Structures of two lectins from the roots of pokeweed (*Phytolacca americana*)

**Tomomi Fujii,[a]‡ Minoru
Hayashida,[a]‡ Mika Hamasu,[b]
Masatsune Ishiguro[c]§ and Yasuo
Hata[a]***

[a]Institute for Chemical Research, Kyoto
University, Uji, Kyoto 611-0011, Japan,
[b]Laboratory of Protein Chemistry and
Engineering, Department of Genetic Resources
Technology, Faculty of Agriculture, Kyushu
University, 6-10-1 Hakozaki, Higashi-ku,
Fukuoka 812-8581, Japan, and [c]Laboratory of
Protein Chemistry and Engineering, Graduate
School of Bioresource and Bioenvironmental
Sciences, Kyushu University, 6-10-1 Hakozaki,
Higashi-ku, Fukuoka 812-8581, Japan

‡ These authors contributed equally to this
work.
§ Present address: Department of Food and
Nutrition, Faculty of Food and Nutrition, Beppu
University, Beppu, Ooita 874-8501, Japan.

Correspondence e-mail: hata@scl.kyoto-u.ac.jp

Pokeweed lectin (PL), a lectin specific for *N*-acetylglucos-
amine-containing saccharides, stimulates peripheral lympho-
cytes to undergo mitosis by binding to their cell surfaces. Four
types of lectins have been isolated from the roots of pokeweed
(*Phytolacca americana*) and shown to contain homologous
domains but to have different molecular sizes and biological
properties. PL-D, the smallest lectin in the group, has two
isolectins, PL-D1 and PL-D2. PL-D1 consists of 84 amino-acid
residues, while PL-D2 is identical to PL-D1 in sequence except
for the lack of two C-terminal residues, Leu83 and Thr84. The
crystal structures of PL-D1 and PL-D2 were solved by the
molecular-replacement method and refined to 1.65 and 1.5 Å
resolution with *R* factors of 17.2 and 17.6%, respectively. The
PL-Ds are composed of two repetitive chitin-binding domains,
each of which has four S—S bridges and one putative
carbohydrate-binding site. The two carbohydrate-binding sites
in PL-D are located on one side of the molecule. The relative
orientation of the two domains in PL-D1 differs from that in
PL-D2. Two C-terminal residues of PL-D1 are invisible in the
present crystal structure, indicating the flexibility of the
region. PL-D2 has a Ca^{2+} ion bound to the C-terminus on the
molecular surface. A wide distribution of acidic residues is
characteristically observed on one side of the C-terminal
region of PL-D.

1. Introduction

Lectins are carbohydrate-binding proteins which cause cell
agglutination and glycoprotein sedimentation and are widely
distributed in nature. Binding of lectins to glycoconjugates on
cell surfaces has been shown to be involved in a variety of
biological processes (Lis & Sharon, 1998). Lectins from the
roots of pokeweed (*Phytolacca americana*) are specific for
N-acetylglucosamine-containing saccharides and stimulate
peripheral lymphocytes to undergo mitosis by binding to their
cell surface (Börjeson *et al.*, 1966; Waxdal, 1974; Yokoyama *et
al.*, 1976). Four different types of lectins with different mole-
cular sizes and biological properties have been isolated so far
from pokeweed roots and have been designated PL-A, PL-B,
PL-C and PL-D (Kino *et al.*, 1995; Yamaguchi *et al.*, 1996).
From analysis of their amino-acid sequences (Fig. 1), it has
been shown that these lectins are composed of several chitin-
binding domains and differ from one another in their number
of domains (Yamaguchi *et al.*, 1995, 1996, 1997). The chitin-
binding domains have characteristic structural motifs that are
commonly found in proteins that bind to chitins, such as some
lectins, antimicrobial peptides and class I chitinases (Raikhel
et al., 1993). However, the case of the PLs is unusual in that
several isolectins differing in their number of domains have

been isolated from one species. Therefore, it is of great interest to analyze the crystal structures of PLs using X-ray diffraction in order to elucidate how the spatial arrangement of the carbohydrate-binding sites depends on the number of domains and how it is related to the function of the PLs. The three-dimensional structures of chitin-binding proteins have been determined for hevein from the rubber tree *Hevea brasiliensis*, which is composed of one domain (Rodríguez-Romero *et al.*, 1991; Andersen *et al.*, 1993), the dimeric molecules of wheat germ agglutinin (WGA) isolectin, which is composed of four domains (Wright, 1987, 1989; Harata *et al.*, 1995), and *Uritica dioica* agglutinin (UDA) isolectin, which is composed of two domains (Harata & Muraki, 2000; Saul *et al.*, 2000).

The smallest lectin of the PLs is PL-D. It has two isolectins named PL-D1 and PL-D2 (Yamaguchi *et al.*, 1996). PL-D1 consists of 84 amino-acid residues and has a molecular mass of 9317. PL-D2 has a sequence that is identical to that of PL-D1 except for the lack of two C-terminal residues Leu83 and Thr84. PL-D is composed of two chitin-binding domains, which share 45% sequence identity with each other (Fig. 1). Both PL-D isolectins show agglutinating activity toward trypsin-treated erythrocytes. A pilot experiment in mytosis has suggested that both PL-Ds show a different extent of mitogenic activity toward lymphocytes. In order to establish the structural basis of the PL group, crystal structure analyses of PL-Ds have been performed as the first stage of structural studies of PLs. Here, we report the crystal structures of two PL-D isolectins and discuss their structural features in carbohydrate-binding sites and C-terminal regions.

2. Materials and methods

2.1. Crystal preparation

PL-D1 and PL-D2 were purified as described previously (Yamaguchi *et al.*, 1996). Lyophilized PL-D2 protein was dissolved in 100 mM Tris–HCl buffer pH 7.5 to a concentration of 10 mg ml^{-1} to prepare the crystallization solution. Undissolved protein was removed by centrifugation. Crystals of PL-D2 were initially grown at 298 K by the hanging-drop vapour-diffusion method using a reservoir solution consisting of 18%(w/v) PEG 8000, 0.2 M calcium acetate and 0.1 M sodium acetate buffer pH 4.6. Crystals for the X-ray experiment were obtained by a batch method supplemented with a seeding technique. A crystal obtained by the hanging-drop vapour-diffusion method was crushed into pieces using a needle to prepare microseeds. 1 μl of reservoir solution containing a small number of microseeds was added to 20 μl of

10 mg ml^{-1} PL-D2 solution in a microtube. The tube was stored at 298 K. Crystals appeared in the precipitate and grew within 1–2 weeks. Although many crystals were poorly shaped and stacked, some well shaped single crystals were found in the crystallization tube. PL-D2 required an appropriate amount of calcium ion for crystallization.

Lyophilized PL-D1 protein was dissolved in 50 mM Tris–HCl buffer pH 7.5 to a concentration of 5 mg ml^{-1} to prepare the crystallization solution. Undissolved protein was removed by centrifugation. PL-D1 crystals were grown at 298 K by the hanging-drop vapour-diffusion method using a reservoir solution consisting of 30%(w/v) PEG 8000, 0.2 M sodium acetate and 0.1 M sodium phosphate buffer pH 6.5. A droplet prepared by mixing 2 μl protein solution and 2 μl reservoir solution was equilibrated against 500 μl reservoir solution at 298 K. Crystals for the X-ray experiment were obtained from a drop to which a few microseeds and 0.3 μl 30%(w/v) 6-aminocaproic acid had been added. PL-D1 did not require any calcium ions for crystallization, unlike PL-D2.

2.2. Data collection and processing

Diffraction data for PL-D2 were collected on a screenless Weissenberg camera equipped with a Fuji imaging plate (400 × 200 mm in size) at station BL18B (Sakabe *et al.*, 1995) of the Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization, Tsukuba, Japan.
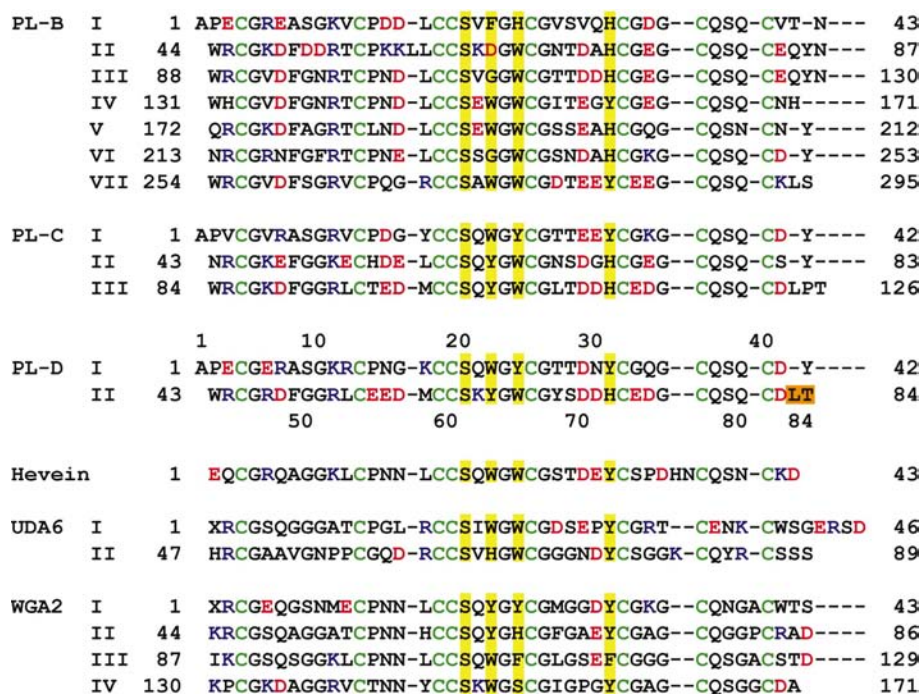
```
PL-B    I     1  APECGREASGKVCPDD-LCCSVFGHCGVSVQHCGDG--CQSQ-CVT-N---    43
        II   44  WRCGKDFDDRTCPKKLLCCSKDGWCGNTDAHCGEG--CQSQ-CEQYN---    87
        III  88  WRCGVDFGNRTCPND-LCCSVGGWCGTTDDHCGEG--CQSQ-CEQYN---   130
        IV  131  WHCGVDFGNRTCPND-LCCSEWGWCGITEGYCGEG--CQSQ-CNH-----   171
        V   172  QRCGKDFAGRTCLND-LCCSEWGWCGSSEAHCGQG--CQSN-CN-Y----   212
        VI  213  NRCGRNFGFRTCPNE-LCCSSGGWCGSNDAHCGKG--CQSQ-CD-Y----   253
        VII 254  WRCGVDFSGRVCPQG-RCCSAWGWCGDTEEYCEEG--CQSQ-CKLS       295

PL-C    I     1  APVCGVRASGRVCPDG-YCCSQWGYCGTTEEYCGKG--CQSQ-CD-Y----    42
        II   43  NRCGKEFGGKECHDE-LCCSQYGWCGNSDGHCGEG--CQSQ-CS-Y----    83
        III  84  WRCGKDFGGRLCTED-MCCSQYGWCGLTDDHCEDG--CQSQ-CDLPT      126

                 1        10        20        30        40
PL-D    I     1  APECGERASGKRCPNG-KCCSQWGYCGTTDNYCGQG--CQSQ-CD-Y----    42
        II   43  WRCGRDFGGRLCEED-MCCSKYGWCGYSDDHCEDG--CQSQ-CDLT        84
                      50        60        70        80  84

Hevein        1  EQCGRQAGGKLCPNN-LCCSQWGWCGSTDEYCSPDHNCQSN-CKD         43

UDA6    I     1  XRCGSQGGGATCPGL-RCCSIWGWCGDSEPYCGRT--CENK-CWSGERSD    46
        II   47  HRCGAAVGNPPCGQD-RCCSVHGWCGGGNDYCSGGK-CQYR-CSSS        89

WGA2    I     1  XRCGEQGSNMECPNN-LCCSQYGYCGMGGDYCGKG--CQNGACWTS----    43
        II   44  KRCGSQAGGATCPNN-HCCSQYGHCGFGAEYCGAG--CQGGPCRAD----    86
        III  87  IKCGSQSGGKLCPNN-LCCSQWGFCGLGSEFCGGG--CQSGACSTD----   129
        IV  130  KPCGKDAGGRVCTNN-YCCSKWGSCGIGPGYCGAG--CQSGGCDA        171
```

Figure 1
Sequence alignment of pokeweed lectins with related proteins. The roman numerals are the names of the domains. The abbreviations are: PL, pokeweed lectin; hevein, antifungal protein from the rubber tree *H. brasiliensis*; UDA6, *U. dioica* agglutinin isolectin 6; WGA2, wheat germ agglutinin isolectin 2. Amino-acid residues are coloured as follows: green, conserved cysteine residues involved in S—S bridges; red, negatively charged residues; blue, positively charged residues. Yellow squares indicate carbohydrate-binding residues conserved in this family. Orange squares indicate the extended residues Leu83–Thr84 in PL-D1, which are truncated in PL-D2.

Two sets of PL-D2 data were collected at 291 K using two crystals with different orientations sealed in glass capillaries. The size of the crystals used was approximately $0.2 \times 0.08 \times 0.02$ mm. One data set was collected with the crystallographic b axis parallel to the spindle axis. It consisted of 11 frames with an oscillation angle of $18.0°$ per frame, a coupling constant of $2.6°$ mm^{-1} and an exposure time of 360 s per frame, covering a range in excess of $180°$. The other data set was collected with the c axis parallel to the spindle axis. It consists of eight frames with an oscillation angle of $15.5°$ per frame, a coupling constant of $1.3°$ mm^{-1} and an exposure time of 372 s per frame, covering a range of over $90°$. The data were processed, scaled and merged with the programs *DENZO* and *SCALE-PACK* (Otwinowski & Minor, 1997). The crystals belong to space group $P2_1$, with unit-cell parameters $a = 23.24$, $b = 56.95$, $c = 29.62$ Å, $\beta = 109.3°$, and contain one molecule in the asymmetric unit with an estimated V_M value of 2.0 Å^3 Da^{-1} (Matthews, 1968).

Diffraction data for PL-D1 were collected on a 2×2 array-type ADSC Quantum-4 CCD detector at station BL6A of the Photon Factory, Tsukuba, Japan. Data collection from PL-D1 was performed at 100 K with a flash-cooled crystal in a liquid-nitrogen gas stream. The size of the crystal used was approximately $0.1 \times 0.1 \times 0.02$ mm. The crystal was soaked in crystallization mother liquor containing $20\%(v/v)$ PEG 400 as a cryoprotectant for about 1 min and mounted in a nylon loop. The data were collected with 180 frames covering a range of $180°$, with an oscillation angle of $1.0°$ per frame and an exposure time of 60 s per frame. The data were processed, scaled and merged with the program *DPS/MOSFLM* (Steller *et al.*, 1997; Leslie, 1992) and the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The crystal belongs to space group $P2_12_12$, with unit-cell parameters $a = 48.67$, $b = 49.01$, $c = 29.93$ Å, and contains one molecule in the asymmetric unit with an estimated V_M value of 1.9 Å^3 Da^{-1} (Matthews, 1968). Data collection and processing statistics for the PL-D crystals are summarized in Table 1.

2.3. Structure determination and refinement

The structure of PL-D2 was determined by the molecular-replacement method using the program *MOLREP* (Vagin & Teplyakov, 1997) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The structure of domain IV (residues 132–169) in wheat germ agglutinin isolectin 3 (WGA3; PDB code 1wgt) was employed as a starting model. From the estimated V_M value it was expected that the asymmetric unit contained one PL-D2 molecule. From comparison of the amino-acid sequences, the PL-D2 molecule was expected to contain two chitin-binding domains. Molecular replacement was carried out on the assumption that two fragments of WGA3 domain IV should exist in the asymmetric unit. The rotation function and the translation function were calculated at 3.5 Å resolution. The first solution was found as the highest peaks in the rotation function and the translation function, yielding a correlation coefficient of 0.201 and an R factor of 0.578. After fixing the position of the fragment for the

Table 1
Data statistics.

Values in parentheses are those for the last shell.

	PL-D1	PL-D2
Data-collection statistics		
Detector (beamline)	CCD (PF-BL6A)	IP (PF-BL18B)
Temperature (K)	100	293
Resolution limit (Å)	40–1.65 (1.75–1.65)	40–1.5 (1.55–1.50)
Observed reflections [$I > 0\sigma(I)$ for PL-D1, $1\sigma(I)$ for PL-D2]	61545	57637
Independent reflections [$I > 0\sigma(I)$ for PL-D1, $1\sigma(I)$ for PL-D2]	9068	11513
Completeness (%)	99.8 (99.8)	98.5 (95.4)
R_{merge}† (%)	4.9 (8.0)	3.1 (12.7)
Refinement statistics		
Resolution limit (Å)	40–1.65 (1.75–1.65)	40–1.5 (1.59–1.50)
No. reflections [$F > 0\sigma(F)$]	9034 (1486)	11513 (1820)
R factor‡ (%)	17.2 (17.6)	17.6 (22.1)
R_{free} factor§ (%)	20.5 (24.7)	20.7 (25.7)
No. atoms		
Protein	622	623
Ca^{2+}	0	1
Water	133	80
Mean B factors (Å^2)		
Main chain	10.3	15.0
Side chain	13.8	19.1
Ca^{2+}	—	12.4
Water	24.7	29.1
R.m.s. deviations from ideal geometry		
Bond lengths (Å)	0.004	0.004
Bond angles (°)	1.3	1.2
Residues in Ramachandran plot¶ (%)		
Most favoured	78.5	86.2
Additionally allowed	21.5	13.8
Generously allowed	0.0	0.0

† $R_{\text{merge}} = \sum_i |I_i - \langle I_i \rangle| / \sum_i I_i$, where $\langle I_i \rangle$ is the average of I_i over all symmetry equivalents. ‡ R factor $= 100 \sum ||F_\text{o}| - |F_\text{c}|| / \sum |F_\text{o}|$, where $|F_\text{o}|$ and $|F_\text{c}|$ are the observed and calculated structure-factor amplitudes, respectively. § The R_{free} factor is the R factor calculated for a selected subset of reflections excluded from the refinement (10% of the total reflections). ¶ Most favored, additionally allowed and generously allowed regions as defined by *PROCHECK* (Laskowski *et al.*, 1993).

first solution, the second solution was found as the third highest peak in the rotation function and the first highest peak in the translation function, yielding a correlation coefficient of 0.370 and an R factor of 0.508. Both fragments positioned for these solutions had adequate room to accommodate the connecting peptide between the C-terminus of one domain and the N-terminus of the other, which confirmed that the fragment models were positioned in a reasonable arrangement. An inspection of the structural model built into the electron-density map was performed with the program *TURBO-FRODO* (Roussel & Cambillau, 1991). Refinement was carried out with the program *X-PLOR* (Brünger, 1992) in the initial stages and with the program *CNS* (Brünger *et al.*, 1998) from the middle of the process onward. Rigid-body refinement was carried out at 10–3.0 Å resolution. The molecular-dynamics method using simulated annealing from 2000 K and manual correction of the model with σ_A-weighted $2F_\text{o} - F_\text{c}$ and $F_\text{o} - F_\text{c}$ maps were applied to the refinement. The resolution was gradually extended from 2.5 to 1.5 Å. In the course of the refinement, water molecules were assigned to

density peaks within hydrogen-bonding distance and at higher than the 3.0σ level in the σ_A-weighted $F_o - F_c$ map and added to the model. Moreover, one large peak, which was larger than those of water molecules, was found at higher than the 5.0σ level in the vicinity of the C-terminus of the protein. It was assigned to a calcium ion based on its coordination. The structural model was further refined by B-factor and positional refinements. The refinement of the structure converged to an R factor of 17.6% for 10 323 reflections in the resolution range 40–1.5 Å. The free R factor was 20.7% for a randomly selected 10% of the data within the same resolution range (1190 reflections).

The structure of PL-D1 was determined by the molecular-replacement method with the program *MOLREP* (Vagin & Teplyakov, 1997) using the refined PL-D2 structure as a starting model. From the estimated V_M value it was expected that the asymmetric unit contained one PL-D1 molecule. The rotation function and the translation function were calculated at 3.5 Å resolution. The solution was obtained as the highest peaks in the rotation function and the translation function. Refinement was carried out using the program *CNS* (Brünger *et al.*, 1998), as in the case of PL-D2. In the rigid-body refinement each domain was treated as a rigid group. Refinement of the structural model converged to an R factor of 17.2% for 8083 reflections within the resolution range 40–1.65 Å. The free R factor was 20.5% for a randomly selected 10% of the data in the same resolution range (951 reflections). The refinement statistics of PL-D1 and PL-D2 are summarized in Table 1.
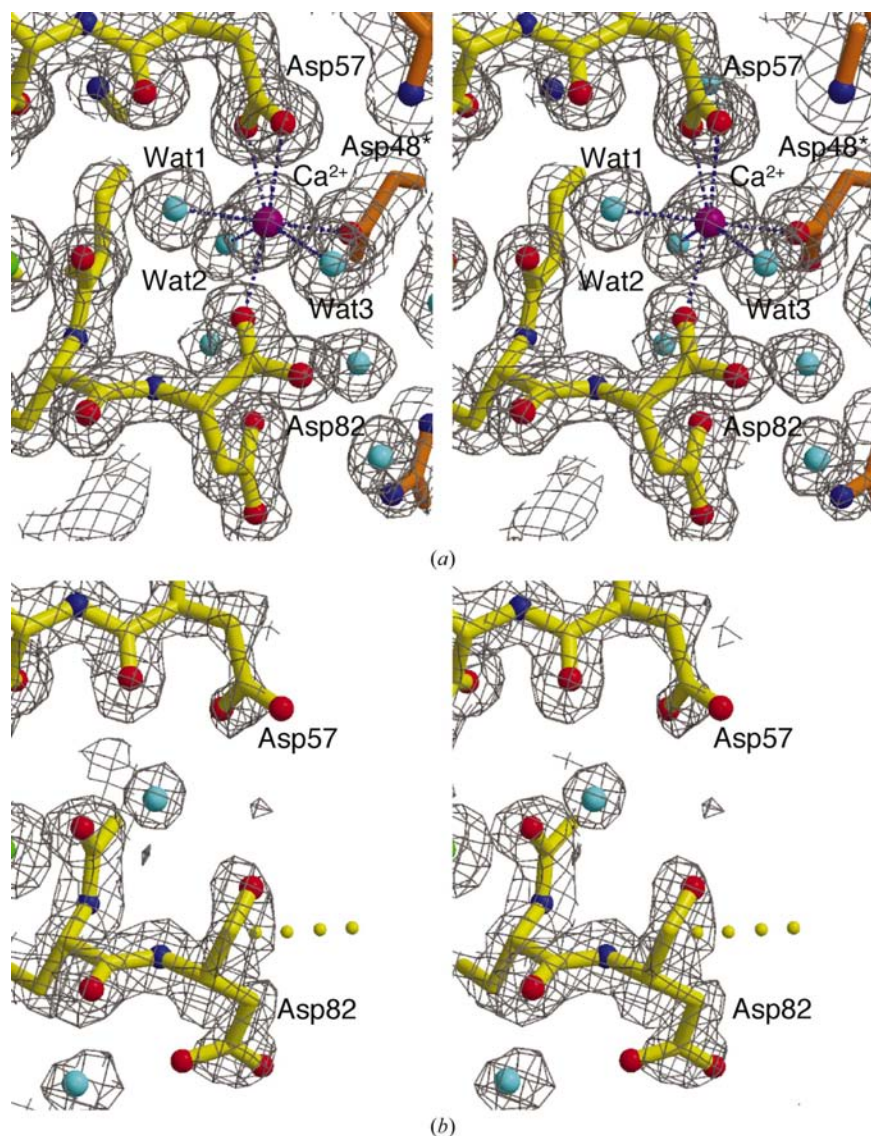
3. Results and discussion

3.1. Overall structure

The final models contain the polypeptide chain of residues 1–82 and 132 water molecules for PL-D1 and the polypeptide chain of residues 1–82, 81 water molecules and one calcium ion for PL-D2. The PL-D2 molecule binds a calcium ion at the C-terminus in the crystal (Fig. 2a). Two C-terminal residues of the PL-D1 molecule, Leu83 and Thr84, do not give clear electron density (Fig. 2b). In contrast to the compact packing through the calcium ion around the C-terminus in the PL-D2 crystal, there is enough space to accommodate the two C-terminal residues which are disordered in the PL-D1 crystal, which adopts a different crystal system to the PL-D2 crystal (Fig. 2). The polypeptide fold of PL-D consists of two chitin-binding domains (Fig. 3). The first domain (domain I) comprises residues 1–42 and the second domain (domain II) residues 43–84 and 43–82 in PL-D1 and PL-D2, respectively. These domains have no characteristic secondary-structure architecture but contain several short secondary-structure elements. There are two short α-helices composed of Asp29–Cys32 in domain I and Asp70–Cys73 in

Figure 2
Electron-density maps in the C-terminal region of PL-Ds. σ_A-Weighted $2F_o - F_c$ electron-density maps contoured at the 3.5σ level are represented in grey. Parts of one polypeptide chain are identified as yellow sticks and those of the other polypeptide related by a crystallographic symmetry operation are drawn as orange sticks. O and N atoms are shown as red and blue balls, respectively. Water molecules are shown as cyan balls. (a) The coordination mode to calcium ion in PL-D2 crystal. The calcium ion is depicted as a large purple ball. Coordination bonds to the calcium ion are depicted as blue broken lines. The residue name labelled with an asterisk indicates that the residue belongs to the crystallographically related molecule. (b) The C-terminal region of PL-D1 lacking density for Leu83–Thr84. Small yellow balls represent the unmodelled peptide region Leu83–Thr84. Fig. 2 was drawn with the program *BOBSCRIPT* (Esnouf, 1997); Figs. 3, 4, 5, 6 and 7a were drawn with *MOLSCRIPT* (Kraulis, 1991) and Fig. 7b was drawn with *GRASP* (Nicholls *et al.*, 1991). All these figures were rendered with the program *Raster3D* (Merritt & Bacon, 1997).

Table 2
Comparison of the domains of PL-Ds, UDA6 and WGA2.

| | R.m.s. C^α distance† | | No. of charged residues‡ | | |
	PL-D2 I	PL-D2 II	Acidic	Basic	Acidic − basic
PL-D2 I	—	0.42	—	—	—
PL-D1 I	0.28	0.40	4	4	0
PL-D1 II	0.43	0.24	9	4	5
Hevein§	1.27	1.33	5	3	2
UDA6 I	0.78	1.02	5	5	0
UDA6 II	1.44	1.42	2	4	−2
WGA2 I	0.62	0.65	3	2	1
WGA2 II	0.58	0.62	2	3	−1
WGA2 III	0.55	0.56	2	2	0
WGA2 IV	0.58	0.55	2	4	−2
WGA3 IV¶	0.60	0.57	2	4	−2
PL-B I	—	—	5	2	3
PL-B II	—	—	7	6	1
PL-B III	—	—	6	2	4
PL-B IV	—	—	4	1	3
PL-B V	—	—	4	3	1
PL-B VI	—	—	3	4	−1
PL-B VII	—	—	6	4	3
PL-C I	—	—	4	3	1
PL-C II	—	—	6	3	3
PL-C III	—	—	8	3	5

† The r.m.s. C^α distance between domain I (residues 4–40) or domain II (residues 45–81) of PL-D2 and the corresponding region of each domain. The PDB coordinates used for this comparison are 1hev (hevein), 1ehd (UDA6), 9wga (WGA2) and 1wgt (WGA3). ‡ Acidic, the number of acidic residues; basic, the number of basic residues; acidic − basic, the number of acidic residues minus the number of basic residues. § The r.m.s. C^α distance of hevein is the average over those of six NMR models. ¶ WGA3 IV is the starting model for the molecular replacement of PL-D2.
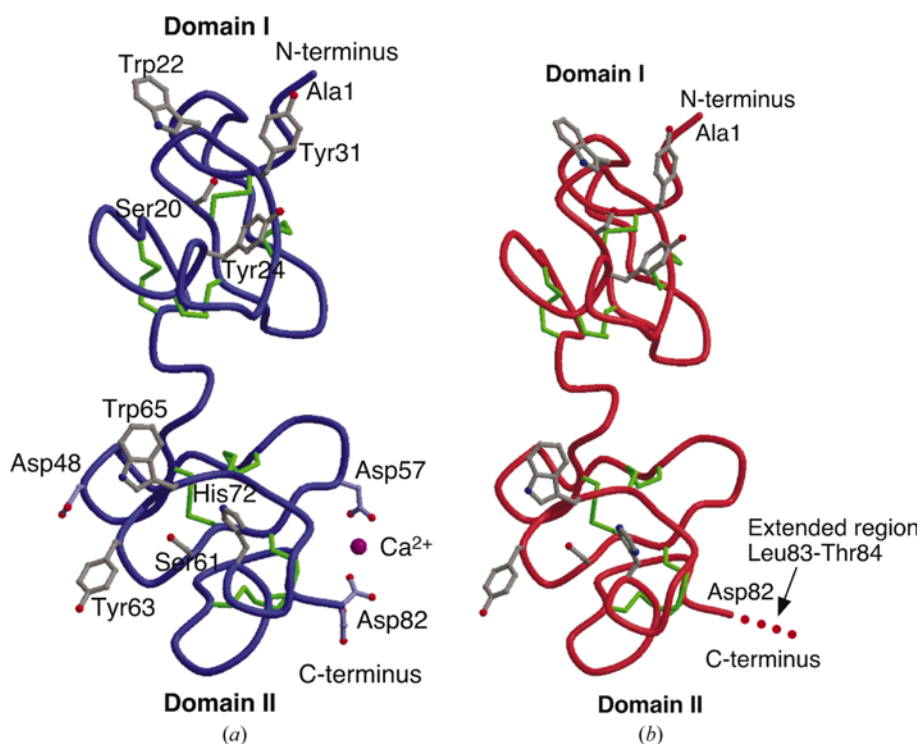
Figure 3
Overall structures of PL-D2 (*a*) and PL-D1 (*b*). The carbohydrate-binding residues conserved in this lectin family are shown as grey ball-and-stick models. Disulfide bridges are shown as green sticks. The calcium ion and residues interacting with the ion in the crystal are shown as a large pink ball and as light blue sticks, respectively.

domain II. Three 3_{10}-helical features composed of three residues also exist in the regions Glu6–Ala8 and Pro14–Gly16 in domain I and Arg47–Phe49 in domain II. The regions Cys18–Cys19, Cys25–Gly26, Cys59–Cys60 and Cys66–Gly67 have β-strand features and the former two and the latter two strands adopt antiparallel β-sheets in domain I and in domain II, respectively. PL-D has four S—S bridges in each domain: Cys4–Cys19, Cys13–Cys25, Cys18–Cys32 and Cys36–Cys40 in domain I and Cys45–Cys60, Cys54–Cys66, Cys59–Cys73 and Cys77–Cys81 in domain II. These S—S bridge sites are the same as those found in the domains of other chitin-binding proteins. These structural elements make each domain rigid. Structural comparisons between the domains of PL-D2 and those of related chitin-binding proteins are listed in Table 2. The length of the linker region connecting the domains in PL-D is the shortest among the related chitin-binding proteins; the linkers, each of which is encompassed by the last Cys residue in one domain and the first Cys residue in the next domain, are composed of four, nine and five residues for PL-D, *U. dioica* agglutinin isolectin 6 (UDA6) and wheat germ agglutinin isolectin 2 (WGA2), respectively (Fig. 1).

3.2. Carbohydrate-binding site

To date, several tertiary structures of chitin-binding proteins complexed with saccharide have been reported for UDA (Saul *et al.*, 2000; Harata & Muraki, 2000), hevein (Asensio *et al.*, 1995) and WGA (Wright, 1990; Wright & Jaeger, 1993). Structural information on the putative carbohydrate-binding residues, sites and mode can be obtained from comparison between the structures of the related proteins.

In PL-D, the carbohydrate-binding residues corresponding to those reported for UDA, hevein and WGA are Ser20, Trp22, Tyr24 and Tyr31 in domain I and Ser61, Trp63, Tyr65 and His72 in domain II (Fig. 1). Structural comparisons between the domains of UDA6 complexed with tri-*N*-acetyl-chitotriose (Harata & Muraki, 2000) and PL-D2 show that the arrangement of these carbohydrate-binding residues is almost the same in all domains in both UDA6 and PL-D2 (Fig. 4). In the crystal of the UDA complex with tri-*N*-acetyl-chitotriose, the saccharide molecule is sandwiched between the carbohydrate-binding site of domain I of one UDA6 molecule and that of domain II of the other UDA6 molecule (Harata & Muraki, 2000). The main interactions between the saccharide and UDA6 molecules are face-to-face stacking interactions between the pyranose rings of the saccharide and the aromatic side

chains of Trp21 and Trp23 in domain I and His67 and Trp69 in domain II of UDA6. A stacking interaction also exists between the acetylamino group of the saccharide and the phenol ring of Tyr30 in domain I and Tyr76 in domain II of UDA6. Several of these carbohydrate-binding residues in UDA6, however, are of different types to the corresponding residues in PL-D. Tyr24 in domain I of PL-D replaces Trp23 in UDA6. In WGA2, a variety of aromatic residues, Tyr, His and Phe, are found at this site (Fig. 1) and interact with saccharides in the complexes of wheat germ agglutinin isolectin 1 (WGA1) and WGA2 (Wright, 1990; Wright & Jaeger, 1993). Tyr63 in domain II of PL-D replaces His67 in UDA6. The residue at this site of WGA2 is also Tyr in domains I and II (Fig. 1). His72 in domain II of PL-D replaces Tyr76 in UDA6, as does Phe in domain III of WGA2. Although a few aromatic residues for carbohydrate binding in PL-D differ from the corresponding

aromatic residues in UDA6, the arrangement of the carbohydrate-binding residues in PL-D is almost the same as that in the UDA6 complexes. Therefore, PL-D would probably interact with tri-N-acetylchitotriose in a face-to-face stacking manner, similar to that in UDA6, between the pyranose rings of the saccharide and the aromatic side chains of the proteins.

In the carbohydrate-binding site of the UDA complex, the phenolic hydroxyl groups of Tyr30 in domain I and Tyr76 in domain II of UDA6 hydrogen bond to the pyranose O3 hydroxyl groups of the saccharide. Hydrogen bonds also exist between the acetylamino groups of the saccharide and the hydroxyl groups of Ser19 in domain I and Ser65 in domain II. These interactions may be important in the recognition of the N-acetylglucosamine residues. The latter hydrogen bonds are probably preserved in the complex of PL-D with saccharide because these Ser residues are conserved in PL-D. The former hydrogen bonds are not similarly preserved in the complex of PL-D. This is because PL-D substitutes His72 for Tyr76 in domain II of UDA6, although the Tyr residue corresponding to Tyr30 in domain I of UDA6 is conserved in PL-D. This difference is the most interesting difference between the carbohydrate-binding residues of PL-D and those of other chitin-binding proteins, including UDA6. The Tyr residue at the site of Tyr76 in domain II of UDA6 is almost completely conserved in hevein, UDA6 and WGA2, except for Phe116 in domain III of WGA2 (Fig. 1). The substitution of His for Tyr at this site is found not only in PL-D but also in PL-B and PL-C (Fig. 1), which is probably characteristic of PLs. The $N^{\varepsilon 2}$ atom of the imidazole ring of His72 in PL-D can interact with the O3 hydroxyl group of saccharide through the hydrogen bond, as found in the phenolic hydroxyl group of Tyr30 or Tyr76 in the UDA6–saccharide complex. Structural comparison of the UDA6–saccharide complex with PL-D2 shows that the distance (4.1 Å) between the $N^{\varepsilon 2}$ atom of the imidazole ring of His72 in PL-D2 and the pyranose O3 hydroxyl group of the saccharide in the UDA6 complex is longer than that (2.7 Å) between the phenolic O atom of Tyr76 in UDA6 and the same group of the saccharide (Fig. 4). If the position of the saccharide in the PL-D complex is slightly closer to His72 than in UDA6, PL-D may bind the saccharide molecule in the same manner as found in the complexes of related proteins.
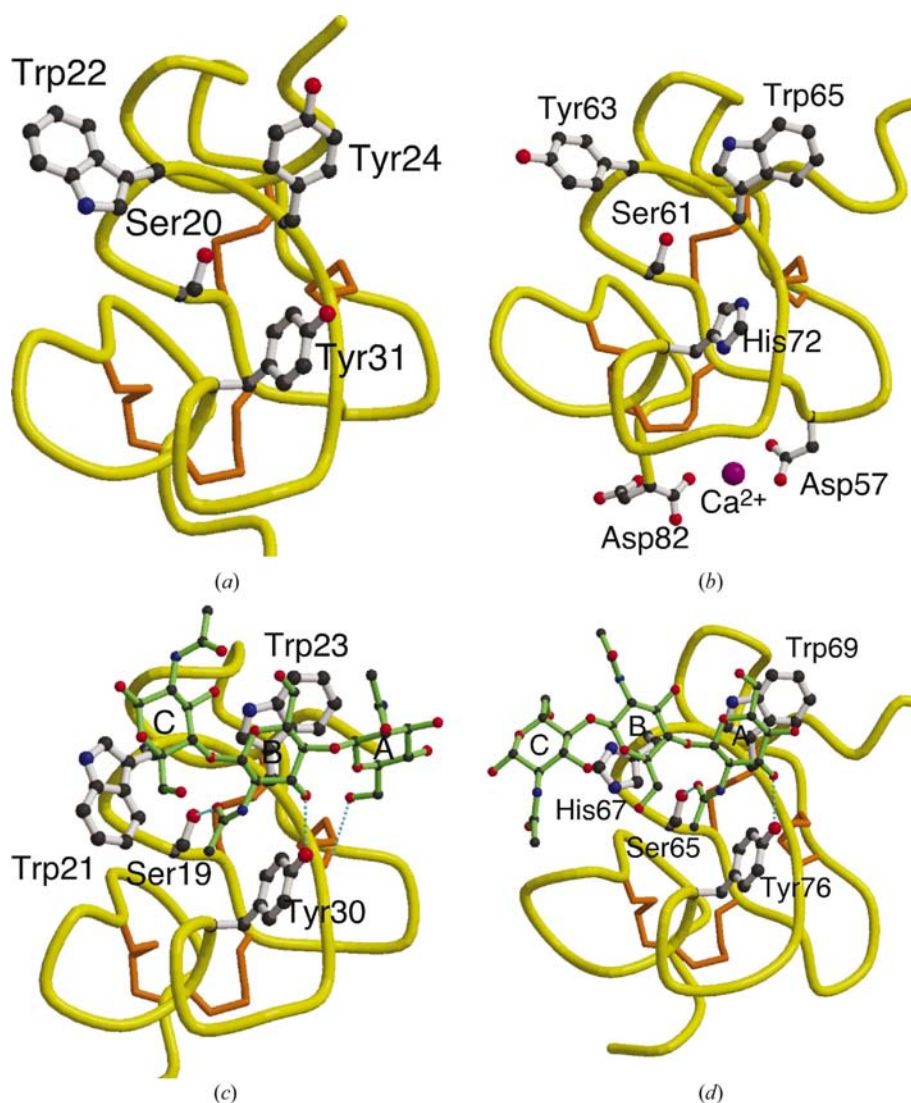

Figure 4
Structural comparison between carbohydrate-binding sites: (*a*) domain I of PL-D2, (*b*) domain II of PL-D2, (*c*) domain I of UDA6 complexed with tri-N-acetylchitotriose and (*d*) domain II of UDA6 complexed with tri-N-acetylchitotriose. Each site is shown in almost the same orientation. Thick ball-and-stick models show carbohydrate-binding residues. Disulfide bridges are shown as orange sticks. Thin ball-and-stick models show tri-N-acetylchitotriose. The hydrogen bonds between the saccharide molecule and the protein are shown by dotted lines coloured cyan.
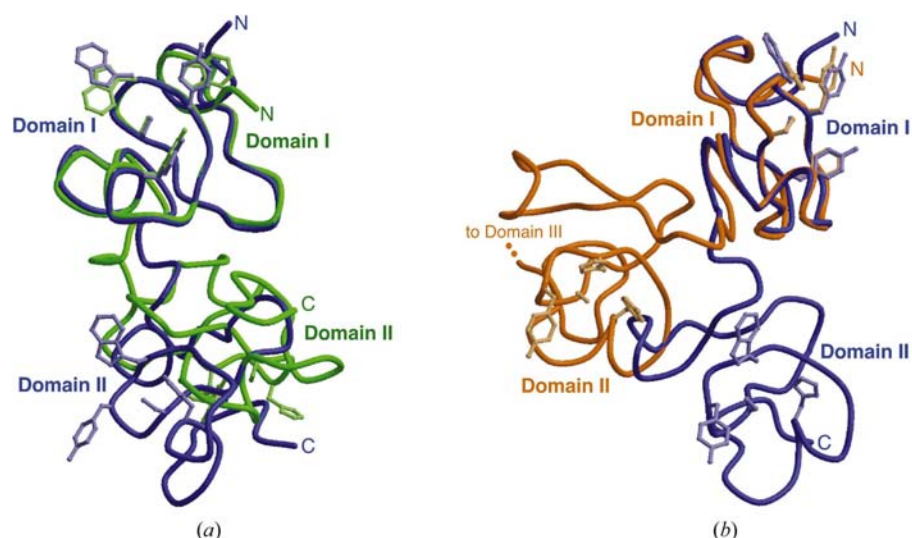
Figure 5
Superpositions of domains I of UDA6 (*a*) and WGA2 (*b*) on that of PL-D2. All corresponding C$^\alpha$ atoms between these domains I were fitted by a least-squares method. Blue line, PL-D2; green line, UDA6; orange line, WGA2. The carbohydrate-binding residues are shown by light blue, light green and light orange sticks for PL-D2, UDA6 and WGA2, respectively.
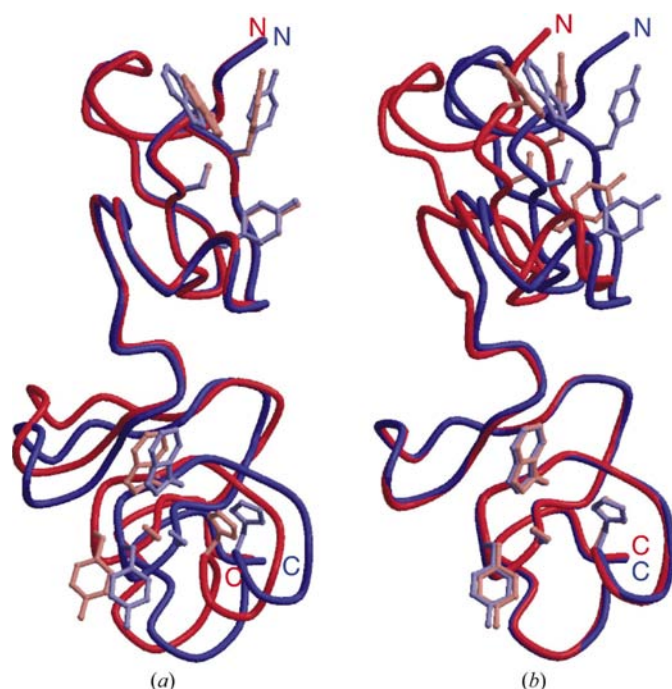

Figure 6
Superposition of each domain of PL-D1 on the corresponding domain of PL-D2: (*a*) domain I and (*b*) domain II. All corresponding C$^\alpha$ atoms between the domains of PL-Ds were fitted by the least-squares method. Blue line, PL-D2; red line, PL-D1. The carbohydrate-binding residues are shown as light blue and light red sticks for PL-D2 and PL-D1, respectively.

3.3. Arrangement of domains

One of the structural features of PL-D lies in the arrangement of the domains, which defines the positional relationship between two putative carbohydrate-binding sites. In PL-D, both putative carbohydrate-binding sites are located on one side of the PL-D molecule (Figs. 3 and 5). In contrast, the two

domains in UDA6 are linked by a linker peptide that is five residues longer than that of PL-D (Fig. 1) and the two carbohydrate-binding sites are oriented on opposite sides of the molecule (Fig. 5*a*). In UDA, it was proposed that the arrangement of the carbohydrate-binding sites would define the activation type of the T-cell receptor on lymphocytes (Saul *et al.*, 2000). A subunit of the WGA dimeric molecule consists of four chitin-binding domains (Fig. 1). When domain I of the PL-D2 molecule is superposed on that of the WGA2 subunit, domain II of PL-D2 is oriented in a completely different direction from that in WGA2 (Fig. 5*b*). As observed in these comparisons, the difference in linker conformation between the related lectins leads to that in domain arrangement among them. The linker conformation of each lectin seems to be uniquely defined depending on the linker length and sequence, the oligomerization of the molecule and so on. This phenomenon might lead to the creation of properties that are characteristic of each chitin-binding lectin.

The relative orientation of the two domains in PL-D1 is different from that in PL-D2, with a 14° extra inclination of one domain realtive to the other (Fig. 6). The angle was calculated with the program *DYNDOM* (Hayward & Berendsen, 1998). The slight differences in the φ and ψ angles of the linker peptide Asp41–Trp43 between PL-D1 and PL-D2 give rise to the difference in domain arrangement between them. The r.m.s. deviations of the corresponding C$^\alpha$ atoms in each domain between PL-D1 and PL-D2 are 0.28 and 0.24 Å for domain I and domain II, respectively (Table 2). No significant structural differences are observed between the corresponding domains of PL-D1 and PL-D2. Because the substantial difference between PL-D1 and PL-D2 lies only in the addition of two C-terminal residues in PL-D1, the different arrangement of domains observed above may be caused by the difference in crystal packing. These facts indicate the flexibility of the linker between the two domains in PL-D. The flexibility of the linker peptide between the domains may be important for the chitin-binding lectin in adjusting the relative orientation of the carbohydrate-binding sites to the relative position of the target carbohydrates for its easy binding.

3.4. Calcium ion-binding site

In the crystal structure of PL-D2, a calcium ion contained in the crystallization solution binds to the C-terminal region of the lectin (Fig. 3). The β-carboxyl group of Asp57, the α-carboxyl group of Asp82 (the C-terminal residue), the β-carboxyl group of Asp48 in the neighbouring molecule and three water molecules coordinate to the ion in an octahedral

geometry (Fig. 2). The calcium–ligand distances range from 2.36 to 2.61 Å (2.47 Å on average). The side-chain carboxylate of Asp57 appears to interact with the calcium ion in the deprotonated resonance form. The neighbouring molecule providing Asp48 as a ligand for the calcium ion belongs to the next unit cell along the *a* axis. In addition to these interactions, Asp57 and Asp82 also make salt bridges with Arg44 and Arg47 of the neighbouring molecule, respectively. Although these interactions appear to participate in the formation of the crystal lattice, the physiological role of these interactions is unknown. Calcium ions are essential for carbohydrate binding in C-type lectins from animals (Drickamer, 1999). The legume lectin concanavalin A binds divalent cations that are involved in carbohydrate-binding activity (Bouckaert *et al.*, 2000). In PL-D2, however, the calcium ion does not appear to be involved in carbohydrate binding because the calcium-binding site is located on the opposite side to the putative carbohydrate-binding site of domain II (Fig. 4*b*).

Figure 7
Distribution of acidic residues in PL-D2. (*a*) C^{α}-trace model with charged residues. The carbohydrate-binding residues, acidic residues and basic residues are shown as grey, light red and light blue ball-and-stick representations, respectively. Disulfide bridges are shown as orange sticks. (*b*) Electrostatic potential surface. The molecule is shown in the same orientation as that in (*a*). The potentials are coloured from $-10kT$ (red) to $+10kT$ (blue).

In PL-D1, the lack of the α-carboxyl group of Asp82 owing to the extension at C-terminus may prevent the binding of a calcium ion in the same manner as that found in the PL-D2 crystal (Fig. 2). This may be partly supported by the experimental result that PL-D1 crystals were not produced under the PL-D2 crystallization conditions, which contained calcium ions.

3.5. Distribution of acidic residues

Another structural feature of PL-D is the existence of a cluster of acidic residues on the molecular surface at the C-terminal side. In the vicinity of the putative carbohydrate-binding site composed of Ser61, Trp63, Tyr65 and His72 in domain II, there is an acidic patch consisting of the carboxyl side chains of Asp70, Asp71, Glu74 and Asp75 (Fig. 7*a*). The α-carboxyl group of the C-terminal Asp82 in PL-D2 and the β-carboxyl groups of Asp82 and Asp57 in PL-D are located slightly further from the acidic patch. Furthermore, the carboxyl side chains of Glu55 and Glu56 are found next to those of Asp57 and Asp82 (Fig. 7*a*). These eight residues form a cluster of acidic residues corresponding to a relatively large area of highly negative potential on the molecular surface at the C-terminal side of PL-D (Fig. 7*b*).

A comparison of the amino-acid sequences among members of the PL family revealed that PL-B and PL-C probably have acidic patches (Fig. 1). The acidic patch corresponding to th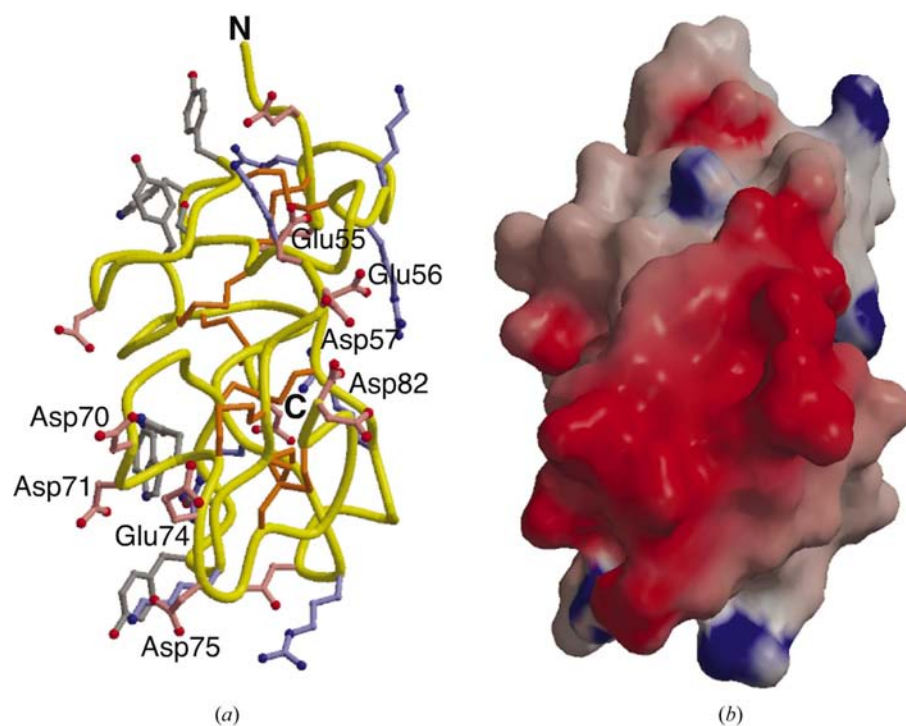at composed of Asp70, Asp71, Glu74 and Asp75 in domain II of PL-D is conserved in domain III of PL-C and domain VII of PL-B, although several aspartate residues are replaced by glutamate residues. Moreover, domain III of PL-B has a patch composed of three acidic residues in this site. Hevein has an acidic patch composed of three residues as in PLs. However, both UDA6 and WGA2 have only one acidic residue in this site (Fig. 1). A larger number of acidic residues are distributed in PLs than in WGA2 and UDA6 (Table 2).

3.6. Structural features that possibly lead to functional difference

A recent pilot mitosis experiment has suggested that the mitogenic activity of PL-D1 toward lymphocytes differs from that of PL-D2: PL-D1 does not show this activity, while PL-D2 does (K. Yamaguchi, M. Uechi & M. Ishiguro, unpublished data). Several structural differences between PL-D1 and PL-D2 which may lead to this difference in their mitogenic activity have been observed in the C-terminal region. One of the differences is that PL-D1 has two C-terminal residues, Leu83–Thr84, that extend from the C-terminus of PL-D2. These two residues were invisible in the electron-density map and seem to be disordered in the crystal. Another structural difference is that PL-D2 binds one Ca^{2+} ion around the C-terminus, in contrast to PL-D1. Moreover, a cluster of acidic residues exists in the vicinity of the C-terminal region of PL-D. It is unknown how these structural features are involved in mitogenic

activity. Further experiments in mitogenic activity should be carried out in order to demonstrate these functional differences and to elucidate the structural requirements for mitogenic activation.

References

Andersen, N. H., Cao, B., Rodríguez-Romero, A. & Arreguin, B. (1993). *Biochemistry*, **32**, 1407–1422.

Asensio, J. L., Cañada, F. J., Bruix, M., Rodriguez-Romero, A. & Jimenez-Barbero, J. (1995). *Eur. J. Biochem.* **230**, 621–633.

Börjeson, J., Reisfeld, R. A., Chessin, L. N., Welsh, P. D. & Douglas, S. D. (1966). *J. Exp. Med.* **124**, 859–872.

Bouckaert, J., Loris, R. & Wyns, L. (2000). *Acta Cryst.* D**56**, 1569–1576.

Brünger, A. T. (1992). *X-PLOR Version 3.1. A System for Crystallography and NMR.* Yale University Press, New Haven, CT, USA.

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

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Drickamer, K. (1999). *Curr. Opin. Struct. Biol.* **9**, 585–590.

Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.

Harata, K. & Muraki, M. (2000). *J. Mol. Biol.* **297**, 673–681.

Harata, K., Nagahora, H. & Jigami, Y. (1995). *Acta Cryst.* D**51**, 1013–1019.

Hayward, S. & Berendsen, H. J. C. (1998). *Proteins*, **30**, 144–154.

Kino, M., Yagaguchi, K., Umekawa, H. & Funatsu, G. (1995). *Biosci. Biotechnol. Biochem.* **59**, 683–688.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Laskowski, R. A. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leslie, A. G. W. (1992). *Jnt CCP4–ESRF/EACMB Newsl. Protein Crystallogr.* **26**, 27–33.

Lis, H. & Sharon, N. (1998). *Chem. Rev.* **98**, 637–674.

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

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.

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

Raikhel, N. V., Lee, H.-I. & Broekaert, W. F. (1993). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**, 591–615.

Rodríguez-Romero, A., Ravichandran, K. G. & Soriano-Garcia, M. (1991). *FEBS Lett.* **291**, 307–309.

Roussel, A. & Cambillau, C. (1991). *Silicon Graphics Geometry Partners Directory*, pp. 86. Mountain View, CA, USA: Silicon Graphics.

Sakabe, N., Ikemizu, S., Sakabe, K., Higashi, T., Nakagawa, A., Watanabe, N., Adachi, S. & Sasaki, K. (1995). *Rev. Sci. Instrum.* **66**, 1276–1281.

Saul, F. A., Rovira, P., Boulot, G., Van Damme, E. J. M., Peumans, W. J., Truffa-Bachi, P. & Bentley, G. A. (2000). *Structure*, **8**, 593–603.

Steller, I., Bolotovsky, R. & Rossman, M. G. (1997). *J. Appl. Cryst.* **30**, 1036–1040.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Waxdal, M. (1974). *Biochemistry*, **13**, 3671–3677.

Wright, C. S. (1987). *J. Mol. Biol.* **194**, 501–529.

Wright, C. S. (1989). *J. Mol. Biol.* **209**, 475–487.

Wright, C. S. (1990). *J. Mol. Biol.* **215**, 635–651.

Wright, C. S. & Jaeger, J. (1993). *J. Mol. Biol.* **232**, 620–638.

Yamaguchi, K., Mori, A. & Funatsu, G. (1995). *Biosci. Biotechnol. Biochem.* **59**, 1384–1385.

Yamaguchi, K., Mori, A. & Funatsu, G. (1996). *Biosci. Biotechnol. Biochem.* **60**, 1380–1382.

Yamaguchi, K., Yurino, N., Kino, M., Ishiguro, M. & Funatsu, G. (1997). *Biosci. Biotechnol. Biochem.* **61**, 690–698.

Yokoyama, K., Yano, O., Terao, T. & Osawa, T. (1976). *Biochim. Biophys. Acta*, **427**, 443–452.