Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Combining *in situ* proteolysis and mass spectrometry to crystallize *Escherichia coli* PgaB

Dustin J. Little,[a,b] John C. Whitney,[a,b] Howard Robinson,[c] Patrick Yip,[a] Mark Nitz[d] and P. Lynne Howell[a,b]*

[a]Program in Molecular Structure and Function, The Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada, [b]Department of Biochemistry, Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada, [c]Biology Department, Brookhaven National Laboratory, Upton, NY 11973-5000, USA, and [d]Department of Chemistry, Faculty of Arts and Sciences, University of Toronto, Toronto, ON M5S 3H6, Canada

Correspondence e-mail: howell@sickkids.ca

Received 3 April 2012
Accepted 16 May 2012

The periplasmic poly-β-1,6-*N*-acetyl-D-glucosamine (PNAG) de-*N*-acetylase PgaB from *Escherichia coli* was overexpressed and purified, but was recalcitrant to crystallization. Use of the *in situ* proteolysis technique produced crystals of PgaB, but these crystals could not be optimized for diffraction studies. By analyzing the initial crystal hits using SDS–PAGE and mass spectrometry, the boundaries of the protein species that crystallized were determined. The re-engineered protein target crystallized reproducibly without the addition of protease and with significantly increased crystal quality. Crystals of the selenomethionine-incorporated protein exhibited the symmetry of space group $P2_12_12_1$ and diffracted to 2.1 Å resolution.

1. Introduction

Bacteria growing in matrix-embedded biofilms cause device-related infections and are responsible for between 65 and 80% of all chronic infections (Potera, 1999). Bacterial biofilms represent a significant medical problem because once established they are difficult to eradicate, as the bacteria are protected from antibiotics, the environment and the innate immune system (Donlan & Costerton, 2002). Recent studies suggest that a key exopolysaccharide required for the structural development and integrity of the biofilm in *Escherichia coli* is poly-β-1,6-*N*-acetyl-D-glucosamine (PNAG; Wang *et al.*, 2004; Itoh *et al.*, 2008). PNAG production is dependent on the *pgaABCD* operon, which encodes four proteins, PgaA/B/C/D, responsible for the synthesis, modification and export of the polymer (Itoh *et al.*, 2008). PgaB is an ~77 kDa outer membrane lipoprotein that contains an N-terminal domain responsible for PNAG de-*N*-acetylation and a C-terminal domain of unknown function (Itoh *et al.*, 2008). Currently, there is no structural information available for PgaB or for related PNAG de-*N*-acetylases, thus prompting us to initiate structural studies to understand the molecular basis of PNAG de-*N*-acetylation and its requirement for polymer export and subsequent biofilm formation.

Here, we describe the crystallization and preliminary X-ray characterization of PgaB, which required *in situ* proteolysis for crystallization. *In situ* proteolysis, which involves adding a trace amount of protease to the protein solution during the crystallization experiment to cleave region(s) of the protein that may be inhibiting crystal formation, has been gaining prominence as a means of rescuing stalled crystallization projects (Dong *et al.*, 2007; Wernimont & Edwards, 2009). However, the *in situ* proteolysis crystallization conditions obtained for PgaB could not be used to reproducibly grow diffraction-quality crystals, which prevented us from obtaining phase information using selenomethionine (SeMet)-incorporated protein. We show how mass spectrometry of PgaB crystals generated by *in situ* proteolysis allowed us to determine the boundaries of the protein species that crystallized and how re-engineering the protein construct eliminated the need for protease during the crystallization process. This dramatically increased crystal quality, size and reproducibility, and allowed phase determination using the single-wavelength anomalous dispersion (SAD) technique (Hendrickson, 1991). The protocol used and described here could have widespread application given its potential to increase the success rates for stalled *in situ*

proteolysis crystallization projects and allow the continuation of structural studies on the target protein.

2. Materials and methods

2.1. Cloning and protein expression

The plasmid pCR*pgaB* (Itoh *et al.*, 2008) was used as the template and *pgaB*-specific primers were designed to include residues 22–672 and 42–672 for cloning into the pET28a expression vector (Novagen) using forward primers that contained an *Nde*I site (5′-GGG**CATATG**ATTAGCCAGTCAAGA-3′ and 5′-GGG**CATATG**CAACCGTGGCCGCAT-3′, respectively) and a reverse primer that contained a *Xho*I site (5′-GG**CTCGAG**TTAATCATTTTTCGGATA-3′), resulting in the expression plasmids pET28-PgaB$_{22–672}$ and pET28-PgaB$_{42–672}$. pET28-PgaB$_{42–672}$ was used to generate the expression plasmid pET28-PgaB$_{42–655}$ by changing Ile656 to a stop codon with the QuikChange Site-Directed Mutagenesis Kit (Stratagene) using 5′-GCATAACCAACCTGAATGAGACCTTATTCGTCCTG-3′ and 5′-CAGGACGAATAAGGTCTCATTCAGGTTGGTTATGC-3′ as the forward and reverse primers, respectively. The fidelity of the sequences was verified in all cases (ACGT Inc., Toronto, Canada).

E. coli BL21 (DE3) cells transformed with the appropriate expression plasmid were grown in 2 l Luria–Bertani (LB) broth containing 50 µg ml^{-1} kanamycin at 310 K until the OD$_{600}$ of the cell culture reached 0.5–0.6, at which point protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1.0 m*M*. The cells were incubated post-induction for an additional 18 h at 291 K before being harvested by centrifugation at 5000*g* for 20 min. All protein constructs used in this study were expressed using this protocol except for the SeMet-incorporated protein, which was expressed as per the protocol of Lee *et al.* (2001) using B834 Met$^-$ *E. coli* cells (Novagen).

2.2. Protein purification

The following protocol was used to purify all proteins used in this study. Cell pellets were resuspended in 50 ml lysis buffer [50 m*M* HEPES–NaOH pH 8.0, 500 m*M* NaCl, 5%(*v*/*v*) glycerol and one protease tablet (Sigma)] and the cells were disrupted by three passes through an Emulsiflex C3 at 103 MPa (Avestin Inc.). Insoluble cellular debris was separated by centrifugation for 30 min at 31 000*g*. The supernatant was applied onto a 5 ml Ni–NTA Superflow cartridge (Qiagen) pre-equilibrated with buffer *A* [20 m*M* HEPES–NaOH pH 8.0, 300 m*M* NaCl, 10 m*M* imidazole, 5%(*v*/*v*) glycerol]. The column was washed with ten column volumes of buffer *A* and the bound protein was eluted using a linear 10–250 m*M* gradient of imidazole in buffer *A*. The eluted fractions (5–10 ml) were pooled and dialyzed against 2 l buffer *B* [20 m*M* HEPES pH 8.0, 150 m*M* NaCl, 5%(*v*/*v*) glycerol] for 16 h at 277 K. The hexahistidine tag was removed by incubating the protein at 298 K for 3 h with one unit of thrombin (Novagen) per 4 mg protein. Untagged protein was separated from tagged protein by purification on a 5 ml Ni–NTA Superflow cartridge (Qiagen) pre-equilibrated with buffer *A* supplemented with 20 m*M* imidazole. The untagged protein was collected, further purified and buffer-exchanged into buffer *B* by size-exclusion chromatography using a HiLoad 16/60 Superdex 200 prep-grade gel-filtration column (GE Healthcare). The purity of each of the protein samples was judged to be >95% by SDS–PAGE and the protein could be concentrated to 8–10 mg ml^{-1} and stored at 277 K for one month without precipitation or degradation.

2.3. Limited proteolysis, *in situ* proteolysis and mass spectrometry

Limited proteolysis of PgaB$_{22–672}$ was performed using trypsin and chymotrypsin (Sigma). PgaB$_{22–672}$ (1 mg ml^{-1}) was incubated at protease:protein ratios of 1:100, 1:500, 1:1000 and 1:5000(*w*/*w*) for 24 h at 310 K. Samples were taken periodically and analyzed by SDS–PAGE. The stable proteolytic fragments were further analyzed using in-gel tryptic digestion matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS; Advance Protein Technology Center, The Hospital for Sick Children) and the resulting spectra were viewed using *Scaffold* 2 (Proteome Software, Inc.).

The Proti-Ace kit (Hampton Research) was used as per the manufacturer's instructions to determine appropriate protease:protein ratios for initial *in situ* proteolysis crystallization trials to produce a stable core or minimal proteolysis. The ratios (*w*/*w*) of proteases to protein used for the *in situ* proteolysis crystallization trials were as follows: trypsin, 1:5000; chymotrypsin, 1:5000; subtilisin, 1:5000; elastase, 1:1000; papain, 1:1000; endoproteinase Glu-C, 1:100.

To determine the protein species that crystallized in the PgaB$_{42–672}$ *in situ* proteolysis crystallization experiments, approximately 20 crystals were transferred to a 4 µl drop of reservoir solution. The crystals were repeatedly washed by successively transferring them into fresh 4 µl drops a total of three times, once in well solution and twice in water. The final drop was transferred to a 500 µl Eppendorf tube, diluted to 15 µl with water and maintained at 277 K for 1 d to dissolve the crystals. 5 µl of the solution was analyzed by SDS–PAGE and the remaining 10 µl was analyzed using electrospray ionization mass spectrometry (ESI-MS; Advance Protein Technology Center, The Hospital for Sick Children).

2.4. Crystallization

Initial crystallization trials were performed with 8–10 mg ml^{-1} PgaB$_{22–672}$ and PgaB$_{42–672}$ using a Mosquito robot with 96-well Art Robbins Instruments Intelli-Plates (Hampton Research) and commercially available screens from Hampton Research (Crystal Screen and Crystal Screen 2) and Emerald BioSystems (Wizards I–IV). Protein (400 nl) was mixed with precipitant in a 1:1 ratio and equilibrated against 100 µl precipitant at 293 K. For the PgaB$_{42–672}$ *in situ* proteolysis trials, the Proti-Ace kit protease stocks at 1 mg ml^{-1} were used. 10, 1 or 0.2 µl protease was added to 100 µl of 10 mg ml^{-1} protein to give approximate protease:protein ratios of 1:100(*w*/*w*), 1:1000(*w*/*w*) or 1:5000(*w*/*w*), respectively. The protease was added to the protein solution and mixed by inverting just prior to crystallization setup, which was performed as described above for PgaB$_{22–672}$.

Optimal PgaB$_{42–672}$ *in situ* proteolysis crystals were grown in 48-well VDX plates (Hampton Research) using streak-seeding. 1.5 µl protein solution [8 mg ml^{-1} with 1:50(*w*/*w*) endoproteinase Glu-C] was mixed with an equal volume of precipitant [14%(*w*/*v*) polyethylene glycol (PEG) 8000, 0.2 *M* calcium acetate, 0.1 *M* 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 5.8] and equilibrated against 200 µl precipitant at 289 K. Thin rod-like crystals grew to maximum dimensions of 100 × 50 × 10 µm after three weeks.

SeMet-incorporated PgaB$_{42–655}$ was crystallized in 48-well VDX plates (Hampton Research) using streak-seeding. 1.5 µl protein solution (15 mg ml^{-1}) was mixed with an equal volume of precipitant [14–16%(*w*/*v*) PEG 8000, 0.2 *M* calcium acetate, 0.1 *M* MES pH 6.0] and equilibrated against 200 µl precipitant at 293 K. Large rod-shaped crystals grew to maximum dimensions of 500 × 100 × 75 µm after one week.

Table 1
Data-collection statistics for PgaB crystals.

Values in parentheses are for the highest resolution shell.

Crystal	Native $PgaB_{42–672}$, *in situ* proteolysis (Fig. 2*d*)	SeMet $PgaB_{42–655}$ (Fig. 2*e*)
Wavelength (Å)	1.08	0.98
Temperature (K)	100	100
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 91.0, b = 102.9, c = 151.9, $\alpha = \beta = \gamma = 90$	a = 91.1, b = 102.4, c = 150.9, $\alpha = \beta = \gamma = 90$
Resolution (Å)	50.0–2.50 (2.59–2.50)	50.0–2.10 (2.18–2.10)
No. of reflections	713318	1108638
No. of unique reflections	50019	82493
Multiplicity	14.3 (14.3)	13.5 (12.6)
Completeness (%)	100.0	99.9 (99.3)
Average $I/\sigma(I)$	10.0 (2.6)	7.50 (2.8)
R_{merge}† (%)	10.8 (54.3)	11.7 (59.7)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl)\rangle$ represent the diffraction-intensity values of the individual measurements and the corresponding mean values, respectively. The summation is over all unique measurements.

2.5. Data collection

$PgaB_{42–672}$ *in situ* proteolysis crystals were cryoprotected for 60 s in precipitant solution supplemented with 25%(*v*/*v*) glycerol prior to vitrification in liquid nitrogen. Diffraction data were collected at 100 K on beamline X29, National Synchrotron Light Source (NSLS). A 0.16 mm collimated beam was used to collect a total of 360 images of 1° oscillation on an ADSC Quantum 315 detector with a 250 mm crystal-to-detector distance and an exposure time of 0.5 s per image. Diffraction data for SeMet $PgaB_{42–655}$ crystals cryoprotected for 10 s in precipitant solution supplemented with 10%(*v*/*v*) glycerol and 10%(*v*/*v*) ethylene glycol prior to vitrification in liquid nitrogen were collected as described above for $PgaB_{42–672}$ except that a 270 mm crystal-to-detector distance was used. The data were integrated, reduced and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The data-collection statistics are summarized in Table 1.

3. Results and discussion

Wernimont and Edwards have demonstrated the efficacy of *in situ* proteolysis and its ability to rescue stalled crystallization projects (Wernimont & Edwards, 2009). Presented here is the protocol that we used to overcome two pitfalls encountered during *in situ* proteolysis crystallization: specifically, an approach to take when crystals obtained using *in situ* proteolysis cannot be optimized for diffraction studies or reproducibility problems affect the ability to obtain phase information for structure determination (Fig. 1).

Initial studies using the mature form of PgaB lacking its predicted signal sequence and putative lipidation site revealed that $PgaB_{22–672}$ was recalcitrant to crystallization. A stable core encompassing residues 42–672, $PgaB_{42–672}$, was identified using limited proteolysis, secondary-structure prediction and sequence-conservation analysis. $PgaB_{42–672}$ behaved similarly in solution to $PgaB_{22–672}$. This protein produced phase separation in a number of crystallization conditions (Fig. 2*a*), but these conditions could not be optimized to produce crystals. Therefore, *in situ* proteolysis crystallization trials were conducted after first pre-screening $PgaB_{42–672}$ with the Proti-Ace kit to determine appropriate protease:protein ratios for the trials. Overall, six different proteases were screened using six commercially available sparse-matrix screens. Clusters of poor-quality crystals (Fig. 2*b*) formed in two conditions, Crystal Screen condition No. 46 and Wizard II condition No. 28, when the protein drop was supplemented with endoproteinase Glu-C as the protease. These crystals were reproducible; however, the crystals could not be optimized into a singular form suitable for diffraction studies. Numerous optimization techniques were carried out, including varying the protease concentration, temperature, precipitant concentration, salt concentration and buffer pH and the use of additives and seeding methods. After extensive optimization, we serendipitously obtained one three-dimensional crystal that diffracted to 2.5 Å resolution, but crystal growth in this condition was not reproducible (Fig. 2*d* and Table 1). As no reasonable search model was available to solve the structure of PgaB using molecular replacement, phase information was still required for structure determination. SeMet $PgaB_{42–672}$ was prepared but failed to produce crystals using *in situ* proteolysis in sparse-matrix or grid-optimized screens. Owing to crystal irreproducibility of the native protein, heavy-atom soaking methods did not present a reasonable approach for phasing, as a large number of high-quality crystals would be required to achieve successful derivatization. Thus, to obtain phase information using the SAD technique with SeMet incorporation, the endoproteinase Glu-C cut site on $PgaB_{42–672}$ was determined with the hypothesis that this would eliminate the need to use *in situ* proteolysis for crystallization.

The crystals produced by *in situ* proteolysis were dissolved and analyzed by SDS–PAGE. A single species was present at ∼72 kDa, suggesting that cleavage was occurring at the N-terminus and/or the C-terminus. The dissolved crystals were then analyzed by ESI-MS. If SDS–PAGE had revealed multiple bands, suggesting that proteolysis was occurring internally, in-gel tryptic digest followed by LC/MS

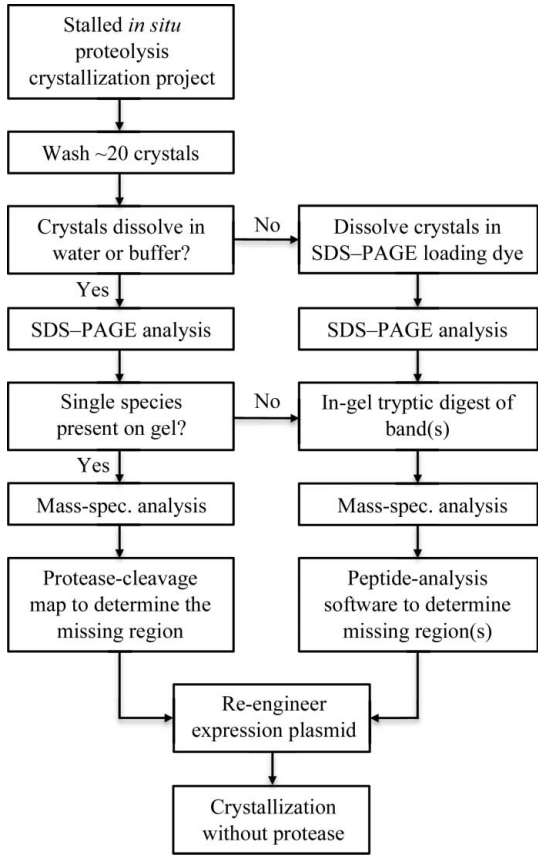

Figure 1
Workflow used to rescue stalled *in situ* proteolysis PgaB crystallization. *In situ* proteolysis crystals can be washed and analyzed by SDS–PAGE and mass spectrometry. The results can determine where the protease cleavage site(s) is located so that the expression plasmid can be re-engineered to produce recombinant protein for crystallization without the use of protease.

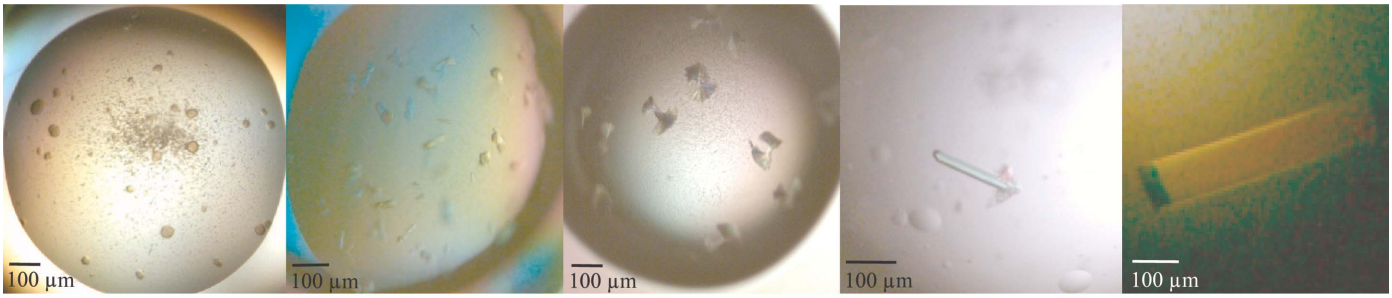

Figure 2
Progress of PgaB crystallization trials. (*a*) $PgaB_{42–672}$ produced phase separation in several crystallization conditions. (*b*) Initial $PgaB_{42–672}$ *in situ* proteolysis crystal hit using 1:100(*w*/*w*) endoproteinase Glu-C. (*c*) Typical $PgaB_{42–672}$ *in situ* proteolysis crystals produced during extensive optimization. (*d*) The only three-dimensional crystal obtained using *in situ* proteolysis, which diffracted to 2.5 Å resolution. (*e*) A crystal of SeMet $PgaB_{42–655}$ obtained after grid-optimized screening of condition No. 28 from Wizard II, which diffracted to 2.1 Å resolution.

peptide mass fingerprinting would have been used to determine the cleavage site(s). In this scenario, the internal region being proteolyzed would have been deleted in subsequent construct design. The in-gel tryptic digest LC/MS method would also have been used if the protein crystals had not dissolved in water or buffer conditions suitable for ESI-MS analysis. The mass-spectrometric results revealed that PgaB in the crystals had a mass of 71 120 Da. As $PgaB_{42–672}$ has a theoretical mass of 73 165 Da, this indicated that 2045 Da of PgaB had been removed during *in situ* proteolysis. Examination of the endoproteinase Glu-C cleavage map for PgaB predicted a proteolysis site after Glu655 that would correspond to a 2047 Da truncation. This suggested that $PgaB_{42–655}$ might be a more suitable construct for crystallization. The pET28-$PgaB_{42–672}$ vector was re-engineered to generate pET28-$PgaB_{42–655}$ and was used to express $PgaB_{42–655}$. Purified $PgaB_{42–655}$ crystallized readily in multiple sparse-matrix conditions: the two crystallization conditions previously identified and three new conditions (Crystal Screen condition No. 18, Wizard III condition No. 19 and Wizard IV condition No. 47), demonstrating that *in situ* proteolysis was no longer required for crystallization. SeMet $PgaB_{42–655}$ was prepared and produced crystals suitable for structure determination without requiring any protease using standard grid-optimized screens around condition No. 28 of Wizard II (Fig. 2*e*). The SeMet-incorporated crystal diffracted to 2.1 Å resolution and belonged to space group $P2_12_12_1$, with unit-cell parameters a = 91.1, b = 102.4, c = 150.9 Å, $\alpha = \beta = \gamma = 90^\circ$ (Table 1). The calculated solvent content with two molecules in the asymmetric unit is 50.3% (2.47 $Å^3$ Da^{-1}; Matthews, 1968). The Se-SAD data produced excellent quality electron-density maps; model building and refinement is currently in progress.

4. Conclusions

The crystallization of PgaB revealed two potential challenges that can occur when using *in situ* proteolysis, namely crystal irreproducibility and the associated difficulties in phase determination using *de novo* techniques. However, the method that we used to determine the protein species crystallizing allowed construct re-engineering and the generation of reproducible diffraction-quality crystals that were suitable for structure determination. Obtaining the ideal PgaB construct for crystallization not only helped to expedite the structure-determination process, thereby saving valuable time and resources, but will also allow further structural studies. The protocol outlined here (Fig. 1) should be generally applicable to other stalled crystallization projects where crystals obtained using the *in situ* proteolysis technique prove to be recalcitrant to optimization.

The authors would like to thank Dr Tony Romeo for the gift of the pCR*pgaB* plasmid and Drs Yura Lobsanov, Joel Weadge and Trevor Moraes for helpful discussions. This work was supported by research grants from the Canadian Institutes of Health Research (CIHR; Nos. 43998 and 259362 to PLH and MN, respectively). DJL and JCW have been supported by graduate scholarships from the University of Toronto and from the Natural Sciences and Engineering Research Council of Canada, Cystic Fibrosis Canada, the Ontario Graduate Scholarship Program and The Hospital for Sick Children Foundation Student Scholarship Program, respectively. PLH is the recipient of a Canada Research Chair. Beamline X29 at the National Synchrotron Light Source is supported by the United States Department of Energy Office of Biological and Environmental Research and the National Institutes of Health National Centre for Research Resources.

References

Dong, A. *et al.* (2007). *Nature Methods*, **4**, 1019–1021.
Donlan, R. M. & Costerton, J. W. (2002). *Clin. Microbiol. Rev.* **15**, 167–193.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Itoh, Y., Rice, J. D., Goller, C., Pannuri, A., Taylor, J., Meisner, J., Beveridge, T. J., Preston, J. F. III & Romeo, T. (2008). *J. Bacteriol.* **190**, 3670–3680.
Lee, J. E., Cornell, K. A., Riscoe, M. K. & Howell, P. L. (2001). *Structure*, **9**, 941–953.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Potera, C. (1999). *Science*, **283**, 1837–1839.
Wang, X., Preston, J. F. III & Romeo, T. (2004). *J. Bacteriol.* **186**, 2724–2734.
Wernimont, A. & Edwards, A. (2009). *PLoS One*, **4**, e5094.