



# The structure of a pectin-active family 1 polysaccharide lyase from the marine bacterium *Pseudoalteromonas fuliginea*

Joanne K. Hobbs<sup>‡</sup> and Alisdair B. Boraston\*

Department of Biochemistry and Microbiology, University of Victoria, PO Box 1700 STN CSC, Victoria, BC V8W 2Y2, Canada. \*Correspondence e-mail: boraston@uvic.ca

Received 27 May 2024

Accepted 19 June 2024

Edited by J. Agirre, University of York, United Kingdom

<sup>‡</sup> Current address: School of Biology, University of St Andrews, St Andrews KY16 9ST, United Kingdom.**Keywords:** polysaccharide lyases; pectinases; marine bacteria; *Pseudoalteromonas*; pectinate.**PDB reference:** PfPL1, 9buj**Supporting information:** this article has supporting information at journals.iucr.org/f

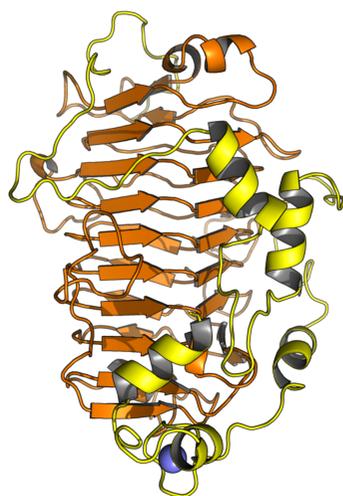
*Pseudoalteromonas fuliginea* sp. PS47 is a recently identified marine bacterium that has extensive enzymatic machinery to metabolize polysaccharides, including a locus that targets pectin-like substrates. This locus contains a gene (locus tag EU509\_03255) that encodes a pectin-degrading lyase, called PfPL1, that belongs to polysaccharide lyase family 1 (PL1). The 2.2 Å resolution X-ray crystal structure of PfPL1 reveals the compact parallel  $\beta$ -helix fold of the PL1 family. The back side of the core parallel  $\beta$ -helix opposite to the active site is a meandering set of five  $\alpha$ -helices joined by lengthy loops. A comparison of the active site with those of other PL1 enzymes suggests a catalytic mechanism that is independent of metal ions, such as  $\text{Ca}^{2+}$ , but that substrate recognition may require metal ions. Overall, this work provides the first structural insight into a pectinase of marine origin and the first structure of a PL1 enzyme in subfamily 2.

## 1. Introduction

*Pseudoalteromonas* is a globally distributed genus that is typically associated with marine environments, including the colonization of macroalgal surfaces (Martin *et al.*, 2014; Bowman *et al.*, 1997). *P. fuliginea* sp. PS47 is a recently identified bacterium that was isolated from the surface of macroalgae harvested from the marine environment of the Pacific Northwest. This bacterium has the ability to process carrageenan, agarose and pectin-like polysaccharides (Pluvinage *et al.*, 2020; Hettle *et al.*, 2019; Hobbs *et al.*, 2019).

Pectin is a complex polysaccharide found in the cell walls of plants, particularly in fruits such as apples, citrus fruits and berries. It is primarily composed of chains of  $\alpha$ -1,4-linked D-galacturonic acid residues (homogalacturonan), although some forms also contain L-rhamnose in the backbone (rhamnogalacturonan). The complexity of pectins is increased by varying degrees of branching, acetylation and/or methylesterification, such as in rhamnogalacturonan II (RGII), which is considered to be one of the most complex polysaccharides in nature. Pectin is not common to macroalgae, but aquatic plants do have pectin-like polysaccharides. Zosterin, which is an apiogalacturonan, is a pectin-like molecule from the aquatic seagrasses in the *Zostera* genus. It has a homogalacturonan backbone substituted with O3-linked  $\beta$ -D-apiofuranose residues, or short chains thereof. Acetyl and methylester modifications have also been detected (Lv *et al.*, 2015). This 'marine pectin' is, therefore, a possible substrate for bacteria with the appropriate enzymatic machinery.

Polysaccharide lyase family 1, PL1, is a family of primarily polygalacturonan/pectin lyases that catalyse depolymerization through a  $\beta$ -elimination mechanism, resulting in the formation



OPEN ACCESS

Published under a CC BY 4.0 licence

**Table 1**

Macromolecule-production information.

Source organism	<i>Pseudoalteromonas fuliginea</i> sp. PS47
DNA source	Genomic DNA
Forward primer†	<u>GCCGCGCGGAGCCAACTCGACTCAAAT</u> TTAGCCTTTAAAAATGC
Reverse primer†	<u>GCTCGAATTCGGATCGATTACTCCGTAA</u> TCGAATTTATATAAGC
Expression vector	pET-28a
Expression host	<i>Escherichia coli</i>
Complete amino-acid sequence of the construct produced‡	<b>MGSSHHHHSSGLVPRGSHMAS</b> LDSNL AFKNADGYGKYTQGGRDGKIYIVNSLE DNPKNPAKGLRLHALKRYKRTVVFNI SGVIHLKEPIIVKSGFLTIAGQTSPPG ITVAGAPVQVSDADHIIIRYMRFLGT FKLAEDSMSVRNSRDIIDHCSFSWSV DETASFYNNQRF <sup>†</sup> TLQNSIVAASLNHSI HPKGHHGGYGGIWWGNKASFINNVIAHH NSRTPRLNGSRLKPPYDEQEFVEFSN NIIFNWGSNNVYGSSENGRPNLNNIYK PGPASKAIQLVLDLWYSPNITKSQAYIS GNYFVGDDEKITADNRLGVNRYRTSKDAK RKNISMDDKRLSRVKLEPINGAVNSAT INSTQKTYSTLIKEKNVGFANFNANGMF LDNIDTQVLNQVDGSTPINGKGLINSE LEMIKSWEYERQFLGF <sup>‡</sup> PDIIDKNKDG INDRWAAKNPTNQHNINAYINSITE

† The vector sequences used for In-Fusion cloning are underlined. ‡ The His<sub>6</sub>-tag sequence is shown in bold.

of a double bond between the C4 and C5 atoms of the sugar residue at the nonreducing end of the polysaccharide chain (Zheng *et al.*, 2021). This family is further stratified into 13 subfamilies (Lombard *et al.*, 2010). *P. fuliginea* sp. PS47 has a pectin-degradation locus that encodes a PL1 referred to as *PfPL1* (previously known as *PsPL1*), in addition to 19 other proteins comprising the machinery for pectin depolymerization, transport and assimilation (Hobbs *et al.*, 2019). *PfPL1* is crucial to initiate pectin depolymerization and has demonstrated activity on apiogalacturonan (Hobbs *et al.*, 2019). This enzyme belongs to the largest PL1 subfamily, subfamily 2 (denoted PL1<sub>2</sub>), which is one of only two subfamilies with characterized members but no available structures (subfamily 8 is the other). Four additional members of PL1<sub>2</sub> have been functionally characterized as pectin lyases: three from the homogalacturonan polysaccharide-utilization locus of *Bacteroides thetaiotaomicron* (BT\_4115, BT\_4116 and BT\_4119) and one from *Dickeya dadantii* (PelZ) (Luis *et al.*, 2018; Kita *et al.*, 1996). Here, we report the structure of *PfPL1*, which is the first structure of a PL1<sub>2</sub> protein. A comparison of the *PfPL1* structure with the structures of PL1 enzymes from other subfamilies reveals similarities in substrate recognition but potential differences in the catalytic machinery.

## 2. Materials and methods

### 2.1. Macromolecule production

The gene encoding *PfPL1* from *P. fuliginea* sp. PS47 (locus tag EU509\_03255; GenBank KAA1163790.1), lacking the predicted secretion signal peptide (amino-acid boundaries 29–464), was cloned as described previously into pET-28a as an in-frame fusion with an N-terminal His<sub>6</sub> tag to generate pET-28a-

**Table 2**

Crystallization.

Method	Vapour diffusion, hanging drop
Temperature (K)	291
Protein concentration (mg ml <sup>-1</sup> )	11
Buffer composition of protein solution	20 mM Tris–HCl pH 8.0, 150 mM sodium chloride
Composition of reservoir solution	20% PEG 3350, 0.1 M sodium malonate pH 7.0
Volume and ratio of drop	2 µl, 1:1
Volume of reservoir (µl)	500

*PfPL1* (Hobbs *et al.*, 2019). Protein production and purification were performed as outlined previously with the addition of a final size-exclusion chromatography step using a Sephacryl S-200 column (GE Healthcare) pre-equilibrated with 20 mM Tris–HCl pH 8.0, 150 mM sodium chloride (Hobbs *et al.*, 2019). The purified protein was concentrated to ~8–11 mg ml<sup>-1</sup> using a stirred-cell ultrafiltration device with a 10 000 Da molecular-weight cutoff membrane (Millipore) for crystallization. Macromolecule-production information is summarized in Table 1.

### 2.2. Crystallization

Initial crystallization trials were carried out using the commercially available screens Index (Hampton Research) and MSCG1 (Anatrace) by the sitting-drop vapour-diffusion method at 291 K with a 1:1 ratio of protein (8 mg ml<sup>-1</sup> in 20 mM Tris–HCl pH 8.0, 150 mM sodium chloride) and crystallization solutions (0.5 µl each). An initial hit was optimized using the hanging-drop vapour-diffusion method and protein at 11 mg ml<sup>-1</sup>, resulting in improved crystals using 20% (w/v) PEG 3350, 0.1 M sodium malonate pH 7.0. Crystallization information is summarized in Table 2.

### 2.3. Data collection and processing

Only a single suitable crystal was obtained and this was cryoprotected in crystallization solution supplemented with 20% (v/v) 2-methyl-2,4-pentanediol prior to mounting directly in a nitrogen stream at 100 K. Diffraction data were collected on an instrument comprising a PILATUS 200K 2D detector coupled to a MicroMax-007 HF X-ray generator with a VariMax-HF ArcSec Confocal Optical System and an Oxford Cryostream 800. The data were integrated, scaled and merged using *HKL-2000*. Data-collection and processing statistics are summarized in Table 3.

### 2.4. Structure solution and refinement

The crystal structure of *PfPL1* was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using a search model of *PfPL1* generated with *AlphaFold* (Jumper *et al.*, 2021; Varadi *et al.*, 2022). The model was corrected by manual building with *Coot* (Emsley *et al.*, 2010) and refinement with *phenix.refine* (Liebschner *et al.*, 2019). The addition of water molecules was performed with Find Waters in *Coot* and manually checked after refinement. Refinement procedures were monitored by flagging 5% of all observations as ‘free’

**Table 3**

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	Rigaku MicroMax-007 HF rotating anode
Wavelength (Å)	1.5418
Temperature (K)	100
Detector	Dectris PILATUS 200K
Rotation range per image (°)	0.25
Total rotation range (°)	128
Exposure time per image (s)	180
Space group	$P2_12_12_1$
$a, b, c$ (Å)	48.50, 58.14, 149.08
$\alpha, \beta, \gamma$ (°)	90, 90, 90
Mosaicity (°)	0.77–1.01
Resolution range (Å)	20.00–2.20 (2.24–2.20)
Total No. of reflections	91830
No. of unique reflections	21844
Completeness (%)	97.8 (82.0)
Multiplicity	4.2 (3.1)
$\langle I/\sigma(I) \rangle$	14.4 (3.2)
$R_{\text{p.i.m.}}$	0.048 (0.168)
$R_{\text{meas}}$	0.103 (0.300)
$CC_{1/2}$	0.984 (0.866)
Overall $B$ factor from Wilson plot (Å <sup>2</sup> )	31.6

(Brünger, 1992). Model validation was performed with *MolProbity* (Chen *et al.*, 2010). All model statistics are shown in Table 4.

### 3. Results and discussion

#### 3.1. Overall structure of *PfPL1*

*PfPL1* crystallized in space group  $P2_12_12_1$ , with unit-cell parameters  $a = 48.50$ ,  $b = 58.13$ ,  $c = 149.08$  Å. *MATTHEWS\_COEF* analysis (Kantardjieff & Rupp, 2003) indicated the presence of a single protein molecule in the asymmetric unit, with a Matthews coefficient of  $2.16 \text{ Å}^3 \text{ Da}^{-1}$  and a solvent content of 43.14%. Analysis with *PISA* (Krissinel & Henrick, 2007) indicated the absence of noncrystallographic dimers that would be stable in solution. Residues 25–458 of the recombinant protein could be modelled and refined with no gaps in the protein backbone.

The structure of *PfPL1* (PDB entry 9buj) comprises at its core the parallel  $\beta$ -helix common to family 1 polysaccharide lyases (Fig. 1*a*). This compact fold of ~300 amino acids at the N-terminus is adorned by an ~125-amino-acid C-terminal region made up of  $\alpha$ -helices and large loops that meander across the parallel  $\beta$ -helix surface (Fig. 1). The C-terminal region terminates in a structural motif resembling an EF-hand; however, in this case the calcium ( $\text{Ca}^{2+}$ ) ion is bound on what would be the back of the thumb rather than in the palm at the junction of the thumb and forefinger (Fig. 1*b*). The  $B$  factor of the  $\text{Ca}^{2+}$  ion is comparable to the average  $B$  factor of the protein, indicating that it is relatively well ordered and thus may represent a structural component that pins together these secondary-structure elements at the C-terminus.

A structural similarity search using the *DALI* server returned members of PL1 as the top hits. The two most similar structures were VexL from *Achromobacter denitrificans* [PDB entry 6fi2;  $Z$ -score and root-mean-square deviation (r.m.s.d.)

**Table 4**

Structure solution and refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	19.50–2.19 (2.29–2.19)
Completeness (%)	96.0
$\sigma$ Cutoff	$F > 0.000\sigma(F)$
No. of reflections, working set	21445 (2181)
No. of reflections, test set	1057 (106)
Final $R_{\text{cryst}}$	0.185 (0.210)
Final $R_{\text{free}}$	0.252 (0.268)
Cruickshank DPI	0.24
No. of non-H atoms	
Protein	3417
Ligand ( $\text{Ca}^{2+}$ )	1
Water	202
Total	3620
R.m.s. deviations	
Bond lengths (Å)	0.009
Angles (°)	1.075
Average $B$ factors (Å <sup>2</sup> )	
Protein	32.45
Ligand ( $\text{Ca}^{2+}$ )	33.17
Water	36.29
Ramachandran plot	
Favoured regions (%)	93.3
Allowed (%)	6.7
Outliers (%)	0.00
PDB code	9buj

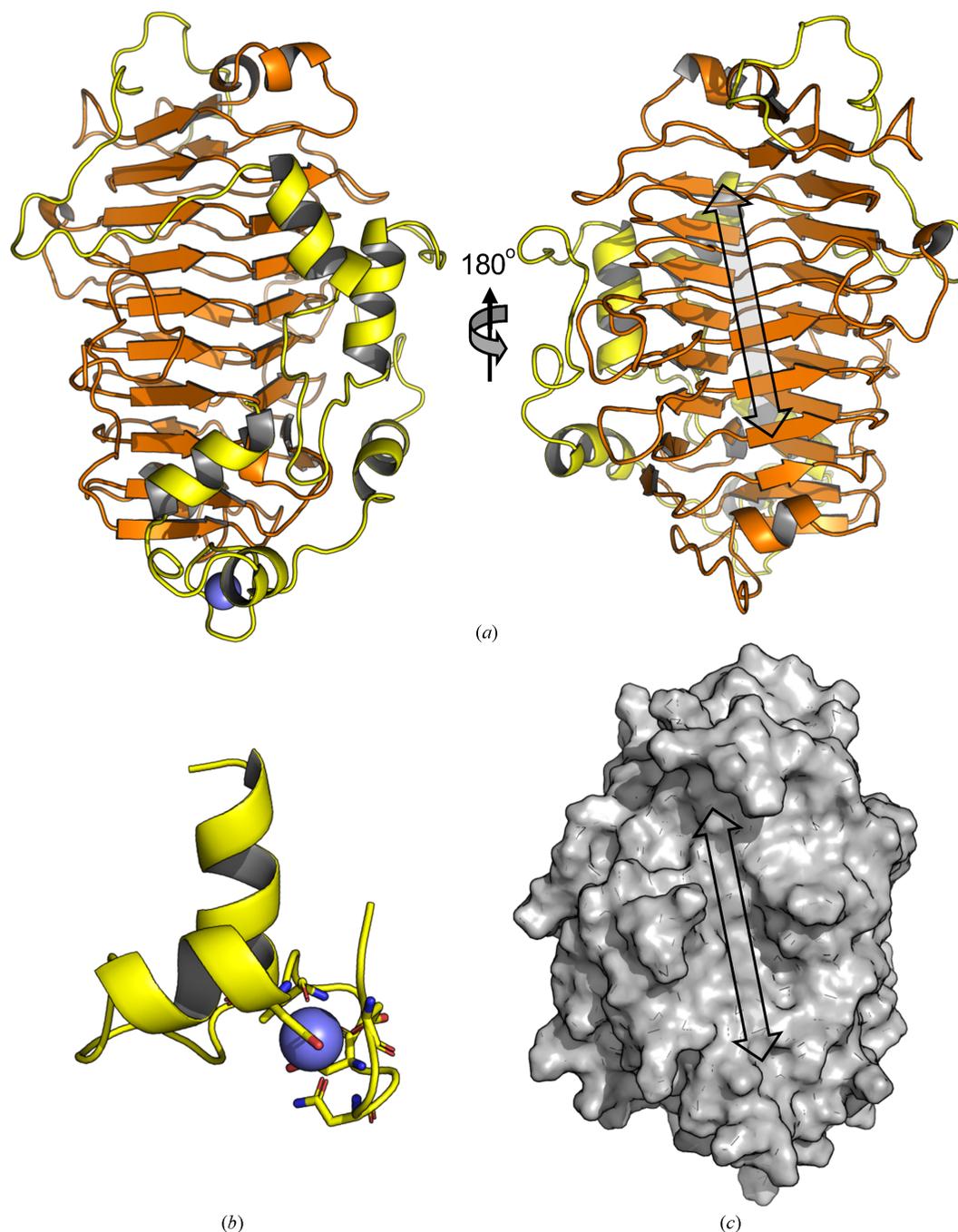
of 31.7 and 2.4 Å, respectively, over 284 matched  $\text{C}^\alpha$  residues; 21% amino-acid sequence identity] and Jun a 1, the major pollen allergen from *Juniperus ashei* (PDB entry 1pxz;  $Z$ -score and r.m.s.d. of 31.3 and 2.8 Å, respectively, over 290 matched  $\text{C}^\alpha$  residues; 23% amino-acid sequence identity) (Fig. 2*a*; Liston *et al.*, 2018; Czerwinski *et al.*, 2005). VexL belongs to subfamily 13 of PL1, while Jun a 1 belongs to subfamily 1 (Fig. 2*a*). Amongst the next four most similar structures was PelC from *Dickeya chrysanthemi* EC16 (PDB entry 1air;  $Z$ -score and r.m.s.d. of 24.8 and 3.0 Å, respectively, over 269 matched  $\text{C}^\alpha$  residues; 17% amino-acid sequence identity; Fig. 2*b*), which belongs to subfamily 6 and a mutant of which has notably been captured in complex with intact substrate (PDB entry 2ewe; Scavetta *et al.*, 1999). The catalytic groove of *PfPL1* was identified by comparison to these other PL1 structures and runs along the face of one  $\beta$ -sheet (Fig. 1*a*, right). The groove is relatively shallow (Fig. 1*c*), perhaps to assist in the accommodation of decorated pectins, such as the apioagalacturonan found in the marine environment, on which we have shown *PfPL1* to be active (Hobbs *et al.*, 2019).

#### 3.2. Structure of the catalytic groove

To provide some insight into the potential interactions of *PfPL1* with substrate, we compared its structure with the complexed structures of VexL (PDB entry 6fi2) and PelC (PDB entry 1ewe) in some detail. The VexL structure is a product complex with a trisaccharide of  $\alpha$ -1,4-linked *N*-acetyl-galacturonic acid occupying subsites +1 to +3 (Liston *et al.*, 2018). An overlap of this structure with *PfPL1* clearly reveals conservation of the catalytic arginine residues, Arg220 and Arg235, in *PfPL1* (Fig. 2*c*). Based on their interaction with VexL, the side chain of Arg220 would abstract the proton from

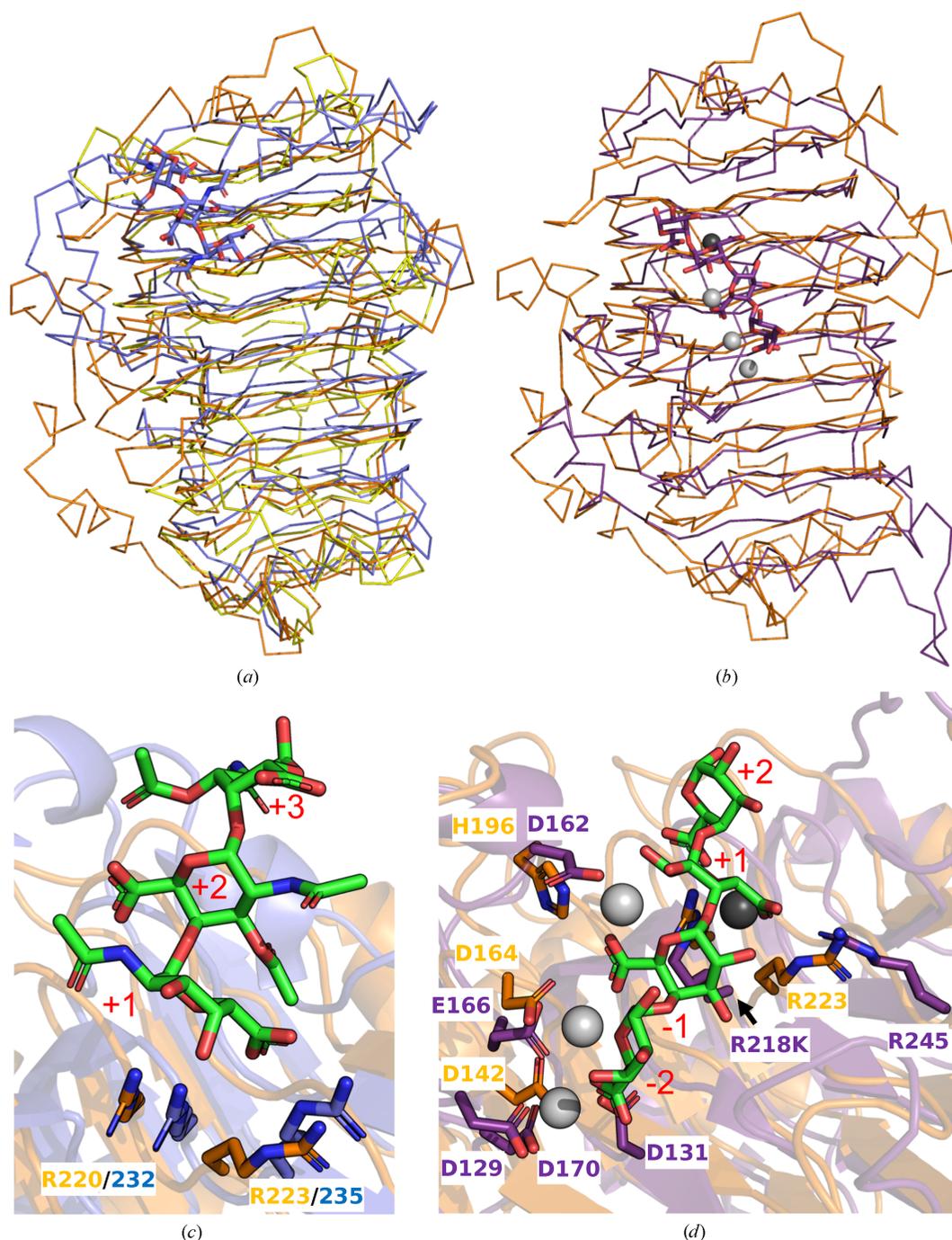
C5 and the side chain of Arg235 would help to neutralize the negative charge on the C6 carboxylate. Some pectin lyases rely on calcium ions for their catalytic mechanism; however, VexL is reported to be metal-independent. The similarity in the arrangement of the catalytic machineries of the two proteins suggests that *Pf*PL1 also does not rely on metals for its catalytic mechanism. This is supported by the overlay of *Pf*PL1 with the PelC complex, an enzyme whose catalytic mechanism is proposed to rely on  $\text{Ca}^{2+}$ . The complex of PelC, obtained by mutating the catalytic arginine to lysine, has  $\alpha$ -1,4-

tetragalacturonic acid spanning subsites  $-2$  to  $+2$  (Fig. 2*d*; Scavetta *et al.*, 1999). A catalytic  $\text{Ca}^{2+}$  ion in PelC is bound in the pocket and is sandwiched between the substrate and the protein. *Pf*PL1 does not seem to possess an analogous pocket (Figs. 1*c* and 2*d*). However, in PelC additional  $\text{Ca}^{2+}$  ions in the minus ( $-$ ) subsites are involved in bridging the interaction of the substrate with residues in the active site. Some of the side chains involved in coordinating these 'binding'  $\text{Ca}^{2+}$  ions are structurally conserved in *Pf*PL1 (Fig. 2*d*). Taken together, these comparisons suggest a  $\text{Ca}^{2+}$ -dependent mode of



**Figure 1**

Overall structure of *Pf*PL1. (a) Cartoon representation of *Pf*PL1 shown from both sides. The core parallel  $\beta$ -helix is shown in orange and the C-terminal meandering adornment is shown in yellow. (b) An enlargement of the pseudo-EF-hand motif found at the C-terminus of *Pf*PL1. Residues involved in coordinating the  $\text{Ca}^{2+}$  ion are shown as sticks. (c) Surface representation of *Pf*PL1 shown as a solvent-accessible surface. The arrows in (a) and (c) represent the location of the catalytic groove.



**Figure 2**  
 Comparison of *Pf*PL1 with other PL1 enzymes reveals insight into its active site. (a) Overlay of *Pf*PL1 (orange), VexL (PDB entry 6fi2, blue) and Jun a 1 (PDB entry 1pxz, yellow) shown as C $\alpha$  ribbons. (b) Overlay of *Pf*PL1 (orange) and PelC (PDB entry 1ewe, purple) shown as C $\alpha$  ribbons. (c) Active site of *Pf*PL1 (orange) compared with that of VexL (PDB entry 6fi2, blue). (d) Active site of *Pf*PL1 (orange) compared with that of PelC (PDB entry 1ewe, purple). Ca<sup>2+</sup> ions involved in substrate binding in PelC are shown as light grey spheres. The proposed catalytic Ca<sup>2+</sup> ion involved in the catalytic mechanism of PelC is shown as a dark grey sphere.

substrate binding for *Pf*PL1 that is similar to PelC but a catalytic mechanism that does not rely on metals.

#### 4. Conclusion

The crystal structure of *Pf*PL1 from the marine bacterium *P. fuliginea* sp. PS47 is the first for the largest subfamily of

PL1, subfamily 2. The structure reveals relatively simple loops forming the walls of the active site to create a relatively shallow catalytic groove when compared with other PL1 enzymes. The arrangement of the active site suggests that substrate binding relies on metal atoms to bridge the substrate to active-site residues, but that the catalytic mechanism does not require metals.

## Funding information

This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

## References

- Bowman, J. P., McCammon, S. A., Brown, M. V., Nichols, D. S. & McMeekin, T. A. (1997). *Appl. Environ. Microbiol.* **63**, 3068–3078.
- Brünger, A. T. (1992). *Nature*, **355**, 472–475.
- 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.* **D66**, 12–21.
- Czerwinski, E. W., Midoro-Horiuti, T., White, M. A., Brooks, E. G. & Goldblum, R. M. (2005). *J. Biol. Chem.* **280**, 3740–3746.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Hettle, A. G., Hobbs, J. K., Pluvinage, B., Vickers, C., Abe, K. T., Salama-Alber, O., McGuire, B. E., Hehemann, J.-H., Hui, J. P. M., Berrue, F., Banskota, A., Zhang, J., Bottos, E. M., Van Hamme, J. & Boraston, A. B. (2019). *Commun. Biol.* **2**, 474.
- Hobbs, J. K., Hettle, A. G., Vickers, C. & Boraston, A. B. (2019). *Appl. Environ. Microbiol.* **85**, e02114-18.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature*, **596**, 583–589.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Kita, N., Boyd, C. M., Garrett, M. R., Jurnak, F. & Keen, N. T. (1996). *J. Biol. Chem.* **271**, 26529–26535.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.
- Liston, S. D., McMahon, S. A., Le Bas, A., Suits, M. D. L., Naismith, J. H. & Whitfield, C. (2018). *Proc. Natl Acad. Sci. USA*, **115**, E4870–E4879.
- Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M. & Henrissat, B. (2010). *Biochem. J.* **432**, 437–444.
- Luis, A. S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., Baslé, A., Cartmell, A., Terrapon, N., Stott, K., Lowe, E. C., McLean, R., Shearer, K., Schückel, J., Venditto, I., Ralet, M.-C., Henrissat, B., Martens, E. C., Mosimann, S. C., Abbott, D. W. & Gilbert, H. J. (2018). *Nat. Microbiol.* **3**, 210–219.
- Lv, Y., Shan, X., Zhao, X., Cai, C., Zhao, X., Lang, Y., Zhu, H. & Yu, G. (2015). *Mar. Drugs*, **13**, 3710–3731.
- Martin, M., Portetelle, D., Michel, G. & Vandenberg, M. (2014). *Appl. Microbiol. Biotechnol.* **98**, 2917–2935.
- 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.
- Pluvinage, B., Robb, C. S., Jeffries, R. & Boraston, A. B. (2020). *Acta Cryst.* **F76**, 422–427.
- Scavetta, R. D., Herron, S. R., Hotchkiss, A. T., Kita, N., Keen, N. T., Benen, J. A. E., Kester, H. C. M., Visser, J. & Jurnak, F. (1999). *Plant Cell*, **11**, 1081–1092.
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D. & Velankar, S. (2022). *Nucleic Acids Res.* **50**, D439–D444.
- Zheng, L., Xu, Y., Li, Q. & Zhu, B. (2021). *Bioresour. Bioprocess.* **8**, 79.