Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Azospirillum irakense pectate lyase displays a toroidal fold

Hector Novoa de Armas,[a] Christel Verboven,[a] Camiel De Ranter,[a] Jos Desair,[b] Ann Vande Broek,[b] Jos Vanderleyden[b] and Anja Rabijns[a]*

[a]Laboratorium voor Analytische Chemie en Medicinale Fysicochemie, Faculteit Farmaceutische Wetenschappen, K.U. Leuven, E. Van Evenstraat 4, B-3000 Leuven, Belgium, and [b]Centrum voor Microbiële en Plantengenetica, K.U. Leuven, Kasteelpark Arenberg 20, B-3001 Heverlee, Belgium

Correspondence e-mail: anja.rabijns@pharm.kuleuven.ac.be

Received 3 December 2003
Accepted 15 March 2004

PDB Reference: pectate lyase, 1r76, r1r76sf.

The three-dimensional structure of *Azospirillum irakense* pectate lyase (PelA) has been determined at a resolution of 2.65 Å. The crystals are hexagonal, belonging to space group $P6_522$, with unit-cell parameters $a = b = 85.37$, $c = 231.32$ Å. Phase information was derived from a multiple-wavelength anomalous dispersion (MAD) experiment using a Hg derivative. Refinement of the model converged to $R_{cryst} = 20.08\%$ and $R_{free} = 25.87\%$. The overall structure of PelA does not adopt the characteristic parallel β-helix fold displayed by pectate lyases from polysaccharide lyase (PL) families PL1, PL3 and PL9. Instead, it displays a predominantly α-helical structure with irregular coils and short β-strands, similar to the recently reported structure of the catalytic module of the *Cellvibrio japonicus* pectate lyase Pel10Acm. The topologies of the two structures have been compared. They show two 'domains' with the interface between them being a wide-open central groove in which the active site is located. The active sites of the crystal structures are also compared and their similarities and differences are discussed.

1. Introduction

Polysaccharide lyases (PLs; E.C. 4.2.2.*x*) are a class of glycoside bond-cleaving enzymes that are particularly important in plant biochemistry in the degradation of pectin, one of the main components of the plant cell wall (Henrissat *et al.*, 2001). Most of the PL enzymes are of microbial origin.

The enzymes characterized to date have enabled certain general mechanistic features to be identified; they have a requirement for sugar uronic acids. This dependence on free or methyl-esterified substrate C5 uronic acid groups and divalent cations, typically Ca^{2+}, seems essential to allow 'acidification' and hence abstraction of the C5 proton of the +1 subsite sugar residue. This leads to the release of a 4,5-unsaturated product: oligosaccharides with 4-deoxy-α-D-glucuronosyl groups at their non-reducing termini, as illustrated in Fig. 1. The intricacies of the lyase catalytic mechanism remain elusive, even in the light of biochemical and structural analyses (Scavetta *et al.*, 1999).

Pectate lyases (PeLs; EC 4.2.2.2), alternatively known as pectate transeliminases, are secreted microbial enzymes that degrade pectin and are therefore important in plant pathogenesis (Anderson, 1998). The absence of homology between the primary sequences of PeLs from distinct families (PL1, PL2, PL3, PL9 and PL10 from the PL classification of carbohydrate-active enzymes; CAZymes; http://afmb.cnrs-mrs.fr/CAZY/) suggests that they have evolved from different lineages. The fact that such catalytically similar enzymes have evolved independently may reflect their

different functions in nature. Strikingly, homology was observed between the bacterial *Erwinia chrysanthemi* Pell and four PeL proteins from a phytopathogenic fungus, *Neciria haemotococca* (*Fusarium solani*; Gonzalez-Candelas & Kolattukudy, 1992; Guo *et al.*, 1995*a*,*b*; Shevchik *et al.*, 1997), demonstrating that PeLs belonging to the same family are present in distantly related organisms. On the other hand, the lyase mechanisms of members of different families appear to be of similar efficiency (Brown *et al.*, 2001); this means it is unclear what evolutionary advantage any of these classes of enzyme offers to their host.

Understanding of PeLs has been considerably advanced by structural determinations of PeLs from families PL1 and PL3: *E. chrysanthemi* PelC (Yoder, Keen *et al.*, 1993), *E. chrysanthemi* PelE (Lietzke *et al.*, 1994), *B. subtilis* Pel (Pickersgill *et al.*, 1994), *B. subtilis* strain Ksm15 pectate lyase (Akita *et al.*, 2000) and *E. chrysanthemi* PelA (Thomas *et al.*, 2002). These results led to a novel protein structural fold and the assignment of a unique catalytic role to an arginine residue. The polypeptide backbones of all these enzymes form a single domain consisting of parallel β-strands that are wound into a large right-handed coil termed a parallel β-helix. In addition to this novel fold they share other unique structural features, including a highly organized core (Yoder, Lietzke *et al.*, 1993). The loops fold over the exterior surface of the parallel β-helix and confer a unique shape and charge on each PeL. From the perspective of effectiveness as a plant virulence factor, the most important feature of the parallel β-helix fold is the stability that it confers on an enzyme that must function in the hostile extracellular environment. In the past decade, the genetics of bacterial pectinase biosynthesis have been extensively studied in phytopathogens, especially in the soft-rotting *Erwinia* species *E. carotovora* and *E. chrysanthemi* (Lietzke *et al.*, 1994). Both species were found to produce a set of pectin-depolymerizing activities such as PeLs, polygalacturonases, pectin methyl esterases and a pectin acetyl esterase (Barras *et al.*, 1994; Hugouvieux-Cotte-Pattat *et al.*, 1996; Pissavin *et al.*, 1996; Shevchik *et al.*, 1997; Shevchik & Hugouvieux-Cotte-Pattat, 1997). Of these, the PeLs are the major pectinases and play a key role in the development of soft-rot disease.

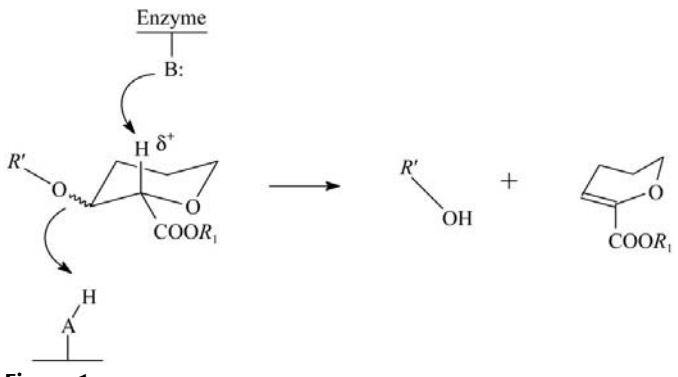


Figure 1
Generic catalytic mechanism (β-elimination mechanism) for a polysaccharide lyase. The generally accepted mechanism involves proton abstraction from C5 of the +1 subsite sugar residue adjacent to the C6 carboxyl moiety to release the 4,5-unsaturated product *via* a β-elimination mechanism (Henrissat *et al.*, 2001). R', monosaccharide unit, R_1, alkyl or H.

Azospirillum irakense KBC1 (Bekri *et al.*, 1999) grows and fixes N_2 in media containing pectin or polygalacturonate (PGA) as the sole carbon source, indicating the presence of a catabolic pathway for further breakdown of the oligogalacturonides generated by the pectinolytic activity. Degradation of PGA yielded products with an absorbance peak at 550 nm upon incubation with thiobarbituric acid (TBA), which is characteristic of the unsaturated oligogalacturonides generated by a sugar lyase. Subsequently, a peL4 gene coding for a 44 kDa acidic endo-PeL (PelA) was identified and was further characterized at the genetic and biochemical levels by means of heterologous expression in *Escherichia coli* (Bekri *et al.*, 1999).

PelA originally defined a new family of PL enzymes (family 10; PL10 on the CAZymes server), since it displayed no homology to other known bacterial, fungal or plant pectinases according to sequence-comparison programs. *A. irakense* has never been reported to be pathogenic in plants, so the structure of PelA would be of particularly interest in order to identify potential differences from phytopathogenic PeLs for which three-dimensional structures have been reported.

New members of family 10, originally defined by *A. irakense* PelA, have been reported: Pel-15E from *Bacillus* sp. strain KSM-P15 (Sawada *et al.*, 2000) and Pel10A from *Cellvibrio japonicus* (formerly *Pseudomonas cellulose*; Brown *et al.*, 2001). Pel-15E possesses some regions that are homologous to PelA, with an overall 39.8% amino-acid identity. The 35.8 kDa C-terminal module of Pel10A (named Pel10Acm) was shown to have 30 and 36% amino-acid identity to PelA and Pel-15E, respectively. During the last stages of the refinement of our PelA crystal structure, the three-dimensional structure of the catalytic module of Pel10A (Pel10Acm) was reported (Charnock *et al.*, 2002).

The optimum activity of PelA was observed at pH 9, which is comparable to the values of 10 for the optimum activity of Pel10Acm from *P. cellulosa* and 10.5 for Pel-15E from *Bacillus* sp. This is a characteristic feature of PeLs, their optimum pH always being in the range pH 8.5–10.5 independent of the microorganism that produces them. In addition to a highly alkaline pH for its activity, like other PL enzymes PelA requires the presence of Ca^{2+} ions. Three sequence patterns that are conserved among the major extracellular PL enzymes of *E. chrysanthemi* and some PL enzymes from non-*Erwinia* phytopathogens and are suspected of being involved in Ca^{2+} binding and catalytic activity (Henrissat *et al.*, 1995) could not be identified in the PelA protein. Unlike the *E. chrysanthemi* PeL enzymes (Lojkowska *et al.*, 1995; Shevchik *et al.*, 1997; Tardy *et al.*, 1997), a high activity of purified *A. irakense* PelA was observed over a wide range of Ca^{2+} concentrations (from 0.2 to 2 m*M*). Other cations cannot substitute for Ca^{2+} and some were even found to inhibit PelA activity. Similar to observations made for the five major *E. chrysanthemi* PL enzymes (Tardy *et al.*, 1997), Zn^{2+} was the strongest inhibitor of PelA activity.

To date, only 15 sequence-based families of PLs have been discovered, in contrast to the 94 families of glycoside hydrolases (GHs; Coutinho & Henrissat, 1999). At the beginning of this study there was no structure reported in the PDB (Sussman *et al.*, 1998) of a protein belonging to family PL10 of the CAZymes. The main objective of this study was to solve the three-dimensional structure of PelA. Along with the recently reported structure of Pel10Acm (Charnock *et al.*, 2002), the crystal structure of PelA will provide detailed three-dimensional information on this most recently discovered family of lyases, leading to the identification of the key residues in the substrate binding and the lyase catalytic mechanism.

2. Materials and methods

2.1. Purification and crystallization

PelA was produced and purified according to a modification of a protocol initially described by Bekri *et al.* (1999). An *Escherichia coli* transformant with the pFAJ0612 plasmid encoding the *A. irakense* PelA enzyme was grown at 310 K in LB broth until the stationary growth phase. The culture was centrifuged at 6000*g* and the pellet was resuspended to a final concentration of 20%(*w*/*v*) in 50 m*M* sodium acetate buffer pH 5 containing 400 U DNaseI per millilitre (Sigma). Bacteria in this suspension were lysed in FastRNA tubes (BIO101) with a FastPrep FP120 device (Savant). The FastRNA tubes were centrifuged at 12 000*g* in a microcentrifuge, the clear lysate was transferred to 50 ml tubes (Falcon) and the pH was again adjusted to pH 5. The lysate was stored on ice for several hours or at 253 K overnight. After centrifugation for 20 min at 4000*g* at 277 K and filtration through a Millex 0.22 µm pore filter (Millipore), the cleared lysate was loaded on a cation-exchange column (SP Sepharose HP 16/10, Amersham Biosciences). Proteins were eluted with a linear gradient of sodium chloride in 50 m*M* sodium acetate pH 5 (from 0 to 0.3 *M* NaCl in 30 min; flow rate 2 ml min^{-1}; 1 ml fractions were collected). The pectinolytic activity of the fractions was assayed by monitoring the increase in absorbance at 232 nm owing to the formation of unsaturated uronide upon incubation with PGA. Fractions exhibiting pectinolytic activity were pooled and 1.5 volumes of HIC buffer [1.7 *M* $(NH_4)_2SO_4$, 20 m*M* Tris–HCl pH 7.4] were added. The mixture was loaded onto a hydrophobic interaction column (Phenyl Sepharose HP 16/10, Amersham Biosciences). Elution was carried out using a decreasing linear gradient of $(NH_4)_2SO_4$ in Tris–HCl pH 7.4 [from 1 to 0 *M* $(NH_4)_2SO_4$ in 40 min]. Fractions exhibiting pectinolytic activity were pooled and concentrated on Vivascience concentrators (Vivaspin 6 ml, 5000 MWCO). The concentrated solution was loaded onto a gel-filtration column (Superdex 200 prep-grade 16/60, Amersham Biosciences). The protein was eluted with GFC buffer (50 m*M* NaCl, 20 m*M* Tris–HCl pH 7.2). Positive fractions were pooled and stored on ice or at 277 K. The purity of the resulting protein was checked on 12.5% SDS–PAGE PhastGels using the PhastSystem (Amersham Biosciences).

The protein solution was concentrated to approximately 10 mg ml^{-1} by ultrafiltration using a Microcon concentrator (Amicon) with a 3 kDa cutoff and was crystallized as described in Novoa de Armas *et al.* (2002). Diffracting rod-shaped crystals were obtained by the vapour-diffusion method from a solution containing 0.2 *M* ammonium sulfate, 15% 2-propanol, 20% polyethylene glycol 4000 and 0.1 *M* imidazole pH 7.8. A 2:1 ratio of protein to reservoir solution in the hanging drop was used. The reservoir solution consisted of 700 µl crystallization solution.

2.2. Data collection

A native data set was collected from these crystals as described previously (Novoa de Armas *et al.*, 2002). Heavy-atom derivative screening identified the formation of a mercury isomorphous derivative when native PelA crystals were soaked in 0.5 m*M* ethylmercury phosphate for 24 h and the heavy-atom salt solution was prepared with 0.2 *M* lithium sulfate instead of 0.2 *M* ammonium sulfate containing 25% glycerol as cryoprotectant. A PelA Hg-derivative crystal was subjected to a three-wavelength MAD data collection (Hendrickson, 1991). Data were collected at wavelengths near the Hg L_{III} absorption edge on the tunable beamline of the ELETTRA synchrotron, Trieste, Italy. The wavelengths corresponding to the peak and inflection points were determined from the changes in X-ray fluorescence of the PelA crystals as a function of the wavelength. From the fluorescence measurements, the inflection-point wavelength was found to be 1.0107 Å (λ_1). The peak-wavelength data were collected at 0.9919 Å (λ_2). A remote wavelength of 0.9184 Å was chosen as λ_3. For each wavelength, a complete data set was measured and the same crystal was rotated through the same angular interval. The statistics for this MAD data set are shown in Table 1. The crystals were mounted in cryo-loops and then plunged into liquid nitrogen. The cryo-loops were placed on a goniometer head and maintained at 100 K during diffraction data collection. The three-wavelength MAD data set was collected using a MAR CCD 165 mm detector. The *DENZO* and *SCALEPACK* packages (Otwinowski & Minor, 1997) were used for determination of the unit-cell parameters, data indexing, reduction and scaling.

2.3. Structure determination and refinement

From the data set collected from a Hg-derivative crystal it was possible to locate one Hg site in the anomalous Patterson map and to refine this site (isomorphous and anomalous phasing powers after refinement were 2.53 and 1.19, respectively, to 2.65 Å resolution). Patterson maps were initially generated with the *MAPVIEW* routine of the *PHASES* program (Furey & Swaminathan, 1990). The program *SHARP* (de La Fortelle & Bricogne, 1997) was used to confirm and refine heavy-atom sites as well as to find possible additional sites. Quality indicators of the MAD phase refinement are also shown in Table 1. *SOLOMON* (Abrahams & Leslie, 1996)

Table 1
Statistics of MAD data collection, phasing and refinement.

Values in parentheses are for the highest resolution shells. The remote-wavelength (λ_3) data set was chosen as a reference for scaling purposes.

Data set	λ_1 (inflection)	λ_2 (peak)	λ_3 (remote)
Data-collection and phasing statistics			
Space group	$P6_522$		
Unit-cell parameters (Å, °)	$a = b = 85.37$, $c = 231.32$, $\alpha = \beta = 90$, $\gamma = 120$		
λ (Å)	1.0107	0.9919	0.9184
Resolution (Å)	20–2.65	20–2.65	20–2.65
Completeness (%)	96.9 (98.4)	98.5 (98.2)	99.8 (100)
Multiplicity†	4.7 (4.7)	5.8 (4.7)	7.9 (7.1)
$\langle I/\sigma(I)\rangle$†	8.3 (2.4)	9.6 (1.2)	9.5 (2.3)
R_{sym}‡ (%)	7.2 (28.9)	5.2 (24.0)	6.7 (29.5)
R_{meas}‡ (%)	8.4 (33.7)	6.0 (28.4)	7.6 (33.6)
PCV‡ (%)	9.9 (41.5)	7.4 (39.9)	9.3 (43.9)
Overall R_{sym}‡ (%)	7.8	—	—
Overall R_{meas}‡ (%)	8.0	—	—
Overall PCV‡ (%)	10.8	—	—
R_{merge}† (%)	7.0 (18.4)	6.0 (15.3)	0.0
R_{ano} (%)	3.3 (15.9)	3.9 (14.1)	5.5 (20.9)
Starting f'	−23.04	−10.64	−8.10
Starting f''	7.04	9.85	8.51
Refined f'	−26.14	−19.62	−8.10
Refined f''	6.72	9.68	8.51
Phasing power			
PP_{iso}	2.53	2.70	—
PP_{ano}	1.19	1.43	1.02
Overall FOM at 2.65 Å	0.44		
Refinement statistics			
Resolution range (Å)			20–2.65
Reflections			12 662
Final R_{work} (%)			20.08
Final R_{free} (%)			25.87
Average atomic B factors (Å^2)			
Main-chain atoms			41.9
Side-chain atoms			42.9
Water molecules			35.4
2-Propanol and glycerol molecules			63.7
Cl^- ion			58.1
Hg^+ ion			60.6
R.m.s. deviation of the model			
Bond length (Å)			0.007
Bond angles (°)			1.21
B factors, bonded main chain (Å^2)			1.41
B factors, bonded side chain (Å^2)			1.89

† Extracted from *SCALEIT* (Collaborative Computational Project, Number 4, 1994). ‡ Extracted from *SCALA* (Collaborative Computational Project, Number 4, 1994).

solvent flattening from the *SHARP* program resulted in electron-density maps showing several α-helical motifs.

About 200 out of 408 residues were automatically traced into the electron-density maps using the programs *ARP/wARP* (Lamzin *et al.*, 2001) and *MAID* (Levitt, 2001). The remaining residues were manually built using the *O* graphical interface (Jones *et al.*, 1991). They are located in regions of the PelA electron-density maps in which the electron density is poorer and comprise several long and disordered loops.

At this stage of the PelA structure determination, a PelA model containing 351 amino acids (85% of the total of 408 residues) was subjected to one round of refinement against the native 1.97 Å data. The *CNS* program (v.1.1) was used for refinement (Brünger *et al.*, 1998). Prior to the minimization procedure, a test set was created for the R_{free} calculations by randomly choosing 10% of the observed structure factors. The actual minimization of the PelA structure consisted of a conventional conjugate-gradient least-squares refinement (Powell, 1977) followed by a simulated-annealing run using the slow-cooling protocol with the system heated to 3000 K (Brünger *et al.*, 1987, 1990; Brünger, 1991). Bulk-solvent corrections (Jiang & Brünger, 1994) were applied in order to correct for solvent scattering at low resolution. Target values and weights for bond lengths, bond angles *etc.* were extracted from the Engh and Huber parameter set (Engh & Huber, 1991).

At the start of the refinement of our initial PelA model, comprising 351 amino acids, some residues were traced as Gly owing to uncertainty in the side-chain tracing. After this first round of refinement the values of R_{cryst} and R_{free} dropped to 39.4 and 43.5%, respectively. Further model-rebuilding and refinement steps were performed in order to obtain a more accurate and more complete PelA structure. When about 70% of the whole model had been refined, water molecules were introduced based on objective criteria. Water molecules were progressively inserted during refinement when (i) a minimum 3σ peak was present in the $|F_{obs}| - |F_{calc}|$ difference map, (ii) a peak was visible in the $2|F_{obs}| - |F_{calc}|$ map, (iii) the B value for a water molecule did not attain a high and unreasonable value during refinement (B factor accepted <70 Å^2) and (iv) the water molecule was stabilized by hydrogen bonding. At all stages, σ_A-weighted OMIT $2|F_{obs}| - |F_{calc}|$ and $|F_{obs}| - |F_{calc}|$ Fourier electron-density maps were calculated and inspected with the *O* program (Jones *et al.*, 1991)

The model at this stage clearly showed that the protein structure has two distinctive 'domains': one formed mainly of well ordered α-helices (residues 50–150 and 160–282) and a second one formed mainly of small β-sheets and loops (282–354 and 370–432).

Refinement with the 1.97 Å native data converged to $R_{cryst} = 27.18\%$ and $R_{free} = 32.09\%$, rather high values. During the last stages of the refinement, it was also noticed that some loops and regions in the second 'domain' of the protein had extremely high B values as a result of its high flexibility. In the $2|F_{obs}| - |F_{calc}|$ maps in the refinement at 1.97 Å these disordered residues located in loops showed up with poor and diffuse electron density as a result of their flexibility, while they were clearly present in the experimental map calculated from the phases obtained by *SHARP* after solvent flattening.

These facts led us to reconsider the refinement steps. The model, as obtained at this stage of refinement with the 1.97 Å data set, was further refined using the 2.65 Å data from the 'remote' wavelength used in the MAD data collection. The refinement with these data proceeded well and many of the disordered residues in some of the loops could be traced without the ambiguity shown in the $2|F_{obs}| - |F_{calc}|$ map from the 1.97 Å data. The B factors also refined to more reasonable values. This phenomenon could be the result of a lack of isomorphism between the high-resolution native data and that from the derivative crystal, particularly as a result of structural differences in the flexible parts of the protein structure within

the native crystal with respect to the Hg-derivative crystal. The *B* factor of the rigid part of the structure, formed mainly of α-helices, remains the same when refined using either data set at different resolutions.

With the 2.65 Å data set the refinement converged to R_{cryst} = 20.08% and R_{free} = 25.87%. This model contained 384 residues and 254 water molecules. Some large peaks which could not be explained by the presence of water molecules remained in the electron-density maps. According to their shape and orientation it was found that these peaks correspond to three molecules of glycerol, one molecule of 2-propanol and a Cl^- ion. The Hg^+ atom bound to Cys277 could also be clearly seen in the difference map. The stereochemistry of the model was analyzed with the programs *CNS* and *PROCHECK* (Laskowski *et al.*, 1993). For the comparison of PelA with Pel10Acm, the program *LSQMAN* (Kleywegt & Jones, 1994) was used.

Molecular graphics were generated with *MOLSCRIPT* (Kraulis, 1991), *BOBSCRIPT* (Esnouf, 1999) and *Raster3D* (Merritt & Bacon, 1997).

3. Results and discussion

3.1. Quality of the structure

The resulting model of the PelA structure at 2.65 Å contains 384 residues, 254 waters, three molecules of glycerol, one molecule of 2-propanol, a Cl^- ion and the Hg^+ ion present in the derivatized crystal. Properties of the final model are summarized in Table 1. The missing residues (151–159 and 355–369) are located in regions of high flexibility and show no visible density in the electron-density maps. Contrary to the missing loop residues 151–159 that are located in the predominantly α-helical 'domain' of the structure, residues 355–369 are located in a long flexible loop in the more flexible 'domain' of the structure. The PelA model has a good geometry, having 99.4% of residues in the most preferred regions in the Ramachandran plot (Ramachandran *et al.*, 1963). One residue, Ala306, is located in a generously allowed region. Another residue, Thr175, is located in the disallowed region. This residue is in a classic γ-turn and has well defined electron density in the $2|F_{obs}| - |F_{calc}|$ map.

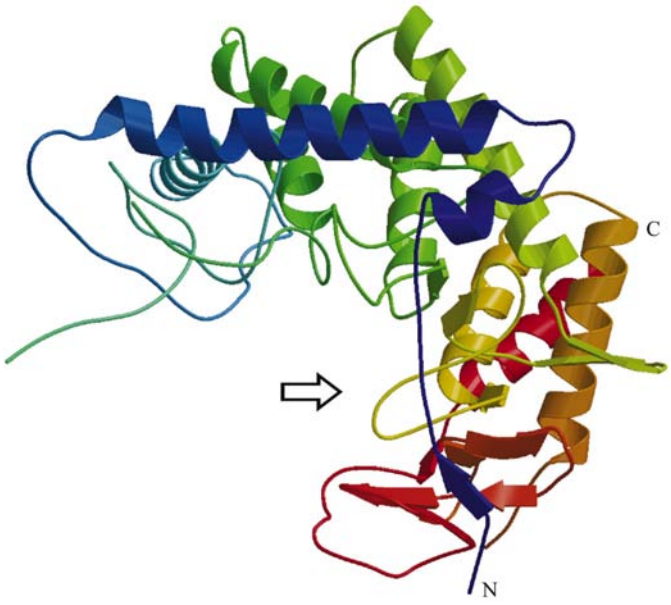


Figure 2
Schematic diagram of the PelA fold. The α-helices are presented as spirals, β-strands are shown as thick arrows in the amino-to-carboxyl direction and connecting loops are depicted as thin lines. This figure was produced using *BOBSCRIPT* (Esnouf, 1999) and *Raster3D* (Merritt & Bacon, 1997). A colour ramp from the N-terminus (blue) to the C-terminus (red) was used. The arrow indicates the region in which the catalytic site is located.

3.2. Overall description

The three-dimensional structure of PelA extends from residues 25 to 432 (Fig. 2) with some disordered loops that could not be found in the electron-density maps (residues Ser151–Ala159 and Arg355–Glu369). The PelA structure has 384 residues distributed in 11 antiparallel β-strands (6%), 11 helices (38.8%) and six 3_{10}-helices (5.5%). The remaining 49.7% of the residues are located in loops and coils.

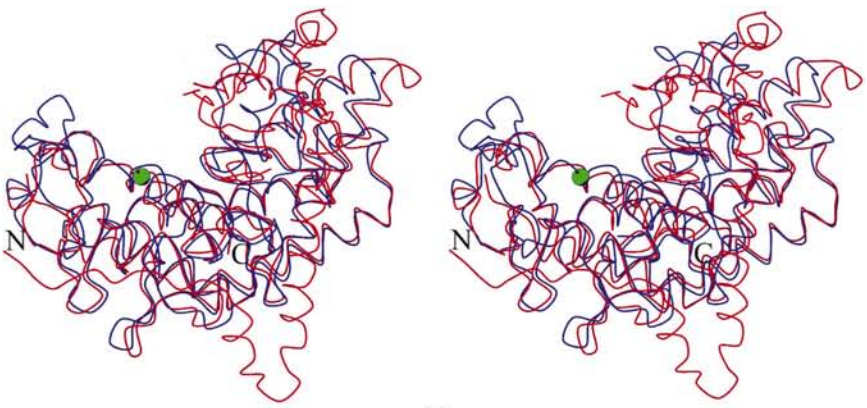
This three-dimensional structure does not adopt the parallel β-helix fold displayed by PeLs from families PL1, PL3 and PL9. Instead, PelA displays a predominant α-helical structure with irregular coils and short β-strands. Recently, the structure of Pel10Acm has also been reported to possess this fold (Charnock *et al.*, 2002).

The structural topology of the PelA protein shows two 'domains' with the interface between them being a wide-open central groove. This groove is supposed to host the substrate-binding and catalytic site, as reported in other endo-acting PLs such as xylanases. This was confirmed by the structure determination of the mutant D189A of Pel10Acm in order to obtain a 'Michaelis' complex with $GalA_3$ (PDB code 1gxo). The N-terminal 'domain' of PelA starts with a short β-strand and an irregular coil and is then followed by a helix (residues 37–42), some β-turns and the largest α-helix in the structure (49–74) of 41 Å in length. These first two helices cross over the last helix (residues 262–282) of an α-toroid (helical domain) formed mainly of α-helices and displaying a compact $(\alpha/\alpha)_3$-barrel.

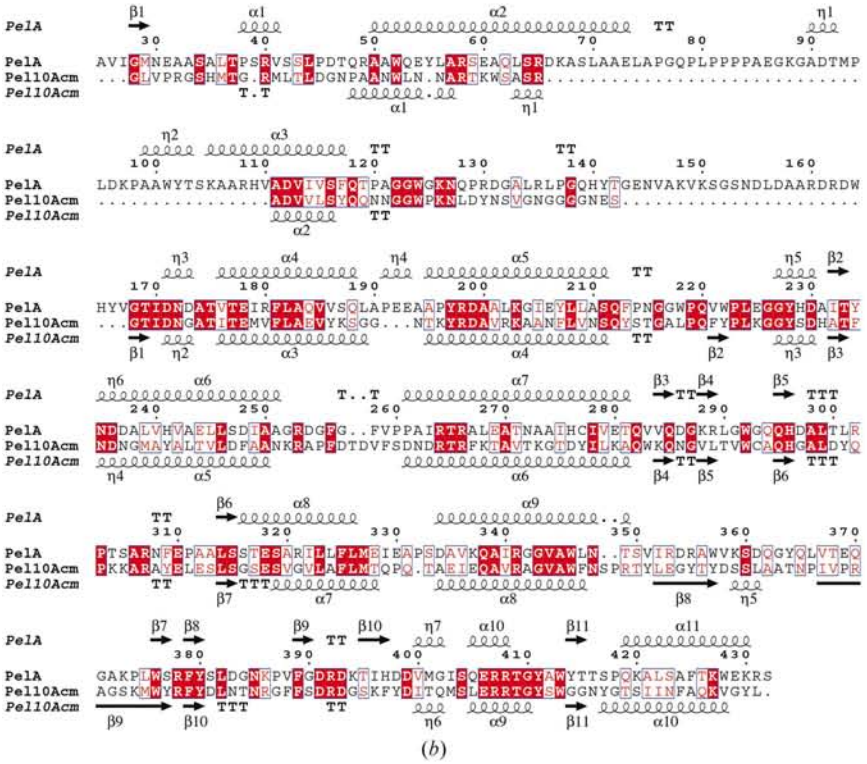
The C-terminal 'domain' is mainly formed of small β-strands, irregular loops and a final α-helix (residues 418–431). This part of the structure is more flexible and has higher *B* factors than the N-terminal domain.

3.3. Comparative studies

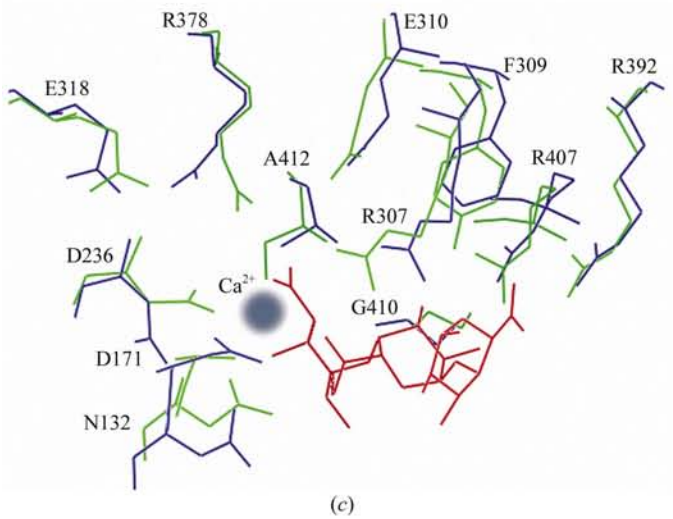
PelA has a predominantly α-helical structure like the recently reported structure of *C. japonicus* Pel10Acm (Charnock *et al.*, 2002; PDB codes 1gxm, 1gxn and 1gxo). Although this type of structure bears no relationship to those of other PeLs, some limited topological similarity to other α-toroidal folds exists, including the $(\alpha/\alpha)_6$- and $(\alpha/\alpha)_7$-toroids displayed by other classes of CAZymes, *e.g.* GH family 47 (α-1,2-mannosidases; PDB code 1hcu). Many PLs acting on other uronic acid polymers, such as the PL5 (alginate lyases) and



(*a*)



(*b*)



(*c*)

PL8 (chondroitin and hyaluronan lyases) families, display variants on the $(\alpha/\alpha)_6$-fold. This suggests that the toroid fold is a powerful and adaptable scaffold for CAZymes, as is the 'classic' β-helix for the PL1 family.

Superimposing the C^{α} atoms of PelA (residues Gln118–Gln128, Val167–Gln188, Ala195–Phe256, Val259–Asp334, Val336–Thr349, Ser350–Asp354 and Lys373–Glu429) with the equivalent C^{α} atoms of Pel10Acm (residues Gln361–Leu371, Ser385–Ser406, Thr410–Phe471, Phe476–Glu551, Ile552–Ser565, Thr568–Gly572 and Ser591–Gly647) gives an r.m.s. deviation of 1.31 Å in *LSQMAN* (Kleywegt & Jones, 1994) and shows that the overall fold in the two structures is quite similar in the selected regions (Fig. 3*a*). Larger sequence variations were found in the N-terminal region than in the C-terminal region.

Fig. 3(*b*) shows the sequence alignment, generated by *CLUSTALW* (Thompson *et al.*, 1994) and prepared using *ESPRIPT* (Gouet *et al.*, 1999), of PelA with the catalytic module of Pel10A. From the N-terminal to residue 105 of PelA the two structures are completely different: this includes a small β-strand, an irregular coil and two α-helices, one of which is the longest helix in the PelA structure (41 Å). From residues 106 to 133 the similarity starts to show until the appearance of a long disordered loop in PelA (residues 134–168 containing the missing residues 151–159); in contrast to this Pel10Acm shows a small loop here. This loop is located above a wall of the substrate-binding cleft and its disorder may be evidence that PelA is in an open conformation. Subsequently, the fold of both structures remains more or less the

Figure 3
Comparison of PelA and Pel10Acm. (*a*) Stereoview of the superimposed C^{α} traces of the PelA (red) and Pel10Acm (blue) structures. The green spot represents the catalytic site. The figure was drawn using *MOLSCRIPT* (Kraulis, 1991). (*b*) Structure-based sequence alignment; numbering is given for PelA. Secondary-structure elements are shown by arrows and helices. Conserved and homologous amino acids are boxed in red and white, respectively. The alignment was produced by *CLUSTALW* (Thompson *et al.*, 1994). The sequence of Pel10Acm was obtained from the SwissProt/TrEMBL database. The figure was produced using *ESPRIPT* (Gouet *et al.*, 1999). (*c*) Superposition of the active-site residues of PelA and Pel10Acm in complex with $GalA_3/Ca^{2+}$. The PelA residues are in blue, Pel10Acm residues in green and the $GalA_3$ molecule in red. The grey sphere represents the position of the Ca^{2+} ion. Single-letter residue codes were used in order to obtain a clearer picture. The molecules were superimposed using *LSQMAN* (Kleywegt & Jones, 1994) and the figure was generated in *O* (Jones *et al.*, 1991). See text for further details regarding equivalent residues.

same except for a long loop that is disordered in PelA (residues 355–369) and is also believed to close the substrate-binding cleft, as does the other disordered loop. The so-called C-terminal 'domain' is also quite similar in both structures, showing a longer terminal α-helix of 22 Å in PelA (residues 418–431).

3.4. Active site

In order to study the active site of PelA, different co-crystallization experiments and soaks with different substrates (GalA, $GalA_2$ and $GalA_3$) were tried. These substrates were selected because they constitute the natural degradation products of the enzyme. These trials were also combined with $CaCl_2$ in the crystallization conditions. None of these trials revealed ligands or Ca^{2+} in the electron-density maps. It is noteworthy that co-crystallization experiments of native Pel10Acm with substrates did not reveal the presence of ligand or Ca^{2+} (Charnock *et al.*, 2002).

After refinement of the PelA structure, a glycerol molecule (an additive of the cryoprotectant) was found bound to the key residue Asn172. This glycerol mimics the sugar moiety and is obviously more favoured than solvent water molecules. This finding suggests minor site-directed mutagenesis in order to optimize this region to obtain a more specific binding for GalA. Charnock and coworkers succeeded in crystallizing and solving the structure of the 'Michaelis' complex of the D189A mutant of Pel10Acm (Asp189 corresponds to Asp171 in PelA) with $GalA_3$ (PDB code 1gxo). In this case, the trisaccharide occupies subsites −1 to +2, consistent with the unique mode of bond cleavage for this substrate, and allowed description of the enzyme–substrate interaction.

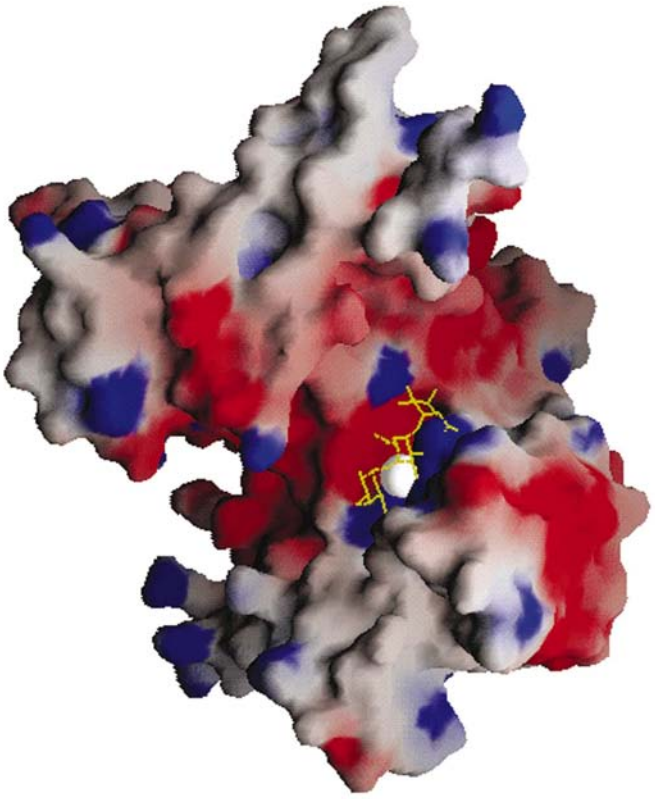


Figure 4
Electrostatic potential surface of PelA with fitted substrate. Red represents negative potential, blue positive and white neutral. The fitted $GalA_3$ is shown in yellow and behind it is the Ca^{2+}, shown as a white sphere. The figure was prepared with *GRASP* (Nicholls *et al.*, 1991).

The active site in Pel10Acm is located in a buried wide-open central groove between the two 'domains' of the protein. Because of the topological similarities between PelA and Pel10Acm, the C^{α} atoms comprising the active site of both proteins were superimposed and best-fitted using *LSQMAN* (Kleywegt & Jones, 1994), minimizing the C^{α} atomic distances of the amino acids between both structures (r.m.s. deviation of 0.94 Å). The $GalA_3$ model and Ca^{2+} were placed in the PelA active site in the same orientation as the $GalA_3$ and Ca^{2+} in Pel10Acm (Fig. 3*c*). The active-site residues in PelA10cm are Asp389 (mutated to Ala389), Asn390, Asp451, Arg524, Tyr526, Glu527, Glu535, Arg596, Arg610, Arg625, Gly628 and Ser630. These residues are highly conserved with their counterparts in PelA (Asp171, Asn172, Asp236, Arg307, Phe309, Glu310, Glu318, Arg378, Arg392, Arg407, Gly410 and Ala412).

At the −1 (leaving-group) subsite, Arg596 in Pel10Acm forms a salt bridge with the substrate carboxylate group and a coordination bond with the Ca^{2+} ion. In the PelA structure, the equivalent Arg378 is present in another rotamer conformation. However, it is likely that upon binding of the substrate the rotamer present in Pel10Acm is preferred, favouring the formation of this salt bridge. In Pel10Acm, the Ca^{2+} ion is also coordinated by an O atom from the +1 subsite sugar carboxylate and one carboxylate O atom from Asp451, with three water molecules completing the hexacoordination. This coordination of the Ca^{2+} ion by the carboxylate O atom from Asp451 is obviously not present in PelA, which displays another rotamer for this residue (Asp236). Charnock and coworkers showed that substitution of Asp451 led to a total loss of activity towards polygalacturonic acids (Charnock *et al.*, 2002). The side chain of Asp236 of PelA lies at the end of a helical segment that is held in its appropriate main-chain conformation through direct hydrogen bonds from both its flanking main-chain amide groups through to the side-chain carboxylate Glu318, which is present as the same rotamer in both structures. Mutation of Glu535, which is equivalent to Glu318 in PelA, reduces the activity about 200-fold (Charnock *et al.*, 2002) even towards $GalA_3$, from which it lies at a distance of some 7–10 Å, suggesting that it plays a role in maintaining the structural integrity of the catalytic centre groups.

The chemistry of these enzymes happens in the +1 galacturonic acid moiety, in which the Arg625 in PelA10cm forms a hydrogen bond with both the C2 and C3 hydroxyl groups of this subsite. In the PelA case the interaction with the C2 has a distance of 3.9 Å. These interactions are also observed in family 11 xylanases (Sabini *et al.*, 1999). The possible interaction between the O3 and the main-chain carbonyl of Gly628 is also not present in Gly410 of PelA.

The side chain of Arg307 (Arg524 in Pel10Acm) is within hydrogen-bonding distance of Glu310 (Glu527 in Pel10Acm),

forming two hydrogen bonds in both protein structures. This arginine is the only potential catalytic base in close proximity to C5—H5 in the +1 subunit, with a distance of 2.7 Å between its side chain and C5 (2.5 Å in the case of Pel10Acm).

The hydrogen bond formed between the hydroxyl group of Tyr526 and the +2 sugar carboxylate in Pel10Acm is obviously not present in PelA, which has a phenylalanine instead. Regardless of this difference, both types of residues superimpose well (see Fig. 3*c*) in their aromatic rings.

Arg610 forms a salt bridge with the +2 sugar carboxylate group at C6 in Pel10Acm that is also possible for Arg392 of PelA with $GalA_3$ modelled in the active site. According to Charnock *et al.* (2002), the involvement of the −1 and +2 subsite carboxylate groups in the salt-bridge formation and the coordination of the −1 and +1 subsite carboxylate O atoms with Ca^{2+} may contribute to the specificity of these enzymes for homopolygalacturonic acid. The Asp389 in Pel10Acm located between the +1 and +2 subsites was shown to be catalytically inactive when mutated to Ala. However, its position is isostructural with that of the native Pel10Acm and PelA.

The electrostatic potential surface of the model complex with $GalA_3$ (Fig. 4) was calculated with *GRASP* (Nicholls *et al.*, 1991). PelA has an acidic active-site groove consistent with the measured pI value of 5.74.

Since the orientation of the conserved catalytic residues in PelA is similar to that in PelA10cm, we infer that the enzymatic mechanisms of the enzymes may be similar. Fig. 5 describes this catalytic mechanism.

The PL10 family reveals an example of convergent evolution of catalytic mechanism. Comparison of the Pel10Acm $GalA_3/Ca^{2+}$ complex with the inactive mutant R218K $GalA_4/Ca^{2+}$ complex of Pel1C, a member of the PL1 family (Charnock *et al.*, 2002; Herron *et al.*, 2000), shows an identical disposition of six active-centre groups despite the fact that these two families of enzymes have different topologies (the 'β-helix' motif for PL1).

Both families of enzymes perform anti β-elimination. This means that the two leaving groups are on opposite sides of the incipient double bond. The active-centre similarity of this enzyme with structurally unrelated anti-eliminases from family PL1 allows some description of the likely reaction trajectory (Charnock *et al.*, 2002). The only potential catalytic base in proximity to a substrate C5 H atom is Arg524 in Pel10Acm (Arg307 in PelA), which is located at the putative +1 subsite (see Fig. 3*c*). This location is identical to the putative arginine base in PL1 enzymes (Herron *et al.*, 2000). The conclusion of Charnock and coworkers is that a deprotonated arginine functions as the catalytic base in both systems. Arginine is a unusual base in enzymatic reactions, but it is perhaps not surprising given the high pH optima of pectate lyases.

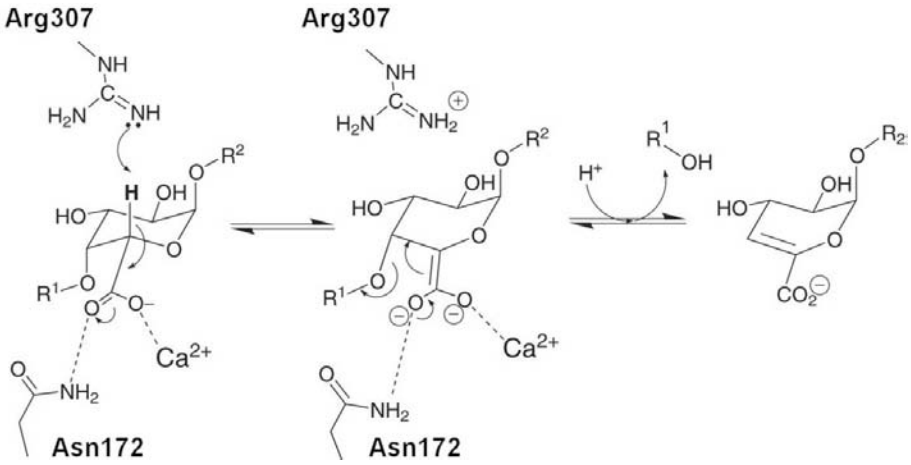


Figure 5
Putative reaction mechanism for PL10 enzymes, as proposed by Charnock *et al.* (2002). Proton abstraction by arginine is followed by leaving-group elimination.

4. Conclusions

The structure of PelA was solved to a resolution of 2.65 Å. The overall structure of PelA does not adopt the characteristic parallel β-helix fold displayed by PeLs from families PL1, PL3 and PL9. Instead, it displays a predominant α-helical structure with irregular coils and short β-strands, as in the recently reported structure of *C. japonicus* pectate lyase Pel10Acm. The structural topologies of the PelA and Pel10Acm structures were compared. They show two 'domains' (a rigid N-terminal 'domain' and a more flexible C-terminal 'domain'), with the interface between them being a wide-open central groove in which the active site is located. The active sites of the crystal structures were also compared and their similarities and differences were discussed. The evidence for convergent evolution of the catalytic mechanism in the PL1 and the PL10 families arises from the fact that there is an identical disposition of the six active-centre groups despite the two enzyme families having different topologies and low homology similarities.

HNDA would like to thank K. U. Leuven, Belgium for an IRO scholarship. AR is a Postdoctoral Research Fellow of the Fund for Scientific Research-Flanders (Belgium). ChV is a Postdoctoral Research Fellow of the Research Fund K. U. Leuven (Belgium). We acknowledge the support of the protein crystallography project team at the EMBL BW7B beamline at the DORIS storage ring, DESY, Hamburg and the staff of beamline 5.2R, ELETTRA, Trieste. The European Union is acknowledged for supporting the work at EMBL Hamburg (HCMP to Large Installations Projects, contract No. HPRI-CT-1999-00017) and at Sinchrotrone Trieste (Human Potential Programme Transnational Access to Major Research Infrastructures, contract No. HPRI-CT-1999-00033).

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Akita, M., Suzuki, A., Kobayashi, T., Ito, S. & Yamane, T. (2000). *Acta Cryst.* D**56**, 749–750.
Anderson, V. E. (1998). *Comprehensive Biological Catalysis: A Mechanistic Reference*, edited by M. Sinnott, Vol. 2, pp. 115–133. London: Academic Press.
Barras, F., van Gijsegem, F. & Chatterjee, A. K. (1994). *Annu. Rev. Phytopathol.* **32**, 201–234.
Bekri, M. A., Desair, J., Keijers, V., Proost, P., Searle-van Leeuwen, M., Vanderleyden, J. & Vande Broek, A. (1999). *J. Bacteriol.* **181**, 2440–2447.
Brown, I. E., Mallen, M. H., Charnock, S. J., Davies, G. J. & Black, G. W. (2001). *Biochem. J.* **335**, 155–165.
Brünger, A. T. (1991). *Ann. Rev. Phys. Chem.* **42**, 197–223.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Acta Cryst.* A**46**, 585–593.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
Charnock, S. J., Brown, I. E., Turkenburg, J. P., Black, G. W. & Davies, G. J. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 12067–12072.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Coutinho, P. M. & Henrissat, B. (1999). *Recent Advances in Carbohydrate Bioengineering*, edited by H. J. Gilbert, G. Davies, B. Henrissat & D. Svenssons, pp. 3–12. Cambridge: The Royal Society of Chemistry.
Engh, R. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Furey, W. & Swaminathan, S. (1990). *Am. Crystallogr. Assoc. Meet. Abstr.* **18**, 73.
Gonzalez-Candelas, L. & Kolattukudy, P. E. (1992). *J. Bacteriol.* **174**, 6343–6349.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Guo, W. L., Gonzalez-Candelas, L. & Kolattukudy, P. E. (1995*a*). *Arch. Biochem. Biophys.* **232**, 352–360.
Guo, W. L., Gonzalez-Candelas, L. & Kolattukudy, P. E. (1995*b*). *J. Bacteriol.* **177**, 7070–7077.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Henrissat, B., Coutinho, P. M. & Davies, G. J. (2001). *Plant Mol. Biol.* **47**, 55–72.
Henrissat, B., Heffron, S. E., Yoder, M. D., Lietzke, S. E. & Jurnak, F. (1995). *Plant Physiol.* **107**, 963–976.
Herron, S. R., Benen, J. A. E., Scavetta, R. D., Visser, J. & Jurnak, F. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 8762–8769.
Hugouvieux-Cotte-Pattat, N., Condemine, G., Nasser, W. & Reverchon, S. (1996). *Annu. Rev. Microbiol.* **50**, 213–257.
Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.
Jones, T. A., Zou, J. Y., Cavan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Jones, T. A. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **30**, 9–14.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lamzin, V. S., Perrakis, A. & Wilson, K. S. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 720–722. Dordrecht: Kluwer Academic Publishers.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Levitt, D. G. (2001). *Acta Cryst.* D**57**, 1013–1019.
Lietzke, S. E., Keen, N. T., Yoder, M. D. & Jurnak, F. (1994). *Plant Physiol.* **106**, 849–862.
Lojkowska, E., Masclaux, C., Boccara, M., Robert-Baudouy, J. & Hugouvieux-Cotte-Pattat, N. (1995). *Mol. Microbiol.* **16**, 1183–1195.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Nicholls, A., Sharp, K. & Honing, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.
Novoa de Armas, H., Rabijns, A., Verboven, C., Desair, J., Vande Broek, A., Vanderleyden, J. & De Ranter, C. (2002). *Acta Cryst.* D**58**, 292–295.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pickersgill, R., Jenkins, J., Harris, G., Nassar, W. & Robert-Baudouy, J. (1994). *Nature Struct. Biol.* **1**, 717–723.
Pissavin, C., Robert-Baudouy, J. & Hugouvieux-Cotte-Pattat, N. (1996). *J. Bacteriol.* **178**, 7187–7196.
Powell, M. J. D. (1977). *Math. Prog.* **12**, 241–254.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jørgensen, P. L., Schülein, M., Dupont, C., Davies, G. J. & Wilson, K. S. (1999). *Chem Biol.* **6**, 483–492.
Sawada, K., Ogawa, A., Ozawa, T., Sumitomo, N., Hatada, Y., Kobazyashi, T. & Ito, S. (2000). *Eur. J. Biochem.* **267**, 1510–1515.
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.
Shevchik, V. E. & Hugouvieux-Cotte-Pattat, N. (1997). *Mol. Microbiol.* **24** 1285–1302.
Shevchik, V. E., Robert-Baudouy, J. & Hugouvieux-Cotte-Pattat, N. (1997). *J. Bacteriol.* **179**, 7321–7330.
Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. (1998). *Acta Cryst.* D**54**, 1078–1084.
Tardy, F., Nasser, W., Robert-Baudouy, J. & Hugouvieux-Cotte-Pattat, N. (1997). *J. Bacteriol.* **179**, 2503–2511.
Thomas, L. M., Doan, C. N., Oliver, R. L. & Yoder, M. D. (2002). *Acta Cryst.* D**58**, 1008–1015.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Yoder, M. D., Keen, N. T. & Jurnak, F. (1993). *Science*, **260**, 1503–1507.
Yoder, M. D., Lietzke, S. E. & Jurnak, F. (1993). *Structure*, **1**, 241–251.