



Detecting the nature and solving the crystal structure of a contaminant protein from an opportunistic pathogen

Riccardo Pederzoli,^a Delia Tarantino,^b Louise J. Gourlay,^b Antonio Chaves-Sanjuan^{b,c} and Martino Bolognesi^{b,c*}

Received 25 January 2020

Accepted 2 August 2020

Edited by M. J. Romao, Universidade Nova de Lisboa, Portugal

Keywords: contaminant proteins; cyanase hydratase; *Serratia*.

PDB reference: cyanase hydratase, 6tv0

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

^aHamburg Unit c/o DESY, European Molecular Biology Laboratory, Notkestrasse 85, 22603 Hamburg, Germany,

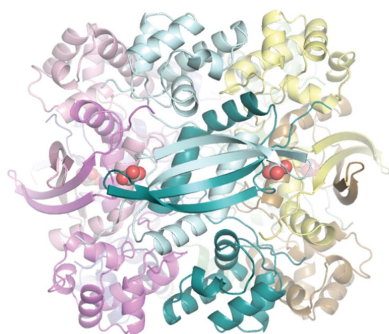
^bDepartment of Biosciences, University of Milano, Via Celoria 26, 20133 Milano, Italy, and ^cInvernizzi Pediatric Research Center, University of Milano, Via Celoria 26, 20133 Milano, Italy. *Correspondence e-mail: martino.bolognesi@unimi.it

The unintentional crystallization of contaminant proteins in the place of target recombinant proteins is sporadically reported, despite the availability of stringent expression/purification protocols and of software for the detection of contaminants. Typically, the contaminant protein originates from the expression organism (for example *Escherichia coli*), but in rare circumstances contaminants from different sources have been reported. Here, a case of contamination from a *Serratia* bacterial strain that occurred while attempting to crystallize an unrelated protein from *Burkholderia pseudomallei* (overexpressed in *E. coli*) is presented. The contamination led to the unintended crystallization and structure analysis of a cyanase hydratase from a bacterial strain of the *Serratia* genus, an opportunistic enterobacterium that grows under conditions similar to those of *E. coli* and that is found in a variety of habitats, including the laboratory environment. In this context, the procedures that were adopted to identify the contaminant based on crystallographic data only are presented and the crystal structure of *Serratia* spp. cyanase hydratase is briefly discussed.

1. Introduction

In biochemical and biotechnological contexts, it is often useful to validate the roles of specific amino-acid residues by analyzing the crystal structures of several mutants of a parent protein. Structure determination of the mutant protein is usually a straightforward task, and is usually tackled by molecular replacement (MR) using the wild-type (WT) protein as a search model. Occasionally, however, critical residue mutations, or the binding of specific ligands, can trigger unexpected structural rearrangements that can prevent the successful application of MR. It is also possible that a contaminant protein can crystallize in place of the targeted protein. In both cases, substantial efforts might be spent on unsuccessful MR phasing attempts before it is realized that either the conformation has significantly changed from that of the WT protein or that the analyzed crystal does not contain the intended target (Niedzialkowska *et al.*, 2016).

The case of a contaminant protein is likely to be worse since, from a crystallographic viewpoint, contamination may not be easy to spot, and such experimental events may be underestimated. Several factors contribute to this picture: (i) the search for structures with similar unit-cell parameters might not return any matches, (ii) the molecular weight of the target and of the contaminant proteins may be similar enough to prevent their discrimination by SDS-PAGE and (iii) powerful software for the detection of contaminants has only



recently become available (Ramraj *et al.*, 2012; Hungler *et al.*, 2016; Simpkin *et al.*, 2018).

Contamination is an unlikely event, and is even less likely to arise from an organism that differs from that used for overexpression; indeed, very few such cases have been reported in the literature (Musille & Ortlund, 2014; Butryn *et al.*, 2015). Here, we present a case of contamination from a *Serratia* strain that occurred while attempting to crystallize an engineered protein from *Burkholderia pseudomallei* that was overexpressed in *Escherichia coli*. We mostly focus on the strategy that was devised to solve the structure of the unsought protein, together with the steps that led to the identification of the contaminant and its bacterial source. Various hypotheses on the origins of the contamination and the factors that helped in the identification process are discussed.

2. Materials and methods

2.1. Protein purification and crystallization

Crystals of the (contaminant) *Serratia* cyanase hydratase were obtained during efforts to crystallize a mutant of the known antigen and peptidoglycan-associated lipoprotein BPSL2765 from *B. pseudomallei* (PDB entry 4b5c; Gourlay *et al.*, 2013), engineered to host a 20-residue epitope graft derived from a distinct antigen (17.1 kDa, hereafter referred to as BPSL2765^{MUT}; unpublished data). The engineered gene was cloned into the pET-14b (Novagen) bacterial expression vector and transformed into *E. coli* BL21 (DE3) pLysS competent cells (Invitrogen). Overexpression was performed in Superior Broth (AthenaES) supplemented with ampicillin (100 µg ml⁻¹) and chloramphenicol (34 µg ml⁻¹); protein expression was induced upon the addition of 0.5 mM isopropyl β-D-1-thiogalactoside (IPTG), incubating overnight at 20°C. The bacterial cells were harvested and suspended in IMAC buffer A (5 mM imidazole, 300 mM KCl, 50 mM KH₂PO₄ pH 7.4), mechanically lysed at 170 MPa with a cell disruptor (Constant Systems Ltd) and centrifuged for 1 h at 18 000g at 4°C; the supernatant was purified using a 5 ml Bio-Scale Mini Profinity IMAC Cartridge (Bio-Rad) before direct desalting into 10 mM Tris pH 8.0, 100 mM NaCl, 10% (v/v) glycerol using a 50 ml Bio-Scale Mini Bio-Gel P-6 Desalting Cartridge (Bio-Rad). Standard Bio-Rad Profinity protein-purification protocols were adopted. Overall, the yield was 8 mg of purified protein per litre of culture; the protein migrated as a single clear band on SDS-PAGE at the expected molecular weight of 17 kDa (data not shown).

The purified BPSL2765^{MUT} was concentrated to 30 mg ml⁻¹ for crystallization screening. Drops were assembled in sitting-drop vapor-diffusion wells using an Oryx4 crystallization robot (Douglas Instruments) and Greiner CrystalQuick 96-well plates (Greiner Bio-One). Different protein concentrations (30, 50 and 70%) were prepared by mixing with mother liquor, maintaining a final drop volume of 0.4 µl and a reservoir volume of 100 µl. After approximately two months, tiny crystals (<20 µm) grew at 297 K from condition A6 [15% (w/v)

Table 1

Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data collection	
Space group	<i>P</i> 12 ₁ 1
<i>a</i> , <i>b</i> , <i>c</i> (Å)	86.79, 93.14, 111.19
α , β , γ (°)	90.0, 95.7, 90.0
Wavelength (Å)	1.000
Resolution range (Å)	
Upper limit along reciprocal axes†	2.25, 2.23, 1.95 (2.17–1.95)
Lower limit	71.64
<i>R</i> _{p.i.m.} ‡	0.089 (0.560)
CC _{1/2} §	0.991 (0.614)
$\langle I/\sigma(I) \rangle$	5.6 (1.6)
Multiplicity	6.6 (6.4)
Completeness (%)	
Spherical	72.3 (13.9)
Ellipsoidal	94.2 (63.8)
Wilson <i>B</i> factor along reciprocal axes†	29.74, 28.94, 24.03
Refinement	
Resolution (Å)	64.99–1.95 (2.03–1.95)
No. of reflections	91801 (403)
<i>R</i> _{work} / <i>R</i> _{free}	0.176/0.220 (0.243/0.299)
No. of protein molecules in asymmetric unit	10
No. of protein residues	1550
No. of oxalate ions	5
No. of glycerol molecules	45
No. of water molecules	881
Average <i>B</i> factor (Å ²)	37.64
R.m.s.d., bond lengths (Å)	0.005
R.m.s.d., bond angles (°)	0.68
Ramachandran plot statistics	
Favored regions (%)	98.76
Outliers (%)	0.00

† Reciprocal axes = $0.733a^* + 0.680c^*$, b^* , $-0.514a^* + 0.858c^*$. ‡ Anisotropically truncated data. § CC_{1/2} is the correlation coefficient of the mean intensities between two random half sets of data.

polyethylene glycol 6000, 5% (v/v) glycerol] of JBScreen Classic 4 (Jena Biosciences).

2.2. Data collection and processing

A single crystal was soaked in a cryoprotectant solution [25% (v/v) glycerol, 15% (w/v) polyethylene glycol 6000] and flash-cooled in liquid nitrogen. X-ray diffraction data were collected at 100 K on the ID29 beamline at the ESRF, Grenoble, France (de Sanctis *et al.*, 2012) using a PILATUS 6M-F detector (Dectris, Zurich, Switzerland). A native data set was collected at a wavelength of 1.000 Å, indexed in the monoclinic space group *P*2₁ and anisotropically truncated at 1.95 Å resolution. Diffraction data were processed using *XDS* (Kabsch, 2010) and scaled and merged with *STARANISO* (Tickle *et al.*, 2018; the high-resolution data cutoff was based on the $\langle I/\sigma(I) \rangle = 1.5$ criterion (Evans, 2011). Data-collection statistics are reported in Table 1.

2.3. Structure solution and refinement

All attempts at structure solution by MR using a number of search models based on chain A of the WT BPSL2765 model (PDB entry 4b5c) were unsuccessful. A thorough search for contaminant proteins was undertaken as described below, which eventually led to the identification of the crystallized protein as a cyanate hydratase. MR with a cyanate hydratase

search model (PDB entry 4y42; Butryn *et al.*, 2015) yielded a clear solution, which was subjected to model building in *ARP/wARP* (Langer *et al.*, 2008), after which the contaminant protein was confirmed to originate from a bacterium of the *Serratia* genus. Cycles of refinement with *REFMAC5* (Murshudov *et al.*, 2011) and *phenix.refine* (with the application of NCS restraints; Adams *et al.*, 2010) and manual model building in *Coot* (Emsley *et al.*, 2010) completed the model refinement. The final structure contains one oxalate anion in each enzyme active site. Refinement statistics are reported in Table 1. The coordinates and structure factors for the final model (refined at 1.95 Å resolution; R_{work} and R_{free} of 0.176 and 0.220, respectively) have been deposited in the PDB with accession code 6tv0.

3. Results and discussion

3.1. Analysis of the unit cell and solvent content

Solvent-content calculation based on the Matthews coefficient (V_M ; Matthews, 1968, 1976) suggested the presence of between ten ($V_M = 3.01 \text{ Å}^3 \text{ Da}^{-1}$, ~59% solvent) and 15 ($V_M = 2.01 \text{ Å}^3 \text{ Da}^{-1}$, ~39% solvent) protomers of the expected BPSL2765^{MUT} in the crystal asymmetric unit. The self-rotation function (SRF) was computed using the native data with *MOLREP* (Vagin & Teplyakov, 2010) and *POLARRFN* (<https://services.mbi.ucla.edu/selfrot/>) for different values of the integration radius and data resolution. In all cases, the SRF showed similar features (Fig. 1), notably five twofold-axis peaks at $\kappa = 180^\circ$ and a single fivefold-axis peak at $\kappa = 72^\circ$. These features suggested a protein assembly with D_5 symmetry, *i.e.* a double pentameric structure with the two rings stacked on top of each other. This hypothesis was in keeping with the solvent-content analysis and was later used to plan and guide the phasing strategy.

3.2. Initial MR attempts

Since attempts to solve the sought BPSL2765^{MUT} structure by MR proved to be unsuccessful, an initial search for contaminants was performed by screening the entire PDB for structures with unit-cell parameters similar to those of the

collected data set; however, no hits were found. More sophisticated MR tests were then performed based on the oligomeric assembly deduced from the SRF and the solvent content (a summary flow chart for the automated pipeline is presented in Supplementary Fig. S1). A number of docking programs [*SAM* (Ritchie & Grudin, 2016), *HSYMDOCK* (Yan *et al.*, 2018), *Rosetta* symmetry docking (André *et al.*, 2007), *GalaxyWEB* (Ko *et al.*, 2012), *SymmDock* (Schneidman-Duhovny *et al.*, 2005) and *M-ZDOCK* (Pierce *et al.*, 2014)] were used to generate models with D_5 symmetry starting from chain A of the BPSL2765 search model (PDB entry 4b5c). The models (~2000) were then tested using an automated pipeline prepared as a Bash script with a Python wrapper; the pipeline employs *Phaser* (McCoy *et al.*, 2007) and *MOLREP* with default settings. The most promising solutions, as judged by the MR indicators (TFZ equivalent, LLG and packing for *Phaser* and Score, contrast, TF/sigma and wR_{fac} for *MOLREP*) and the R factors, were retained and used in the second step. The selected models were tested with a different automated pipeline that uses *Phaser* and *MOLREP* and varies some of the parameters that are known to be critical for the success of MR (data resolution and expected r.m.s.d. for *Phaser* and data resolution, similarity, completeness and number of rotation peaks for *MOLREP*). The most promising models (based on the same criteria as were used at the end of the first step) were retained and subjected to *REFMAC5* refinement, *SHELXE* expansion (Thorn & Sheldrick, 2013) and/or NCS averaging with density modification. Nevertheless, none of the MR solutions could be successfully refined or expanded, nor could the density be improved using any of the methods listed above, suggesting that the MR solutions selected were in fact false positives.

3.3. Contaminant search and source identification

A second, more thorough check for contaminants was then carried out using the recently developed program *SIMBAD* (Simpkin *et al.*, 2018). As expected from our preliminary search, the program could not find a solution during the first step (*i.e.* searching the PDB for unit-cell parameters), but *SIMBAD* identified cyanase from *S. proteamaculans* (PDB

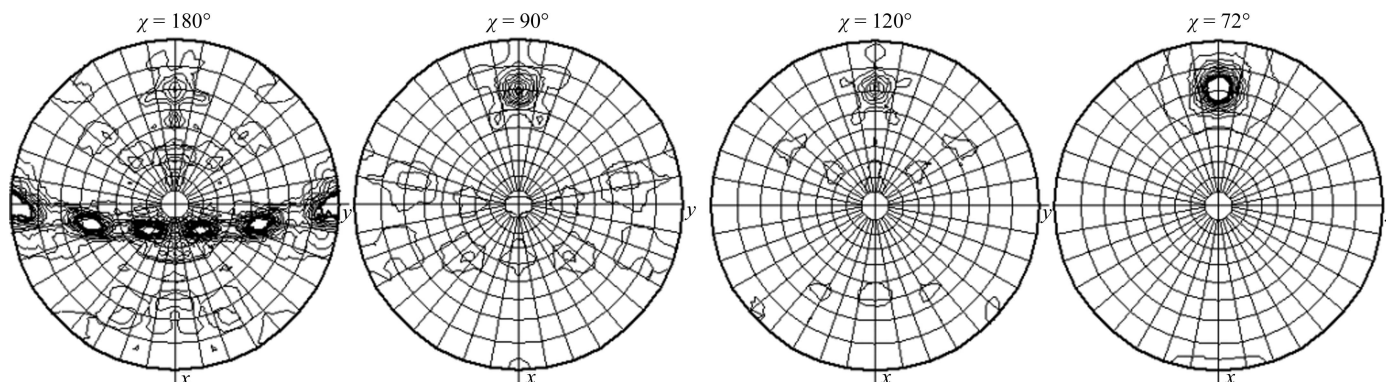


Figure 1

Self-rotation function computed with *MOLREP* for the native data set at different κ sections (radius of integration of 15 Å, resolution of 2.32 Å)

entry 4y42; Butryn *et al.*, 2015) as the best protein model matching the contaminant protein in the subsequent MR search. In fact, a solution was initially found with the *AMoRE* rotation function (Z -score = 54.6), which was confirmed by a full MR search with *MOLREP* (Score = 0.6798, TF-score = 19.16). Based on the *SIMBAD* result, a full MR search with *MOLREP* using PDB entry 4y42 as the search model was repeated independently, followed by *REFMAC5* restrained refinement and by one cycle of model building in *ARP/wARP*, which confirmed the crystallized protein to be a cyanate hydratase, likely from the Gram-negative bacterium *Serratia*. Following refinement, the r.m.s.d. value calculated for 771 matching C^α pairs in the refined structure and the search model (PDB entry 4y42) was 0.51 Å.

To further explore the origin of the contaminant, the following method was used: a main-chain-only model was built using *ARP/wARP* containing dummy atoms in the place of side chains. *PSI-BLAST* and *HMMER* (Finn *et al.*, 2015) were then used to generate a position-specific scoring matrix (PSSM) for the contaminant sequence. In this step, the contaminant sequence is searched against a database representing known protein families. During the first iteration, this procedure yielded a matching sequence from an *S. proteamaculans* strain (E -value of 3.8×10^{-41}). This sequence was used to build a full model in *ARP/wARP* starting from the main-chain-only model. Further cycles of refinement with *REFMAC5* and *phenix.refine* (applying NCS restraints) and model editing in *Coot* led to the reassignment of a few residues based on the optimal electron-density fit. Analysis of the sequence extracted from the final model confirms that the crystallized cyanase hydratase pertains to the *Serratia* genus, without identifying the exact strain.

3.4. Hypotheses about the source of contamination

Several hypotheses have been considered to explain the origin of the contamination, among which contamination at the protein-expression stage appears to be the most plausible. In fact, several antibiotic-resistant *Serratia* strains can grow in the presence of ampicillin (Stock *et al.*, 2003). Carbenicillin, which is normally used as a reliable alternative, was not employed during the fermentation stage, as contamination problems related to the use of ampicillin had not occurred in our past experience. In addition, *Serratia* grows under conditions similar to those for *E. coli* and can be found in the environment, in hospitals, in the urinary tract and even in the laboratory space, particularly in floor grooves. A combination of such factors (dated or degraded ampicillin stocks combined with ubiquitous antibiotic-resistant *Serratia* strains and a poorly sterile laboratory environment) are likely to be responsible for the contamination during the expression stage. With regard to the chance of growing crystals from a contaminated protein solution, it should be noted that a contaminant protein concentration of as low as 5% in the crystallization solution has been reported to yield crystals of an unintended target (Veesler *et al.*, 2008).

3.5. The *Serratia* spp. cyanase hydratase decamer

The *Serratia* spp. cyanase hydratase crystal structure was refined at 1.95 Å resolution with an R_{work} and R_{free} of 0.176 and 0.220, respectively. All individual chains in the crystallized homodecamer are virtually identical, with an excellent structural match of their C^α traces (maximum r.m.s.d. of 0.203 Å over 155 C^α pairs; Fig. 2*a*). The cyanase hydratase protomer structure very closely resembles those of cyanases deposited in the PDB, and consists of two domains: an N-terminal domain, which forms a five-helix bundle, and a C-terminal catalytic domain, which has a unique fold (Fig. 2*b*). Pairs of protomers are organized to form dimers through intricate contacts between two C-terminal domains, and five dimers assemble into a decamer with D_5 symmetry. The interfaces between dimers generate the five symmetrically disposed active sites of the enzyme, which are comprised of residues from four adjacent chains (*i.e.* two contacting dimers) in the decamer; here, the residues forming the catalytic triad (Arg96, Glu99, Ser122 and their NCS-related equivalents at the interface of adjacent dimers) are responsible for substrate binding (Fig. 2*c*).

In our case, clear density for an active-site ligand molecule could be observed. In fact, after a few refinement cycles, additional positive and symmetrical electron density appeared in the $2F_o - F_c$ and $F_o - F_c$ electron-density maps in all five active sites. Further refinement improved the density and, based on its characteristic shape and in agreement with active-site ligands found in the cyanase structure with PDB code 1dwk (Walsh *et al.*, 2000), suggested fitting the ligand as an oxalate anion. The presence of an oxalate ion in the five active sites was confirmed by calculating a polder map (Liebschner *et al.*, 2017; Fig. 2*c*). Oxalate, together with other low-molecular-weight dicarboxylic acids and monoanions, is a known inhibitor of *E. coli* cyanase (PDB entry 1dwk; Walsh *et al.*, 2000) and can be found in the culture medium, possibly as a product of bacterial metabolism. Beside the oxalate ions, the structure also contains a number of glycerol molecules from the crystallization drop and from the cryoprotectant solution.

4. Conclusions

We describe here the chance crystallization of a cyanase hydratase from an unknown strain belonging to the *Serratia* genus in complex with a natural inhibitor and in a novel crystal form, a case that adds to two previous reports (Musille & Ortlund, 2014; Butryn *et al.*, 2015) of unexpected *Serratia* contamination. Our results suggest that contamination from microorganisms other than those used for recombinant protein overexpression is a possibility that may be more likely than expected.

Some lessons can be learnt from this and previous contamination reports. Laboratory practice, with proper and regular cleaning of the laboratory space, including instrumentation, is of primary significance and is probably the most important way to reduce any unintended contamination. This adds to the necessary checks (SDS-PAGE, mass spectrometry and chromatographic analyses in particular) that must be

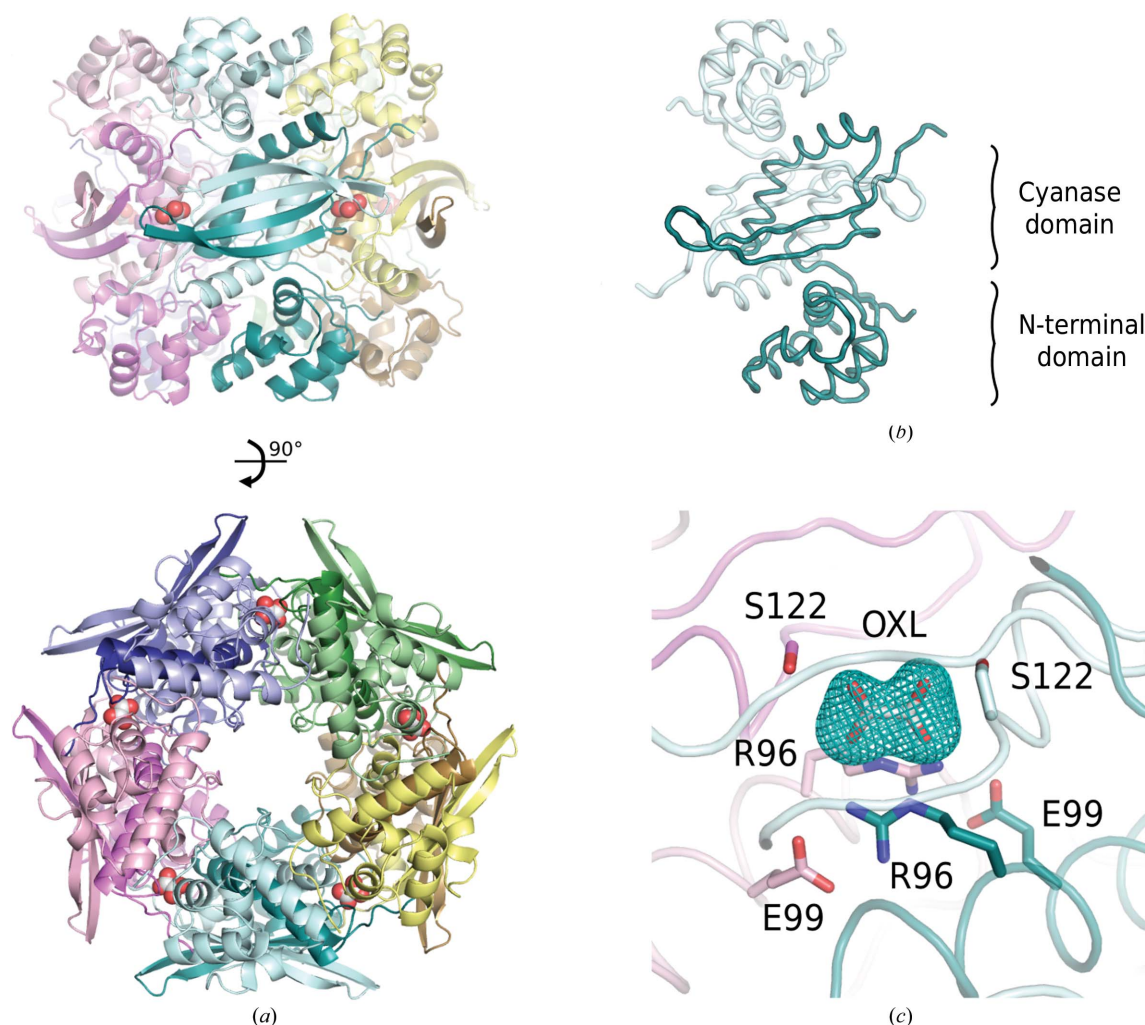


Figure 2

Structure of *Serratia* spp. cyanate hydratase. (a) Lateral and top views of the homodecameric assembly of the protein are shown in the upper and lower panels, respectively. Protein chains are represented in ribbon cartoon format, with the oxalate anions shown as spheres. (b) Cyanate hydratase dimer domain structure. The cyanase catalytic domain comprises elements from both protomers. (c) The active site of *Serratia* spp. cyanase at the interface of four protomers. The catalytic triad residues are represented as sticks and labeled. The oxalate ligand is shown as sticks together with its polder map as a mesh contoured at 3.0σ .

carried out during the expression and purification of the target protein. All such measures will probably not exclude the possibility of contamination, but will certainly decrease it more than any other practice. In the case presented here, various factors contributed to delaying the detection of the contaminant: on one hand, the negative results obtained from screening the PDB for structures with similar unit-cell parameters (this limitation has been recently overcome by the implementation of programs such as *SIMBAD* for the rapid screening of large databases of structures by MR), and on the other the false-positive MR results. Moreover, in the case of BPSL2765^{MUT} a contaminant was initially considered to be unlikely because the contaminants reported so far usually have a small size. Despite the fact that the molecular weights of known contaminants range from 10 to 140 kDa, most of them are in the range between 20 and 40 kDa, which also corresponds to the weight range containing 70% of all crystallized proteins reported in the PDB (Hungler *et al.*, 2016).

This observation confirms the rarity of our case, as the molecular weight of the *Serratia* cyanase hydratase is approximately 171 kDa. Moreover, because of the long time that was required for crystal growth, partial proteolysis of the BPSL2765^{MUT} protein was suspected and was assumed to be one of the factors complicating the MR search.

A check for contaminants should become an essential part of data analysis following data collection and processing, and immediately after initial unsuccessful attempts at structure solution. The simplicity and the rapidity of a contamination check should convince crystallographers to include this step in the routine process of data analysis, thus preventing time-consuming cases similar to that reported here. In parallel, crystallographers should be encouraged to report similar contamination cases and to deposit the contaminant structures, even when they are already known, thus aiding the identification of such contaminants by the crystallographic community.

Acknowledgements

This work was carried out as part of RP's PhD thesis in a collaborative agreement between the University of Milano (PhD School in Molecular and Cellular Biology) and The European Molecular Biology Laboratory (Hamburg Outstation). We are very grateful to the late Delia Tarantino for her valuable support in the initial stages of this study.

Funding information

The contribution from the Italian Ministry for Education, Universities and Research project 'Epitope grafting on bacterial vesicles to develop a novel *Burkholderia* vaccine (EGV)' (grant 2015JTL4HL) is acknowledged.

References

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* **D66**, 213–221.
- André, I., Bradley, P., Wang, C. & Baker, D. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 17656–17661.
- Butryn, A., Stoeck, G., Linke-Winnebeck, C. & Hopfner, K.-P. (2015). *Acta Cryst.* **F71**, 471–476.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. R. (2011). *Acta Cryst.* **D67**, 282–292.
- Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A. & Eddy, S. R. (2015). *Nucleic Acids Res.* **43**, W30–W38.
- Gourlay, L., Peri, C., Ferrer-Navarro, M., Conchillo-Solé, O., Gori, A., Rinchai, D., Thomas, R. J., Champion, O. L., Michell, S. L., Kewcharoenwong, C., Nithichanon, A., Lassaux, P., Perletti, L., Longhi, R., Lertmemongkolkhai, G., Titball, R. W., Daura, X., Colombo, G. & Bolognesi, M. (2013). *Chem. Biol.* **20**, 1147–1156.
- Hungler, A., Momin, A., Diederichs, K. & Arold, S. T. (2016). *J. Appl. Cryst.* **49**, 2252–2258.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Ko, J., Park, H., Heo, L. & Seok, C. (2012). *Nucleic Acids Res.* **40**, W294–W297.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nat. Protoc.* **3**, 1171–1179.
- Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. & Adams, P. D. (2017). *Acta Cryst.* **D73**, 148–157.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Matthews, B. W. (1976). *Annu. Rev. Phys. Chem.* **27**, 493–523.
- 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.
- 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.* **D67**, 355–367.
- Musille, P. & Ortlund, E. (2014). *Acta Cryst.* **F70**, 166–172.
- Niedzialkowska, E., Gasiorowska, O., Handing, K. B., Majorek, K. A., Porebski, P. J., Shabalin, I. G., Zasadzinska, E., Cymborowski, M. & Minor, W. (2016). *Protein Sci.* **25**, 720–733.
- Pierce, B. G., Wiehe, K., Hwang, H., Kim, B., Vreven, T. & Weng, Z. (2014). *Bioinformatics*, **30**, 1771–1773.
- Ramraj, V., Evans, G., Diprose, J. M. & Esnouf, R. M. (2012). *Acta Cryst.* **D68**, 1697–1700.
- Ritchie, D. W. & Grudin, S. (2016). *J. Appl. Cryst.* **49**, 158–167.
- Sanctis, D. de, Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., Mairs, T., McSweeney, S., Petitdemange, S., Rey-Bakaikoa, V., Surr, J., Theveneau, P., Leonard, G. A. & Mueller-Dieckmann, C. (2012). *J. Synchrotron Rad.* **19**, 455–461.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. (2005). *Nucleic Acids Res.* **33**, W363–W367.
- Simpkin, A. J., Simkovic, F., Thomas, J. M. H., Savko, M., Lebedev, A., Uski, V., Ballard, C., Wojdyr, M., Wu, R., Sanishvili, R., Xu, Y., Lisa, M.-N., Buschiazzi, A., Shepard, W., Rigden, D. J. & Keegan, R. M. (2018). *Acta Cryst.* **D74**, 595–605.
- Stock, I., Burak, S., Sherwood, K. J., Gröger, T. & Wiedemann, B. (2003). *J. Antimicrob. Chemother.* **51**, 865–885.
- Thorn, A. & Sheldrick, G. M. (2013). *Acta Cryst.* **D69**, 2251–2256.
- Tickle, I. J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C. & Bricogne, G. (2018). *STARANISO*. <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Veesler, D., Blangy, S., Cambillau, C. & Sciara, G. (2008). *Acta Cryst.* **F64**, 880–885.
- Walsh, M. A., Otwinowski, Z., Perrakis, A., Anderson, P. M. & Joachimiak, A. (2000). *Structure*, **8**, 505–514.
- Yan, Y., Tao, H. & Huang, S. (2018). *Nucleic Acids Res.* **46**, W423–W431.