STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

# Crystallization and biochemical characterization of an archaeal lectin from *Methanococcus voltae* A3

**N. Sivaji, K. V. Abhinav and M. Vijayan***

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. *Correspondence e-mail: mv@mbu.iisc.ernet.in

Received 23 March 2017
Accepted 24 April 2017

Edited by Z. Dauter, Argonne National Laboratory, USA

**Keywords:** archaeal lectins; $\beta$-prism I fold; X-ray studies; *Methanococcus voltae*.

A lectin from *Methanococcus voltae* A3 has been cloned, expressed, purified and characterized. The lectin appears to be specific for complex sugars. The protein crystallized in a tetragonal space group, with around 16 subunits in the asymmetric unit. Sequence comparisons indicate the lectin to have a $\beta$-prism I fold, with poor homology to lectins of known three-dimensional structure.

## 1. Introduction

Lectins are multivalent proteins of non-immune origin which specifically bind various carbohydrates (Weis & Drickamer, 1996; Loris, 2002; Sharon, 2007; Pérez *et al.*, 2015). They have received considerable attention, particularly in view of the importance of protein–carbohydrate interactions in biological recognition, especially on the cell surface. Lectins were first isolated from plants, with their best known property being the ability to agglutinate red blood cells. Subsequently, they have been found to occur in other forms of life, such as animals, bacteria and viruses (Hamblin & Kent, 1973; Chandra *et al.*, 2006). Lectins are known to be involved in symbiosis, cell–cell interactions and innate immunity, and have antifungal, antiviral, antiproliferative and mitogenic activities (Rini & Lobsanov, 1999; Feizi, 2000; Wong *et al.*, 2006; Ngai & Ng, 2007; Zhang *et al.*, 2009). They assume widely different folds and quaternary structures (Chandra *et al.*, 2006; Pérez *et al.*, 2015); the only feature that they have in common is the ability to specifically bind sugar structures. Plant and animal lectins have been studied extensively (Loris, 2002; Taylor & Drickamer, 2007; Abhinav & Vijayan, 2014), while work on microbial lectins has been substantial but much less extensive (Wiley *et al.*, 1981; Sixma *et al.*, 1991; Swaminathan & Eswaramoorthy, 2000). Efforts in our laboratory have primarily been on plant lectins (Banerjee *et al.*, 1994; Sankaranarayanan *et al.*, 1996; Vijayan & Chandra, 1999; Natchiar *et al.*, 2007; Chandran *et al.*, 2013; Abhinav & Vijayan, 2014; Abhinav *et al.*, 2017), and have recently been extended to include mycobacterial lectins (Abhinav *et al.*, 2013; Patra *et al.*, 2014).

One domain of life, namely the archaea, had remained a virgin area in terms of lectin research. Recently, we identified 46 putative lectin/lectin-like domains through a bioinformatics search of 165 archaeal genomes (Abhinav *et al.*, 2016). Criteria based on sequence and structure indicate that they are distributed among six structural families involving seven-bladed $\beta$-propeller, $\beta$-trefoil, legume lectin, C-type lectin, $\beta$-prism I and tachylectin folds. Among the six folds, the $\beta$-prism I fold was first identified as a lectin fold in our

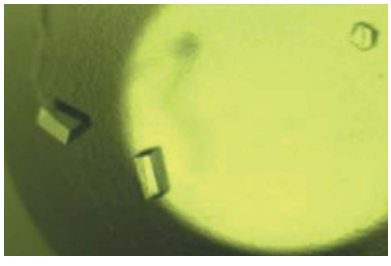

**Table 1**
Macromolecule-production information.

| | |
|---|---|
| Source organism | *M. voltae* A3 |
| DNA source | *M. voltae* A3 genomic DNA |
| Forward primer | 5′-CCCATATGATGGCACAAAATGACAACTATAT-ATACAGCACTGAG-3′ |
| Reverse primer | 5′-CCAAGCTTTTTTTTTAGGAATATTAAACCTA-ATCTATCAATATCAGACCCC-3′ |
| Cloning vector | pET-21b |
| Expression vector | pET-21b |
| Expression host | *E. coli* BL21 (DE3) |
| Complete amino-acid sequence of the construct produced† | MAQNDNYIYSTEVGGVGGTPFTFMQESGTITSIK-FNWSDQYKLLHHIEVKFINNANIYATGDPKGN-HEVILEIDDDETIIGSVIGYKKGNDGRCTGVK-LTTSKGKSIMAGYFEESLITTYTGKLAGIKGG-AGSDIDRLGLIFLKK*VEHHHHHH* |

† Extra amino acids owing to the vector used are in italics. Residues that are possibly involved in sugar binding are in underlined.

laboratory (Sankaranarayanan *et al.*, 1996). Since then, much of the reported work on $\beta$-prism fold lectins has been carried out here (Chandran *et al.*, 2013; Abhinav *et al.*, 2017). When entering a virgin area, it is prudent to choose a familiar system for initial study; hence the choice of a putative $\beta$-prism I fold archaeal lectin for the present investigation. Homologues of the well characterized $\beta$-prism I fold could be identified in *Methanococcus maripaludis* C7 and *M. voltae* A3. In the former, it exists as a domain in a two-domain protein. In the latter, a single gene (Mvol_0737) encodes the lectin, which is unaccompanied and uninfluenced by any other domain. We have now cloned, expressed, purified and crystallized Mvol_0737 (here after referred to as *Mevo* lectin). These experiments and the X-ray characterization of the crystals, as reported here, constitute the first successful investigation of its kind on an archaeal lectin. The identification of a number of lectins in archaea through a genomic search and the experimental study of one of them indicate that lectins exist in all three domains of life and probably evolved into functional proteins before the three domains diverged.

## 2. Materials and methods

### 2.1. Macromolecule production

The gene (Mvol_0737) encoding a 145-residue protein from *M. voltae* A3 was PCR-amplified using Phusion polymerase (NEB). The PCR product was then digested and cloned into pET-21b vector with restriction sites for NdeI and HindIII to obtain C-terminally His-tagged protein. Cloning was confirmed by primer-based sequencing (Table 1). The construct was transformed into *Escherichia coli* BL21(DE3) cells for overexpression. The transformed cells were grown in LB broth at 37°C until the $OD_{595}$ reached 0.6. The cultures were then induced with 1 m*M* isopropyl $\beta$-D-1-thiogalactopyranoside and incubated separately at 16°C for 16 h. The cells were subsequently harvested by centrifugation at 5000*g* for 20 min followed by resuspension in a buffer consisting of 30 m*M* Tris–HCl pH 7.4, 300 m*M* NaCl, 5 m*M* imidazole, 10%(*v*/*v*) glycerol. Cell lysis was performed by sonication. The cell lysate was spun at 14 000*g* for 30 min. Ni–NTA purification was performed and the fractions containing the desired protein were pooled, concentrated and dialysed into 30 m*M* Tris–HCl pH 7.4, 300 m*M* NaCl, 10 m*M* imidazole. The protein purity was confirmed using SDS–PAGE (Fig. 1*a*). The mass spectrum of the purified protein was also recorded using MALDI–TOF (Fig. 1*b*). The purified protein was concentrated to 7 mg ml$^{-1}$ for crystallization experiments.

**2.1.1. Haemagglutination assay**. The activity of the protein was confirmed by haemagglutination (Fig. 2*a*). The protein was serially diluted in PBS and 100 µl of the solution was added to a haemagglutination titre plate (U-bottom). 100 µl of 4% red blood cells (RBC; washed in PBS) was added to each well and incubated at room temperature for 2–3 h.

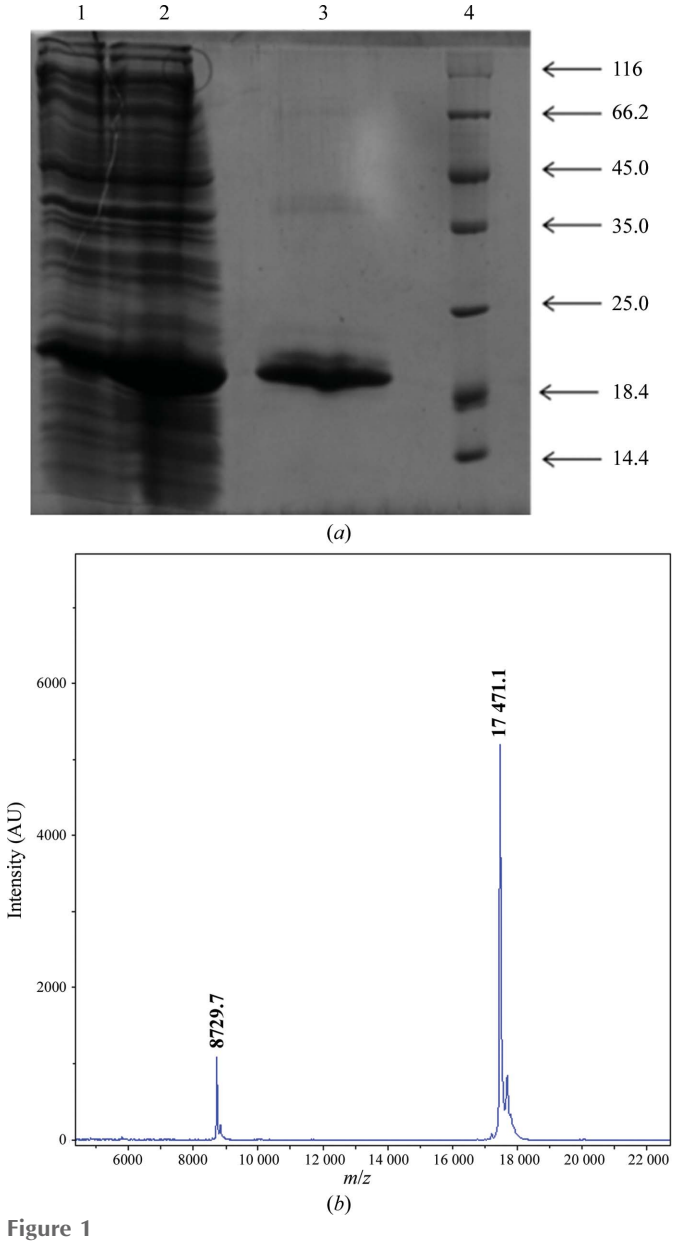


**Figure 1**
(*a*) SDS–PAGE profile of *Mevo* lectin. Lane 1, uninduced culture; lane 2, induced culture; lane 3, purified *Mevo* lectin; lane 4, molecular-mass marker (labelled in kDa). (*b*) Mass-spectroscopic (MALDI–TOF) analysis of *Mevo* lectin.

**Table 2**
Crystallization.

| | |
|---|---|
| Method | Microbatch under oil |
| Plate type | 72-well microbatch plates |
| Temperature (K) | 295 |
| Protein concentration (mg ml$^{-1}$) | 7 |
| Buffer composition of protein solution | 30 m*M* Tris–HCl pH 7.4, 300 m*M* NaCl, 10 m*M* imidazole, 10%(*v*/*v*) glycerol |
| Composition of reservoir solution | 6%(*v*/*v*) Tacsimate pH 6.0, 0.1 *M* MES monohydrate pH 6.0, 25%(*w*/*v*) polyethylene glycol 4000 |
| Volume and ratio of drop | 4 µl drop: 2 µl protein solution + 2 µl crystallization condition |

Haemagglutination was assessed visually by checking for cell clumping. The highest dilution of lectin causing visible agglutination was identified as a titre value corresponding to 5 mg ml$^{-1}$ protein concentration. The initial haemagglutination-inhibition study indicated that simple sugars do not inhibit the haemagglutination caused by the protein.

**2.1.2. Dynamic light scattering**. Dynamic light-scattering (DLS) experiments were performed on a SpectroSize 300 instrument (Molecular Dimensions). Protein samples were analysed at concentrations of 0.5, 1 and 7 mg ml$^{-1}$. 20 measurements were made on each sample. The data were analysed using the SpectroSize 300 software, resulting in average hydrodynamic size-distribution profiles (Fig. 2*b*).

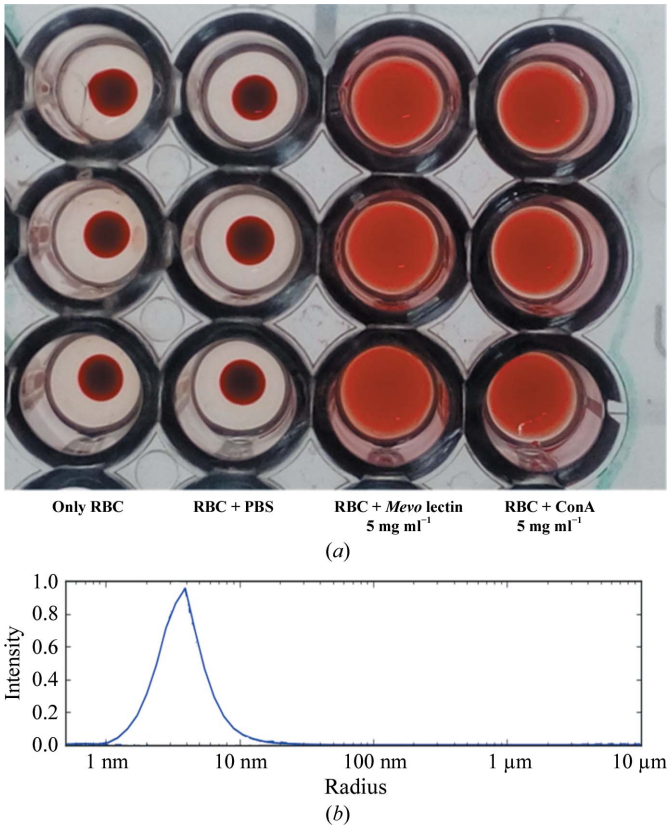


**Figure 2**
(*a*) Haemagglutination activity of *Mevo* lectin. (*b*) Scattered intensity *versus* hydrodynamic radius in DLS experiments at a protein concentration of 1 mg ml$^{-1}$.

**Table 3**
Data collection and processing.

Values in parentheses are for the outer shell.

| | |
|---|---|
| Diffraction source | Rigaku ultraX 18 Cu $K\alpha$ rotating anode |
| Wavelength (Å) | 1.5418 |
| Temperature (K) | 100 |
| Detector | MAR345 image-plate system |
| Crystal-to-detector distance (mm) | 200 |
| Rotation range per image (°) | 1 |
| Total rotation range (°) | 198 |
| Exposure time per image (s) | 360 |
| Space group | $P4_12_12$ or $P4_32_12$ |
| Unit-cell parameters (Å, °) | $a = b = 169.45$, $c = 192.23$, $\alpha = \beta = \gamma = 90$ |
| Mosaicity (°) | 0.30 |
| Resolution range (Å) | 96.12–2.50 (2.64–2.50) |
| Total No. of reflections | 1250638 (142201) |
| No. of unique reflections | 96842 (13946) |
| Completeness (%) | 100.0 (99.9) |
| Multiplicity | 12.9 (10.2) |
| $\langle I/\sigma(I)\rangle$ | 12.5 (2.5) |
| $R_{meas}$ | 0.145 (1.013) |
| $R_{merge}$ | 0.139 (0.963) |
| Overall *B* factor from Wilson plot (Å$^2$) | 31.8 |

Deconvolution of the autocorrelation function was performed using the *CONTIN* algorithm (Provencher, 1982).

### 2.2. Crystallization

Crystallization trials were carried out at 295 K by the microbatch-under-oil method in a 72-well plate, mixing 2 µl protein solution at a protein concentration of 7 mg ml$^{-1}$ in Tris buffer pH 7.4 containing 30 m*M* Tris, 300 m*M* NaCl, 10 m*M* imidazole and 10%(*v*/*v*) glycerol with 2 µl of commercially available crystallization solutions from Hampton Research. Crystals appeared in PEGRx 2 condition No. 29 consisting of 6%(*v*/*v*) Tacsimate pH 6.0, 0.1 *M* MES monohydrate pH 6.0, 25%(*w*/*v*) polyethylene glycol 4000 after 14 d (Fig. 3, Table 2).

### 2.3. Data collection and processing

The crystal diffracted to a resolution of 2.5 Å at the home source. The data were collected on an in-house MAR345 image plate mounted on a Rigaku ultraX 18 rotating-anode X-ray generator operating at 40 kV and 80 mA with a copper anode. The data were processed using *MOSFLM* (Leslie,

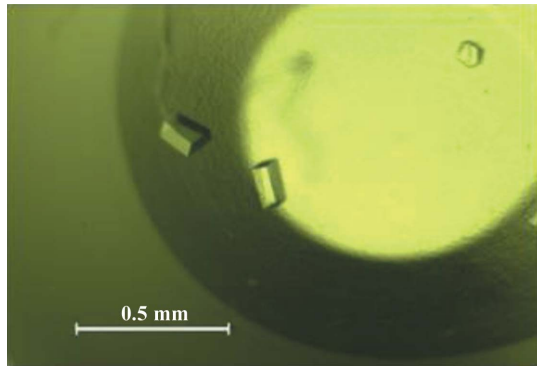


**Figure 3**
Crystals of *Mevo* lectin.

1992) and scaled using *SCALA* from the *CCP*4 program suite (Winn *et al.*, 2011). The crystal parameters and X-ray data-collection and processing statistics are summarized in Table 3.

## 3. Results and discussion

A lectin from *M. voltae* A3 (*Mevo* lectin) with a C-terminal His tag has been cloned, expressed and purified to homogeneity (Fig. 1*a*). Although the molecular weight of the expressed protein calculated from the sequence including the His tag is 16.81 kDa, the mass-spectrometric data indicated the molecular weight to be 17.47 kDa (Fig. 1*b*), possibly on account of post-translational modifications. The circular-dichroism spectrum indicates an abundance of $\beta$-strands (data not shown). The lectin agglutinates red blood cells (Fig. 2*a*), but agglutination is not inhibited by simple sugars, indicating that the lectin is specific for complex sugars. The protein crystallized (Fig. 3) in space group $P4_12_12$ or $P4_32_12$, with unit-cell parameters $a = b = 169.45$, $c = 192.23$ Å. The Matthews coefficient ($V_M = 2.46$ Å$^3$ Da$^{-1}$, corresponding to a solvent content of 50.1%; Matthews, 1968) and probability and solvent-content estimates (Kantardjieff & Rupp, 2003) suggest the presence of 16 or 17 subunits in the crystallographic assymetric unit. The number is likely to be 16 as lectins usually exist as oligomers.

DLS experiments indicate that the samples are substantially monodisperse, with the predominant species at all concentrations having a hydrodynamic radius (Fig. 2*b*) that would roughly correspond to the molecular weight of a tetramer. If *Mevo* lectin is a tetramer, as indicated by the results of the DLS experiments, the crystal would contain four crystallographically independent tetramers. There appeared to be no ambiguity in the space group and unit-cell parameters or in the quality of the intensity data. A twinning test performed using *TRUNCATE* from *CCP*4 did not indicate any twinning. None of the non-origin peaks in Patterson maps computed with data in the resolution ranges 10–5, 10–4 and 10–3 Å had heights greater than 20% of that of the origin peak. Thus, the Patterson maps do not indicate significant translational noncrystallographic symmetry in the crystal (Read *et al.*, 2013). The self-rotation function contains one prominent peak, indicating noncrystallographic twofold symmetry. It is not clear whether this represents a single noncrystallographic twofold axis or has resulted from combination of more than one axis. Assuming that each tetrameric molecule has 222 symmetry, the crystallographic asymmetric unit would contain a minimum of 12 noncrystallographic twofold symmetry elements. It is perhaps too ambitious to try to determine the orientations of each of them from the self-rotation function.

Among the plant lectins with known three-dimensional structures, tetrameric ipomoelin (PDB entry 3r50; Chang *et al.*, 2012) exhibits maximum sequence identity (22%) to *Mevo* lectin, indicating that the molecule probably has a $\beta$-prism I fold. A global search showed a slightly higher sequence identity (26%) of the lectin to the $\beta$-prism I fold lectin domain of a zebrafish aerolysin-like protein (PDB entry 4zno; Jia *et al.*, 2016). The structure-prediction software *I-TASSER* (Zhang, 2008) gave an acceptable model of the lectin subunit, which superposed on the lectin domain of the aerolysin-like protein and on the a subunit of ipomoelin with root-mean-square deviations in $C^\alpha$ positions of 1.7 and 2.0 Å, respectively. A similar result was obtained when *Robetta* (Kim *et al.*, 2004) was used for structure prediction. These results additionally support the assignment of the $\beta$-prism I fold to the archaeal lectin. The carbohydrate-binding sites of $\beta$-prism I fold lectins are characterized by a G . . . G*XXX*D motif. Lectins in which one, two or all three of the Greek keys in the fold carry a binding site are known (Sharma *et al.*, 2007; Chandran *et al.*, 2013). From careful sequence comparisons, it was suggested that *Mevo* lectin has only one binding site (Abhinav *et al.*, 2016) situated on the Greek key comprising the N- and C-terminal stretches of the polypeptide chain (Table 1). Repeated attempts to solve the structure by molecular replacement using various $\beta$-prism I lectin subunits and their oligomers as search models did not succeed, presumably on account of poor homology or inadequate structural similarity. While these attempts will be continued, efforts will also be made to solve the structure using *ab initio* methods such as those involving anomalous dispersion.

## Acknowledgements

The data sets used in the present work were collected at the facility for X-ray crystal structure determination and protein design at the Institute, supported by the Department of Science and Technology (DST). NS is a UGC senior research fellow. MV is Albert Einstein Professor of the Indian National Science Academy.

## Funding information

Funding for this research was provided by: Department of Science and Technology, Ministry of Science and Technology, Government of India.

## References

Abhinav, K. V., Samuel, E. & Vijayan, M. (2016). *Proteins*, **84**, 21–30.
Abhinav, K. V., Sharma, K., Surolia, A. & Vijayan, M. (2017). *IUBMB Life*, **69**, 72–78.
Abhinav, K. V., Sharma, A. & Vijayan, M. (2013). *Proteins*, **81**, 644–657.
Abhinav, K. V. & Vijayan, M. (2014). *Pure Appl. Chem.* **86**, 1335–1355.
Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A. & Vijayan, M. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 227–231.
Chandra, N. R., Kumar, N., Jeyakani, J., Singh, D. D., Gowda, S. B. & Prathima, M. N. (2006). *Glycobiology*, **16**, 938–946.
Chandran, T., Sharma, A. & Vijayan, M. (2013). *Adv. Protein Chem. Struct. Biol.* **92**, 135–178.
Chang, W., Liu, K., Hsu, F., Jeng, S. & Cheng, Y. (2012). *PLoS One*, **7**, e40618.
Feizi, T. (2000). *Immunol. Rev.* **173**, 79–88.
Hamblin, J. & Kent, S. P. (1973). *Nature New Biol.* **245**, 28–30.
Jia, N. *et al.* (2016). *EMBO Rep.* **17**, 235–248.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.

Kim, D. E., Chivian, D. & Baker, D. (2004). *Nucleic Acids Res.* **32**, W526–W531.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Loris, R. (2002). *Biochim. Biophys. Acta*, **1572**, 198–208.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Natchiar, S. K., Suguna, K., Surolia, A. & Vijayan, M. (2007). *Crystallogr. Rev.* **13**, 3–28.
Ngai, P. H. K. & Ng, T. B. (2007). *Appl. Microbiol. Biotechnol.* **74**, 366–371.
Patra, D., Mishra, P., Surolia, A. & Vijayan, M. (2014). *Glycobiology*, **24**, 956–965.
Pérez, S., Sarkar, A., Rivet, A., Breton, C. & Imberty, A. (2015). *Methods Mol. Biol.* **1273**, 241–258.
Provencher, S. W. (1982). *Comput. Phys. Commun.* **27**, 229–242.
Read, R. J., Adams, P. D. & McCoy, A. J. (2013). *Acta Cryst.* D**69**, 176–183.
Rini, J. M. & Lobsanov, Y. D. (1999). *Curr. Opin. Struct. Biol.* **9**, 578–584.
Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. & Vijayan, M. (1996). *Nature Struct. Biol.* **3**, 596–603.
Sharma, A., Chandran, D., Singh, D. D. & Vijayan, M. (2007). *J. Biosci.* **32**, 1089–1110.
Sharon, N. (2007). *J. Biol. Chem.* **282**, 2753–2764.
Sixma, T. K., Pronk, S. E., Kalk, K. H., Wartna, E. S., van Zanten, B. A., Witholt, B. & Hol, W. G. J. (1991). *Nature (London)*, **351**, 371–377.
Swaminathan, S. & Eswaramoorthy, S. (2000). *Nature Struct. Biol.* **7**, 693–699.
Taylor, M. E. & Drickamer, K. (2007). *Curr. Opin. Cell Biol.* **19**, 572–577.
Vijayan, M. & Chandra, N. (1999). *Curr. Opin. Struct. Biol.* **9**, 707–714.
Weis, W. I. & Drickamer, K. (1996). *Annu. Rev. Biochem.* **65**, 441–473.
Wiley, D. C., Wilson, A. I. & Skehel, J. J. (1981). *Nature (London)*, **289**, 373–378.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wong, J. H., Wong, C. C. T. & Ng, T. B. (2006). *Biochim. Biophys. Acta*, **1760**, 808–813.
Zhang, Y. (2008). *BMC Bioinformatics*, **9**, 40.
Zhang, G. Q., Sun, J., Wang, H. X. & Ng, T. B. (2009). *Acta Biochim. Pol.* **56**, 415–421.