. A. (1997). *Nature Struct. Biol.* **4**, 519–522.
Waterman, D. G., Winter, G., Gildea, R. J., Parkhurst, J. M., Brewster, A. S., Sauter, N. K. & Evans, G. (2016). *Acta Cryst.* D**72**, 558–575.
Waterman, D. G., Winter, G., Parkhurst, J. M., Fuentes-Montero, L., Hattne, J., Brewster, A. S., Sauter, N. K. & Evans, G. (2013). *CCP4 Newsl. Protein Crystallogr.* **49**, 16–19.
Webb, D. C., Rosenberg, H. & Cox, G. B. (1992). *J. Biol. Chem.* **267**, 24661–24668.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
Zaborina, O., Holbrook, C., Chen, Y., Long, J., Zaborin, A., Morozova, I., Fernandez, H., Wang, Y., Turner, J. R. & Alverdy, J. C. (2008). *PLoS Pathog.* **4**, e43.