Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a yellow lupin pathogenesis-related PR-10 protein belonging to a novel subclass

Oliwia Pasternak,[a] Jacek Biesiadka,[b] Rafal Dolot,[c] Luiza Handschuh,[a] Grzegorz Bujacz,[a,c] Michal M. Sikorski[a] and Mariusz Jaskolski[a,d]*

[a]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, [b]Department of Crystallography, Free University, Berlin, Germany, [c]Institute of Technical Biochemistry, Technical University of Lodz, Poland, and [d]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland

Correspondence e-mail: mariuszj@amu.edu.pl

Received 7 September 2004
Accepted 2 November 2004

PDB Reference:
pathogenesis-related PR-10 protein, 1xdf, r1xdfsf.

Pathogenesis-related (PR) proteins of class 10 are abundant in higher plants. Some of these proteins are induced under stress conditions as part of the plant defence mechanism. Other homologues are developmentally regulated and their expression varies in different plant organs. The PR-10 proteins are encoded by multigene families, have a weight of about 17 kDa and are found in the cytosol. In yellow lupin, nine different homologues have been identified and divided into two subclasses, LlPR-10.1 and LlPR-10.2. Within each subclass the sequence identity is about 75–91%, while across the subclasses it is only 59–60%. Here, the crystal structure of a yellow lupin PR-10 protein from the second subclass, LlPR-10.2A, is presented. The structure was solved by molecular replacement and refined to R = 0.205 using 1.9 Å resolution data. The general fold of LlPR-10.2A resembles that of the other PR-10 proteins and consists of a long C-terminal α-helix surrounded by a seven-stranded antiparallel β-sheet, with two shorter α-helices located between strands β1 and β2. The most variable part of the structure, the C-terminal helix, is strongly kinked towards the β-sheet core in both LlPR-10.2A molecules present in the asymmetric unit. This unexpected feature reduces the size of the hydrophobic cavity observed in other PR-10 proteins that is reported to be the ligand-binding site. As in other PR-10 structures, a surface loop located near the entrance to the cavity shows very high structural conservation and stability despite the high glycine content in its sequence.

1. Introduction

Plants respond to pathogen infection or environmental stress by increasing the expression of a number of genes that encode pathogenesis-related (PR) proteins. These are operationally defined as inducible factors expressed as part of the active plant defence mechanism. The biological functions of several classes of defence proteins have been recognized, including chitinases, β-glucanases, peroxidases and protein inhibitors (van Loon *et al.*, 1994). The role of other PR proteins in the defence response still remains to be elucidated. To date, PR proteins have been classified into 14 families (van Loon & van Strien, 1999). PR-10 class proteins are ubiquitous in the plant kingdom and are encoded by multigene families. They are small (155–163 amino acids) slightly acidic intracellular proteins and are resistant to proteases. They were first discovered in the plant response to pathogen infection and described as intracellural pathogenesis-related (IPR) proteins. IPR proteins have allergenic properties and can be found in pollen grains (major pollen allergens; Breiteneder *et al.*, 1989, 1992, 1993; Larsen *et al.*, 1992), fruits and vegetables (major food allergens; Vanek-Krebitz *et al.*, 1995; Breiteneder *et al.*,

1995). Another subfamily of PR-10 proteins, called major latex proteins (MLP), are present in the latex of some plants, including the opium poppy (Nessler & Burnett, 1992; Osmark *et al.*, 1998) and bell pepper (Pozueta-Romero *et al.*, 1995). Almost all these proteins cause IgE-mediated type I allergy. In this group, the best studied allergen is the birch (*Betula verrucosa*) pollen protein Bet v 1, the crystal structure of which was described by Gajhede *et al.* (1996). The allergenic aspect has been studied by crystallography for a Bet v 1–IgG Fab complex (Spangfort *et al.*, 1999) and in competition experiments with the human IgE fraction of allergic patients (Mirza *et al.*, 2000). In that work, the monoclonal murine IgG antibody against the conservative glycine-rich loop of Bet v 1 and other PR-10 proteins was used. On the basis of weak sequence homology and secondary-structure prediction, an additional distinct group of proteins, known as cytokinin-specific binding proteins (CSBPs; Fujimoto *et al.*, 1998), has been tentatively included in the PR-10 class.

Although the biological function of PR-10 proteins has not yet been determined, their gene-expression patterns, accumulation in different plant organs and physicochemical properties are well recognized. Many genes encoding PR-10 proteins show organ- or tissue-specific expression and appear to be developmentally regulated (Crowell *et al.*, 1992; Barratt & Clark, 1993; Mylona *et al.*, 1994; Constabel & Brisson, 1995; Walter *et al.*, 1996; Sikorski *et al.*, 1999). Some of them are expressed constitutively (Sikorski *et al.*, 1999), while others are induced in aerial parts of the plant by various stress factors, *e.g.* wounding (Warner *et al.*, 1992, 1993), UV radiation (Pinto & Ricardo, 1995), chemical pollutants (Utriainen *et al.*, 1998), or in response to pathogen invasion (Somssich *et al.*, 1986, 1988; Breda *et al.*, 1996; Pinto & Ricardo, 1995; McGee *et al.*, 2001). Presumably, PR-10 proteins constitutively expressed in the roots are involved in self-defence and those inducible in various organs are involved in the plant developmental programme. However, the coexistence of many homologous genes in one plant species makes functional studies very difficult.

One of the hypotheses concerning the biological function of PR-10 proteins is based on their sequence similarity to a ribonuclease isolated from phosphate-starved ginseng cells (Moiseyev *et al.*, 1994, 1997). Ribonuclease activity was also found for the birch pollen allergen Bet v 1 (Bufe *et al.*, 1996; Swoboda *et al.*, 1996) as well as for PR-10 proteins from white (Bantignies *et al.*, 2000) and yellow lupin (Sikorski, unpublished results) and from hot pepper (Park *et al.*, 2004). However, the high abundance of PR-10 proteins and various expression models with no relation to pathogen response make the RNase hypothesis controversial. Another explanation of a possible biological function of PR-10 proteins comes from structural studies of their ligand complexes. Recently, Mogensen *et al.* (2002) have used an ANS (8-anilino-1-naphthalenesulfonic acid) displacement assay to identify a range of physiologically relevant ligands, such as fatty acids, flavonoids and cytokinins, that may bind to the Bet v 1 allergen. ANS binds to Bet v 1 with 1:1 stoichiometry and NMR data indicate that the binding takes place in an internal cavity. The ability of the tested ligands to displace ANS indicates that they also bind in the cavity, although the exact binding sites seem to vary among different ligands. A crystallographic study of a complex between Bet v 1 and deoxycholate reported by Markovic-Housley *et al.* (2003) shows that the Y-shaped cavity is partly occupied by two deoxycholate molecules that are bound in tandem. The structural similarity of deoxycholate and brassinosteroids, which are ubiquitous plant steroid hormones, may indicate that the pollen-allergen protein is responsible for transport of phytosteroid hormones. Further mass-spectrometric analysis has revealed a specific non-covalent interaction of Bet v 1 with a brassinolide or 24-epicastasterone. This, together with the discovery of the very high affinity of CSBP proteins for another class of plant hormones, cytokinins (Fujimoto *et al.*, 1998), may indicate a general plant-hormone-carrier function for PR-10 proteins in the plant defence response to pathogen invasion, as well as in growth and developmental processes.

Our research is focused on the characterization of yellow lupin (*Lupinus luteus*) PR-10 proteins. We have identified nine homologous *pr-10* clones in a yellow lupin cDNA library encoding proteins that can be classified into two different subclasses: LlPR-10.1 (three homologues) and LlPR-10.2 (six homologues). The subclassification is based on different levels of gene transcripts and protein accumulation and on differences in the length and similarity of the polypeptide sequences. The suffixes A, B, C... after the decimal digit designating the subclasses are added to distinguish the different homologues. We have demonstrated previously (Sikorski *et al.*, 1999) that the lupin PR-10 proteins of subclass LlPR-10.1 are abundant in roots and are developmentally and differentially regulated during plant growth in various organs. Our recent data (Handschuh, 2004) show that the lupin *pr-10.2* gene homologues are also involved in plant developmental processes; however, their expression level is lower in comparison with the *pr-10.1* subfamily. The level of amino-acid identity across the subclasses does not exceed 60%, while it is as high as 91% within a subclass. The most variable region among all PR-10 proteins is the 25-amino-acid C-terminal fragment that folds into an α-helix (α3). There is only 50% amino-acid identity in this C-terminal fragment between the two lupin subclasses. The degree of variability in this region may be crucial for protein–ligand interactions and stability of biologically relevant complexes, as helix α3 forms one wall of the ligand-binding site. So far, the three-dimensional structure of two homologues representing yellow lupin subclass PR-10.1 have been determined by X-ray crystallography (Biesiadka *et al.*, 2002). Moreover, the structure of PR-10 proteins from two other plant species, white birch (*B. verrucosa*; Gajhede *et al.*, 1996) and cherry (*Prunus avium*; Neudecker *et al.*, 2001), have been described (Table 1). Although the PR-10 proteins are restricted to the plant world, they bear remarkable structural similarity (but not sequence similarity) to mammalian lipid-transfer domains (START domains) such as phosphatidylcholine-transfer protein (Roderick *et al.*, 2002) or the START domain of MLN64 (Tsujishita & Hurley, 2000). The present paper describes the crystal structure of a yellow lupin

Table 1
PR-10 protein structures in the PDB.

Source	Name	Method	Resolution (Å)	No.†	PDB code	Reference
White birch (*B. verrucosa*)	Bet v 1	X-ray	2.0	1	1bv1	Gajhede *et al.* (1996)
	Bet v 1	NMR	—	20	1btv	Gajhede *et al.* (1996)
	Bet v 1A	NMR	—	23	1b6f	Schweimer *et al.* (1999)
	Bet v 1A complex with antibody	X-ray	2.9	—	1fsk	Mirza *et al.* (2000)
	Bet v 1 deoxycholate complex	X-ray	1.9	1	1fm4	Markovic-Housley *et al.* (2003)
	Bet v 1 E45S mutant	X-ray	3.1	—	1llt	Spangfort *et al.* (2003)
	Bet v 1A N28T, K32Q, E45S, P108G mutant	X-ray	2.15	—	1qmr	To be published
Cherry (*P. avium*)	Pru av 1	NMR	—	22	1e09	Neudecker *et al.* (2001)
	Pru av 1 E45W mutant	NMR	—	24	1h2o	Neudecker *et al.* (2003)
Yellow lupin (*L. luteus*)	LlPR-10.1A	X-ray	1.95	1	1icx	Biesiadka *et al.* (2002)
	LlPR-10.1B	X-ray	2.25	2	1ifv	Biesiadka *et al.* (2002)
	LlPR-10.2A	X-ray	1.9	2	1xdf	This work

† Number of molecules in the asymmetric unit (X-ray) or number of model structures (NMR).

PR-10 protein from the second subclass (LlPR-10.2A) and discusses its comparison with the previously reported PR-10 structures. Since the current subclassification of the lupin PR-10 proteins reflects the high degree of sequence variation in the C-terminal fragment, it is of interest to determine whether these differences are translated into structural variations and ultimately into ligand-binding specificity.

2. Materials and methods

2.1. Protein preparation and purification

The protein was overexpressed in *Escherichia coli* cells under the T7 promoter (Studier *et al.*, 1990). The cDNA encoding LlPR-10.2A was subcloned into a pET-3a plasmid and expressed in bacterial strain BL21(DE3)pLysS using a modified variant of an earlier protocol (Sikorski, 1997). The cells were collected and kept overnight at 253 K. After freezing and thawing in buffer *L* (20 m*M* sodium phosphate pH 7.5, 5% glycerol, 10 m*M* β-mercaptoethanol), the lysed bacteria were sonicated for shearing chromosomal DNA and the inclusion bodies were pelleted. The inclusion bodies were treated with 7.2 *M* urea in buffer *L* at 310 K for 1 h. The lysate was centrifuged at 30 000*g* for 15 min at 277 K and the supernatant was fractionated with ammonium sulfate. The protein precipitate was resuspended in buffer *L*. After dialysis against buffer *L*, the soluble protein fraction was applied onto a DE52-cellulose column for removal of lipids and nucleic acids. The recombinant LlPR-10.2A protein was eluted using a stepwise NaCl gradient in buffer *L*. Fractions eluted between 50 and 100 m*M* NaCl were pooled, dialyzed against buffer *L* and submitted for FPLC separation. The first step was ion-exchange chromatography on a MonoQ column followed by size-exclusion chromatography on Superdex 75. Homogenous protein fractions were collected and fractions that were contaminated with a 30 kDa bacterial protein were dialyzed against buffer *D* (20 m*M* Tris–HCl pH 8.0, 5% glycerol, 10 m*M* β-mercaptoethanol) and purified as above. All homogenous protein fractions were pooled together and dialyzed against 3 m*M* sodium citrate buffer pH 6.3 and concentrated to 20–30 mg ml^{-1} using Millipore Centricon 10 filters.

Table 2
Statistics of data collection and processing.

Values in parentheses are for the last shell.

Space group	$P2_12_12_1$
Unit-cell parameters	
a (Å)	46.9
b (Å)	64.9
c (Å)	107.9
Molecules in AU	2
Radiation source	BNL NSLS, beamline X9B
Temperature (K)	100
No. measured reflections	133637
No. unique reflections	30830
Resolution limits (Å)	25.0–1.9 (1.97–1.90)
$R_{\rm int}$	0.052 (0.412)
Completeness (%)	99.2 (97.8)
$\langle I/\sigma(I)\rangle$	28.3 (2.7)
Redundancy	4.3 (3.2)

Table 3
Refinement statistics.

Programs used	*CNS* v.1.1/*REFMAC*5
Resolution limits (Å)	20–1.9
No. reflections	28579
No. reflections in test set	1193
Protein atoms	2346
Metal cations	1
Solvent atoms	296
$R/R_{\rm free}$	0.205/0.251
$\langle B\rangle$ for protein atoms (Å^2)	26.1
$\langle B\rangle$ for solvent atoms (Å^2)	56.4
R.m.s.d. from ideal bond lengths (Å)	0.019
R.m.s.d. from ideal bond angles (°)	1.78
Most favoured Ramachandran φ/ψ angles (%)	92.5

2.2. Crystallization

Crystallization was carried out using the hanging-drop vapour-diffusion technique. The protein formed crystals over a reservoir containing 1.2 *M* sodium citrate and 0.2 *M* HEPES buffer pH 7.5. The crystallization droplet contained 2 µl of a 6 mg ml^{-1} protein solution and 2 µl reservoir solution. The crystals grew as aggregated plates. To obtain a single well formed crystal, the microseeding technique was used. The specimen selected for data collection had dimensions of 0.4 × 0.25 × 0.1 mm. Prior to data collection, the crystal was soaked in a cryoprotectant solution containing 20% glycerol and 80%

reservoir solution. The crystal was mounted in a fibre loop and flash-frozen (Teng, 1990) directly in the cold nitrogen cryostream at the synchrotron beamline.

2.3. X-ray data collection and processing

X-ray diffraction data extending to 1.9 Å resolution were collected at beamline X9B of the Brookhaven Synchrotron using an image-plate MAR 18 cm detector. The images were indexed and integrated using *DENZO* and scaled in *SCALEPACK* (Otwinowski & Minor, 1997), yielding a unique data set characterized by an $R_{\rm int}$ of 0.052. The space group is $P2_12_12_1$, with unit-cell parameters $a = 46.9$, $b = 64.9$, $c = 107.9$ Å. The Matthews coefficient (Matthews, 1968) of 2.71 Å^3 Da^{-1} indicated the presence of two LlPR-10.2A molecules in the asymmetric unit and a solvent content of 54.2%. The data-collection and processing statistics are summarized in Table 2.

2.4. Structure solution

The structure was solved by molecular replacement using the program *AMoRe* (Navaza, 1994). The coordinates of an LlPR-10.1B molecule (PDB code 1ifv) were used as a search model (the sequence identity between the search and target protein was 58%). The translation function located two copies of the search probe within the asymmetric unit with the same rotation parameters, corresponding to the top peak of the rotation function. After rigid-body refinement, the solution was characterized by a correlation coefficient of 0.55 and an R factor of 0.49.

2.5. Structure refinement

The initial refinement was carried out using the *CNS* program package (Brünger *et al.*, 1998) and a standard protocol consisting of simulated annealing followed by an optimization of the atomic positions and B factors. The graphics program *O* (Jones *et al.*, 1991) was used for manual modelling according to electron-density maps. In the early phase of the refinement, NCS restraints were applied to the calculations. In the final stage of the refinement, *REFMAC*5 (Murshudov *et al.*, 1997) was used with the TLS parameters (Winn *et al.*, 2001) defined for each of the two LlPR-10.2A molecules in the asymmetric unit. Both molecules required substantial rebuilding, mainly in the region of loops L5, L7 and L9 and at the end of the C-terminal helix. The inclusion of the TLS parameters in the refinement significantly improved the R and $R_{\rm free}$ factors by 0.045 and 0.034, respectively. Water molecules were located using the *PEAKMAX* and *WATPEAK* programs from the *CCP*4 suite. They were retained if they formed at least one hydrogen bond (between 2.6 and 3.5 Å) and corresponded to a peak at least three times the σ level of an $F_{\rm o} - F_{\rm c}$

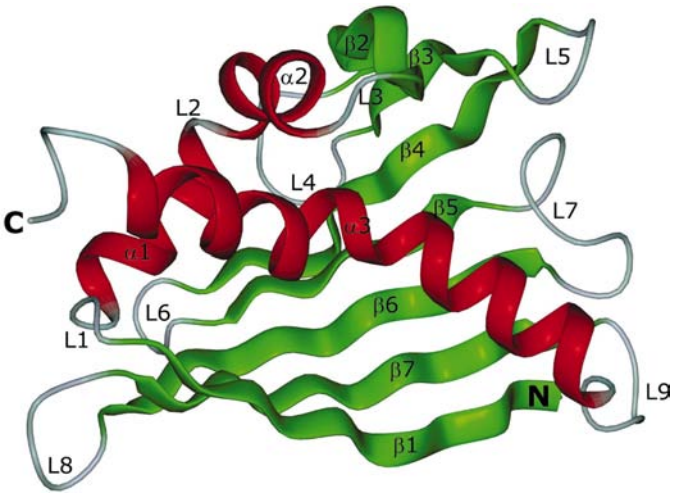

Figure 1
The overall fold of the LlPR-10.2A protein. The annotated secondary-structure elements are indicated in red (α-helices), green (β-strands) and grey (loops). All structural figures were prepared using *DINO* (http://www.dino3d.org)

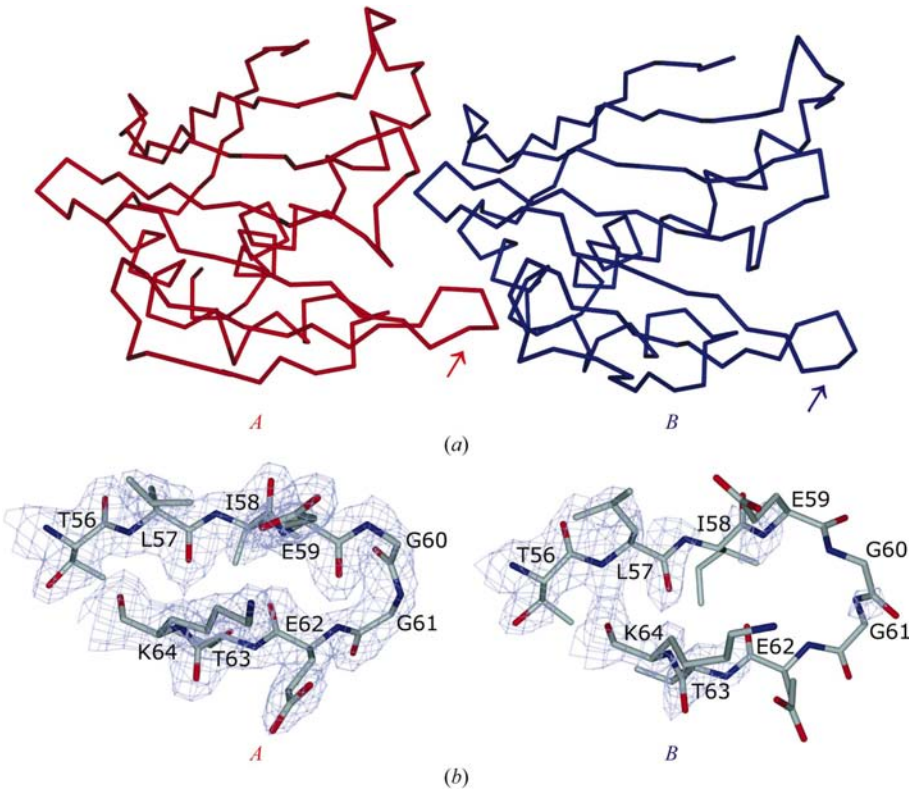

Figure 2
The asymmetric unit of LlPR-10.2A. (*a*) C$^\alpha$ trace of the two LlPR-10.2A molecules in the asymmetric unit. Molecule *A* is shown in red and molecule *B* in blue. Arrows indicate the position of loop L5. (*b*) Loops L5 from both LlPR-10.2A molecules in the asymmetric unit, with the corresponding $2F_{\rm o} - F_{\rm c}$ map contoured at 1.5σ. The figure illustrates that the ordered nature of loop L5 of molecule *A* (in contrast to that of molecule *B*) is caused by crystal contacts between the two molecules.

map and at least the σ level of a $2F_o - F_c$ map. All automatically added water molecules were checked manually in *O*. One of the highest difference peaks was located within the L3 loop of molecule *A* and had five close $X\cdots O$ contacts (about 2.6 Å) with the protein (four carbonyl groups) or solvent (one water molecule). Based on the coordination scheme (incomplete and deformed octahedron) and the composition of the purification and crystallization buffers (elution with NaCl and precipitation with 1.2 *M* sodium citrate), this peak was interpreted as a sodium cation. The refinement statistics are summarized in Table 3.

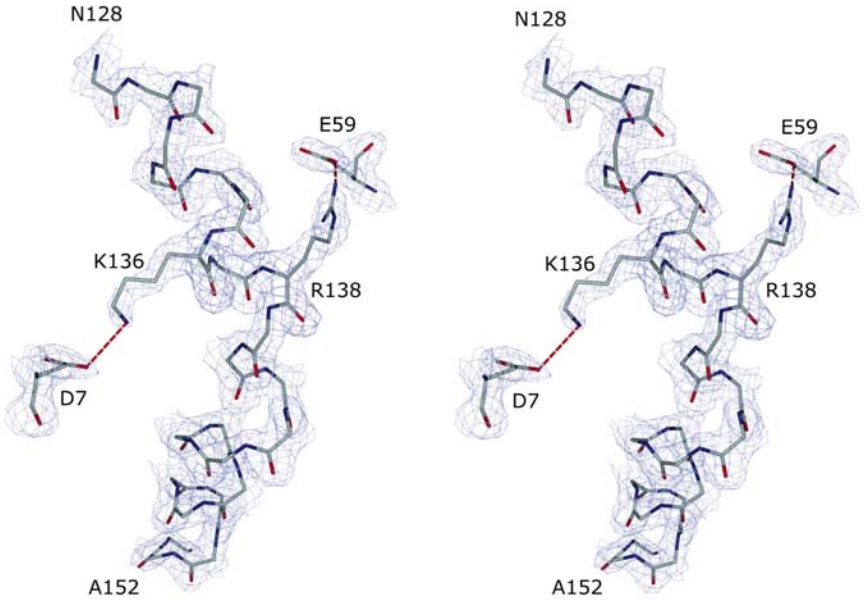

Figure 3
A stereoview illustrating the conformation of helix α3 (shown without side chains, except Lys136 and Arg138) in LlPR-10.2A (molecule *A*). A kink of almost 60° in the middle of the helix is well supported by the $2F_o - F_c$ electron-density map (contoured at the 1.0σ level). The helix is stabilized by salt bridges formed by Lys136 and Arg138 and by hydrophobic interactions with the β-sheet (not shown).

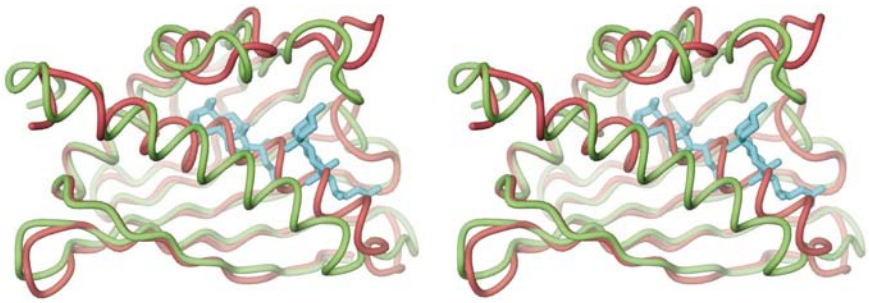

Figure 4
A stereoview of the superposition of LlPR-10.2A molecule *A* (red) on the Bet v 1 molecule (green) in complex with two deoxycholate ligands (cyan) (Markovic-Housley *et al.*, 2003). It is clearly seen that the kink in the α3 helix in LlPR-10.2A decreases the volume of the internal cavity and in consequence excludes the binding of ligands as large as deoxycholate.

3. Results and discussion

3.1. Model quality and overall fold

The refined model of LlPR-10.2A has good overall geometry (Table 3) and the Ramachandran plot statistics (Ramachandran *et al.*, 1963) show that 92.5% of the peptide dihedral angles are in the most favoured regions, 5.6% are in additionally allowed regions and only 1.9% (five residues) are in generously allowed regions. The quality of the electron-density maps is high and there are only a few less clear areas, observed in loop regions.

The two LlPR-10.2A molecules have the same overall folding scheme as observed in other members of the PR-10 family: a long C-terminal helix (α3) is embedded in a protein core formed by a seven-stranded antiparallel β-sheet (Fig. 1). Two additional short helices separate the strands β1 and β2, which form the two edges of the β-sheet. There are nine loops in the polypeptide fold, five of which are β-hairpin loops (L4–L8) connecting the consecutive β-strands of the β-sheet. Between the two main structural elements of the PR-10 fold, the β-sheet and the C-terminal helix α3, there is a hydrophobic cavity running into the protein core.

3.2. The asymmetric unit

The crystallographic asymmetric unit contains two monomers of LlPR-10.2A, designated *A* and *B*. They are related by a translation of about 33 Å approximately along the *b* direction, with a rotation of about 5°. The monomers have almost the same fold, with the r.m.s.d. calculated in *ALIGN* (Cohen, 1997) for 147 C^{α} pairs being 0.58 Å. The highest deviations, amounting to 6.02 Å for Gln126, are observed in loops L5 (Glu59–Thr63), L7 (Gly86–Val88) and L9 (Lys122–Pro127). The L5 and L9 loop regions show some degree of disorder visible in molecule *B* (loop L5) or in both molecules (loop L9). Apart from the two protein monomers, each consisting of 157 amino acids, 281 water molecules, a HEPES molecule and one Na^+ cation were found in the asymmetric unit. The HEPES molecule is located close to loop L6 of molecule *A* and is characterized by a relatively high average *B* factor of 68.8 $Å^2$ when refined with 0.5 occupancy. The correctness of the interpretation of the metal site as a sodium cation is confirmed by the satisfactory refinement of the *B* factor (47.6 $Å^2$) and by the final $Na^+\cdots O$ distances (2.3–2.9 Å). No analogous metal site could be found in molecule *B*, indicating that the binding constant is not very high.

Contacts between the monomers within the asymmetric unit are formed by loops L5, L7 and the N-terminal fragment of helix α3 from molecule *A* and by loops L4, L6 and helix α1 from molecule *B*. The disorder of loop L5 in molecule *B* is

manifested by fragmented electron density. This region is also poorly defined in all previously described structures of PR-10 proteins. Surprisingly, in molecule *A* of LlPR-10.2A, loop L5 appears to be well ordered and is characterized by lower *B* factors (about 20 Å^2 compared with 30 Å^2 in chain *B*) and high electron-density quality (Fig. 2). The stabilization of this loop is a consequence of crystal interactions between the two protein chains. Loop L5 from molecule *A* is in close contact with helix α1, chain β4 and loop L4 from molecule *B*. It is stabilized by Glu59, which forms both intramolecular (with Arg138 NH2 and Lys64 NZ) and intermolecular (with Glu75 N) interactions. It is also hydrogen bonded by numerous well ordered water molecules. Except for the intramolecular hydrogen bond of the Glu59 side chain with Arg138, no such interactions are observed in molecule *B*.

3.3. The β-sheet

The molecular β-sheet consists of seven antiparallel β-strands of connectivity +6x, −1, −1, −1, −1, −1 (Richardson, 1981). The polypeptide chain runs from strand β1 to β2 (which form the edges of the β-sheet) *via* a connecting segment consisting of two short α-helices (α1 and α2). Five hairpin loops join the consecutive strands β2–β7 and bring strand β7 into hydrogen-bonding contact with β1. The entire β-sheet has very regular stable conformation, even though it has a pronounced curvature. In other PR-10 structures, a conserved pattern of β-bulges at the main chain is observed (at Leu70–Asp71, Ile84–Ile85 and Leu94–Glu95 in LlPR-10.1A; Biesiadka *et al.*, 2002). In LlPR-10.2A, the Ile84–Ile85 segment of the β5 strand has regular β-conformation of the main chain in both molecules in the asymmetric unit.

3.4. The C-terminal helix α3

The C-terminal helix α3 contains a severe distortion in the central region, which is clearly visible in the electron density (Fig. 3). A kink of almost 60°, introduced around Phe142, which is present only in the LlPR-10.2A sequence, divides the helix into two distinct fragments. The kinked region of the helix is inserted into a groove in the core of the β-sheet. It is anchored on the N-terminal side by Lys136 and Arg138, which form salt bridges to Asp7 (in β1) and Glu59 (in loop L5). It is also stabilized by hydrophobic interactions of Phe142 and Phe143, which are part of an aromatic cluster also formed by Tyr19 (α1), Tyr80 (β5), Tyr82 (β5) and Phe99 (β6). As a consequence of this arrangement, the internal cavity is much less pronounced than in other PR-10 structures and fewer water molecules are found inside the protein. The two deepest water molecules (Wat1 and Wat11 in molecule *A*) that are present in both LlPR-10.2A monomers form the same hydrogen bonds and are characterized by low *B* factors. However, they are not conserved in the other PR-10 structures. The present observations confirm the predicted flexibility of helix α3 (Biesiadka *et al.*, 2002).

3.5. The hydrophobic cavity

The hydrophobic pocket located inside the protein between the residues of the interior face of the β-sheet and the three α-helices was reported to have potential for binding steroid and other hydrophobic ligands (Mogensen *et al.*, 2002; Markovic-Housley *et al.*, 2003). The narrowing of the internal cavity near the kink of the α3 helix in LlPR-10.2A excludes the binding of ligands as large as deoxycholate (Fig. 4).

To compare the internal cavities in different PR-10 proteins, the *SURFNET* program (Laskowski, 1995) has been used. For the Bet v 1 protein, it yielded a volume of 3500 Å^3 for the cavity as calculated in the original report (Gajhede *et al.*, 1996). A very similar volume has been determined for the LlPR-10.1 proteins: 3500 Å^3 for LlPR-10.1A and 3000 and 3100 Å^3 for the two molecules of LlPR-10.1B (Biesiadka *et al.*, 2002). In contrast, in the present protein the volume is not only significantly reduced but is also divided into two discontinuous segments: 1200 + 1100 Å^3 in molecule *A* and 1100 + 700 Å^3 in molecule *B*. As expected, the discontinuity, or drastic narrowing, of the tunnel is a consequence of the kinked conformation of helix α3.

The presence of many PR-10 homologues in one plant species is a strong indication of their diversified physiological role and suggests variable ligand specificity. One mechanism of controlling such diversity and ligand specificity could be related to modulation of the size and shape of the binding pocket by differences in conformation of helix α3.

It is of interest that a kink in the C-terminal α-helix has also been observed in the START domains, which are involved in lipid transfer in mammalian cells (Roderick *et al.*, 2002). The START domains show remarkable topological similarity to the plant PR-10 proteins (see §3.8) and in particular also have an internal cavity, formed within the C-terminal helix β-grip motif, where their lipid ligands are bound. It has been postulated that these ligands might enter the cavity through a mechanism involving a conformational change of the α-helix and/or of a structurally important loop (Tsujishita & Hurley, 2000). An analysis of the atomic displacement parameters in the present structure (not shown) indicates that in both molecules they are the highest in the L9 loop (about 50 Å^2), which is the structural hinge for helix α3, and in the C-terminal part of helix α3 itself (above 60 Å^2). This pattern, which is quite clear in comparison with the rather low protein $\langle B \rangle$ value (Table 3), suggests that mobility of these structural fragments (especially of the α3 helix) might also play a role in ligand binding by PR-10 proteins.

3.6. The glycine-rich loop

The glycine-rich loop L4 (Gly45–Thr51) with the sequence motif GNGGPGT is one of the most conserved fragments of the PR-10 sequences. Superposition of the L4 loops from all known PR-10 structures shows that it is not only conserved in sequence but that the conformation is almost identical. This loop shows unusual rigidity, which is surprising because this fragment contains four glycine residues that typically cause structural lability. The rigidity of loop L4 can be explained by a characteristic pattern of hydrogen bonds which includes antiparallel β-sheet interactions at its stem (Glu44···Ile52) and three 'internal' hydrogen bonds formed by the

Table 4
Superposition statistics of LlPR-10.2A (molecule *A*) *versus* other crystallographic models of PR-10 proteins.

Calculations were made in *ALIGN* (Cohen, 1997) for C^{α} atoms (auto mode).

Molecule	R.m.s.d. (Å)	Max. distance (Å)	No. pairs	Sequence length	Sequence identity (%)
LlPR-10.2A (*B*)	0.58	6.02	147	157	100
LlPR-10.1A	1.42	9.63	145	155	58
LlPR-10.1B (*A*)	1.28	7.38	149	155	59
LlPR-10.1B (*B*)	1.25	8.26	150	155	59
Bet v 1	1.39	9.37	133	159	48
Bet v 1 (deoxycholate complex)	1.57	9.68	134	159	48
Bet v 1 (E45S mutant)	1.59	9.10	137	159	48
Bet v 1 (antibody complex) (*A*)	1.56	9.05	138	159	48
Bet v 1 (N28T, K32Q, E45S, P108G mutant)	1.44	9.46	135	159	48

Thr51 OG1 atom (to Gly47 N, Gly48 O and Gly48 N) (Fig. 5). In addition, the conformation of loop L4 is stabilized by main-chain···main-chain hydrogen bonding to the adjacent Ile70 residue (β4), which interacts with Gly50 N and Thr51 O. It seems that Thr51, which is almost absolutely conserved in all PR-10 sequences, is crucial for the stability of loop L4. These unusual features of loop L4 suggest that it plays an important structural or functional role. However, no experimental proof has yet been provided.

3.7. The N- and C-termini

The amino-terminus of the LlPR-10.2A protein is highly ordered, forming the β1 strand of the β-sheet. Apart from the intramolecular hydrogen bonding with strand β7, the β1 chain of molecule *A* forms antiparallel β-type interactions with strand β1 of a crystallographically related molecule *B* (Val2 N···Thr4 O, Val2 O···Thr4 N, Thr4 N···Val2 O, Thr4 O···Val2 N). In all PR-10 crystal structures, the N-terminal NH_3^+ group, in addition to being involved in the β1–β7 interaction, is also stabilized by conserved hydrogen bonds to residues in loop L9 (Thr121 O, Gly123 O, Thr121 OG1) (Biesiadka *et al.*, 2002).

Although the protein was expressed in a prokaryotic system, the N-terminal methionine has been removed by a bacterial methionyl aminopeptidase (MAP), as reported in other PR-10 cases (Biesiadka *et al.*, 2002). This is in agreement with the results of Hirel *et al.* (1989), who found that the extent of cleavage of an N-terminal methionine by MAP in *E. coli* is inversely related to the size of the side chain of the amino acid in the subsequent position. Since in recombinant LlPR-10.2A protein the residue following the methionine (*i.e.* the first amino acid in the genomic sequence) is glycine, the excision ratio is 100%.

The carboxyl terminus is ordered and very well defined in the electron density. Its stabilization is mainly provided by intramolecular hydrogen bonds formed by the C-terminal Asn157 residue (Asn157 N···Pro154 O, Asn157 ND2 ···Ser151 O, Asn157 ND2···His153 O), with an additional link between Glu155 and a symmetry-related Ala107. These interactions are present in both LlPR-10.2A molecules. The conformation of the C-terminal fragment is quite different when compared with the other PR-10 proteins, since it is bent in the opposite direction (Fig. 6). Although residue Arg17 at the beginning of helix α1 is strictly conserved between Bet v 1 and LlPR-10.2A, in the lupin protein it is not involved in the stabilization of the C-terminus, in contrast to the situation observed in Bet v 1.[1]

3.8. Structural comparison of PR-10 molecules

So far, the structures of four PR-10 proteins have been reported. Among them are three structures provided by X-ray crystallography, white birch pollen allergen Bet v 1 (Gajhede *et al.*, 1996; Mirza *et al.*, 2000; Spangfort *et al.*, 2003) and two homologues from yellow lupin, LlPR-10.1A and LlPR-10.1B (Biesiadka *et al.*, 2002), and two structures solved by NMR spectroscopy, white birch pollen allergen Bet v 1 (Gajhede *et al.*, 1996; Schweimer *et al.*, 1999) and cherry fruit allergen Pru av 1 (Neudecker *et al.*, 2001, 2003). Superposition of the LlPR-10.2A protein on the other PR-10 structures shows that it adopts the canonical PR-10 fold. However, some significant structural differences are also observed (Fig. 6). In Table 4, the r.m.s.d. values for the superpositions of C^{α} atoms between LlPR-10.2A and all the other PR-10 proteins are listed. A detailed analysis shows that the β-sheet is highly conserved and rigid. The most fundamental differences are located within the C-terminal helix α3. The location and conformation of the α3 helix is different in all known structures of PR-10 proteins, especially in its N-terminal fragment. This is partly a consequence of the flexibility of loop L9, which results in different positioning of the helix at its very beginning. Interestingly, in Bet v 1 the helix is shifted about one coil along the helical axis in comparison with the yellow lupin proteins and Pru av 1. The helix is almost straight in Bet v 1, Pru av 1 and LlPR-10.1B. In LlPR-10.1A, a fragment of this region (residues 135–137) is disordered, as the corresponding electron density was fragmented even in the main chain (Biesiadka *et al.*, 2002). The largest deformation, manifested in a kink of about 60°, is observed in the present LlPR-10.2A protein. The different conformation of helix α3 results in variation of the volume and shape of the hydrophobic cavity, as discussed above.

The LlPR-10.2A protein has also been compared with the START domains of mammalian lipid-binding proteins. A *DALI* (Holm & Sander, 1998) search has identified five PDB entries (1jss, Romanowski *et al.*, 2002; 1em2, Tsujishita & Hurley, 2000; 1ln1, 1ln2, 1ln3, Roderick *et al.*, 2002) with PR-10 topology, which could be superposed on the LlPR-10.2A molecules (C^{α} atoms) with an r.m.s.d. of about 3.3–3.9 Å and with negligible sequence identity (5–9%). This confirms that the START and PR-10 domains belong to the

[1] The interactions of the Arg17 side chain in Bet v 1 involve Ala157 O and not the C-terminal COO group, as mistakenly hinted at by Biesiadka *et al.* (2002).

same structural class, but is not sufficient for definitive conclusions about the physiological role of the latter proteins.

4. Conclusions

Here, we have reported the crystal structure of LlPR-10.2A determined to a resolution of 1.9 Å. LlPR-10.2 is a subclass of pathogenesis-related proteins from yellow lupin, identified on the basis of the clustering of numerous sequences of LlPR-10 proteins. The structure reveals that the two independent molecules of LlPR-10.2A have the same folding pattern as all other PR-10 class members, confirming the overall scheme of a rigid β-sheet wrapped around a variable helix α3. These two structural elements seem to be crucial for the functioning of PR-10 molecules, as they form a cavity within the protein core in which various small-molecule ligands could be bound. The ligand-carrier role of the PR-10 proteins is suggested by their topological similarity to the START domains (r.m.s.d. in C^{α} superpositions of 3.3–3.9 Å), which bind lipids within a similarly formed cavity. The variability of helix α3 has its source in the amino-acid sequence, which shows very little conservation in the N-terminal half of this helix. This sequence variability is reflected in the structural variability, seen as different modes of deformation of helix α3. The degree of this deformation culminates in the present LlPR-10.2A structure, where the α3 helix is kinked in the middle, with the two fragments forming an angle of about 120°. This deformation of helix α3 leads to a decrease of the volume of the cavity, seen *inter alia* as a reduction of the number of water molecules occluded within the protein molecule. We believe that the modulations of the shape and size of the cavity, controlled by the variable sequence of helix α3, have evolved as a way of regulating the specificity of the numerous PR-10 proteins towards their ligands. For instance, it is quite impossible to place the deoxycholate ligands complexed by a PR-10 protein from birch pollen within the cavity of the LlPR-10.2A molecules. In addition, the present structure reveals increased mobility at the hinge and C-terminus of helix α3. In analogy to the START domains, this could be interpreted as an indication of the mechanism (involving a conformational change of the helix) of ligand entry. However, a systematic biochemical study will be necessary for a more general understanding of the rules governing the affinity of different PR-10 proteins towards their small-molecule ligands.

The present structure demonstrates for the first time that at an elevated concentration of Na^{+} the cation is coordinated within one of the loops of the PR-10 fold. However, the biological significance, if any, of this observation is unknown.

We wish to thank Alina Kasperska for skillful help with protein expression and purification. This work was supported by grants from the International Centre for Genetic Engi-

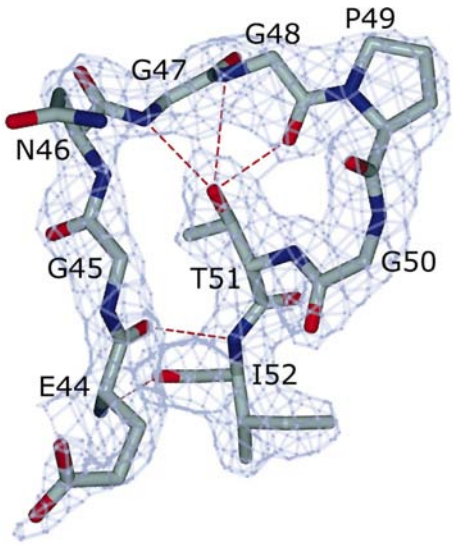

Figure 5
Hydrogen-bonding interactions within the glycine-rich loop L4. The figure shows loop L4 of molecule *A*, but an identical hydrogen-bonding pattern is present in molecule *B*. The $2F_o - F_c$ electron-density map is contoured at the 1.3σ level.

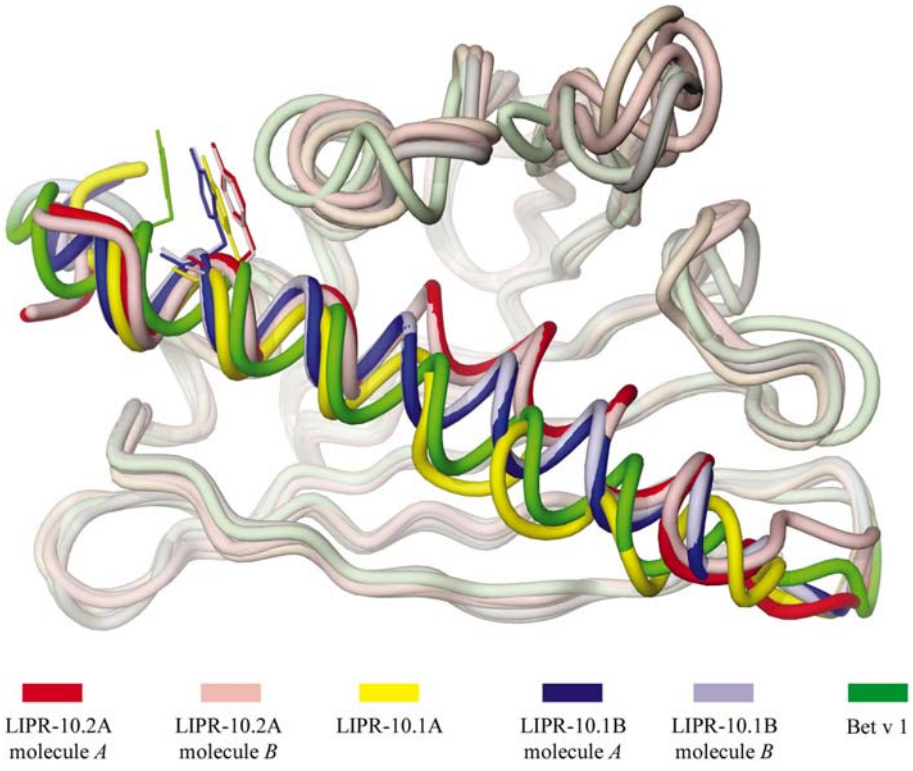

Figure 6
Superposition of all PR-10 molecules of known X-ray structure. The position of Tyr149 (Tyr148 in LlPR-10.1 and Tyr150 in Bet v 1) is indicated to show the axial shift of helix α3 in Bet v 1. Calculations were performed in *ALIGN* (Cohen, 1997) using C^{α}-atom positions.

neering and Biotechnology (ICGEB) to MJ (CRP/POL00-06) and from the State Committee for Scientific Research (KBN) to MMS (6 P04B 004 21) as well as by a subsidy from the Foundation for Polish Science (to MJ). Some of the calculations were carried out in the Poznan Metropolitan Supercomputing and Networking Center.

References

Bantignies, B., Seguin, J., Muzac, I., Dedaldechamp, F., Gulick, P. & Ibrahim, R. (2000). *Plant Mol. Biol.* **42**, 871–881.

Barratt, D. H. P. & Clark, J. A. (1993). *Planta*, **184**, 14–23.

Biesiadka, J., Bujacz, G., Sikorski, M. M. & Jaskolski, M. (2002). *J. Mol. Biol.* **319**, 1223–1234.

Breda, C., Sallaud, C., El-Turk, J., Buffard, D., de Kozak, I., Esnault, R. & Kondorosi, A. (1996). *Mol. Plant-Microbe Interact.* **9**, 713–719.

Breiteneder, H., Ferreira, F., Hoffmann-Sommergruber, K., Ebner, C., Breitenbach, M., Rumpold, H., Kraft, D. & Scheiner, O. (1993). *Eur. J. Biochem.* **212**, 355–362.

Breiteneder, H., Ferreira, F., Reikerstorfer, A., Duchene, M., Valenta, R., Hoffmann-Sommergruber, K., Ebner, C., Breitenbach, M., Kraft, D. & Scheiner, O. (1992). *J. Allergy Clin. Immunol.* **90**, 909–917.

Breiteneder, H., Hoffman-Sommergruber, K., Oriordain, G., Susani, M., Ahorn, H., Ebner, C., Kraft, D. & Scheiner, O. (1995). *Eur. J. Biochem.* **223**, 484–489.

Breiteneder, H., Pettenburger, K., Bito, A., Valenta, R., Kraft, D., Rumpold, H., Scheiner O. & Breitenbach, M. (1989). *EMBO J.* **8**, 1935–1938.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kruszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Bufe, A., Spangfort, M. D., Kahlert, H., Schlaak, M. & Becker, W.-M. (1996). *Planta*, **199**, 413–415.

Cohen, G. R. (1997). *J. Appl. Cryst.* **30**, 1160–1161.

Constabel, C. P. & Brisson, N. (1995). *Mol. Plant-Microbe Interact.* **8**, 104–113.

Crowell, D. N., John, M. E., Russel, D. & Amasino, R. M. (1992). *Plant Mol. Biol.* **18**, 459–466.

Fujimoto, Y., Nagata, R., Fukasawa, H., Yano, K., Azuma, M., Iida, A., Sugimoto, S. & Shudo, K. (1998). *Eur. J. Biochem.* **258**, 794–802.

Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., Larsen, J. N, van Neerven, R. J. J., Schou, C., Lowenstein, H. & Spangford, M. D. (1996). *Nature Struct. Biol.* **3**, 1040–1045.

Handschuh, L. (2004). PhD Thesis. Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland.

Hirel, Ph.-H., Schmitter, J.-M., Dessen, P., Fayat, G. & Blanquet, S. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 8247–8251.

Holm, L. & Sander, C. (1998). *Nucleic Acids Res.* **26**, 316–319.

Jones, T. A., Cowan, S., Zou, J. Y. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Larsen, J. N., Stroman, P. & Ipsen, H. (1992). *Mol. Immunol.* **29**, 703–711.

Laskowski, R. A. (1995). *J. Mol. Graph.* **13**, 323–330.

Loon, L. C. van & van Strien, E. A. (1999). *Phys. Mol. Plant Pathol.* **55**, 85–97.

Loon, L. C. van, Pierpoint, W. S., Boller, T. & Conejero, V. (1994). *Plant Mol. Biol. Rep.* **12**, 245–264.

McGee, J. D., Hamer, J. E. & Hodges, T. K. (2001). *Mol. Plant-Microbe Interact.* **14**, 877–886.

Markovic-Housley, Z., Degano, M., Lamba, D., von Roepenack-Lahaye, E., Clemens, S., Susani, M., Ferreira, F., Scheiner, O. & Breiteneder, H. (2003). *J. Mol. Biol.* **325**, 123–133.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Mirza, O., Henriksen, A., Ipsen, H., Larsen, J. N., Wissenbach, M., Spangford, M. D. & Gajhede, M. (2000). *J. Immunol.* **165**, 331–338.

Mogensen, J. E., Wimmer, R., Larsen, J. N., Spangfort, M. D. & Otzen, D. E. (2002). *J. Biol. Chem.* **277**, 23684–23692.

Moiseyev, G. P., Beintema, J. J., Fedoreyeva, L. I. & Yakovlev, G. I. (1994). *Planta*, **193**, 470–472.

Moiseyev, G. P., Fedoreyeva, L. I., Zhuravlev, Y. N., Yasnetskaya, E., Jekel, P. A. & Beintema, J. J. (1997). *FEBS Lett.* **407**, 207–210.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Mylona, P., Moerman, M., Yang, W.-C., Gloudemans, T., van De Kerckhove, J., van Kammen, A., Bisseling, T. & Franssen, H. J. (1994). *Plant Mol. Biol.* **26**, 39–50.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Nessler, C. L. & Burnett, R. J. (1992). *Plant Mol. Biol.* **20**, 749–752.

Neudecker, P., Lehmann, K., Nerkamp, J., Haase, T., Wangorsch, A., Fotisch, K., Hoffmann, S., Roesch, P., Vieths, S. & Scheurer, S. (2003). *Biochem. J.* **376**, 97–107.

Neudecker, P., Schweimer, K., Nerkamp, J., Scheurer, S., Vieths, S., Sticht, H. & Rosh, P. (2001). *J. Biol. Chem.* **276**, 22756–22763.

Osmark, P., Boyle, B. & Brisson, N. (1998). *Plant Mol. Biol.* **38**, 1243–1246.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Park, C.-J., Kim, K.-J., Shin, R., Park, J. M., Shin, Y.-C. & Paek, K.-H. (2004). *Plant J.* **37**, 186–198.

Pinto, P. M. & Ricardo, C. P. P. (1995). *Plant Physiol.* **109**, 1345–1351.

Pozueta-Romero, J., Klein, M., Houlne, G., Schantz, M.-L., Meyer, B. & Schantz, R. (1995). *Plant Mol. Biol.* **28**, 1011–1025.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.

Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 164–339.

Roderick, S. L., Chan, W. W., Agate, D. S., Olsen, L. R., Vetting, M. W., Rajashankar, K. R. & Cohen, D. E. (2002). *Nature Struct. Biol.* **9**, 507–511.

Romanowski, M. J., Soccio, R. E., Breslow, J. L. & Burley, S. K. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 6949–6954.

Schweimer, K., Sticht, H., Boehm, M. & Roesch, P. (1999). *Appl. Magn. Reson.* **17**, 449–464.

Sikorski, M. M. (1997). *Acta Biochim. Pol.* **44**, 565–578.

Sikorski, M. M., Biesiadka, J., Kasperska, A. E., Kopcinska, J., Lotocka, B., Golinowski, W. & Legocki, A. B. (1999). *Plant Sci.* **149**, 125–137.

Somssich, I. E., Schmelzer, E., Bollmann, J. & Hahlbrock, K. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 2427–2430.

Somssich, I. E., Schmelzer, E., Kawalleck, P. & Hahlbrock, K. (1988). *Mol. Gen. Genet.* **213**, 93–98.

Spangfort, M. D., Mirza, O., Ipsen, H., Van Neerven, R. J., Gajhede, M. & Larsen, J. N. (2003). *J. Immunol.* **171**, 3084–3090.

Spangfort, M. D., Mirza, O., Svensson, L. A., Larsen, J. N., Gajhede, M. & Ipsen, H. (1999). *Acta Cryst.* D**55**, 2035–2036.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods Enzymol.* **185**, 60–89.

Swoboda, I., Hoffman-Sommergruber, K., O'Riordain, G., Scheiner, O., Heberle-Bors, E. & Vicente, O. (1996). *Physiol. Plant.* **96**, 433–438.

Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.

Tsujishita, Y. & Hurley, J. H. (2000). *Nature Struct. Biol.* **7**, 408–414.

Utriainen, M., Kokko, H., Auriola, S., Sarrazin, O. & Karenlampi, S. (1998). *Plant Cell Environ.* **21**, 821–828.

Vanek-Krebitz, M., Hoffman-Sommergruber, K., Machado, M. L. D., Susani, M., Ebner, C., Kraft, D., Scheiner, O. & Breiteneder, H. (1995). *Biochem. Biophys. Res. Commun.* **214**, 538–551.

Walter, M. H., Liu, J.-W., Wunn, J. & Hess, D. (1996). *Eur. J. Biochem.* **239**, 281–293.

Warner, S. A. J., Scott, R. & Draper, J. (1992). *Plant Mol. Biol.* **19**, 555–561.

Warner, S. A. J., Scott, R. & Draper, J. (1993). *Plant J.* **3**, 191–201.

Winn, M., Isupov, M. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.